



**Politecnico  
di Torino**

## Politecnico di Torino

Corso di Laurea Magistrale in Ingegneria Gestionale

A.a. 2023/2024

Sessione di Laurea Dicembre 2024

# **Machine Data in campo Automotive**

Relatori:

Carlo Cambini

Candidati:

Alessia Barbagallo



## Indice

ABSTRACT .....	5
1. DATA SCIENCE .....	7
1.1. EVOLUZIONE DEI DATI E DELLE TECNICHE DI ANALISI.....	7
1.2. BIG DATA DRIVEN .....	9
1.2.1. LE TRE (CINQUE) V DEI BIG DATA .....	9
1.2.2. CASI D'USO: TRASFORMARE I BIG DATA IN UN VANTAGGIO COMPETITIVO .....	12
MANUFACTURING .....	13
RETAIL .....	15
HEALTHCARE .....	16
OLI&GAS .....	16
FINANCIAL.....	17
1.3. IL CICLO DEI BIG DATA .....	18
1.3.1. BIG DATA MANAGEMENT.....	20
1.3.2. BIG DATA ANALYTICS .....	23
1.4. INFRASTRUTTURA IT PER I BIG DATA .....	25
1.4.1. CLOUD COMPUTING .....	26
1.5. STRUMENTI ANALITICI PER I BIG DATA .....	28
2. AUTOMOTIVE DATA MANAGEMENT .....	34
2.1. LA RIVOLUZIONE DEI BIG DATA NELL'AUTOMOTIVE: DALLE AUTO CONNESSE ALLA TRASFORMAZIONE DELLA CUSTOMER EXPERIENCE .....	34
2.1.1. "PROJECT 5.0": THE DIGITAL WAVE.....	39
2.2. MOBILISIGHT: STELLANTIS'S DATA BUSINESS.....	41
2.2.1. RIVOLUZIONARE LA MOBILITÀ CON I DATI IN TEMPO REALE.....	43
2.3. CONNECTED VEHICLES CV.....	45
2.3.1. ARCHITETTURA DI BIG DATA APPLICATIVA IN CAMPO AUTOMOTIVE .....	48
2.3.2. APPLICAZIONE: SERVIZIO METEOROLOGICO E SERVIZIO ECO-DRIVING .....	54
2.3.3. VALUTAZIONE PRESTAZIONI: QUALITÀ DEI DATI.....	55
2.4. METODOLOGIA DATA-DRIVEN PER LA VALUTAZIONE DELLA PROGETTAZIONE HMI.....	58
2.5. BIG DATA NEL GREEN CAMPUS .....	63
2.5.1. HYDROGEN FUEL CELL VEHICLE, HFCV.....	67

2.5.2. STELLANTIS CIRCULAR ECONOMY: "SUSTAINera" .....	79
CONCLUSIONE .....	81
BIBLIOGRAFIA .....	83
SITOGRAFIA .....	84



# *ABSTRACT*

L'elaborato esplora l'importanza e l'evoluzione dei Big Data nel supporto decisionale e nel vantaggio competitivo per le aziende, con particolare attenzione all'automotive. Viene analizzato il ciclo di vita dei Big Data, che include la raccolta, gestione, archiviazione e analisi dei dati: si descrive come la trasformazione dei dati grezzi in informazioni utili passi attraverso tecniche avanzate e sia possibile solo grazie alla presenza di un'adeguata architettura infrastruttura cloud. L'obiettivo principale è mostrare come i dati possano essere convertiti in valore e migliorare la gestione aziendale, trasformandosi in un vantaggio competitivo in ambiti quali manifatturiero, retail, sanità e finanza.

Un approfondimento è dedicato al settore dell'automotive. Si fornisce una descrizione dettagliata di una corrente architettura infrastrutturale di machine data sviluppata dall'ex Group PSA, attuale Stellantis N.V. Si parla di veicoli connessi, dotati di sensori avanzati e sistemi di connettività, conosciuti come Internet of Things, che raccolgono dati su vari aspetti del veicolo, come le condizioni del motore, lo stile di guida e l'ambiente circostante. Questo flusso costante di dati consente alle case automobilistiche di analizzare le prestazioni in tempo reale, ottimizzando la gestione dei veicoli e fornendo nuove funzionalità quale la descritta metodologia data driven per la valutazione della progettazione HMI. L'elaborato esplora studi scientifici su applicazione di machine data su tecnologie ecocompatibili, come i veicoli a idrogeno HFCV e il relativo quadro per la manutenzione circolare abilitata dall'IoT.

Stellantis è il caso di studio centrale, con i progetti "Mobilisight" e "SUSTAINera".



# CAPITOLO 1

## 1. DATA SCIENCE

### 1.1. EVOLUZIONE DEI DATI E DELLE TECNICHE DI ANALISI

Nel corso degli anni, la gestione dei dati e i processi di analisi per trasformare i dati in informazioni hanno registrato un progresso esponenziale.

Le prime tecnologie disponibili permettevano di registrare su supporti magnetici informazioni relative alle attività aziendali e di condurre solo analisi statistiche e limitate alla semplice estrazione dei dati memorizzati.

Negli anni Ottanta, con l'avvento dei database relazionali e del relativo linguaggio Structured Query Language (SQL), si inaugura l'era di un'analisi dati estremamente dinamica e sofisticata.

SQL ha permesso di estrarre dati in modo più efficiente, consentendo sia analisi aggregate che dettagliate. Tuttavia, le analisi furono condotte su database operativi, noti come sistemi OLTP (Online Transaction Processing), ottimizzati principalmente per le transazioni (aggiunta, eliminazione e modifica dei dati) piuttosto che per l'analisi di grandi volumi di dati. Inoltre, la storicizzazione dei dati era limitata o assente in molti sistemi OLTP, rendendo difficile ricostruire il contesto storico dei dati. A ciò si aggiungeva la presenza di numerose applicazioni con dati replicati e manipolati da ciascuna che portava ad una mancanza di uniformità e coerenza dei dati.

Per superare questa limitazione, negli anni '90 sono stati sviluppati i Data Warehouse, che integrano e conservano dati affidabili su processi aziendali. Questi database contengono dati integrati e affidabili relativi ai processi aziendali e sono alla base per le attività analitiche della Business Intelligence (BI). La BI comprende modelli, metodi, processi, persone e strumenti per raccogliere e organizzare i dati aziendali, trasformandoli in informazioni accessibili, conservabili e presentabili in modo efficace per supportare le decisioni aziendali. Nonostante ciò, la BI si concentra principalmente su una visione storica dei dati, consentendo solo una valutazione retrospettiva degli eventi passati o in corso.

Nel corso degli anni 2000, è emersa la necessità di analisi previsionali per anticipare eventi e ottenere vantaggi commerciali, portando allo sviluppo delle tecniche di Data Mining per estrarre informazioni, pattern e relazioni dai dati.



Di seguito le tecnologie sviluppate in tale ambito.

- **Business Analytics:** utilizzo di modelli matematici e statistici per analisi e il data mining, spesso con funzionalità di analisi visuale dei dati
- **il Mobile BI e Reporting:** sviluppo di applicazioni per la visualizzazione dei report sui dispositivi mobili come smartphone e tablet
- **Self-Service BI:** software intuitivi e di semplice utilizzo che permettono agli utenti finali di creare report, analisi e modelli dati.
- **Cloud computing:** offerta di risorse hardware e software come servizi su Internet per analisi dei dati e la BI, riducendo la dipendenza dalle risorse locali
- **Big Data:** tecnologie per analisi avanzata di grandi volumi di dati eterogenei e/o non strutturati,

Infine, l'era del Big Data o Data Driver ha portato a una nuova frontiera nell'analisi dei dati, consentendo di estrarre ed elaborare informazioni da enormi volumi di dati eterogenei e non strutturati. Questo ha aperto nuove opportunità per le aziende nell'identificare modelli, trend, e relazioni altrimenti nascosti nei dati: analisi più accessibile e flessibile, consentendo agli utilizzatori finali di poter visualizzare report e analisi in modo autonomo. In sintesi, l'evoluzione continua nell'ambito dell'analisi dei dati riflette la crescente importanza e complessità del ruolo dei dati nell'ambito decisionale e strategico delle aziende moderne.

## 1.2. BIG DATA DRIVEN

### 1.2.1. LE TRE (CINQUE) V DEI BIG DATA

La nozione di Big Data fa riferimento a informazioni caratterizzate da una vasta varietà, un incremento nei volumi e una maggiore velocità di flusso. Comunemente ci si riferisce a queste caratteristiche usando il concetto delle tre V. In termini più semplici, i big data rappresentano insiemi di informazioni di dimensioni notevoli e complessi. Questi insiemi sono così massicci che le tradizionali soluzioni software per l'elaborazione dati non riescono a gestirli adeguatamente.

In dettaglio le tre "V":

- Volume → La quantità dei dati è cruciale. Nel contesto dei Big Data, ci si trova a dover gestire grandi quantità di dati non strutturati a bassa densità. Questi possono comprendere dati di valore sconosciuto, quali flussi di dati da Twitter, clickstream su siti web o app mobili, oppure informazioni da dispositivi dotati di sensori. Basti pensare che ci sono 4,6 miliardi di smartphone attivi e circa 2 miliardi di persone hanno accesso a Internet. Dalla fine del ventesimo secolo ad oggi, il volume dei dati in circolazione è aumentato in maniera esponenziale (rif. Figura 2):

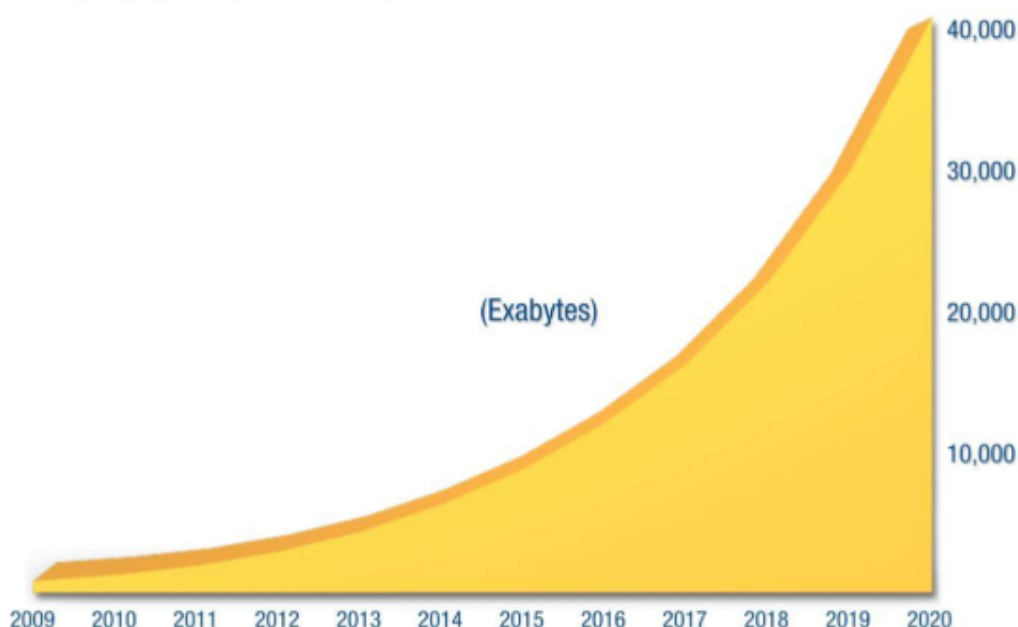


Figura 1 – Fonte [https://it.wikipedia.org/wiki/Big\\_data#/media/File:BigDataGrowth.png](https://it.wikipedia.org/wiki/Big_data#/media/File:BigDataGrowth.png)

- **Velocità** → Rapidità con cui i dati vengono ricevuti e, eventualmente, elaborati. Normalmente, la velocità più elevata dei dati fluisce direttamente nella memoria invece di essere scritta sul disco. Alcuni prodotti intelligenti abilitati per Internet funzionano in tempo reale e richiedono valutazioni e azioni in tempo reale.
- **Varietà** → Diversità dei tipi di dati disponibili. In passato, i tradizionali dati erano strutturati e ad hoc per un database relazionale. Con l'avvento dei Big Data, emergono nuovi tipi di dati non strutturati: testi, audio e video che richiedono un'elaborazione preliminare aggiuntiva per estrarre significato e supportare i metadati.

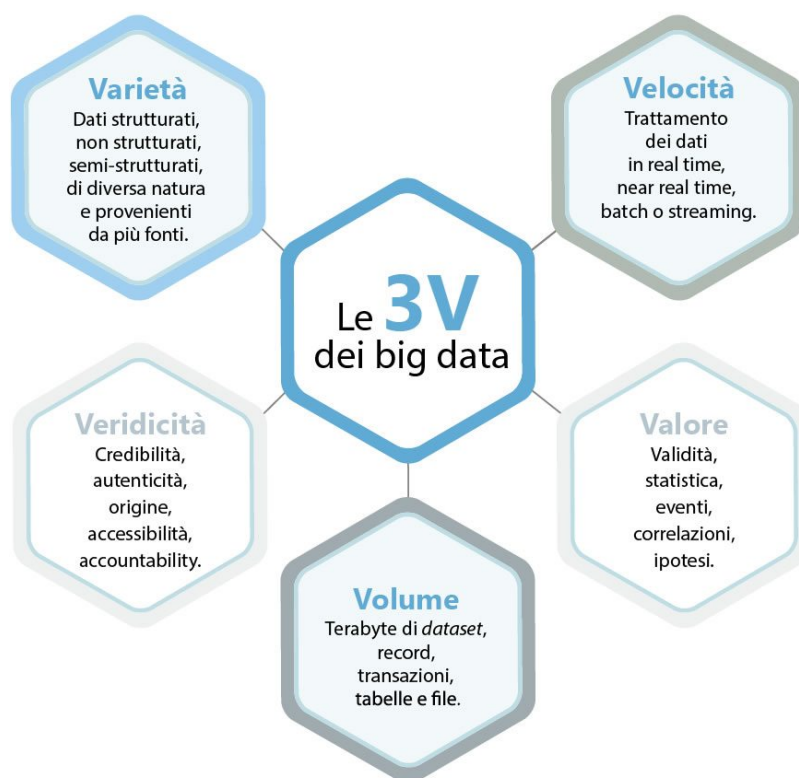


Figura 2 – Fonte <https://forward.recentiprogressi.it/wp-content/uploads/2016/11/le3v-dei-big-data.jpg>

Negli ultimi anni sono emersi due ulteriori caratteristiche fondamentali: Valore e Veridicità. I dati portano con sé un valore intrinseco, ma rimangono inutili fino a quando questo valore non viene riconosciuto. Altrettanto significativa è la questione della verità dei dati: quanto sono affidabili e attendibili?

I recenti avanzamenti tecnologici hanno notevolmente abbassato i costi relativi all'archiviazione e all'elaborazione dei dati, rendendo più conveniente e accessibile l'archiviazione di grandi quantità di informazioni rispetto al passato. Con un volume maggiore di Big Data disponibile a costi più bassi e accessibili, diventa possibile prendere decisioni aziendali accurate e precise. Non a caso, oggi, i big data sono diventati una forma di capitale preziosa. Si pensa alle principali aziende tecnologiche globali: gran parte del loro valore deriva dall'analisi costante dei dati, che viene sfruttata per migliorare l'efficienza operativa e sviluppare nuovi prodotti.

Ma sfruttare il valore dei Big Data non si limita alla loro semplice analisi, si tratta piuttosto di un processo completo di scoperta che coinvolge analisti, utenti finali e dirigenti, i quali devono porre le domande appropriate, riconoscere i pattern, formulare ipotesi e anticipare comportamenti.

È evidente come il concetto di Big Data porta con sé diversi elementi e componenti che bisogna considerare per poter realizzare il processo ciclico di raccolta, gestione ed analisi dati che vedremo in dettaglio nel prossimo paragrafo.

In particolare, si possono delineare i seguenti step da attenzionare che sono essenziali alla riuscita di qualsiasi progetto di Big Data:

- 1.** Un reale caso d'uso aziendale in cui i Big Data possono generare valore aggiunto
- 2.** Organizzazione e Gestione del ciclo di vita di Big Data
- 3.** Infrastruttura IT per i Big Data
- 4.** Strumenti Analitici per i Big Data

### *1.2.2. CASI D'USO: TRASFORMARE I BIG DATA IN UN VANTAGGIO COMPETITIVO*

Si cominciò a prendere coscienza della quantità di dati generati dagli utenti, attraverso piattaforme come Social Media, Youtube, e altri servizi online gratuitamente accessibili. Non a caso, la stessa consapevolezza ha condotto allo sviluppo di Hadoop, un framework open source progettato specificatamente per l'archiviazione e l'analisi dei Big Data. Nello stesso periodo NoSQL ha iniziato a guadagnare popolarità.

L'avvento di framework open source, come Hadoop e più recentemente Spark, ha permesso l'ascesa dell'era Big Data Driver, poiché li rendono più gestibili e meno costosi da archiviare. Infatti, negli anni successivi, il volume dei Big Data ha raggiunto livelli stratosferici: non sono solo gli esseri umani a generare esponenziali quantità di dati ma con la diffusione dell'Internet delle cose "IOT", sempre più oggetti e dispositivi sono connessi a Internet, permettendo così di raccogliere istantaneamente e con continuità dati su comportamenti dei consumatori e sulle performance dei prodotti.

Inoltre, se da una parte questa enorme quantità di dati si delinea come input di algoritmi di Machine learning, dall'altro, lo stesso Machine Learning ha contribuito a generare una quantità ancora maggiore di dati.

Il Cloud Computing ha ulteriormente ampliato la possibilità dei Big Data, offrendo una scalabilità flessibile, dove gli sviluppatori possono facilmente creare cluster dedicati per testare specifiche porzioni di dati. In tale ottica, i database grafici stanno diventando sempre più rilevanti, grazie alla loro capacità di visualizzare grandi quantità di dati in modo rapido ed esaustivo facilitando le analisi.

Nei prossimi capitoli, vedremo in dettaglio il ciclo di vita per l'estrazione di conoscenza dai Big Data e l'impiego della stessa per gestire e trasformare in modo efficiente dataset che crescono rapidamente in volume e varietà: ogni fase di ciascun processo modifica lo stato ed il loro contenuto, contribuendo a convertire moli di dati ancora grezzi in valore e dunque arricchendo il modello stesso. Questi processi costituiscono un riferimento per la creazione e scelta di un'architettura, nonché di una struttura logica e fisica che definisce come i Big Data devono essere memorizzati, acceduti e gestiti all'interno di un flusso ciclico.

In particolare, è importante capire come i Big Data possano essere lo strumento per affrontare una serie di attività aziendali, dallo sviluppo prodotto agli analytics.

Di seguito elencate le attività maggiormente sviluppate oggi nei settori Manufacturing, Retail, Healthcare, Financial Services, e Oil&Gas.

## MANUFACTURING

### Manutenzione predittiva



Figura 3 – Fonte [file:///C:/Users/Alessia/Downloads/PIIT\\_Landing\\_page\\_C2V\\_approfondimento\\_V1.4%20\(1\).pdf](file:///C:/Users/Alessia/Downloads/PIIT_Landing_page_C2V_approfondimento_V1.4%20(1).pdf)

L'obiettivo è attuare un piano di strategia Predittiva alla Manutenzione, prendendo decisioni sulla base dei dati rilevati in tempo reale. In tal modo è possibile ridurre i guasti del 75%, gli interventi manutentivi del 25%, i tempi di fermo produzione, le inefficienze su tutta la catena produttiva e quindi, di conseguenza, permettere un incremento complessivo di produzione fino al 25%.

In ambito di analisi statistiche, un dato significativo è che solo il 18% dei guasti è imputabile all'età degli asset, mentre in otto casi su dieci il loro deterioramento precoce è dovuto a condizioni di lavoro non idonee per il macchinario.

Secondo un rapporto Mckinsey, attuare un piano di manutenzione predittiva incrementa il ciclo di vita delle attrezzature fino al 40%, per cui, la linea di produzione ha un ritorno finanziario nettamente maggiore.

In particolare, la manutenzione predittiva si riferisce all'utilizzo di dati generati da una macchina, sia dati strutturati (anno, marca, modello di attrezzatura) che dati multi-strutturati (registrazione di log, dati dei sensori, messaggi di errore, temperatura del motore etc.) al fine di prevedere la probabilità di guasto prima che questa si verifichi per evitare i tempi di fermo macchina non pianificati. Quest'ultimi, infatti, possono avere

impatti economici significativi in termini di produzione e costi per l'End User, basti pensare che ci sono aziende il cui costo di un'ora di fermo macchina si aggira intorno ai 250.000€. Per molti processi manifatturieri, è anche importante prevedere la vita residua dei sistemi e dei componenti per garantire l'operatività secondo specifiche: uscire dal range di tolleranza può essere tanto grave quanto un fermo linea.

Quindi, un primo passo è che la macchina sia dotata di sensori e strumenti che acquisiscono le grandezze critiche della macchina ; in seguito i dati vengono elaborati e interpretati da dispositivi che effettuano diagnostica ( picchi di corrente, copie, temperature, pressione...) mediante modelli di analisi statistica o evoluti algoritmi di Machine Learning per determinare lo stato di salute dei componenti e di lavoro della macchina (Condition Monitoring, analisi del motore e della meccanica collegata). Per cui, ad esempio, un sistema diagnostico basato su reti neurali, riconoscerà un certo valore di corrente anomalo in relazione a quello che stanno facendo tutte le altre variabili in un dato istante della produzione. Lo stesso valore di corrente, in altri momenti e condizioni, potrebbe essere normale.

La raccolta di informazioni e dati avviene con l'introduzione di soluzioni specifiche di connettività avanzata, come l'IIoT Gateway FP-I4C che risulta essere di facile integrazione, interconnessione ed integrabilità con altri macchinari e sistemi IT/aziendali (gestionali); Si tratta di soluzioni smart poiché dotate di dispositivi in grado di produrre, raccogliere, gestire e fornire dati utili alla digitalizzazione della fabbrica e alla sua ottimizzazione tramite l'IoT (Industrial Internet of Things). Oltre ad offrire un accesso in teleassistenza (VPN/Cloud) al macchinario, offre la possibilità "strategica", sia per il costruttore di macchina e sia per l'utilizzatore finale, di portare i dati dal campo ai sistemi IT, realizzando l'interconnessione tra i mondi OT/IT utilizzando le tecnologie OPC UA, MQTT, SQL.

### Sfida

Le aziende devono integrare dati provenienti da formati diversi e identificare i segnali che porteranno all'ottimizzazione della manutenzione.

### *Efficienza Operativa e Ottimizzazione Produzione*

L'ottimizzazione delle linee di produzione può ridurre i costi e aumentare i ricavi: l'analisi dei big data permette di individuare le aree da cui si può trarre beneficio, in dettaglio quali sono gli step della catena produttiva che portano ad un aumento del tempo di produzione e quali causano ritardi (es. colli di bottiglia)

L'analisi e valutazione dei processi aziendali mediante tali dati garantisce un'efficienza operativa, in particolare permette di rispondere in modo proattivo ai feedback dei clienti, anticipandone le future richieste.

### Sfida

Saper gestire il volume dei dati con il crescente numero di fonti, utenti e applicazioni.

Nel campo dell'ottimizzazione di produzione si richiede ai produttori di analizzare i dati delle attrezzature, l'uso di materiali e altri fattori specifici, combinando i diversi tipi di dati

## *RETAIL*

### Sviluppo Prodotto

L'obiettivo è anticipare la domanda dei clienti.

L'analisi dei big data si conduce classificando attributi chiave dei prodotti passati ed attuali e modellando la relazione tra questi attributi e il successo commerciale delle offerte, al fine di cercare modelli predittivi alla base di una pianificazione/ produzione / lancio di un nuovo sviluppo prodotto / servizio. Un buon punto di partenza è accedere a dati provenienti da workflow, social media, primi lanci aziendali etc.

Sfida:

Analizzare un elevato volume di dati provenienti da formati variabili al fine di creare segmenti in base al comportamento del cliente, in particolare identificare modelli / comportamenti d'uso che guidano potenziali nuove offerte in grado di anticipare e quindi soddisfare la domanda dei clienti stessi.

### Esperienza e valore del cliente nel tempo

Si parla di customer center, nonché di una vera e propria corsa al cliente.

Raccogliendo dati dai social media, dalle visite web, registri delle chiamate e da altre fonti di dati, le aziende possono migliorare le interazioni con i clienti con l'obiettivo di massimizzare il valore offerto: l'analisi dei big data permette di fornire offerte personalizzate, ridurre Churn Rate (o Customer Churn) e gestire proattivamente i problemi. I Big Data forniscono informazioni sui modelli di spesa, in modo da poter identificare una data categoria di clienti ad alto valore a cui il marketing può mirare con offerte speciali, o a cui il servizio clienti può dedicarsi maggiormente aumentandone la fidelity.

### Sfida

È necessario analizzare un elevato volume di dati sulle transazioni dei clienti e creare modelli AI sofisticati che esaminano il comportamento passato e siano in grado di prevedere azioni future.

### Ottimizzazione prezzi

Mediante analisi di dati provenienti dalle app mobili, dagli acquisti in negozio e dalle geolocalizzazioni è possibile conoscere la vera redditività dei clienti, nonché come possono essere segmentati i mercati e il potenziale di opportunità future. L'analisi del profitto e del



margin end to end può identificare aree in cui i profitti potrebbero essere in perdita ottenendo un'ottimizzazione degli stessi prezzi.

### Sfida

Necessità di dati che devono essere correlati e uniti a più set di dati sui prezzi, e capacità di riuscire a gestire e analizzare milioni di transazioni con molti tipi di set dati.

## HEALTHCARE

### Ricerca Genomica

Utilizzando i big data, i ricercatori possono identificare geni e biomarcatori di malattie per aiutare i pazienti a individuare problemi di salute ma soprattutto, prevenire lo sviluppo di malattie in futuro. I risultati possono persino permettere di delineare cure sanitarie personalizzate.

### Sfida

Il volume di dati genomici è enorme, per cui si hanno continue esecuzioni di complessi algoritmi sui dati che richiedono lunghi tempi di elaborazione

### Esperienza dei pazienti

L'obiettivo per gli enti sanitari è fornire un miglior trattamento e qualità di assistenza senza aumentarne i costi. In tale direzione i big data aiutano a migliorare l'esperienza del paziente nel modo più efficiente in termini di costo: l'analisi permette di creare una visione a 360 gradi dell'assistenza fornita mentre il paziente è in cura.

### Sfide

Integrare diverse fonti di dati, alcune dei quali potrebbero essere dati multi-strutturati, altri invece dati da estrapolare nell'analisi dei percorsi sanitari dei pazienti stessi. Il fine è, soprattutto in questo campo, analizzare i dati per evidenziare tendenze ed eventuali anomalie nei modelli di default e riuscire così a creare dei nuovi modelli predittiva.

## OLI&GAS

### Manutenzione preventiva dell'attrezzatura

Le compagnie petrolifere e del gas hanno difficoltà ad avere una completa visibilità sulle condizioni delle loro attrezzature, soprattutto in siti remoti offshore.

I big data permettono di fornire informazioni in modo che le aziende possano prevedere la vita utile residua dei loro componenti, assicurandosi così che i loro asset operino con efficienza produttiva ottimale e in condizioni di totale sicurezza.

### Sfida

I dati delle macchine, sensori e registri vari provenienti dalle attrezzature arrivano in formati variabili, per cui, risulta complesso integrare la variabilità di tutti questi dati rapidamente in un tempo utile che permetta di metterli in funzione per prevenire efficacemente i tempi di inattività.

## *FINANCIAL*

### Frodi

Nel campo dei servizi finanziari, si pone al centro al centro la sicurezza e la conformità dei servizi offerti. Nel dettaglio ci si riferisce alla responsabilità che la società di servizi finanziari ha di conservare, gestire e proteggere il denaro e le informazioni finanziarie dei clienti. Implica l'adesione alle normative federali, statali e locali che determinano i livelli di standard di sicurezza relativa ai dati dei clienti.

Utilizzando i dati, si possono individuare pattern tipici di potenziali frodi.

### Sfida

Si richiede un'analisi incrociata di diversi insiemi di dati di transazione con informazioni supplementari (es. eventi di interazione e comportamento dei clienti)

### Antiriciclaggio

Le leggi Antiriciclaggio impongono alle società dei servizi finanziari di svolgere attività per ottenere la conformità ai requisiti di legge e dimostrare la corretta diligenza segnalando attività sospette di riciclaggio di denaro. In questo settore abbastanza complicato, l'analisi dei big data può aiutare le aziende a monitorare attivamente e poter quindi identificare possibili modelli di operazioni di riciclaggio. Caso analogo al pattern di frodi.

### Sfida

Analisi e identificazione di transazioni AML complesse.

### 1.3. IL CICLO DEI BIG DATA

Abbiamo visto come, dato un reale caso di business, sia possibile sfruttare appieno il potenziale dei dati e quindi ottenerne vantaggio competitivo. Alla base del funzionamento di questo meccanismo analitico che permette di apportare valore, vi è un ciclo iterativo che comprende una sequenza di specifiche fasi, ognuna delle quali contribuisce alla gestione e all'analisi dei dati su larga scala, che si ripetono in loop con un flusso di dati continuo e in costante evoluzione.



Figura 4 – Fonte Kumar Rahul, 2020. *Data Life Cycle Management in Big Data Analytics. Procedia Computer Science 173 (2020) 364–371*

In generale, i Big Data vengono utilizzati per raccogliere, archiviare elaborare e trasformare l'enorme quantità di dati in informazioni significative attraverso strumenti e tecniche strategiche. L'acquisizione di conoscenza dai Big Data e il suo impiego per il miglioramento delle attività decisionali richiedono la definizione di processi efficienti per gestire e trasformare rapidamente dataset di grandi dimensioni e varietà. Ogni fase di tali processi modifica il contenuto dei dati, contribuendo a trasformare dati grezzi in valore, arricchendo il modello analitico dei dati. Questi processi guidano la scelta o la creazione di un'architettura logica e fisica per la memorizzazione, l'accesso e la gestione dei Big Data all'interno di un'organizzazione. Inoltre, specificano soluzioni per affrontare problemi che possono derivare dal trattamento dei Big Data come la qualità dei dati in termini di grado di coerenza dei dati e grado di completezza, la scalabilità delle infrastrutture, le componenti hardware e software come framework e database, il flusso di informazioni, privacy e sicurezza dei dati.

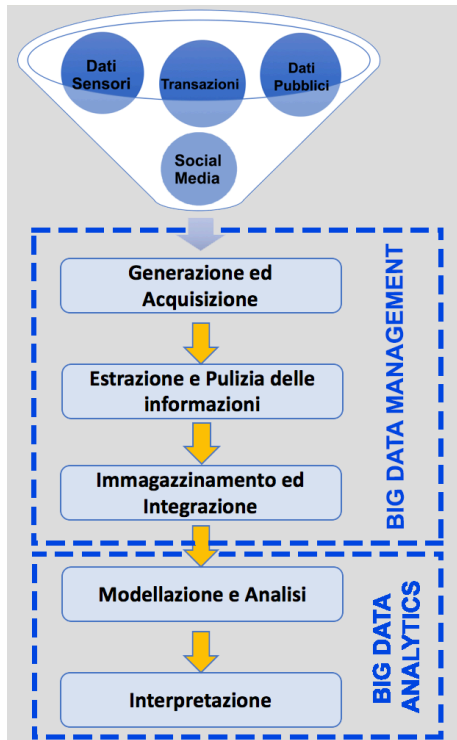


Figura 5-Fonte: [https://it.wikipedia.org/wiki/Big\\_data](https://it.wikipedia.org/wiki/Big_data)

Il Ciclo di vita dei Big data si suddivide in due macroaree:

- ❖ Big Data Management: tecnologie e processi per l'acquisizione e la memorizzazione dei big Data, nonché la preparazione e il recupero di questi.
- ❖ Big Data Analytics: processi utilizzati per analizzare e quindi acquisire conoscenza da grandi dataset al fine di condurre un'analisi di condizioni verificate (descriptive analytics) , di condizioni probabili che si possono verificare in futuro ( predictive analytics) o delineare delle deduzioni su cui basare il processo decisionale ( prescriptive analytics.).

### 1.3.1. BIG DATA MANAGEMENT

La generazione dei big data viene a sua volta catalogata in dati generati dall'essere umano, dati generati dalle macchine e dati derivanti da processi di business.

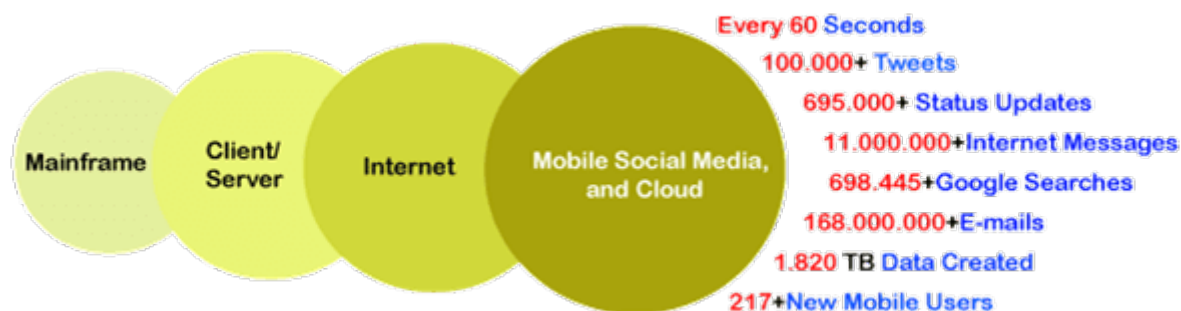


Figura 6 – Fonte: <https://www.javatpoint.com/big-data-characteristics>

I data human generated sono frutto delle attività degli utenti, ad esempio, le informazioni che carichiamo sui social network come Facebook e LinkedIn, sui multimedia sharing quali Instagram e YouTube, o in dettaglio, le recensioni e i feedback pubblicati sui portali di e-commerce, siti di notizie o aggregatori.

I dati machine generated, invece, sono quelli prodotti da sorgenti quali sensori GPS, dispositivi Internet Of Things, centrali di monitoraggio di eventi meteorologici, sistemi di High Frequency Trading di mercati finanziari, o dati elaborati dai macchinari medici e da altri strumenti scientifici.

In generale, i dati human o machine generated, possono appartenere alla categoria di dati business generated nel caso in cui sono stati acquisiti internamente ad un'azienda al fine di registrare tutte le attività data-driven dei processi di business aziendali. Si tratta di dati storici, memorizzati come dati relazionali che rappresentano pagamenti, ordini, dati di produzione e d'inventario, finanziari e di vendite. La generazione di questa tipologia di dato ha reso necessario la presenza di tecnologie e strumenti di analisi in tempo reale che permettano alle aziende di creare valore aggiunto veicolando i processi e le decisioni del proprio business aziendale.

Analogamente al processo di generazione, anche quello di acquisizione, può essere categorizzato in base a modalità specifiche. Il modo più semplice per acquisire dati è quello di accedere ad API messe a disposizione da servizi Web come ad esempio le Twitter API, e le API fornite da motori di ricerca Google e Bing.

È possibile acquisire dati anche attraverso software specifici di web scraping che effettuano raccolte automatiche di informazioni da vari documenti online.

Altresì, è possibile recuperare dati mediante strumenti ETL che eseguono i processi di Extract (Estrazione), Transform (Trasformazione), e Load (Caricamento) sia ai database relazionali che a quelli non relazionali. Ad esempio, uno degli strumenti di ETL più usati nel campo dei Big Data è Apache Sqoop che consente di importare ed esportare enormi quantità di dati da database relazionali e no, verso la piattaforma Apache Hadoop, e viceversa.

Infine, è possibile utilizzare tecnologie che permettono l'acquisizione continua di flussi di dati rapidamente generati; si tratta di sistemi capaci di catturare ed elaborare eventi e condizioni, e di salvarli quasi in tempo reale su database in modo efficiente. Esempi di tecnologie più diffuse: Apache Flume, Microsoft StreamInsight.

Dopo l'acquisizione, i dataset vengono trasferiti a sistemi di archiviazione, locali o remoti, tramite mezzi di trasmissione ad alta velocità. Questo insieme di dati grezzi, strutturato e no, si chiama Data Lake. Quest'ultimo, effettuate le fasi descritte precedentemente, risulta pronto per l'elaborazione. È fondamentale, per evitare l'insorgere di problemi legati alla qualità del dato acquisito, eseguire ulteriori operazioni di precompilazione il cui obiettivo è filtrare i dati da informazioni incomplete e ridonanti al fine di diminuirne l'entropia. A seguire l'immagazzinamento e l'integrazione delle informazioni.

In particolare, la fase di big data management si conclude con lo scrubbing dei dati: processo di estrazione e pulizia dataset che include la modifica di tutti i valori di data in un formato standard comune, la correzione di errori di dati, e la rimozione di eventuali valori anomali.

Infatti, il primo problema da considerare è che i dataset, dopo essere stati raccolti, non sono pronti per essere elaborati; si tratta di dati con rappresentazioni diverse e non standard (es. prescrizioni, letture da sensori, dati da macchine di monitoraggio, dati di immagini etc....) Quindi, è necessario creare un processo di estrazione che prelevi le informazioni richieste e le rappresenti in una forma standard e strutturata finalizzata all'analisi. La progettazione del processo di estrazione dipende dal campo di applicazione di provenienza del dato e dalla sua rappresentazione. Inoltre, in tale parte del ciclo, è importante considerare che i Big Data possono contenere informazioni false; quest'ultime sono contenute nei dati che chiamiamo "Stand Alone" e identificati come valori anomali all'interno del grande set di dati. È indispensabile, di conseguenza, utilizzare tecniche di pulizia dei dati che includono vincoli per controllarne la validità e modelli di errore per garantire la qualità del dato.

La successiva fase di immagazzinamento dei big data ha il principale obiettivo di garantirne la disponibilità nel tempo. Tale fase presenta una doppia sfida: da un lato memorizzare e organizzare enormi dataset non strutturati o semi-strutturati, e dall'altro garantire la disponibilità e l'interattività affidabile, veloce e sicura con essi. Per soddisfare tali esigenze, è stata sviluppata un'infrastruttura, che combina hardware per l'archiviazione con meccanismi di gestione dei dati,

inclusi file system distribuiti e nuove tipologie di database. L'infrastruttura per i Big data consiste in un insieme di nodi di archiviazione ed elaborazione distribuiti in rete, collegati tramite reti ad alta velocità e in grado di scalare orizzontalmente e adattarsi dinamicamente alle esigenze delle applicazioni. I file system distribuiti forniscono una vista unificata dell'infrastruttura di memoria distribuita, consentendo operazioni di lettura e scrittura sequenziali su grandi quantità di dati con alte prestazioni e tolleranza ai guasti. Due dei più noti file system distribuiti sono Google File system (GFS) ed Hadoop Distributed File System (HDFS).

L'immagazzinamento dei dati viene effettuato ricorrendo a linguaggi specifici come il NoSQL. Diversi tipi di database NoSQL sono stati introdotti per gestire dataset caratterizzati da varietà e volume, superando le limitazioni dei tradizionali sistemi RDBMS. Questi database NoSQL abbandonano il modello relazionale SQL, offrendo caratteristiche fondamentali per i Big Data, tra cui la mancanza di uno schema definito e rigido, distribuzione, facile replicazione dei dati, e supporto per l'archiviazione persistente di grandi volumi di dati.

### 1.3.2. BIG DATA ANALYTICS

La fase di big data analytics segue quella di big data management e trasforma definitivamente il dato in informazione di valore. Prevede diversi processi di interrogazione del dato: dall'analisi descrittiva alle predizioni o prescrizioni.

Il Big data analytics vede susseguirsi la modellazione, l'elaborazione e l'analisi del dato col fine di iniziare a trovare informazioni utili e di valore all'interno della complessità dei dataset. La modalità di analisi dipende dal tipo di dato che si considera, ovvero strutturato, semi-strutturato, o non strutturato; e dalla varietà di file presenti all'interno del set.

Tra le tipologie di analisi troviamo:

- Analisi di testi: estrazione di informazioni a partire da testo non strutturato contenuto in documenti, pagine Web, social Network, email ed altri ; è nota come text mining e fa uso di tecniche di machine learning ed analisi statistica, inoltre , grazie a queste metodologie sono stati sviluppati algoritmi per il riconoscimento degli argomenti (*topic modelling*) , la ricerca delle migliori risposte a una domanda (*question answering*), l'individuazione delle opinioni degli utenti su specifiche notizie (*opinion mining*) e altro ancora.
- Analisi di dati multimediali: i contenuti multimediali, come immagini, video e audio, presentano una natura non strutturata e una vasta dimensione, rendendoli adatti all'analisi tramite sistemi di Big Data Analytics; gli algoritmi di machine learning consentono di estrarre informazioni di diversi livelli utili alla descrizione semantica dei file multimediali, tra questi troviamo l'annotazione automatica tramite etichette testuali (*multimedia annotation*) e l'estrazione di caratteristiche visive o sonore (*feature extraction*).
- Analisi del Web: estrazioni di informazioni sulla struttura, contenuti e utilizzo del Web analizzando pagine e collegamenti ipertestuali in modo automatico; questa analisi coinvolge sia testi che contenuti multimediali , facendo uso delle tecniche menzionate in precedenza, in particolare, la topologia del Web può essere ricostruita attraverso algoritmi di crawling che seguono i collegamenti ipertestuali per rivelare relazioni tra pagine o siti Web.

Le tecniche di data mining, machine learning e di analisi statistica come clustering, regressione e correlazione vengono applicati ai Big Data mediante l'uso di modelli di programmazione e framework di elaborazione distribuita. Ciò consente di ottenere rapidamente informazioni aggregate dai database NoSQL o da altre fonti, simile all'utilizzo di SQL nei database relazionali.

La fase di big data analytics si conclude con l'interpretazione dei parametri analizzati. Ciò risulta fondamentale per vari scopi aziendali quali verificare ipotesi empiriche su fenomeni di interesse, prendere decisioni di business più informate, individuare nuovi mercati da esplorare, sviluppare campagne di marketing mirate e migliorare l'efficienza operativa.



A seguito dell'analisi di questo ciclo di gestione, risulta evidente come il Valore dei Big Data risieda nel processo stesso di archiviazione, elaborazione ed analisi, ed è proprio in questo che individuiamo la catena del valore.

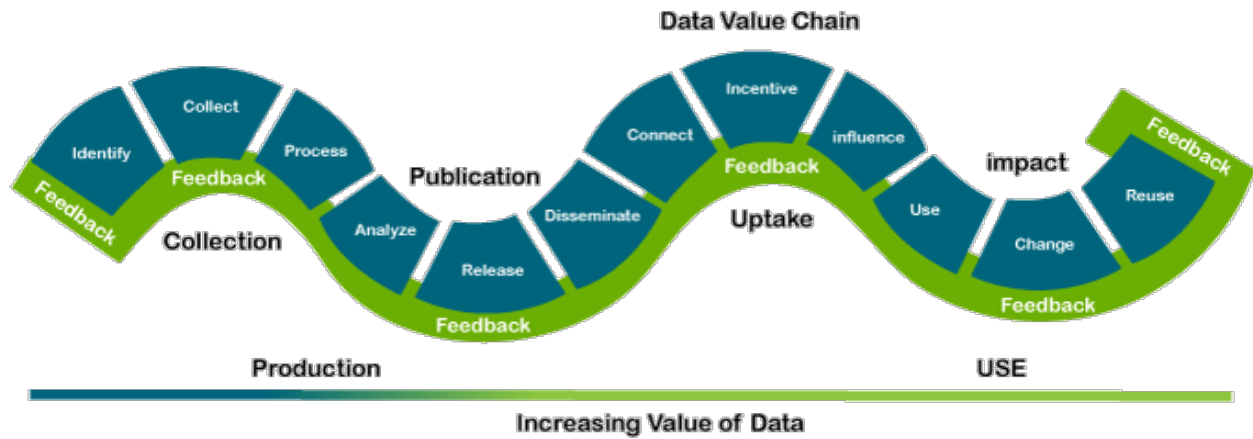


Figura 7 – Fonte <https://www.javatpoint.com/big-data-characteristics>

Allineati i Big Data ad obiettivi aziendali specifici, ovvero definiti in che modo i Big Data supportano e abilitano le principali priorità aziendali e IT, e in seguito, delineata l'organizzazione e la gestione del ciclo di vita dei Big Data, è fondamentale porre l'attenzione sull'Infrastruttura IT, sulla struttura di archiviazione dati e sugli strumenti analitici per i Big Data.

Come si potrà evincere dai prossimi paragrafi, è importante basare nuovi investimenti in competenze, organizzazione o infrastruttura con un forte contesto orientato al business.

## 1.4. *INFRASTRUTTURA IT PER I BIG DATA*

Per garantire il successo di un progetto Big data, le aziende devono investire in un'infrastruttura adeguata e altamente specializzata, in grado di gestire efficacemente la raccolta, l'archiviazione, e l'elaborazione dei dati per renderli utili, assicurandosi al contempo la sicurezza delle informazioni durante l'archiviazione e la trasmissione.

A un livello più ampio, ciò comprende sistemi di archiviazione e server ottimizzati per i Big Data, framework software, database, strumenti di analisi e integrazioni con altre applicazioni.

Spesso questa struttura è implementata tramite macchine fisiche ospitate in un data center remoto. Cloud e virtualizzazione, pur essendo soluzioni pratiche ed efficienti nell'ambito IT, non sempre si adattano al meglio alle esigenze dei Big Data, soprattutto per quando riguarda l'elaborazione dei dati. È evidente che se l'elaborazione e la memoria sono distanti dallo storage, la connessione di rete ne risente, infatti, nei contesti cloud e virtualizzati, il volume dei dati e la velocità di elaborazione possono creare congestioni nella rete.

Per questo motivo spesso si preferisce un'architettura basata su cluster di server fisici, anche economici, ma dotati di elevate capacità di memoria RAM, una o più GPU (schede grafiche) e dischi rigidi veloci, tutti integrati nella stessa scheda madre. A ciò si aggiungono strumenti software progettati per distribuire il carico di lavoro sui singoli server del cluster.

Naturalmente, vi sono eccezioni a questa regola generale: le elaborazioni batch, ad esempio, che non richiedono risposte in tempo reale e possono essere eseguite solo occasionalmente (es. generazione di report finanziari mensili), possono beneficiare dell'utilizzo di servizi cloud attivati solo quando necessario, al fine di ridurre i costi.

Anche la raccolta dei dati può essere complessa, infatti, mentre alcuni dati sono statici e costantemente disponibili, come quelli provenienti da file, log e social media, altri devono essere acquisiti rapidamente e senza ritardi, presentando sfide in termini di prestazioni dello storage e della connettività. I dati dinamici, come i segnali dei sensori e le transizioni finanziarie, devono essere acquisiti in modalità streaming, generando ulteriori complessità.

L'espansione dell'Internet Of Things ha portato a soluzioni di Big Data specificamente progettate per questo obiettivo: sempre più aziende integrano sensori e connettività nei loro prodotti, alimentando così la necessità di gestire grandi quantità di dati.

Tra le opzioni di archiviazione più comuni per i Big Data vi sono i seguenti:

- I data warehouse tradizionali, utilizzati dalle applicazioni aziendali come ERP e CRM, sono una fonte di dati per le applicazioni Big Data.

- I data lake sono contenitori di grandi volumi di dati nel loro formato originale, estratti e utilizzati solo quando necessario per le applicazioni aziendali, e vengono alimentati dalla crescente raccolta di dati mediante utilizzo dei dispositivi IOT.
- I cloud storage per i Big Data.

#### 1.4.1. CLOUD COMPUTING

Il termine “cloud computing” è utilizzato per descrivere la fornitura on demand di calcolo come hardware, archiviazione, database, networking e software, a imprese e individui attraverso la rete Internet. Questo modello consente alle organizzazioni di accedere alle informazioni e di archivarle senza la necessità di gestire dispositivi fisici o infrastrutture IT.

Infatti, con l’aumento continuo della quantità di dati generati e condivisi, e della domanda da parte dei consumatori di maggiori servizi online, le imprese trovano difficile gestire le proprie attività sui server di computing interni.

Più nello specifico per capire questo flusso, basti pensare al modo in cui ad oggi accediamo alla nostra e-mail di posta tramite un browser web, analogamente il cloud computing consente alle imprese di accedere e gestire risorse e applicazioni ovunque ci sia una connessione internet. Inoltre, i servizi cloud sono gestiti e mantenuti da fornitori terzi, consentendo ai team IT di regolare calcolo e archiviazione senza dover affrontare costi iniziali di infrastruttura o configurare e gestire ulteriori sistemi e applicazioni.

I principali modelli di servizio cloud includono:

- ✓ Infrastructure as a Service (IaaS): accesso on demand a calcolo, archiviazione, networking e virtualizzazione
- ✓ Platform as a Service (PaaS): risorse hardware e software necessarie per lo sviluppo di applicazioni cloud.
- ✓ Software as a Service (SaaS): stack di applicazioni complete fornite come servizio cloud, inclusa la manutenzione e la gestione dell’infrastruttura sottostante al software dell’applicazione.

E’ importante delineare i vantaggi e le limitazioni di tale servizio in modo da essere in grado di scegliere il modello più adatto alle richieste ed esigenze aziendali.

Tra i vantaggi troviamo:

- Tempo di commercializzazione più rapido: permette uno sviluppo accelerato poiché sono possibili test di nuove idee e progettazioni di nuove applicazioni senza limitazioni hardware o processi di approvvigionamento lenti.
- Scalabilità e flessibilità: le imprese possono scalare rapidamente risorse e archiviazione in base alle proprie esigenze senza dover investire in infrastrutture aziendali; infatti le aziende non devono pagare o creare l’infrastruttura necessaria per supportare livelli di carico più

elevati, allo stesso modo possono fare lo scale down velocemente se le risorse non vengono utilizzate.

- Risparmio sui costi: si paga solo per le risorse effettivamente utilizzate, evitando sovraccarichi e overprovisioning dei data center.
- Collaborazione più efficace: i dati sono accessibile ovunque, consentendo agli utenti di accedere da qualsiasi dispositivo connesso a Internet.
- Sicurezza avanzata: i provider cloud offrono funzionalità di sicurezza avanzate e gestione centralizzata
- Prevenzione della perdita di dati: backup e ripristino di emergenza offerti dai provider cloud.

Limitazioni che si potrebbero riscontrare:

- Dipendenza dalla connessione internet
- Possibilità di tempi di inattività a causa di calamità naturali o problemi tecnici
- Rischio di vincoli al fornitore e minore controllo sull'infrastruttura sottostante
- Preoccupazioni relative alla sicurezza e alla privacy dei dati
- Complessità di integrazione con i sistemi esistenti e possibili costi imprevisti

Dunque, risulta fondamentale comprendere i servizi offerti, i modelli di prezzo e le responsabilità della sicurezza dei clienti. È importante inoltre, tenere presente, che il passaggio al cloud non richiede necessariamente una migrazione totale, non a caso molte aziende adottano un approccio ibrido per ampliare le capacità esistenti. In quest'ottica, i vantaggi del cloud possono superarne le limitazioni, e permettere di avere flessibilità, affidabilità, migliori prestazioni, riduzione dei costi IT, supporto all'innovazione, all'incorporazione di AI e machine learning.

## 1.5. STRUMENTI ANALITICI PER I BIG DATA

Oltre all'infrastruttura IT appena descritta, occorre porre l'attenzione su specifiche tecnologie alla base del Big Data analytics, nonché della stessa organizzazione e archiviazione dati.

L'enorme proliferazione dei dati provenienti da social media, IoT, dispositivi mobili e interazioni con i clienti, richiede una selezione accurata del tipo di database più idoneo per ogni applicazione, tenendo conto delle prestazioni e dei costi. In particolare, gli sviluppatori devono gestire dati estremamente eterogenei per cui risulta vantaggioso esaminare i principali sistemi di gestione e analisi dei dati disponibili nel mondo del software libero, considerando le principali caratteristiche tecniche da un punto di vista business. Risulta evidente come iniziare con la scelta adeguata diventa sempre più essenziale: la selezione di un sistema di gestione dei dati può anche coinvolgere considerazioni di natura economica, ma la componente tecnologica e la roadmap dei DBMS va messa in primo piano per garantire robustezza e durata del progetto; per questo motivo, anche i dirigenti di alto livello o i responsabili degli investimenti devono avere una conoscenza di base della tecnologia disponibile. In tale contesto, il paradigma Open Source consente di ottenere un ottimo trade off tra le reali esigenze di sviluppo e il relativo adattamento ad una domanda di mercato in continua evoluzione.

Le due principali categorie di database sono quelle relazionale e non relazionale. Un'altra modalità di indicare questa caratteristica è riferirsi al linguaggio di ricerca che nei relazionali è SQL (Structured Query Language) e quindi nei non relazionali è indicato con No-SQL.

Di seguito vengono elencati quelli maggiormente usati.

### Ecosistema Hadoop

La libreria Software Hadoop, un'iniziativa open source della Apache Foundation, costituisce un framework che agevola la gestione di vaste quantità di dati, sia strutturati che non strutturati, permettendo la loro elaborazione distribuita su cluster di computer mediante modelli di programmazione intuitivi.

La sua architettura è pensata per adattarsi da un singolo server fino a migliaia, ciascuno dotato delle necessarie componenti per l'elaborazione e lo storage.

Questo framework comprende diversi moduli:

- ✓ Hadoop Common: una serie di utility fondamentali che supportano gli altri moduli di Hadoop
- ✓ Hadoop distributed File System: garantisce un accesso ad alta velocità ai dati, permettendo di accedere ad una qualsiasi fonte dati raggiungibile tramite un URL
- ✓ Hadoop MapReduce: sistema per elaborazione in parallelo di grandi dataset

Parte integrante di questo ecosistema è Apache Spark, ovvero un framework open source per l'elaborazione distribuita in cluster che funge da motore per la gestione di Big Data

all'interno di Hadoop. Offre supporto per diversi linguaggi di programmazione, tra cui Java, Scala, Python e R, e supporta funzionalità quali SQL, il data streaming, il machine learning e l'elaborazione con database grafico.

### Database SQL

Un database SQL, acronimo di Structured Query Language, è un tipo di database relazionale che utilizza il linguaggio SQL per gestire e manipolare i dati. Questo tipo di database organizza i dati in tabelle con righe e colonne, dove ogni riga rappresenta un record e ogni colonna rappresenta un attributo del record. Il linguaggio SQL fornisce una serie di comandi per creare, leggere, aggiornare e cancellare dati all'interno del database. Con SQL, è possibile eseguire interrogazioni complesse per estrarre informazioni specifiche da una o più tabelle, nonché eseguire operazioni di join per combinare dati provenienti da diverse fonti.

I database SQL sono ampiamente utilizzati in una vasta gamma di applicazioni e settori, dalle piccole imprese alle grandi aziende. Offrono una struttura flessibile e scalabile per gestire dati strutturati in modo efficiente ed affidabile.

Di seguito riportati i database aziendali SQL più conosciuti.

#### ➤ *Database relazionali in cloud*

Nel contesto dei big data, si osserva un cambiamento dal tradizionale database relazionale verso soluzioni di archiviazione non strutturata come NoSQL e piattaforme open source come Apache Hadoop. Tuttavia, questo cambiamento non è altrettanto evidente nell'ambiente cloud, poiché fornitori come Amazon Web Services (AWS), Microsoft e Google stanno facilitando le aziende nella transizione dei loro database relazionali "on premise" verso soluzioni di archiviazione e elaborazione flessibili in cloud. Tra i vantaggi nell'utilizzare un database relazionale in cloud troviamo una riduzione dei costi e un aumento della velocità e scalabilità, permettendo alle aziende di beneficiare della replica multipla per i backup; spesso con almeno sei copie fornite dai principali provider, garantendo un miglioramento significativo della continuità operativa. Inoltre, c'è il vantaggio di non dover gestire e amministrare direttamente il database, poiché i provider cloud offrono servizi di database come servizio; si parla appunto di "Database As A Service" (DAAS).

Attualmente i clienti aziendali mostrano una crescente predisposizione all'adozione cloud, soprattutto attraverso i servizi di cloud pubblico offerti dai principali fornitori, superando le preoccupazioni iniziali di sicurezza e compliance. È da sottolineare che occorre tener conto degli svantaggi che una stessa soluzione cloud porta con sé.

In generale, nell'ambito di una strategia globale di migrazione al cloud, si delinea l'adozione di un approccio multi-cloud, perché le grandi aziende preferiscono non dipendere da un unico provider.

Elencati in seguito, le più diffuse opzioni per utilizzare database relazionali in cloud.

- ❖ AWS AURORA: database relazionale di Amazon Web Service AWS progettato specificamente per il cloud, caratterizzato da alte prestazioni e utenti multipli; tra i clienti di Aurora ci sono Netflix, Expedia e Fnra.
- ❖ AWS RDS: offre il servizio di utilizzo database relazionali o il servizio di trasferimento di un database esistente sul cloud storage di AWS tramite Aurora o altri motori come Oracle Database, Microsoft SQL Server , o open source come MySQL o MariaDB; è offerto come database-as-a-service per cui RDS solleva i team IT dalle problematiche amministrative infatti le risorse non sono installate ed eseguite su server fisici all'interno dell'azienda stessa (database on-premise).
- ❖ Microsoft SQL Server: Microsoft offre il suo database SQL as-a-service sull'infrastruttura cloud Azure garantendo il raggiungimento di zero tempi di inattività e un sistema in grado di apprendere modelli di app per adattarsi in tempo reale.
- ❖ MySQL: progetto open source leader nello spazio dei database relazionali; è un software open source di proprietà di Oracle che può essere eseguito sulle infrastrutture dei principali provider cloud quali AWS, Google o Azure.
- ❖ MariaDB: creato e costruito da alcuni sviluppatori originali di MySQL dopo acquisizione con Oracle per resistere ai guasti dei server e può essere eseguito as-a-service; in particolare è indipendente dall'infrastruttura per cui il cliente può scegliere di distribuire il database on premise, in cloud pubblico o in un ambiente multi-cloud, non a caso è stato progettato per essere facilmente trasferibile evitando quindi la dipendenza da un unico fornitore.
- ❖ Google Cloud SQL: è un servizio di database totalmente gestito che semplifica configurazione, manutenzione, gestione e amministrazione database relazionali sulla Google Cloud Platform; si può utilizzare Cloud SQL con MySQL.

➤ Microsoft SQL

Un punto d'incontro tra performance e costi per le piccole aziende. Si tratta di un database relazionale proprietario di tipo tradizionale disponibile in varie versioni, ognuna dedicata a uno specifico carico di lavoro. Olap (*Online Analytical Processing*) e data mining sono accessibili tramite gli Analysis Services. Microsoft SQL rappresenta probabilmente il miglior compromesso tra performance e costi per le piccole imprese.

Naturalmente esiste anche una versione in cloud, denominata Azure SQL.

- **Kinetica**

È un database relazionale in-memory, ciò significa che elabora i dati direttamente nella veloce RAM invece che nella tradizionale memoria a disco. Per aumentarne la velocità, Kinetica mette a disposizione anche un'accelerazione affidata alle GPU del sistema.

Kinetica ha grande successo nei data warehouse, data lake, e supporta molti ambienti di sviluppo oltre a SQL, tra cui java, JavaScript, C++, Python ed altri. Con Kinetica è possibile esplorare i big data senza dover apprendere nuovi linguaggi di query e di programmazione, né creare nuovi modelli di dati, il che lo rende molto interessante per gli sviluppatori.
  
- **Oracle**

Nata come un semplice database, Oracle è un fornitore di ICT a 360 gradi che offre non solo software (DBMS e applicazioni) ma anche hardware e servizi cloud. Le sue radici risalgono all'approccio relazionale con implementazione nei linguaggi C e C++, che lo hanno reso il più potente e veloce della sua categoria.

Si parla quindi di una piattaforma applicativa completa caratterizzata da diffusione e velocità.

Tra le numerose acquisizioni si ha l'azienda Sun Microsystems che ha portato in dote MySQL, ovvero il database SQL open source più diffuso.
  
- **My SQL**

È il database libero più utilizzato sul web.

In dettaglio, un sistema di database open source contiene un codice sorgente che si può visualizzare, scaricare, modificare, distribuire e riutilizzare gratuitamente. Le licenze open source danno agli sviluppatori la libertà di creare nuove applicazioni utilizzando le tecnologie di database esistenti.

Grazie alla licenza libera e al vasto supporto di client e middleware liberi, MySQL è utilizzato principalmente per alimentare siti web dinamici e per fornire la componente di archiviazione e accesso ai dati in una moltitudine di software, sia server che client.
  
- **Sap Hana**

È una Piattaforma dati OLAP, quindi non solo un database server che integra OLTP, ma anche una completa architettura di real time analytics e un application server. Troviamo Funzionalità per sistemi tradizionali quali ERP, CRM, e SRM. La parte di database può essere descritta come DBMS relazionale in-memory basato su colonne invece che righe, il che consente analisi e elaborazioni più veloci, con



funzione principale di database server per memorizzare e recuperare i dati necessari alle applicazioni aziendali.

### Database in-memory

E' un sistema di gestione di database che principalmente impiega la memoria RAM, anziché l'hard disk, per archiviare i dati. Tale approccio, naturalmente, permette una velocità di elaborazione notevolmente superiore, abilitando applicazioni di analisi in tempo reale su Big Data altrimenti non praticabili.

### Database NoSQL

I database NoSQL sono progettati per superare l'ostacolo della sola risposta a query ad-hoc su dati ben strutturati. Infatti, a differenza dei tradizionali database relazionali che organizzano i dati in tabelle con righe e colonne, i database NoSQL utilizzano modelli di dati diversi, come documenti, colonne, grafi per adattarsi meglio alle esigenze specifiche delle applicazioni moderne. I database NoSQL sono spesso scelti per la loro capacità di scalare orizzontalmente su cluster di server distribuiti, consentendo di gestire grandi volumi di dati e di carichi di lavoro ad alta velocità. Questi sistemi sono particolarmente adatti per applicazioni web, social media, IoT, vasta gamma di casi d'uso, applicazioni mobili, analisi in tempo reale, e altre situazioni in cui flessibilità e le prestazioni sono fondamentali. Ogni tipo di database NoSQL ha i suoi punti di forza e le sue applicazioni ideali; quindi, la scelta del database dipende dalle specifiche esigenze dell'applicazione dei requisiti di prestazioni e scalabilità.

Di seguito alcuni esempi di database aziendali NoSQL.

- MongoDB: database documentale NoSQL che offre una grande flessibilità grazie alla gestione di oggetti JSON è ideale come database generico e particolarmente adatto per applicazioni senza uno schema rigido: ciò consente di memorizzare una vasta gamma di dati, come ad esempio dati di temperatura o velocità di rotazione provenienti da accelerometri e altri dispositivi IoT. L'utilizzo di documenti JSON consente di memorizzare dati di diversi tipi in modo flessibile.
- Amazon DynamoDB: servizio di database fornito e gestito completamente da Amazon Web Services; database NoSQL che si adatta alle esigenze di archiviazione e elaborazione richieste senza la necessità di gestire manualmente l'infrastruttura sottostante pe cui consente di gestire facilmente carichi di lavoro in rapida crescita senza compromettere le prestazioni; quest'ultime si configurano come prestazioni

di lettura e scrittura a bassa latenza garantendo tempi di risposta rapidi anche con grandi volumi di dati e picchi di traffico , il tutto in un contesto di affidabilità e durata garantendo affidabilità dei dati con opzioni di backup e ripristino integrati per proteggere i dati dagli errori umani o dalle interruzioni del sistema; possibilità di integrazioni con altri servizi AWS.

### Database a grafo

Tipologia di database che eccellono nell'esplorazione di dati altamente interconnessi, risultano particolarmente adatti per catturare relazioni complesse all'interno di estese reti di informazioni. A differenza dei database relazionali tradizionali, dove i dati sono organizzati in tabelle, in un database a grafo i dati sono rappresentati come un insieme di nodi interconnessi, ognuno contenente informazioni (proprietà) e relazioni tra i nodi, non a caso, appunto, questo modello si presta bene per gestire dati in cui le relazioni sono di primaria importanza.

Un esempio pratico per comprendere la differenza è immaginare un database musicale con informazioni sugli album, le band, le etichette e gli artisti. Se si desidera visualizzare tutti i musicisti presenti in un determinato album di una band pubblicato su una certa etichetta, con un database relazionale sarebbe necessario esplicitare queste relazioni tramite colonne aggiuntive o nuove tabelle, invece, con un database a grafo, le relazioni tra i dati sarebbero già rappresentate in modo nativo, semplificando notevolmente la gestione delle query complesse.

I vantaggi dei database a grafo diventano evidenti soprattutto quando si gestiscono grandi quantità di relazioni tra i dati. Poiché i database a grafo funzionano memorizzando le relazioni insieme ai dati, l'accesso e le interrogazioni delle relazioni diventano efficienti e veloci: i database a grafo leggono semplicemente la relazione dallo storage.

Le applicazioni dei database a grafo spaziano dalle reti sociali dove permettono di cercare visualizzazioni di dati presenti nei social network come i feed di attività, alle rilevazioni delle frodi, dove invece vi sono dei sistemi che utilizzano questi database per mettere in evidenza relazioni tra entità che altrimenti sarebbe difficile da individuare tramite altra rappresentazione di dati. Similmente, i database a grafo sono perfetti per applicazioni che gestiscono le relazioni o le interdipendenze tra entità, infatti, sono presenti spesso dietro sistemi di gestione dei contenuti e dei beni, sistemi di gestione delle identità e degli accessi e soluzioni di conformità normativa e gestione dei rischi. In sostanza, questi database si distinguono per la loro capacità di gestire dati altamente connessi e di rappresentare le relazioni tra essi.

Poiché i database a grafo hanno casi d'uso relativamente di nicchia, non ce ne sono tanti quanto quelli relazionali, tra i più diffusi troviamo: Neo4J e Microsoft Azure Cosmos DB.

# CAPITOLO 2

## 2. *AUTOMOTIVE DATA MANAGEMENT*

### 2.1. *LA RIVOLUZIONE DEI BIG DATA NELL'AUTOMOTIVE: DALLE AUTO CONNESSE ALLA TRASFORMAZIONE DELLA CUSTOMER EXPERIENCE*

Oggi, grazie alle nuove tecnologie di connessione, le auto comunicano con noi. L'accesso a informazioni diagnostiche e sulle performance del veicolo non è una novità, ma il prossimo futuro promette l'integrazione di questi dati con quelli provenienti dall'ambiente circostante e dal conducente (preferenze di guida, servizi desiderati, contenuti multimediali, etc....), il tutto in tempo reale e con una precisione mai vista prima.

Per accedere a questi dati, che comprendono i cosiddetti Big Data, strutturati e no, sempre più veicoli sono dotati di sensori e tecnologie di connettività integrate. È l'applicazione dell'IoT in campo automotive: le auto connesse generano costantemente dati sui veicoli, motori, stili di guida e condizioni ambientali.

Una volta estratto il significato da questi dati eterogenei e in rapida crescita, la sfida è sfruttarli per offrire alle case automobilistiche, produttori di componenti e concessionari, una visione integrata e in tempo reale delle prestazioni del veicolo in condizioni specifiche. In particolare, tale visione è il punto di partenza nel processo di miglioramento e ottimizzazione delle prestazioni del veicolo.

Quindi dopo la fase di analisi, i dati raccolti diventano informazioni accurate per canalizzare le seguenti attività strategiche.

#### ❖ **OTTIMIZZAZIONI DELLE PRESTAZIONI DEL MOTORE**

Uno dei principali benefici dell'utilizzo dei Big Data per il miglioramento delle prestazioni dei veicoli riguarda l'ottimizzazione del motore. I sensori installati nei veicoli permettono di monitorare in tempo reale il funzionamento del motore, rilevando eventuali inefficienze o problemi. Attraverso l'analisi dei dati, è possibile ottimizzare la miscela aria-carburante, gestire la pressione e controllare la temperatura del motore, aumentando così l'efficienza e riducendo le emissioni inquinanti. Quindi, mediante le auto connesse le aziende automobilistiche sono in grado non solo di monitorare i motori, ma anche di aggiornare i software e controllare le prestazioni del gruppo propulsore.

Questo obiettivo, legato alla sostenibilità ambientale, si delinea come una delle priorità per l'intero settore automobilistico.

❖ **MONITORAGGIO DEI CONSUMI**

I Big Data consentono di analizzare i dati relativi al consumo di carburante, alle condizioni della strada e al comportamento di guida per individuare schemi di utilizzo e inefficienze. Queste informazioni possono essere sfruttate per suggerire modifiche allo stile di guida, percorsi ottimizzati e in particolare, nel caso di veicoli elettrici per migliorare ulteriormente l'efficienza dei motori

❖ **PERSONALIZZAZIONI DELL' ESPERIENZA DI GUIDA**

Le case automobilistiche potrebbero adattare l'esperienza di guida alle preferenze specifiche di ciascun conducente. Analizzando informazioni come le abitudini di guida, le regolazioni dei sedili, l'utilizzo dei sistemi di intrattenimento e l'illuminazione, è possibile creare un ambiente di guida su misura per ogni utente. Questa personalizzazione non solo migliora il comfort, ma può anche ottimizzare l'efficienza e le prestazioni del veicolo. L'obiettivo è integrare il benessere del conducente con l'atto della guida, offrendo un'esperienza più piacevole e meno stressante.

❖ **OTTIMIZZAZIONE DELL'ASSETTO E DELLE SOSPENSIONI**

La raccolta di dati dai sensori delle sospensioni e dei sistemi di assetto permette di adattare il veicolo alle condizioni della strada in tempo reale. Analizzando parametri come accelerazione, velocità, e angolo di sterzata, è possibile regolare l'assetto e le sospensioni del veicolo. Questo processo migliora la stabilità e il controllo durante la guida.

Ad oggi, una delle prime realtà di Big Data Driven applicative è stata introdotta dalla multinazionale *Pirelli & C. S.p.A* con la soluzione *Cyber™ Tyre*. È il primo sistema al mondo basato su pneumatici sensorizzati che raccolgono dati vitali e trasmettono dati elaborati in real time all'auto garantendo maggiore potenza, prestazioni e un'esperienza di guida più sicura. Tutto ciò è possibile grazie ad una trasmissione dei dati rilevati dal sensore hi-tech integrato nello pneumatico alla centralina elettronica ed infine vengono trasmessi ai dispositivi di bordo dell'auto.

*Cyber™ Tyre* permette un setup dell'auto adatto allo pneumatico montato, ABS con frenata ottimizzata, ESP controllo di stabilità potenziato, controllo di trazione grazie ad una possibile stima dell'aderenza. Si garantisce quindi, controllo, sicurezza e performance. Ad esempio, il veicolo è in grado di conoscere quando equipaggiato con pneumatici invernali e potrà avvertire di rallentare per velocità di guida oltre i limiti previsti per tali pneumatici. Analogamente, quando riconosce pneumatici estivi, il sistema può indicare una regolazione di pressione a freddo se rileva alte velocità del veicolo a seguito di elaborazione in real time di misurazioni rilevate di pressione e temperature. In dettaglio, è possibile essere informati sulle soglie di temperature ottimali nelle quali i pneumatici possono fornire prestazioni ottimali; infatti, per migliorare le prestazioni di guida, il sistema è in grado di indicare quando iniziare il riscaldamento dei pneumatici, segnalare quando hanno raggiunto una temperatura ottimale e in caso contrario, quando devono raffreddarsi. Più interessante, in campo performance, è la possibilità dallo pneumatico "intelligente" di rilevare la distribuzione della massa del veicolo permettendo al veicolo di adattarsi e cambiare le sue impostazioni, tra cui assetto sospensioni e sistema frenante.

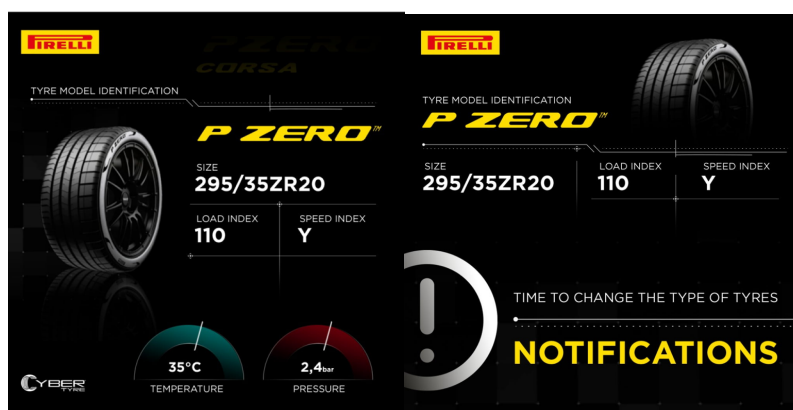


Figura 10–Fonte <https://www.pirelli.com/tyres/it-it/auto/tecnologia/cyber-tyre>

La tecnologia *Cyber™ Tyre* delinea lo sviluppo futuro di *Pirelli* in campo della mobilità e della connettività 5G. A tal proposito *Pirelli* e *Bosch* hanno avviato una collaborazione per sviluppare insieme nuove soluzioni basate su software e innovazioni nella guida, sfruttando i famosi sensori "in-tire". Questo accordo mira ad ottimizzare la dinamica del veicolo approfondendo metodi

avanzati per l'elaborazione dei segnali provenienti dagli pneumatici, utilizzandoli quindi come input per i sistemi di controllo della dinamica del veicolo.

Quindi, da una parte abbiamo il sistema Pirelli Cyber Tyre che rappresenta, come detto precedentemente, la prima soluzione al mondo con sensori integrati negli pneumatici in grado di raccogliere e trasmettere dati in tempo reale al veicolo, dall'altro Bosch ha già sviluppato un'applicazione ESP specifica per i pneumatici Pirelli durante un progetto iniziale con il costruttore di auto ad alte prestazioni Pagani. In particolare, nel contesto di questo nuovo accordo, Bosch apporterà la sua esperienza hardware e software, consolidata come fornitore globale di tecnologie e servizi. Infatti, Bosch è anche un leader nelle tecnologie dei sensori MEMS (Micro-Electro-Mechanical Systems) e sviluppa sensori per la pressione degli pneumatici basati sullo standard BLE ("Bluetooth Low Energy"). Quindi, la sfida consiste nell'integrare questa expertise con le tecnologie di Pirelli, che comprendono hardware, software, algoritmi e modellazione, rendendo possibile la raccolta, elaborazione e trasmissione in tempo reale dei dati dagli pneumatici ai sistemi di controllo elettronico del veicolo attraverso BLE, ottimizzando il consumo energetico.

**TYRES ENTER THE DATA ERA**

INCREASED SAFETY + OPTIMIZED PERFORMANCE + IMPROVED DRIVING PLEASURE

World's first system integrated with car electronics

- ABS Braking Optimization
- ESP Enhanced Stability Control
- TCS Available grip computation

Tyres transmit data to the car processed in real-time

CYBER TYRE

PIRELLI

POWER IS NOTHING WITHOUT CONTROL

Figura 10– Fonte <https://www.bosch-press.it/pressportal/it/it/press-release-76353.html>

Risulta evidente come le tecnologie digitali e l'analisi dei Big Data rivestono un ruolo cruciale nel fornire risposte concrete alle esigenze dei clienti. In dettaglio, incorporando i Big Data in una soluzione CRM, le imprese del settore automobilistico possono anticipare le esigenze dei clienti, ottimizzare il servizio di assistenza e gestire gli investimenti in maniera più strategica e ponderata. Marketing e vendite hanno a disposizione analisi approfondite su tutti i punti di contatto con i clienti, consentendo di segmentare gli stessi clienti con maggiore precisione e di basare le future iniziative su profili più dettagliati: utilizzo dei dati storicizzati per formulare previsioni informate sulle strategie da adottare. Di conseguenza, le tendenze dei diversi segmenti possono essere estratte dai Big Data e utilizzate per prevedere bisogni specifici, indirizzando lo sviluppo del prodotto con maggiore consapevolezza. La sfida è quella di sviluppare una visione integrata del loro target e creare offerte attraenti e diversificate lungo tutto il ciclo di vendita e consumo: l'uso di dati può influire positivamente sulla fidelizzazione dei clienti. Dopo l'acquisto, infatti, comprendere a fondo il comportamento dei clienti e i motivi dell'eventuale abbandono, permetterà alle case automobilistiche di pianificare interventi mirati che favoriscano la fidelizzazione e massimizzano la partecipazione nel mercato dell'aftermarket, riducendo i costi complessivi di marketing

### 2.1.1. "PROJECT 5.0": THE DIGITAL WAVE

Esempio di quanto detto finora è il "**Project 5.0**" di **Magneti Marelli Parts & Service**, un progetto di soluzioni digitali che si pone come obiettivi il Miglioramento della Customer Experience, ottimizzazione e trasformazione dei processi core con arricchimento del Business Model.

*«Project 5.0 rappresenta un cambiamento di cultura aziendale. Grazie alla revisione e alla digitalizzazione dei processi, oltre alla condivisione efficiente delle informazioni, puntiamo a raggiungere una maggiore efficienza interna, un'esperienza dei nostri clienti migliorata, e una maggiore sostenibilità. Attraverso l'implementazione di soluzioni di ultima generazione, inclusa l'intelligenza artificiale generativa, stiamo semplificando la logistica, incrementando la velocità di reazione ai cambiamenti del mercato e rafforzando le relazioni con i nostri partner commerciali. L'iniziativa non si limita ad ottimizzare i flussi di business, ma è volta anche a creare un ambiente di collaborazione e comunicazione tra le persone che faciliti i loro compiti quotidiani. Stiamo abbracciando la trasformazione digitale come una necessità per rafforzare il nostro ruolo nel mercato e posizionarci come un partner ancora più affidabile ed innovativo.»*

**- Franco Ortelli, Head of EMEA Magneti Marelli**

Mediante l'**innovazione** e la **digitalizzazione** nel mondo dell'Independent Aftermarket Automotive, si persegue l'obiettivo di rimanere competitivi in un mercato interconnesso, creando valore tangibile dalla semplificazione dei processi ed operazioni aziendali.

In dettaglio, il progetto mira allo sviluppo delle seguenti aree di competenza: *Extended Supply Chain, CRM, Pricing, Warranty & Claims.*

#### EXTENDED SUPPLY CHAIN

E' il progetto Extended Supply Chain, supportato dalla piattaforma digitale denominata MASP (Marelli Aftermarket Service Portal). Con MASP si rivoluzionano i processi di approvvigionamento automatizzando attività che prima erano gestite manualmente. La piattaforma consente di gestire ordini e documenti digitali, interagire in tempo reale con i fornitori, ricevere avvisi di consegna e accedere a previsioni future di acquisto. Si tratta sostanzialmente di un cruscotto digitale, sempre aggiornato e facilmente accessibile, che rende più semplice ed efficiente la gestione della logistica.



#### CRM

L'obiettivo è valorizzare al massimo la relazione con il cliente, selezionando uno strumento che permetta di conoscere meglio, mantenere un dialogo continuo, comprendere le esigenze e in alcuni casi, anticiparle. Una profilazione accurata permette inoltre di creare proposte personalizzate garantendo un'offerta adatta per ogni domanda nei vari mercati di riferimento. Questa capacità di comprendere, interagire in modo efficiente è resa possibile dall'introduzione di un nuovo sistema CRM (Customer Relationship Management) che centralizza e automatizza le attività di vendita, marketing e assistenza, consentendo di raccogliere e analizzare dati dettagliati tra cui preferenze, storici acquisti e relativi feedback. Il risultato è una partnership solida basata sulla creazione costante di valore.

#### PRICING

Il progetto di pricing di **Magneti Marelli Parts & Services** è un'iniziativa finalizzata a ottimizzare la gestione dei prezzi attraverso l'analisi di mercato, concorrenza e costi di produzione. Utilizzando un software avanzato, vengono sviluppati modelli analitici e sistemi di monitoraggio dei prezzi, permettendo l'adozione di politiche di prezzo dinamiche. È possibile quindi perfezionare le politiche di prezzo e adeguarsi tempestivamente alle fluttuazioni del mercato grazie ad un sistema avanzato di gestione di modelli analitici e all'implementazione di sistemi di monitoraggio su vari mercati sviluppate in collaborazione con un importante azienda leader di settore. In particolare, da oggi, si è in grado di garantire al business la massima profittabilità attraverso strategie di pricing dinamiche e più competitive con i nuovi strumenti di analisi che consentono un posizionamento dei prezzi sempre costante. Si ha così a disposizione un listino aggiornato in tempo reale che tiene conto della concorrenza e dell'oscillazione dei costi di produzione. Si ha maggiore flessibilità sul mercato e pricing proattivo.

#### WARRANTY & CLAIMS

La gestione rapida e semplice delle operazioni di reso, unita alla tracciabilità completa e immediata di ogni pratica, ti garantisce oggi risposte rapide e dettagliate per tutte le richieste legate a garanzie e anomalie di consegna.

Questi vantaggi, che migliorano l'efficienza operativa e semplificano la gestione di garanzie e problematiche di consegna, sono possibili grazie al nuovo progetto Warranty & Claims Digital Management. Si tratta di una piattaforma amministrativa digitale integrata nel portale e-commerce dove è possibile inserire ogni pratica e monitorarne lo stato Real Time, assicurando quindi un controllo costante sulle proprie richieste.

## 2.2. MOBILISIGHT: STELLANTIS'S DATA BUSINESS

Esempio di quanto detto finora è la Mobilisight, il cui progetto in Stellantis nasce nell'ottica di far coesistere ed equilibrare obiettivi di sostenibilità e innovazione tecnologica, stabilendo nuovi standard di mercato e monitorandone l'impatto quotidiano del settore.

Meno di un anno fa, Stellantis ha annunciato la nascita di Mobilisight, quale nuova unità aziendale indipendente dedicata esclusivamente alla crescita del business "Data as a Service" (Daas) della società. Si tratta di una divisione che si concentrerà sullo sviluppo e la commercializzazione di prodotti, applicazioni e servizi B2B innovativi. Le soluzioni di Mobilisights saranno accessibili ad aziende private, servizi pubblici, istituti di istruzione e centri di ricerca, favorendo decisioni basate sui dati e supportando un ampio ventaglio di applicazioni e servizi. Si delinea come una delle sette nuove unità aziendali annunciate nel piano strategico Dare Forward 2030, rappresentando l'impegno di Stellantis nell'avviare e rafforzare la crescita del proprio business legato al software e ai Big Data.



Figura 11– Fonte <https://www.mobilisights.com/en/technology.html>

“La visione di Mobilisights è contribuire a un mondo più intelligente, sfruttando le intuizioni fornite dai dati dei veicoli per ispirare applicazioni e servizi innovativi che possano trasformare e migliorare drasticamente la vita quotidiana degli utenti e delle aziende. Se utilizzati efficacemente, i dati dei sensori e altri dati disponibili dai veicoli connessi possono abilitare una vasta gamma di servizi e applicazioni con benefici significativi, che vanno dall'assicurazione personalizzata basata sull'uso alla rilevazione dei pericoli stradali e gestione del traffico. Con i suoi 14 marchi iconici e milioni di veicoli connessi, Stellantis ha una scala globale di dati senza pari in grado di far avanzare questo business.”

**- Sanjiv Ghate, CEO di Mobilisights**

In dettaglio, Mobilisight permette quindi l'accesso ai dati telematici dei 14 marchi globali di Stellantis, con oltre 13 milioni di veicoli connessi attualmente e una crescita prevista fino a 34 milioni entro il 2030, sia in Europa che in Americhe; con accesso esclusivo e il diritto di concedere in licenza questa elevata mole di dati proveniente dai veicoli di tutti i marchi Stellantis a clienti esterni, inclusi altri produttori di automobili. Questa vasta disponibilità di dati permetterà a Mobilisights di essere meno dipendente da fornitori terzi, garantendo una maggiore autonomia per alimentare le proprie applicazioni che ad oggi, sono ancora in sviluppo. In linea con Dare Forward 2030, Mobilisights sarà un contributore chiave ai 20 miliardi di euro di ricavi annuali incrementali previsti dai servizi software entro il 2030. E' importante sottolineare che in questa sfida di rivoluzionare la mobilità con i dati in tempo reale, Mobilisights e i suoi partner, agiscono nel campo dei Big data, rispettando una rigorosa politica di governance dei dati e privacy. Questo comprende l'uso di dati anonimizzati e aggregati, nonché la condivisione dei dati personali dei clienti solo se ottenuto il loro consenso e per i servizi specifici da loro scelti. Comunque, i clienti possono cambiare idea, e scegliere di non consentire la raccolta, l'uso o la condivisione delle informazioni anche dopo aver dato il loro consenso.

### *2.2.1. RIVOLUZIONARE LA MOBILITÀ CON I DATI IN TEMPO REALE*

Nell'ultimo anno, Mobilisights nell'obiettivo di creare soluzioni innovative, ha ottenuto accordi strategici quali sinergie nella sfida comune di migliorare le esperienze delle flotte e dei consumatori, inclusi i conducenti di veicoli elettrici. In particolare, sono state avviate collaborazioni con diverse decine di clienti, da leader del settore a startup innovative, tra cui Free2move, Geotab, Guidepoint Systems, Echoes, Ocean-Orange Business, Webfleet, MotorQ, CerebrumX, Axa.

In termini di sviluppo dei prodotti, Mobilisights ha inizialmente dato vita a specifici pacchetti di dati per soddisfare le esigenze dei propri clienti. In dettaglio, ha creato pacchetti di dati in evoluzione per flotte per clienti in Europa e Nord America, pacchetti di dati per sviluppare offerte assicurative personalizzate e, infine, un'offerta di streaming di dati per gestire efficientemente la ricarica dei veicoli elettrici.

Ad esempio, i dati forniti in tempo reale dai veicoli connessi, quali il contachilometri e l'accelerazione, vengono analizzate dalle compagnie assicurative, permettendo a quest'ultime di offrire prodotti personalizzati basate sullo stile di guida, premiando la guida sicura.

In campo di veicoli elettrici, Mobilisights delinea come sia possibile sfruttare lo streaming di dati permettendo alle aziende di sviluppare applicazioni intuitive per i proprietari di veicoli elettrici, migliorare la gestione della logistica EV dalla ricarica al monitoraggio dei costi.

Mobilisights ha avviato il servizio Fleet Management Data Pack, integrato in tutti i veicoli commerciali leggeri (LCV) del Model Year 2024 di Stellantis Pro One (Citroën, FIAT Professional, Opel Vauxhall, Peugeot). Si rivoluziona il settore degli LCV con soluzioni di dati connessi all'avanguardia che vengono offerti alle società di noleggio a lungo termine, a breve termine e alle grandi aziende con flotte, con un permesso di accesso estendibile fino a quattro anni al Fleet Management Data Pack. Grazie ai robusti flussi di dati di Mobilisights si migliora la gestione delle flotte: le aziende di noleggio controllano la gestione del carburante e ottimizzano le loro operazioni con dati in tempo reale, migliorando l'efficienza operativa e la soddisfazione dei clienti.

Quindi, Mobilisights propone una soluzione di grande valore per la gestione delle flotte offrendo un accesso privilegiato ai dati dei veicoli connessi del gruppo Stellantis in quasi tempo reale, grazie a un'API versatile e performante. In tal modo i clienti potranno usufruire della ricchezza dei dati, della facilità di integrazione con i fornitori di servizi telematici, e dell'eliminazione dei costi per dispositivi aftermarket. In particolare, la soluzione risulta innovativa e versatile perché si basa su dati incorporati, e garantisce integrazione e un accesso ai dati semplice.

Per capire nel dettaglio i benefici della nuova tecnologia su cui si basa il processo di valore dell'elaborazione dei machine data forniti da Mobilisights , basta analizzare le precedenti soluzioni

#### ✓ **DATI INCORPORATI**

- *Efficienza operativa*: saving di tempo e costo, evitando l'utilizzo dei dispositivi aftermarket.
- *Facilità d'uso*: nessuna richiesta di passaggi aggiuntivi che possono comportare ritardi nell'installazione in quanto è subito attiva.
- *Garanzia di qualità*: dati autentici ed affidabili OEM provenienti da sistemi embedded di automotive. Informazioni accurate, affidabili e coerenti, dal veicolo al cloud.
- *Frequenza e latenza*: dati aggiornati continuamente con la possibilità di essere trasmessi in quasi tempo reale con una latenza minima, garantendo sempre l'accesso agli insight più recenti sul veicolo.
- *Sicurezza dei dati*: flusso di dati sicuro, end-to-end, che garantisce la privacy del proprietario del veicolo e l'integrità delle informazioni.

#### ✓ **INTEGRAZIONE SEMPLICE**

- *Integrazione senza interruzioni*: singola integrazione API per accedere ai dati, senza limiti sulle chiamate API.
- *Ampia scelta*: diversi prodotti di dati a disposizione
- *Opzioni dati flessibili*: possibilità di ricevere flussi di dati continui o dati storici, e anche di cambiare richiesta, in base alle proprie esigenze

#### ✓ **ACCESSO AI DATI**

- *Dati in tempo reale e streaming*: disponibilità di connessione a dati live tramite credenziali sicure e specifici strumenti di configurazione.
- *Dati storici*: API sicura per richiedere dati storici, disponibile sia la chiave API che le linee guida garantendo un facile accesso.

- *Possibilità di provare la piattaforma*: scoprire come funziona la piattaforma con possibilità di una sottoscrizione ai prodotti di dati mediante un processo chiaro e semplice.

### 2.3. *CONNECTED VEHICLES CV*

Dai capitoli precedenti si delineano le opportunità che ad oggi permettono lo sviluppo di nuovi servizi telematici innovativi, tra cui ad esempio la sicurezza del conducente, l'esperienza di guida del cliente, i servizi basati sulla localizzazione, i servizi di manutenzione dei veicoli, e molti altri. Si prevede che entro la fine del 2025 ci saranno circa 2 miliardi di auto connesse sulle strade di tutto il mondo, ognuna delle quali può produrre fino a 30 terabyte di dati al giorno. È importante sottolineare che affinché tutto questo sia possibile occorre sviluppare una piattaforma sia fisica che digitale che si estenda dal veicolo al cloud. In dettaglio, grazie ai sensori integrati e ai dispositivi di comunicazione presenti a bordo, i Connected Vehicles CV, possono eseguire misurazioni quali velocità, temperatura, consumo di carburante, e trasmetterle, in tempo reale tramite connessione 4G/5G a infrastrutture dedicate. Quest'ultime svolgono un ruolo fondamentale per la gestione di dati automobilistici in tempo reale imponendo di fatto vincoli rigorosi sulla piattaforma di gestione dei dati sottostante. In tale ambito, risulta interessante conoscere la struttura di una piattaforma reale che permette di raccogliere, archiviare, elaborare e sfruttare big data in campo automotive (ingestione, archiviazione, elaborazione, sfruttamento, offerta di servizi e condivisione dei dati). Tale architettura di produzione di big data operativa è specificamente quella implementata dal gruppo automobilistico francese Peugeot – Citroen (old PSA Group) che oggi fa parte della holding multinazionale Stellantis N.V.

Si parla di comunicazione Vehicle-to-Infrastructure, noto come "V2I", in quanto i veicoli connessi mediante i loro sensori inviano grandi quantità di dati utili alle piattaforme automobilistiche associate: ad oggi milioni di veicoli connessi possono inviare 170 tipi diversi di dati (numero di identificazione del veicolo, coordinate GPS, giri al minuto, ecc.). Si tratta di dati che, come visto nei capitoli precedenti, sono destinati ad aumentare drasticamente.

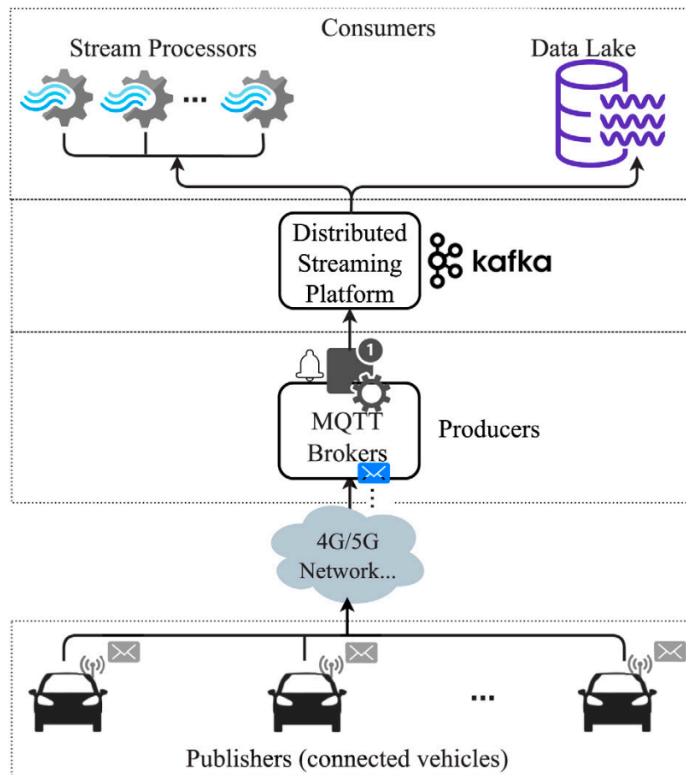


Figura 10– Fonte Mostefaoui, A., Merzoug, M. A., Haroun, A., Nassar, A., & Dessables, F. (2022). Big data architecture for connected vehicles: Feedback and application examples from an automotive group. *Future Generation Computer Systems*, 134, 374-387

I requisiti da cui si parte per la realizzazione di un'architettura di produzione big data sono i seguenti.

- *Privacy dei dati*

Si tratta di garantire una condivisione sicura e flessibile dei dati. I proprietari dei veicoli devono avere la possibilità di condividere le loro informazioni (ad esempio, con i produttori, partner o applicazioni di terze parti), mantenendo però il controllo su quali dati siano accessibili e a chi. In particolare, bisogna essere conformi alle normative e le leggi locali: ad esempio nell'Unione Europea, le aziende devono rispettare il GDPR, che impone rigidi standard di protezione dei dati personali. Il regolamento si applica a qualsiasi organizzazione che gestisce informazioni relative a cittadini dell'UE, indipendentemente dalla sua ubicazione geografica, per cui anche i dati raccolti dai veicoli devono seguire tali direttive. Quindi, è essenziale che i dati siano anonimizzati e trattati in base ai consensi e alle preferenze specificate dai clienti: la piattaforma deve includere strumenti che permettano ai clienti di configurare e aggiornare i propri consensi in qualsiasi momento, inclusi parametri di anonimizzazione e l'elenco di terze parti autorizzate ad accedere ai dati.

- *Efficienza temporale*

Al fine di garantire un'efficace archiviazione ed elaborazione dei dati, privilegiando velocità ed efficienza, l'architettura dovrà supportare due modalità operative principali: Real Time Processing che consente di gestire i flussi di dati mentre vengono generati, e il Batch processing che permette di analizzare offline grandi volumi di dati in modalità differita mediante data lake.

- *Alta disponibilità*

L'architettura del sistema deve essere estremamente robusta ovvero deve avere un'elevata capacità di tolleranza ai guasti. In caso di errori, sia software che hardware, deve essere in grado di riprendersi rapidamente al fine di evitare interruzioni percepibili per l'utente, e soprattutto senza perdita di dati. Questo può essere realizzato attraverso soluzioni come Data Striping che prevede la suddivisione dei dati tra più dischi per aumentare le prestazioni e la protezione, oppure si può optare per la soluzione di ridondanza su più nodi cloud.

- *Scalabilità*

La piattaforma deve essere altamente scalabile per gestire l'aumento previsto di veicoli connessi combinato con la crescente qualità di dati generati. È fondamentale che ogni componente software impiegato sia ottimizzato per l'ingestione, l'archiviazione e l'elaborazione efficiente dei dati, sia in tempo reale che in modalità offline. Inoltre, il design deve essere flessibile, consentendo la scalabilità sia orizzontale con aggiunta di nuovi nodi o server per distribuire il carico di lavoro, sia verticale con miglioramento delle risorse hardware per gestire carichi crescenti.



### 2.3.1. ARCHITETTURA DI BIG DATA APPLICATIVA IN CAMPO AUTOMOTIVE

L'architettura di big data del Group PSA descritta è composta da sei livelli: raccolta dati, ingestione dei dati, accodamento dei dati, gestione dei dispositivi e dati di riferimento, elaborazione dei dati in due differenti modalità, e infine sfruttamento ed offerta dei dati.

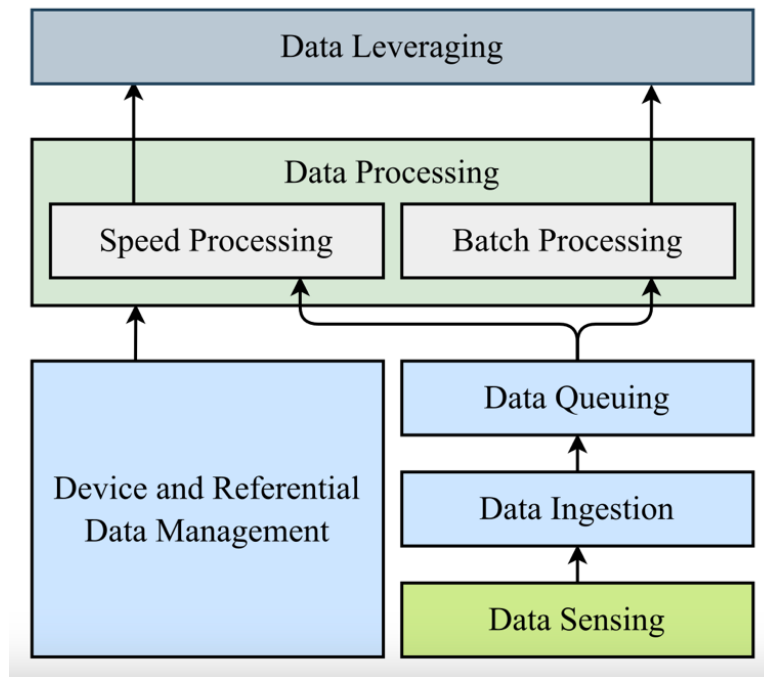


Figura 11– Fonte Mostefaoui, A., Merzoug, M. A., Haroun, A., Nassar, A., & Dessables, F. (2022). Big data architecture for connected vehicles: Feedback and application examples from an automotive group. *Future Generation Computer Systems*, 134, 374-387

Con il termine “Data Sensing” si delinea il primo livello dell’architettura di un sistema di veicoli connessi che, come accennato nei capitoli precedenti, riguarda la raccolta e la trasmissione delle informazioni ambientali dai veicoli verso una stazione base dedicata. Questo passaggio avviene tramite la gestione di una rete di sensori avanzati e unità di servizio telematico TSU installate nei veicoli. I sensori, che si configurano nel campo IoT, raccolgono una serie di dati tra cui informazioni su velocità, posizione GPS, stato dei sistemi interni del veicolo, e li trasmettono utilizzando reti wireless mobili. In dettaglio, le unità TSU sono collegate al bus CAN del veicolo e gestiscono la trasmissione dei dati con una velocità di 1Mbit/s secondo lo standard ISO 11898. Ogni TSU è progettato per scopi specifici, ad esempio, quelle per il monitoraggio sono in grado di trasmettere una grande quantità di dati con alta frequenza, permettendo il monitoraggio in tempo reale di parametri critici da analizzare. Per cui, si spazia da Autonomous Telematic Boxes ATB a scatole telematiche approvate aftermarket, spesso utilizzate su veicoli di fascia bassa e responsabili dell’integrazione di funzionalità aggiuntive come la localizzazione GPS o la gestione dei dati di manutenzione predittiva.

Implementata la strumentazione per la raccolta dati, quest'ultimi vengono trasmessi in formato binario utilizzando il protocollo MQTT, noto per la sua efficienza e bassa latenza. Questo protocollo di comunicazione connette, tramite broker di dati, i veicoli con la loro infrastruttura automobilistica.

Il Data sensing è seguito dal livello intermedio di "Data Ingestion" che ha un ruolo fondamentale in quanto svolge la funzione di acquisire i dati provenienti da fonti definite (come i veicoli connessi) e prepararli affinché possano essere utilizzati dai livelli superiori. Risulta cruciale perché i dati una volta raccolti sono grezzi e non immediatamente pronti per l'analisi; per cui il corretto flusso prevede che questi vengono prima ingeriti, decodificati e formattati correttamente, in modo che i livelli successivi del sistema possono interpretarli e utilizzarli in modo efficace. Come specificato precedentemente, in tale scenario, il protocollo di comunicazione prescelto è MQTT, un protocollo leggero molto diffuso nel campo dell'Internet of Things (IoT). Questo livello intermedio, o front-end, è quindi costituito da broker MQTT, che gestiscono l'intero flusso di dati in quanto ricevono i dati dai veicoli, li decodificano e li riformattano in un formato standardizzato prima di inviarli ai livelli successivi per ulteriori elaborazioni. Grazie alla sua leggerezza, il protocollo MQTT è particolarmente adatto a soluzioni embedded, che richiedono efficienza e ridotto consumo di risorse. In dettaglio, la scelta di un protocollo come MQTT garantisce una gestione affidabile e scalabile dei dati: è stato progettato specificamente per scenari IoT, per cui è ottimizzato per gestire un elevato numero di dispositivi connessi, spesso con risorse limitate e necessità di bassa latenza rendendolo particolarmente indicato per applicazioni dove è necessario elaborare rapidamente grandi volumi di dati, come quelli provenienti dai veicoli connessi, e farlo in modo efficiente. Inoltre, la semplicità di implementazione di MQTT riduce il carico sul sistema, permettendo agli sviluppatori di concentrarsi su altre parti critiche della piattaforma.

Segue il livello di "Data Queuing", ovvero di distribuzione dati che insieme al livello di ingestione svolge un ruolo cruciale nella gestione delle indisponibilità sia dei produttori, quali unità di raccolta dati, sia dei consumatori, ad esempio i processori di flusso o i data lake. Funge da coda distribuita asincrona sicura in quanto caratterizzata da una politica di retention ridefinita per assicurare che in caso di problemi o rallentamenti i messaggi non si perdano ma rimangano disponibili per essere elaborati. In pratica, i produttori e i consumatori non interagiscono direttamente: i messaggi possono restare in coda anche dopo essere stati recuperati da alcuni consumatori, garantendo che tutti i consumatori potenziali possano ricevere i dati necessari in modo resiliente e fluido. Qui entra in gioco Apache Kafka, la piattaforma open-source, che in parole semplici funziona come un sistema di messagistica in cui diversi sistemi possono inviare o ricevere dati in modo continuo. Quindi, il viaggio dei dati inizia con i dispositivi IoT installate nei veicoli, che inviano i dati attraverso i broker MQTT. Occorre tenere presente che il vantaggio principale di MQTT non è la capacità di gestire un elevato volume di messaggi quanto la sua capacità di mantenere milioni di connessioni simultanee con dispositivi IoT; infatti, una volta che i dati sono stati raccolti dai broker MQTT, che fungono da bilanciatori di carico per gestire numerose sessioni TCP in parallelo, essi vengono

trasferiti ai broker Kafka, poiché MQTT ha capacità di archiviazione limitata. In dettaglio, i broker Kafka garantiscono:

- ✓ Gestione di grandi volumi di dati IoT: immaginando di avere migliaia di sensori che inviano dati tramite MQTT, utilizzare Kafka come backend può aiutare a scalare e organizzare questi flussi di dati in modo efficiente, rendendo più semplice l'elaborazione, l'archiviazione e l'analisi in tempo reale.
- ✓ Affidabilità e persistenza dei dati: Kafka è molto affidabile per la gestione dei dati nel lungo periodo, poiché conserva i dati per un tempo definito garantendo che i dati non vadano persi e che possano essere processati anche dopo che sono stati inviati. MQTT, invece, è più orientato a scenari in cui i messaggi vengono consegnati e poi dimenticati subito.
- ✓ Integrazione tra dispositivi IoT e sistemi aziendali: Kafka può fungere da ponte tra dispositivi IoT che comunicano tramite MQTT e i sistemi aziendali che necessitano di analizzare quei dati, come database, dashboard, o strumenti di machine learning.

Quindi, in relazione alla piattaforma analizzata, nel caso di interruzioni tra la produzione e il consumo dei messaggi, Apache Kafka viene utilizzato come buffer. Supponendo che durante la notte si verifichi un picco di 500 milioni di messaggi, MQTT andrà a gestire le connessioni per milioni di dispositivi, inviando rapidamente i messaggi a Kafka. Quest'ultimo sarà in grado di accettare e immagazzinare l'intero carico di dati, assicurando che i consumatori possano accedere alle informazioni quando necessario. Si migliora e si garantisce in tal modo la resilienza e la continuità operativa del sistema.

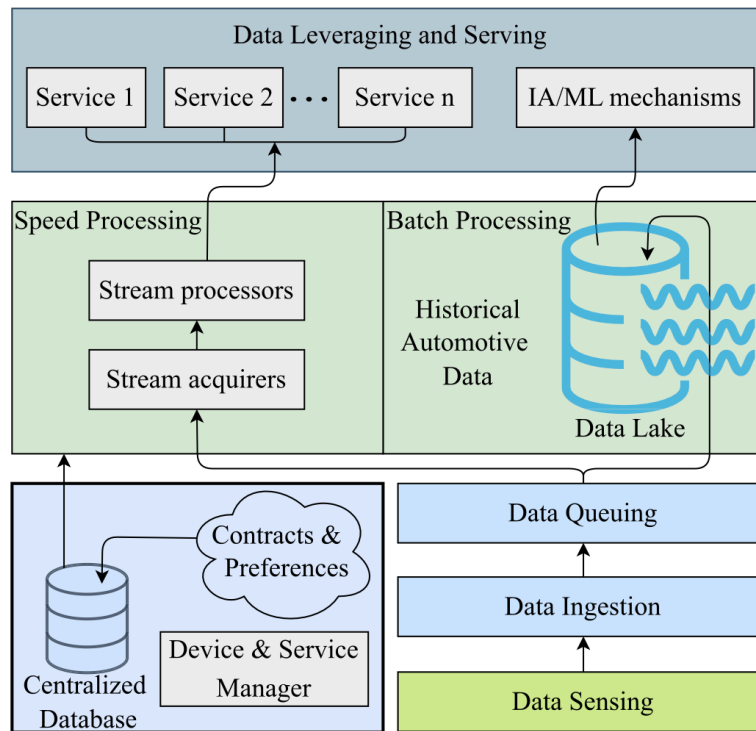


Figura 12– Fonte Mostefaoui, A., Merzoug, M. A., Haroun, A., Nassar, A., & Dessables, F. (2022). Big data architecture for connected vehicles: Feedback and application examples from an automotive group. *Future Generation Computer Systems*, 134, 374-387

Dallo schema illustrato nella Figura 12, i tre livelli principali quali raccolta dei dati, ingestione e accodamento che si concentrano sulla raccolta e sulla preparazione dei dati provenienti dai veicoli, operano in modo indipendente dal livello di riferimento che invece si occupa della gestione dei dispositivi e fornisce le informazioni essenziali per garantire il corretto funzionamento dell'intera piattaforma. Infatti, questo livello funge da base per le decisioni e le operazioni dei livelli superiori, come l'elaborazione avanzata e l'analisi dei dati. Gestisce le politiche di preelaborazione dei dati: ad esempio, può definire quali tecniche di anonimizzazione devono essere applicate ai dati per proteggere la privacy, oppure quali tipi di filtraggio o aggregazione dei dati sono necessari prima che questi vengano inviati ai livelli superiori per l'elaborazione. Queste politiche possono variare a seconda del tipo di dati raccolti, delle normative locali o delle esigenze dei partner commerciali. Altro compito cruciale è la definizione delle politiche di accesso: il livello di riferimento gestisce tutte le funzioni critiche e le informazioni necessarie per la gestione dei big data tra cui informazioni su dispositivi, proprietari dei veicoli, clienti, partner, contratti, preferenze e servizi offerti e, sulla base di questi, determina chi può accedere a specifiche risorse (come informazioni ambientali, geografiche o relative al veicolo) e in quale modalità operativa, sia essa B2B, B2C o entrambe. In pratica, il livello di riferimento stabilisce le regole e i diritti di accesso per i diversi utenti della piattaforma. Dal punto di vista implementativo, è composto da due elementi chiave: un gestore di dispositivi/servizi e un database centralizzato. Il gestore di dispositivi si occupa dell'attivazione, disattivazione e aggiornamento dei dispositivi connessi, garantendo che siano sempre operativi e aggiornati con le ultime configurazioni. Il database centralizzato, invece, funge da repository per tutte le informazioni necessarie, incluse quelle relative a contratti, preferenze dei clienti e politiche di accesso. Questo sistema garantisce che tutti i livelli della piattaforma

lavorino in sinergia e che i dati siano trattati e gestiti in modo conforme alle regolamentazioni e alle esigenze degli utenti. In sintesi, il livello di riferimento non solo gestisce l'infrastruttura tecnica necessaria per il funzionamento dei dispositivi e dei dati, ma stabilisce anche le linee guida operative e di sicurezza che governano il flusso e l'accesso alle informazioni, assicurando che l'intero ecosistema rimanga affidabile, sicuro e scalabile.

Penultimo livello è il “Data Processing”, nonché cuore dell'intera architettura. Si divide in due sottolivelli principali (Figura 12): “Speed Processing” e “Batch Processing”. In parole semplici, lo speed processing è l'elaborazione in tempo reale che si occupa della gestione di flussi di dati continui, mentre con batch processing ci si riferisce al data lake che archivia ed elabora i dati storici attraverso processi offline, come l'applicazione di tecniche di machine learning e deep learning. Le principali funzioni del sottolivello di elaborazione in tempo reale sono svolte attraverso due microservizi di streaming: i consumatori di flussi di dati e i processor. I consumatori di flussi raccolgono dati automobilistici aggiornati dal livello inferiore di accodamento (Apache Kafka) e li trasmettono ai processor di flussi che hanno il compito essenziale di elaborare flussi di dati continui garantendo il rispetto di vincoli stringenti (in tempo reale o quasi) e tenendo conto delle preferenze. Accanto all'elaborazione in tempo reale, si trova il data lake che offre uno storage distribuito efficiente e un'elaborazione offline rapida per i dati automobilistici storici; infatti, per comprendere la bontà di questo particolare storage risulta ottimale studiarne le caratteristiche considerando i seguenti aspetti distintivi: memorizzazione dei dati ed elaborazione dei dati storici. Un data lake è un ampio deposito di dati basato su tecnologie a basso costo, progettato per ottimizzare la raccolta, raffinazione, archiviazione ed estrazione di dati grezzi all'interno di un'azienda. Infatti, invece di collocare i dati in un data warehouse che, per la sua stessa natura deve essere progettato per scopi specifici, è più semplice trasferirli nel data lake nel loro formato originale. Una volta depositati nel data lake, i dati diventano accessibili a tutti i team dell'organizzazione, favorendo una maggiore facilità d'accesso per le analisi. È ovvio che grazie ai vantaggi offerti dai data lake, tra cui la capacità di memorizzare diversi tipi di dati nello stesso repository, la possibilità di definire la struttura dei dati al momento dell'utilizzo (schema-on-read), e il fatto che un data lake possa contenere dati non strutturati o multi-strutturati, si riducono i costi iniziali di ingestione dei dati, come la trasformazione. In altre parole, i dati grezzi possono essere acquisiti e archiviati su larga scala a un costo ridotto. Nella piattaforma analizzata, il data lake implementato è costituito da due componenti principali: la soluzione meno adottata, Hadoop, e la soluzione attualmente preferita, l'archiviazione a oggetti. Inoltre, si ha l'adozione di un sistema ibrido e unificato che sfrutta data lake e data warehouse aziendali (EDW) mirando a creare un ambiente in cui gli utenti possano ottenere risposte a domande complesse con una maggiore quantità di dati e analisi, riducendo al minimo lo sforzo necessario. Se da una parte vi è l'elaborazione e l'archiviazione di dati storici, dall'altro vi è l'elaborazione in tempo reale che attualmente, è implementata con Apache Spark, uno dei framework più noti e utilizzati per l'elaborazione batch. Questo strumento offre numerosi vantaggi, tra cui l'efficienza e l'idoneità per calcoli complessi e iterativi, come quelli richiesti dall'apprendimento automatico. Spark si distingue per la sua facilità d'uso e sviluppo, supportando diversi linguaggi di programmazione come Java, Scala, Python e R. Inoltre, Spark è estremamente veloce nella gestione dei dati in memoria e garantisce tolleranza ai guasti grazie alla sua implementazione dei Dataset Distribuiti Resilienti (RDD). Esulo dalla trattazione specifica di tali soluzioni in quanto al di fuori dello scopo del mio elaborato.

Ultimo livello dell'architettura analizzata è il "Data Leveraging": sfruttamento e offerta dei dati. Sostanzialmente prende i dati prodotti dai livelli inferiori, come il data lake e i sistemi di elaborazione in tempo reale, li indicizza e fornisce risultati delle query agli utenti finali. Il livello di virtualizzazione è una tecnologia che semplifica l'accessibilità ai dati gestendone la complessità sottostante e quindi, permettere di offrire ai consumatori di dati una visione unificata e chiara per manipolare ed interrogare i dati memorizzati in un insieme eterogeneo di archivi. In dettaglio, il software di virtualizzazione dei dati svolge un ruolo fondamentale aggregando fonti di dati sia strutturati che non strutturati attraverso dashboard o strumenti di visualizzazione; ciò non implica la replica dei dati dai sistemi sorgente in quanto essa memorizza solo metadati e logiche necessarie per l'integrazione e la visualizzazione. Questo approccio non solo riduce i costi e il tempo di gestione dei dati, ma consente anche agli utenti di concentrarsi sull'analisi e sulla scoperta di informazioni significative, promuovendo decisioni più informate e tempestive all'interno delle organizzazioni. Riguardo l'indicizzazione di big data, questa risulta necessaria poiché, quando si hanno enormi quantità di dati, l'indicizzazione crea una sorta di "mappa" o "riepilogo" dei dati permettendo di trovare rapidamente le informazioni senza dover analizzare ogni singolo pezzo di dati; nel dettaglio del livello tecnico devono essere considerati database NoSQL. Come spiegato nel capitolo 1 del seguente elaborato, i database NoSQL presentano tre caratteristiche fondamentali, nell'obiettivo di gestire la crescente quantità di dati e aumentare l'affidabilità del sistema contro possibili perdite: alta scalabilità, replicazione dei dati e flessibilità rispetto allo schema, in quanto non predefinito.

L'offerta di dati automobilistici agli utenti finali avviene attraverso due modalità principali: **MQTT** (Message Queuing Telemetry Transport) e **API RESTful**. Nello specifico, MQTT, come visto precedentemente, è un protocollo di messaggistica leggero, ideale per comunicazioni a bassa larghezza di banda e per dispositivi IoT, quindi è specializzato nelle comunicazioni tra dispositivi; mentre API è un'interfaccia che consente alle applicazioni di comunicare tra loro, utilizzando principalmente richieste HTTP per ottenere o inviare dati. Per cui MQTT e API sono due tecnologie utilizzate per far comunicare i dispositivi e le applicazioni, permettendo un accesso flessibile e scalabile ai dati, e rendendo possibile l'integrazione con una vasta gamma di applicazioni. Per garantire la sicurezza e la privacy degli utenti, vi è l'implementazione del protocollo di autorizzazione standard aperto **OAuth**. Questo sistema consente a clienti terzi, come applicazioni mobili e servizi di terze parti, di accedere ai dati personali dei veicoli senza esporre le credenziali degli utenti, garantendo così un elevato livello di protezione.

In aggiunta, il livello di sfruttamento dei dati si avvale di tecniche avanzate di **intelligenza artificiale (AI)** e **machine learning (ML)**. Queste tecnologie non solo migliorano la qualità dei dati memorizzati nel **data lake**, ma li rendono anche pertinenti per una vasta gamma di interrogativi e problematiche. Tra le applicazioni più significative vi è il miglioramento di nuovi prodotti automobilistici e il potenziamento dei servizi ai clienti. Inoltre, l'analisi dei dati consente di identificare tendenze emergenti nel comportamento degli utenti, favorendo l'innovazione continua e l'ottimizzazione delle strategie di marketing. Questo approccio data-driven apre anche

nuove opportunità per sviluppare soluzioni personalizzate, aumentando la soddisfazione del cliente.

### 2.3.2. APPLICAZIONE: SERVIZIO METEOROLOGICO E SERVIZIO ECO-DRIVING

I due servizi implementati nella piattaforma automobilistica del PSA Group sono: servizio meteo e servizio di eco-driving. Quest'ultimo è progettato per consigliare i conducenti al fine di migliorare il loro stile di guida in un'ottica più ecologica. La valutazione adottata si basa su un sistema di punteggi: il punteggio complessivo è dato dalla somma del punteggio relativo allo stile di guida che valuta il comportamento del conducente e del punteggio relativo al profilo della strada che valuta l'itinerario selezionato. Questo sistema analizza il consumo di carburante, le accelerazioni, la velocità, la frenata, i giri motore (RPM) e altri parametri di guida, correlando queste informazioni con dati ambientali come altitudine, pendenza della strada e altre condizioni rilevanti.

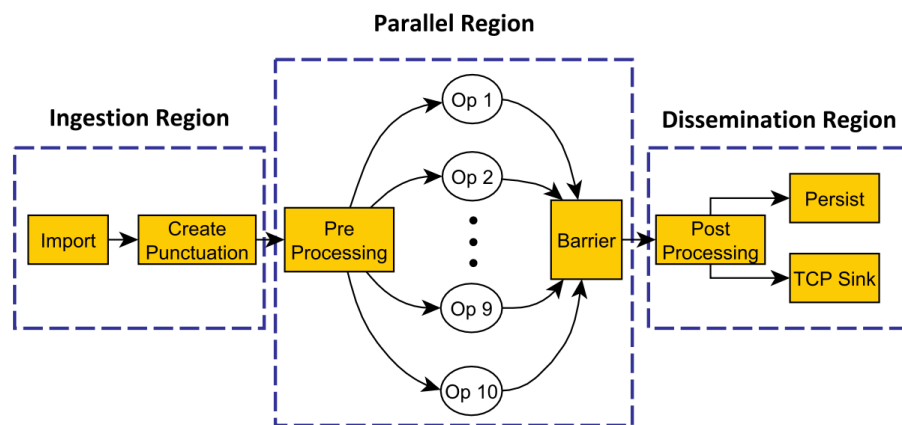


Figura 13– Fonte Mostefaoui, A., Merzoug, M. A., Haroun, A., Nassar, A., & Dessables, F. (2022). Big data architecture for connected vehicles: Feedback and application examples from an automotive group. *Future Generation Computer Systems*, 134, 374-387

La Figura 13 mostra una versione semplificata del grafo del flusso dei dati, evidenziando gli operatori principali e le loro interazioni. L'applicazione è composta da tre sezioni principali: **raccolta**, **elaborazione parallela** e **diffusione**. La maggior parte del lavoro computazionale avviene nella regione parallela, dove le operazioni, essendo indipendenti, possono essere eseguite simultaneamente per accelerare l'elaborazione. Alla fine della fase parallela, una barriera temporale sincronizza i risultati per procedere alla fase successiva.

Per quanto riguarda il servizio meteorologico, sviluppato in collaborazione con Météo-France, l'obiettivo è quello di raccogliere la temperatura esterna ambientale da ogni veicolo connesso, l'etichetta con la posizione geografica attuale del veicolo e inviare tutti i dati all'infrastruttura centrale. In seguito, l'infrastruttura aggrega queste informazioni per aggiornare in tempo reale una mappa della temperatura su scala nazionale. In tal modo si crea una visualizzazione accurata delle condizioni termiche nelle diverse regioni del paese, garantendo una panoramica aggiornata della temperatura in ogni dipartimento.

### 2.3.3. VALUTAZIONE PRESTAZIONI: QUALITÀ DEI DATI

Risulta interessante valutare l'architettura come quella descritta, secondo due criteri principali: prestazioni e qualità. La valutazione della prestazione consiste nell'esaminare il comportamento tecnico della piattaforma di fronte a fluttuazioni del flusso di dati; determinare quanto una piattaforma sia scalabile. Nel seguente elaborato non verrà approfondita questa parte di valutazione poiché richiede una maggiore comprensione delle impostazioni hardware su cui è distribuita l'architettura descritta e un'analisi del trend di evoluzione del numero di pacchetti CV inviati negli ultimi anni.

La dimensione della qualità riguarda l'analisi di due sotto-caratteristiche: qualità dei dati e qualità del servizio "QoS". Quest'ultimo consiste nell'analizzare il livello di comunicazione che collega tutti i componenti: i vari processi e componenti sottostanti devono lavorare sinergicamente per raccogliere, pulire, elaborare e memorizzare i dati. Per cui, è necessaria una piattaforma centrale robusta, tollerante ai guasti e in grado di funzionare come un sistema di messagistica che invia in modo affidabile ed efficiente tutti i dati necessari a ciascun elemento dell'architettura. Relativamente alla piattaforma descritta, si dovrebbe analizzare il sistema di messagistica implementato Apache Kafka esaminando i diversi livelli operativi. Kafka si distingue per la sua capacità di connettere vari **cluster di macchine**, ognuno dedicato a compiti specifici, come l'elaborazione in tempo reale, l'elaborazione batch, l'archiviazione e le piattaforme di visualizzazione.

Più strettamente collegato al seguente elaborato è la qualità dei dati. In termini di qualità dati ci si riferisce a due diverse dimensioni: "intrinseca", riferita ad attributi oggettivi e nativi dei dati, e "contestuale", che riguarda attributi influenzati dal contesto in cui i dati vengono osservati o utilizzati. Gli indicatori delle dimensioni contestuali sono spesso utilizzati in sondaggi e questionari agli utenti, poiché si basano su valutazioni soggettive per quantificare i dati. Di conseguenza, le dimensioni contestuali tendono a trattare informazioni anziché dati, poiché sono costruite quando i dati vengono inseriti in un contesto o in un problema specifico. La valutazione della qualità dei dati "intrinseca" avviene in due fasi della catena del valore dei big data: all'atto della ricezione dei dati e, dopo l'elaborazione.

Di notevole interesse per una maggiore comprensione e chiarezza di quanto detto, risulta l'analisi di qualità dati condotta da un gruppo di ricercatori sull'architettura descritta in collaborazione con Team Engineering di PSA Groupe (*Mostefaoui, A., Merzoug, M. A., Haroun, A., Nassar, A., & Dessables, F. (2022). Big data architecture for connected vehicles: Feedback and application examples from an automotive group. Future Generation Computer Systems, 134, 374-387*)

In dettaglio, si è analizzato il livello di pre-elaborazione ovvero la fase di raccolta immediatamente successiva alla trasmissione dai veicoli IoT all'infrastruttura. In questa architettura, la pre-elaborazione può essere suddivisa in due fasi principali: la **decodifica** dei dati binari grezzi provenienti dalle scatole telematiche, e la fase successiva di pre-elaborazione che avviene all'inizio della catena di elaborazione dei flussi durante la quale vengono filtrati i dati incoerenti. Seguono i



parametri utilizzati per fornire una misura della **qualità intrinseca** dei dati, e i relativi risultati di performance.

1. **Coerenza:** questa dimensione verifica il grado di coerenza dei record di dati correlati con la struttura e il formato predefiniti. Un alto livello di coerenza garantisce che i dati siano integrati e privi di contraddizioni, essenziale per l'affidabilità delle analisi

$$i) \quad CNS_a = \frac{numInconsistentRecords}{totalRecords}$$

Questo indice rappresenta il rapporto tra il numero di record incoerenti e il numero totale di record. In un normale intervallo di 24 ore, l'infrastruttura di PSA Groupe ha elaborato 65 milioni di pacchetti, di cui circa 130.000 generano eccezioni durante il processo. Si ottiene quindi un tasso di errore basso pari circa allo 0,2%

2. **Accuratezza:** Questo parametro si riferisce al grado in cui le informazioni riportate sono esatte e rappresentano valori reali. Dati accurati sono fondamentali per prendere decisioni informate e per costruire fiducia nel sistema, specialmente in applicazioni critiche come la navigazione o la gestione della flotta.

$$i) \quad AM_a = \frac{numCorrectValues}{totalValues}$$

Questa metrica rappresenta il rapporto tra il numero di record corretti ricevuti e il totale dei record. I risultati ottenuti indicano che il 99,9% dei dati è stato ricevuto senza errori. Questo tasso di correttezza è attribuibile al primo passaggio di serializzazione, che filtra quasi tutti i pacchetti anomali.

3. **Completezza:** La completezza riguarda la misura in cui i record possono essere considerati completi nel loro contenuto, ossia senza dati mancanti. La mancanza di dati può compromettere la capacità di analisi e portare a conclusioni errate. Implementare meccanismi per il controllo della completezza è cruciale per garantire un'analisi robusta.

- i) I risultati mostrano che il 100% dei dati trasmessi è completo, senza valori mancanti. Ciò dimostra che l'architettura proposta assicura la completezza dei dati.

4. **Tempestività:** Questa dimensione valuta l'aggiornamento dei dati, assicurando che le informazioni siano attuali e rilevanti al momento dell'analisi. La tempestività è

particolarmente importante in contesti in cui le decisioni devono essere prese rapidamente, come nella gestione del traffico o nella risposta a incidenti.

- i) Nella configurazione operativa descritta, avviene un aggiornamento ogni minuto, durante il quale ogni veicolo connesso invia un pacchetto. La piattaforma è in grado di elaborare (ricevere, decodificare, pulire e memorizzare) questi dati CV in meno di un minuto, tipicamente tra 2 e 25 secondi. Si conclude che la piattaforma automobilistica proposta soddisfa pienamente i requisiti di tempestività relativi alla qualità dei dati.

È importante una valutazione della qualità dei dati a livello di pre-elaborazione perché non solo ottimizza i processi successivi di elaborazione e analisi, ma contribuisce anche a migliorare l'affidabilità generale dell'architettura dati, garantendo che le informazioni siano utili e pertinenti per le applicazioni finali.

## 2.4. *METODOLOGIA DATA-DRIVEN PER LA VALUTAZIONE DELLA PROGETTAZIONE HMI*

Con il termine HMI, o Human-Machine Interface, si intendono tutti quei sistemi che permettono alle persone di interagire con macchine o dispositivi tecnologici; può essere un display, un pannello di controllo o più genericamente un software che mostra informazioni e comandi. In tale ambito, la gestione dell'esperienza utente UEM User Experience Management, si concentra sull'analisi, misurazione e ottimizzazione dell'esperienza che gli utenti hanno quando interagiscono con un prodotto digitale (app, sito web o software). In sintesi, in termini di usabilità, gli obiettivi principali dell'UEM sono: monitorare l'interazione dell'utente, identificare punti critici, migliorare la soddisfazione dell'utente, aumentare l'efficienza. Quindi, i metodi di valutazione dell'UEM sono strumenti progettati per identificare e prevenire problemi di usabilità attraverso indagini come interviste, sondaggi e questionari. Tuttavia, tali metodi si basano principalmente su valutazioni soggettive, le quali non sempre sono sufficientemente specifiche per fornire un quadro completo delle prestazioni del sistema o del comportamento degli utenti; infatti, le domande sono spesso generali e per questo non riescono a riflettere accuratamente i problemi che ogni singolo utente può aver riscontrato. In un approccio tradizionale ciò si rivolge mediante la richiesta di ulteriori approfondimenti da parte degli esperti di usabilità, un processo complicato e dispendioso. L'uso di questionari troppo dettagliati, inoltre, può allungare i tempi di risposta e compromettere la qualità dei risultati, poiché gli utenti potrebbero non ricordare correttamente le loro esperienze o dare risposte incoerenti. Inoltre, occorre considerare che gli utenti tendono a ricordare solo pochi problemi, trascurando informazioni chiave come la durata d'uso o i tempi di risposta del sistema.

Ecco che qui entrano in gioco i Big Data, rappresentando uno strumento sempre più rilevante per la valutazione efficace di un HMI design, in campo automotive, poiché permettono di raccogliere informazioni dettagliate sul comportamento degli utenti in ambienti reali. Un approccio basato sui dati fornisce criteri oggettivi e verificabili, ponendo le basi per metodologie più robuste nella valutazione di una progettazione HMI. Nell'attuale panorama delle interfacce automobilistiche, a causa della crescente complessità dovuta all'integrazione di numerosi sottosistemi tecnologici, prevedere e valutare i comportamenti umani è un'impresa difficile; inoltre, le interazioni dell'utente sono influenzate da numerosi fattori, come la velocità del veicolo, le condizioni stradali e il numero di compiti simultanei che il conducente deve eseguire. Per superare questi limiti, un approccio alternativo è sfruttare i dati disponibili attraverso i sistemi software installati nei veicoli. I moderni sistemi telematici forniscono un flusso continuo di dati che collega le azioni degli utenti con le condizioni stradali, le attività svolte e le caratteristiche del veicolo. L'analisi di questi dati permette di tracciare un quadro dettagliato delle interazioni tra conducente e sistema HMI, offrendo uno strumento potente per identificare problemi di usabilità non rilevabili con metodi

tradizionali. L'integrazione dell'analisi dei big data con la valutazione delle prestazioni del sistema può aiutare a individuare gruppi di utenti che hanno riscontrato problemi simili in condizioni equivalenti. Quindi, questo approccio consente la progettazione di sistemi HMI più efficaci e una valutazione soggettiva più attendibile e precisa in quanto basata su dati reali.

Risulta interessante l'applicazione dell'approccio proposto ai sistemi di supporto alla guida. L'analisi si concentra sulla valutazione HMI di due funzioni del sistema Driver Support: Adaptive Cruise Control ACC e Pilot Assist PA. L'ACC permette di mantenere una distanza di sicurezza dal veicolo precedente regolando automaticamente la velocità, mentre il PA aiuta a mantenere il veicolo in corsia e a adattare la velocità alle condizioni del traffico. Entrambe le funzioni utilizzano telecamere e radar per monitorare l'ambiente circostante e assistere il conducente. Si tratta di sistemi semi-autonomi, per cui il conducente può riprendere il controllo del veicolo in qualsiasi momento.

Nello specifico, sono stati condotti degli studi che hanno confermato la veridicità dell'ipotesi che si basa su quanto detto finora: "se i big data potessero essere utilizzati combinando diversi parametri dei dati, come la durata dell'attivazione, la distanza percorsa, la velocità del veicolo, i dati GPS, i dati dei sensori delle condizioni stradali, ecc., avremo una migliore comprensione di come gli utenti utilizzano il sistema DS e quali problemi di usabilità incontrano". (Fonte *Orlovska, J., Wickman, C., & Söderberg, R. (2018). Big data usage can be a solution for user behavior evaluation: An automotive industry example. Procedia CIRP, 72, 117-122*)

"I risultati del prototipo hanno mostrato che, con l'aiuto dei *big data*, è possibile suddividere gli utenti in base ai parametri scelti, come l'uso della funzione durante le ore diurne/notturne o nelle stagioni estive/invernali, la gamma di velocità e altri fattori. Ciò aiuta a classificare i conducenti in diversi gruppi di utenti per ulteriori indagini. Gli esempi mostrati sono semplici, ma la combinazione di diversi parametri insieme può fornire risultati più avanzati riguardo alla delineazione dei diversi gruppi di utenti. La valutazione soggettiva, basata su una conoscenza particolare di un gruppo specifico di utenti, porta a risultati più accurati nella scoperta delle cause dei problemi di usabilità identificati." "L'analisi del comportamento degli utenti nell'ambiente reale, con l'aiuto dell'analisi dei *big data*, sta diventando sempre più dinamica e utile per i team di ingegneria automobilistica. Ad esempio, analizzando il comportamento degli utenti per le funzioni semi o completamente automatizzate incorporate nei veicoli (PA, ACC, parcheggio automatico, ecc.), possiamo scoprire il livello di apprezzamento e accettazione delle nuove funzionalità introdotte. La conoscenza riguardante il livello di fiducia verso queste funzioni corrisponde a quanto tempo l'utente necessita per adattarsi e fidarsi della nuova tecnologia. Questo può migliorare significativamente la pianificazione strategica riguardante la valutazione del livello di prontezza tecnologica e può contribuire alle implicazioni manageriali riguardanti l'adozione di nuove funzionalità" (Fonte *Orlovska, J., Wickman, C., & Söderberg, R. (2018). Big data usage can be a solution for user behavior evaluation: An automotive industry example. Procedia CIRP, 72, 117-122*)

Di notevole interesse, nello studio presentato, è l'uso del sistema WICE. È l'acronimo di *Web Intelligence and Collaborative Environment*: piattaforma progettata per l'analisi di big data focalizzata sui vari aspetti della gestione e dell'elaborazione dati provenienti da diverse fonti.

Ecco alcune caratteristiche principali del sistema WICE:

1. **Analisi dei dati:** consente di analizzare grandi volumi di dati in tempo reale, fornendo informazioni utili per decisioni strategiche.
2. **Visualizzazione:** offre strumenti per visualizzare i dati in modi intuitivi, come grafici e dashboard, rendendo più facile l'interpretazione delle informazioni.
3. **Collaborazione:** supporta la collaborazione tra diversi utenti, permettendo di condividere dati, report e analisi in modo efficiente.
4. **Integrazione:** può integrare dati provenienti da diverse fonti, come database, servizi web e applicazioni, facilitando un'analisi più completa e approfondita.
5. **Scalabilità:** il sistema è progettato per scalare facilmente, consentendo di gestire volumi di dati sempre crescenti senza compromettere le prestazioni.

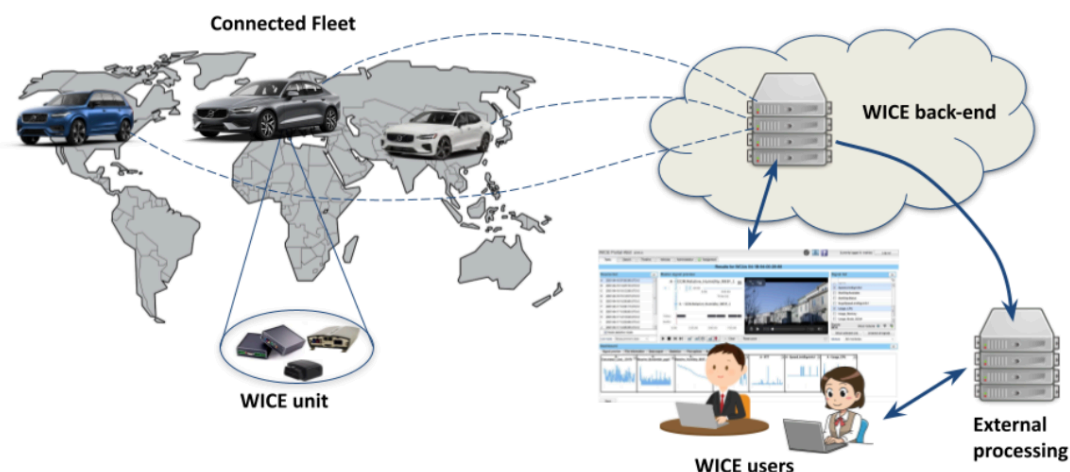
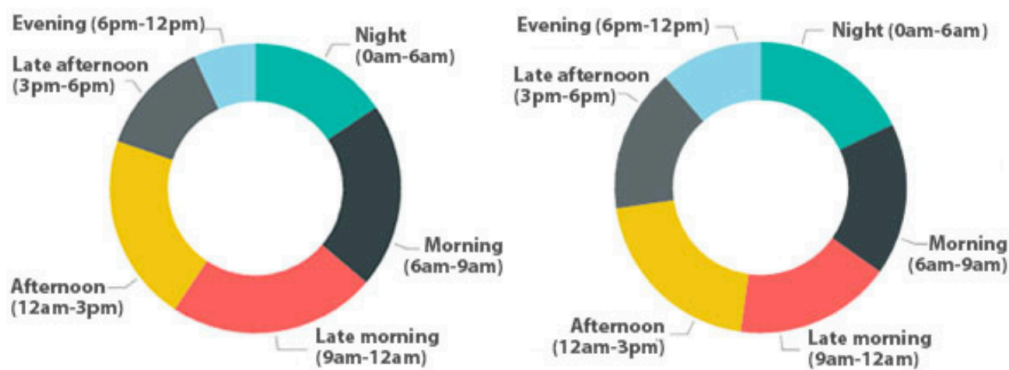


Figura 14– Fonte <http://www.alkit.se/index.php/front-page/industry/industry-projects/wice/>

Quindi, questo sistema consente di raccogliere e analizzare dati da veicoli connessi, supportando funzionalità come la gestione delle flotte o la prototipazione rapida di nuove funzioni per le

centraline elettroniche (ECU) e il download remoto del software(Alkit) In dettaglio, WICE è composto da due parti principali (Figura 14): un'unità di comunicazione wireless (WCU) installata nel veicolo per la registrazione e il monitoraggio dei dati che offre quindi il servizio di cattura di dati dai sensori e la gestione dello stato dei veicoli; e un'infrastruttura di server di back-end che gestisce lo stoccaggio dei dati tra cui segnali diagnostici , posizionamento basato su mappe e chilometraggio, e gestisce anche le interfacce utente.

Lo studio di analisi che ha confermato il miglioramento della valutazione in campo HMI attraverso l'analisi dei big data è stato condotto in due fasi: nella Fase 1 sono state svolte interviste con ingegneri UI/UX per individuare 18 domande di difficile valutazione con i metodi tradizionali di survey, come l'uso delle funzionalità da parte degli utenti ; mentre nella fase 2 ingegneri e analisti di dati hanno identificato i tipi di dati necessari per rispondere a queste domande ottenendo come risultato la progettazione di un caso di studio basato sui data driven.



Usage of ACC (left) and PA (right) during the hours of the day

Figura 15– Fonte Orlovska, J., Wickman, C., & Söderberg, R. (2018). Big data usage can be a solution for user behavior evaluation: An automotive industry example. *Procedia CIRP*, 72, 117-122

I risultati hanno mostrato che il sistema ACC (Adaptive Cruise Control) è usato molto più spesso rispetto al PA (Pilot Assist); in dettaglio tra tutte le attivazioni del sistema DS, il 94,72% riguarda l'ACC e solo il 5,28% il PA. Questo suggerisce che molti conducenti evitano di delegare il controllo dello sterzo, probabilmente per mancanza di fiducia nel sistema o perché il PA non funziona in modo ottimale. Per escludere quest'ultima ipotesi occorre analizzare le prestazioni del PA per escludere problemi di qualità. Lo studio ha anche esaminato se l'uso del sistema dipende da fattori

come le condizioni meteorologiche o l'orario, scoprendo che l'uso è indipendente dall'ora del giorno, ma è necessario verificare la qualità dei sensori in condizioni notturne.

Ne segue che, l'analisi dei big data offre strumenti efficaci per migliorare la valutazione delle interfacce uomo-macchina (HMI) in diversi modi:

- ✓ **Gestione della complessità:** Le interfacce complesse sono difficili da comprendere e valutare poiché di solito, con l'approccio di survey tradizionale, si analizzano solo parti specifiche dell'interfaccia, risultando insufficiente per identificare problemi di usabilità. I big data permettono un'analisi più completa e integrata, rendendo la valutazione più accurata.
- ✓ **Interazione tra comportamento utente e prestazioni del sistema:** Analizzando insieme il comportamento dell'utente e le prestazioni del sistema, i big data consentono di comprendere come questi due elementi si influenzino l'un l'altro. In condizioni normali, gli utenti non possono facilmente immaginare come le variazioni di performance del sistema influenzerebbero il loro comportamento, ma l'analisi dei dati permette di simulare scenari e rispondere a queste domande.
- ✓ **Feedback più realistico sulle prestazioni degli utenti:** In ambienti reali, gli utenti si comportano diversamente rispetto a un laboratorio controllato. L'analisi dei big data consente di catturare questi comportamenti realistici, migliorando la qualità dei feedback rispetto ai test tradizionali.
- ✓ **Diversità geografica:** I big data permettono di raccogliere informazioni da un pubblico globalmente distribuito senza le limitazioni fisiche. Questo è particolarmente vantaggioso per le aziende che operano su mercati internazionali, poiché consente di confrontare l'esperienza utente su diverse regioni geografiche.
- ✓ **Scalabilità:** L'analisi dei big data consente di testare un numero praticamente illimitato di utenti, senza un aumento significativo dei costi o del tempo necessario per la valutazione. Con l'aumento del numero di utenti, l'efficienza economica di ciascun test migliora.

## 2.5. *BIG DATA NEL GREEN CAMPUS*

Quando si parla di economia circolare ci si riferisce al modello economico che mira a ridurre al minimo lo spreco e a massimizzare l'uso efficiente delle risorse, attraverso la progettazione di cicli chiusi in cui i prodotti, e le relative risorse, vengono mantenuti in uso il più a lungo possibile. Questo sistema si oppone al tradizionale modello "lineare" che si configura in produzione-consumo-scarto, promuovendo invece il riutilizzo, la riparazione, il riciclo e la rigenerazione dei materiali e dei prodotti.

In particolare, in un'economia circolare, i prodotti vengono progettati per essere facilmente riparati o riciclati alla fine della loro vita utile, con l'obiettivo di minimizzare l'impatto ambientale. Si tratta di un approccio che coinvolge tutto il ciclo di vita del prodotto, dalla progettazione alla produzione, fino alla gestione del suo smaltimento.

I principi su cui si basa l'economia circolare sono i seguenti:

1. Progettazione sostenibile: prodotti pensati per durare più a lungo, essere riparati e riciclati.
2. Riduzione degli sprechi: minimizzazione degli scarti attraverso il riutilizzo dei materiali.
3. Conservazione del valore: mantenere il più possibile il valore dei materiali, componenti e prodotti nel ciclo economico.
4. Uso di risorse rinnovabili: preferenza per l'uso di risorse che si rigenerano naturalmente.
5. Condivisione e rigenerazione: promuovere modelli di business basati sulla condivisione (come il noleggio e il leasing) e sulla rigenerazione di prodotti e materiali.

Questo modello risulta interessante poiché non solo contribuisce alla protezione ambientale, ma è in grado di generare opportunità economiche, riducendo i costi operativi e creando nuovi mercati e relativi posti di lavoro.

Il ruolo della manutenzione nella gestione della produzione risulta centrale nel campo dell'economia circolare permettendo di perseguire l'obiettivo di progettazione sostenibile, riduzione degli sprechi e conservazione del valore. Ecco che qui entra in gioco il ruolo dell'Internet delle Cose (IoT) come fattore abilitante per migliorare la sostenibilità della produzione.

Attualmente, l'interconnessione delle imprese industriali e manifatturiere viene sempre più realizzata tramite l'uso di sistemi di raccolta dati basati su cloud e analisi avanzate potenziate dall'IoT, con tecnologie come l'intelligenza artificiale e il machine learning che forniscono previsioni sempre più accurate. In tale ambito, il settore automobilistico ha fatto notevoli progressi nell'adozione di approcci di economia circolare, concentrandosi sull'analisi del ciclo di vita completo dei componenti e una progettazione orientata alla riutilizzabilità e rigenerazione dei prodotti.



In particolare, negli ultimi anni, il ruolo della manutenzione predittiva è diventato sempre più centrale con l'adozione del 5G, che ha migliorato la connettività e la capacità di gestire grandi volumi di dati in tempo reale, permettendo una gestione degli asset più agile e sostenibile.

Tra le metodologie su cui si basa una manutenzione sostenibile, si individua la mappatura del flusso del valore che permette di combinare misure di sostenibilità con altre pratiche come il Lean manufacturing, al fine di identificare e implementare azioni di miglioramento nelle attività di manutenzione. Tali approcci sono sempre più integrati anche nell'industria 4.0, che sfrutta dati in tempo reale, intelligenza artificiale e IoT per monitorare e ottimizzare le operazioni di manutenzione. Il tutto permette di prevenire guasti in maniera più efficace, riducendo tempi di inattività e sprechi di risorse.

Di applicazione in campo automotive, accanto alla manutenzione predittiva, vi è una notevole riduzione dell'impatto ambientale grazie all'estensione del ciclo di vita dei prodotti attraverso la rigenerazione. Le opzioni per il trattamento dei componenti a fine vita includono riuso, rigenerazione, riciclaggio e smaltimento. Particolarmente interessante è la rigenerazione in quanto offre opportunità di circolarità. Il fattore chiave per facilitare la rigenerazione è la logistica inversa: processo mediante il quale i componenti usati vengono recuperati e riportati lungo la catena di approvvigionamento, permettendo quindi al sistema di trasportare i prodotti o parti di essi dal consumatore finale al produttore o a centri di rigenerazione per essere riparati, riciclati o smaltiti in modo appropriato. Nel settore automobilistico, questa catena di approvvigionamento è spesso complessa ed inoltre si evidenziano alcune limitazioni nella rigenerazione di componenti elettronici critici per la sicurezza. L'uso di metodologie Lean nella rigenerazione può affrontare sfide come la mancanza di informazioni sui componenti da rigenerare e l'assenza di processi standardizzati per lo smontaggio e la rigenerazione, ma non è abbastanza. Per ovviare a tali problematiche si parla di progettazione per lo smontaggio (Design for Disassembly, DfD) e di progettazione per la manutenzione (Design for Maintenance, DfM), strettamente legati alla rigenerazione. Infatti, i prodotti progettati fin dall'inizio per essere smontati possono essere rigenerati a costi complessivi inferiori. L'obiettivo della manutenzione, orientata al valore, è quello di massimizzare il valore generato dalle operazioni, anziché limitarsi a minimizzare i costi. Ciò comprende vari tipi di manutenzione, come quella correttiva, preventiva e basata sulle condizioni (condition-based maintenance, CBM), evitando interventi solo quando un componente è già danneggiato. È evidente come questo approccio si allinei ai principi dell'economia circolare.

A oggi, la transizione verso un'economia circolare e l'integrazione di soluzioni DfD e DfM continua a crescere e in tale ottica di sviluppo, un ruolo fondamentale può essere svolto dai Big Data Driven.

La sfida in tale campo è integrare le fonti di dati in streaming con quelle statistiche al fine di definire un insieme di parametri sufficienti a descrivere un prodotto nell’ottica della circolarità lungo il suo ciclo di vita.

A tal proposito viene riportato un framework basato su IoT data, che rappresenta i principali flussi di dati e le categorie di parametri presenti nei veicoli moderni, raggruppati per i loro componenti principali. In Figura 16, si individuano elementi diversi al variare del colore. Gli ovali verdi indicano le categorie di parametri generati dai sensori e dalle funzioni del veicolo che processano i dati acquisiti. I rombi gialli rappresentano azioni teoriche sui flussi di dati. Ed infine, I rettangoli blu identificano il gruppo di appartenenza, oltre ai casi d'uso legati ai flussi di dati elaborati.

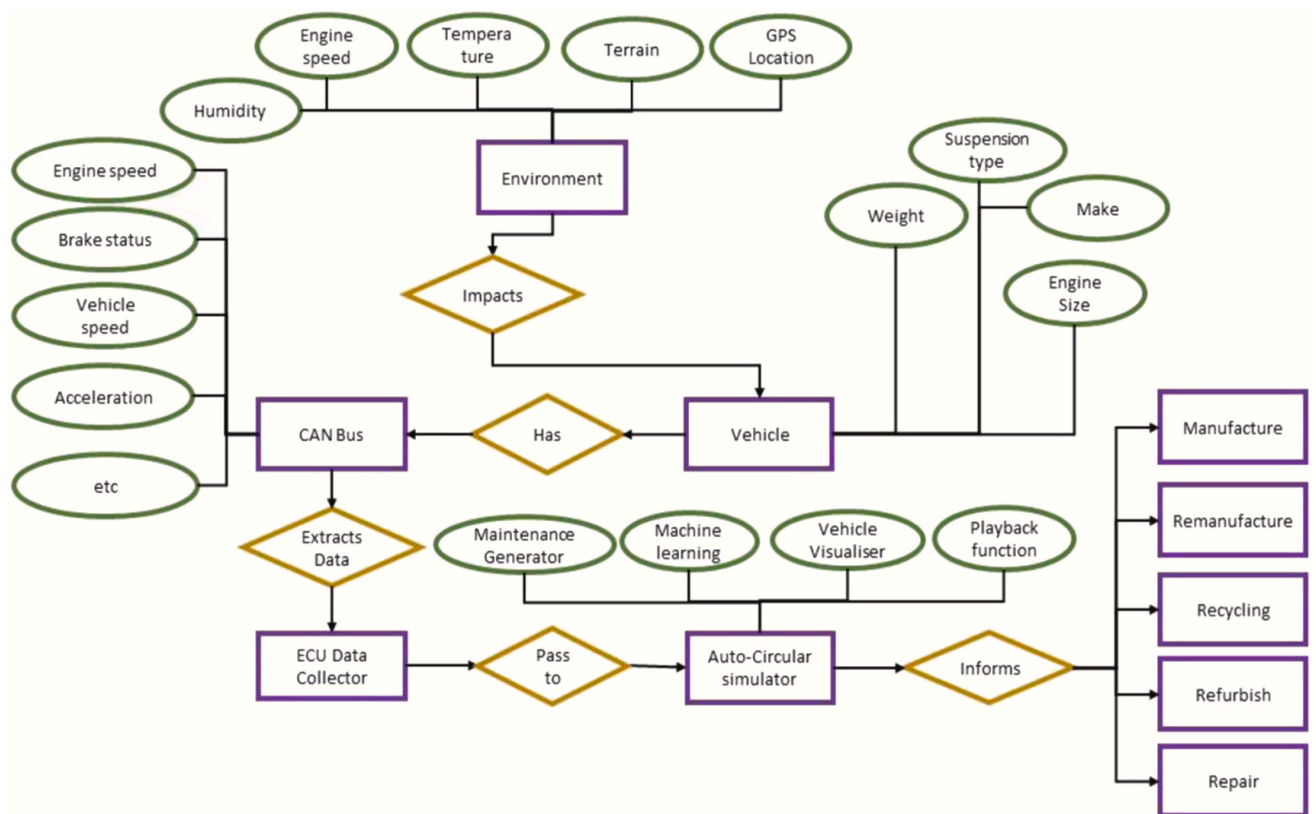


Figura 16– Fonte Turner, C., Okorie, O., Emmanouilidis, C., & Oyekan, J. (2022). Circular production and maintenance of automotive parts: An Internet of Things (IoT) data framework and practice review. *Computers in Industry*, 136, 103593.

In particolare, durante il funzionamento, il quadro permetterà una visualizzazione completa dei complessi flussi di dati provenienti dai singoli componenti e dai gruppi di componenti del veicolo. I dati raccolti dai sensori generali del veicolo, ad esempio sia quelli esterni che rilevano fattori ambientali come temperatura e terreno, che quelli propri del veicolo quale accelerazione, rpm engine e velocità veicolo, potranno essere trasmessi mediante CAN Bus, che come accennato nei precedenti capitoli, è la robusta rete decentralizzata che consente la comunicazione tra i microcontrollori del veicolo. Le unità di controllo elettroniche (ECU) del veicolo offrono un'interfaccia che gestisce i flussi di dati dei sensori collegati tramite il CAN Bus. Tipico del framework proposto è il "Auto-circular simulator", il cui compito è quello di analizzare e fornire flussi di dati elaborati, e scenari ipotetici destinati a supportare i sistemi di manutenzione circolare, rigenerazione e riciclaggio. Infatti, tra le azioni teoriche sui flussi individuate nel quadro descritto vi è l'opzione "Informs" che indica il supporto informativo alle varie aree tecniche di manutenzione circolare tramite interfaccia costituita da una dashboard dati che offre visualizzazioni grafiche tradizionali e strumenti di data mining, oltre a viste grafiche di simulazione che permettono di testare scenari ipotetici tramite tecniche di apprendimento automatico

Gli obiettivi finali del quadro descritto si delineano tra gli output del simulatore Auto-circular:

- ⇒ Schemi specifici per i vari tipi di veicoli, mediante mappe visive che rappresentano i flussi di dati dei sensori per ciascun veicolo.
- ⇒ Scenari che illustrano la configurazione ottimale delle catene di approvvigionamento, con l'obiettivo di promuovere la circolarità nelle operazioni di manutenzione e nel trattamento dei componenti automobilistici usurati.
- ⇒ Diagrammi di flusso che descrivono dettagliatamente le attività di manutenzione, basati sui dati più aggiornati relativi a interventi precedenti.
- ⇒ Riassunti tabellari dei flussi di dati, con la possibilità di creare rappresentazioni grafiche personalizzate.

### 2.5.1. HYDROGEN FUEL CELL VEHICLE, HFCV

Un caso applicativo di quanto detto finora permette di individuare il modo in cui i dati possono essere ricavati dai componenti automobilistici e successivamente impiegati nel processo decisionale automatizzato per la manutenzione e la gestione circolare dei componenti danneggiati o usurati. È importante sottolineare che per agevolare la raccolta dei dati necessari occorre identificare uno specifico set di domande che verrebbero poste da specifici attori all'interno della catena del valore della produzione automobilistica sostenibile. Il caso di studio riportato nel seguente elaborato riguarda il veicolo a celle a combustibile a idrogeno (*Hydrogen Fuel Cell Vehicle*, HFCV).

La cella a combustibile a idrogeno rappresenta un'innovazione tecnologica che impiega l' $H_2$  come carburante per generare elettricità, destinata ad alimentare un motore elettrico, come avviene nei veicoli. Il suo principio di funzionamento si basa su un processo opposto all'elettrolisi: l'idrogeno si combina con l'ossigeno, producendo così energia elettrica e acqua come sottoprodotto.

In dettaglio, il veicolo HFCV è dotato di serbatoi in cui viene stoccato idrogeno ad alta pressione che sarà utilizzato come combustibile per alimentare il sistema a celle a combustibile. Si compone di vari strati, tra cui due elettrodi (anodo e catodo) e un elettrolita che li separa. La cella a combustibile non brucia l'idrogeno ma lo converte in elettricità attraverso una reazione elettrochimica che vede il susseguirsi dei seguenti step.

- I. L'idrogeno  $H_2$  entra nella cella a combustibile attraverso l'anodo dove viene scisso in ioni ( $H^+$ ) ed elettroni ( $e^-$ ) attraverso l'azione di un catalizzatore quale il platino.
- II. La membrana centrale, l'elettrolita, consente il passaggio dei protoni ( $H^+$ ) dall'anodo al catodo, bloccando gli elettroni
- III. Gli elettroni ottenuti dalla scissione dell'idrogeno vengono indirizzati attraverso un circuito esterno, creando un flusso di corrente elettrica che alimenta il motore elettrico e altre componenti elettroniche del veicolo.
- IV. Al catodo, i protoni ( $H^+$ ) che sono passati attraverso l'elettrolita si combinano con l'ossigeno ( $O_2$ ) presente nell'aria e con gli elettroni che arrivano dal circuito esterno, formando acqua ( $H_2O$ ) come prodotto di scarto

Quindi, l'elettricità prodotta dalla cella a combustibile alimenta il motore elettrico del veicolo, che fornisce la trazione necessaria al veicolo. Il più delle volte il veicolo HFCV è dotato di una batteria agli ioni di litio e di super condensatori che aiutano ad immagazzinare l'energia prodotta in eccesso dalla cella a combustibile o a recuperare l'energia durante la frenata (frenata rigenerativa). Infatti, sono utilizzati per superare le limitazioni prestazionali causate dal tempo di risposta ritardato delle celle a combustibile e fornire una potenza extra per accelerazioni rapide.

Analizzando il processo di funzionamento risulta evidente come tra i vantaggi vi sia quello di nessuna emissione nociva: il processo elettrochimico della cella a combustibile emette solo vapore acqueo (H<sub>2</sub>O), rendendo i veicoli HFCV estremamente rispettosi dell'ambiente. Si tratta quindi di veicoli capaci di rifornirsi in pochi minuti anziché richiedere lunghe ore per la ricarica, come le attuali auto elettriche. Nonostante queste potenziali virtù, il mercato sembra mostrare segni di esitazione. Le vendite globali di veicoli a celle di combustibile a idrogeno rimangono marginali rispetto a quelle delle auto elettriche: 14.000 unità vendute nel mondo nel 2023, contro i 10 milioni di veicoli elettrici a batteria e circa 4 milioni di ibridi plug-in. Infatti, la diffusione di questa tipologia di veicoli si trova ancora in una fase iniziale a causa di diverse difficoltà:

- Ritardo Tecnologico nelle Infrastrutture di Supporto: le stazioni di rifornimento per l'idrogeno sono cruciali per la diffusione di questi veicoli, ma rimangono insufficienti.
- Costi Elevati: i veicoli a idrogeno, come il Toyota Mirai 2024, hanno un prezzo di vendita che può raggiungere i \$50.000, superiore a quello delle auto a benzina o elettriche comparabili. Sebbene alcuni paesi come la Corea del Sud offrano incentivi significativi, il costo rimane un fattore limitante.
- Scarsità di Idrogeno Pulito: la maggior parte dell'idrogeno prodotto oggi deriva da combustibili fossili, rendendo la stessa produzione di idrogeno responsabile di inquinamento climatico. Metodi a bassa emissione, come l'elettrolisi alimentata da fonti rinnovabili, sono in sviluppo, ma l'idrogeno pulito rimane in quantità limitata.

Osservando gli investimenti in ricerca e sviluppo per valutare l'impegno delle aziende verso una mobilità sostenibile basata sulla tecnologia ad idrogeno, emerge la maggiore società automobilistica del Giappone e del mondo, Toyota Motor Corporation. Quest'ultima, sta investendo notevolmente in tale tecnologia, sia per le celle a combustibile che per i motori a combustione. Il particolare interesse è rivolto ai veicoli di grandi dimensioni, dove i tempi di rifornimento rapidi sono essenziali. In dettaglio, Toyota sta promuovendo una strategia per aumentare le vendite e promuovere l'adozione di veicoli a idrogeno: ha ridotto significativamente il prezzo della Mirai, la sua auto a celle a combustibile a idrogeno, con uno sconto del 60% negli Stati Uniti, portandola da circa \$60.000 a \$20.000. La riduzione è anche legata agli incentivi ambientali, come i crediti d'imposta previsti dall'*Inflation Reduction Act* che possono ulteriormente ridurre il costo per i consumatori.

Analogamente, BMW prevede di introdurre il suo primo veicolo elettrico a celle a combustibile di serie nel 2028, offrendo ai clienti una nuova opzione di propulsione elettrica a zero emissioni. Questo progetto è frutto di una collaborazione con Toyota Motor Corporation. Quindi, entrambe le aziende condividono l'obiettivo di far progredire l'economia dell'idrogeno, accelerando lo sviluppo di sistemi di propulsione a celle a combustibile di nuova generazione.

Il sistema di propulsione sarà sviluppato congiuntamente e impiegato sia nei veicoli commerciali che in quelli utilitari, permettendo a BMW e Toyota di offrire modelli HFCV distinti ma tecnologicamente avanzati. I costi di sviluppo e approvvigionamento saranno ridotti grazie alle sinergie tra le due aziende. Il primo modello HFCV di BMW sarà basato su un veicolo esistente, ma con una variante a celle a combustibile, e sarà complementare alle altre tecnologie elettriche dell'azienda.

BMW e Toyota, sostengono l'espansione delle infrastrutture di rifornimento di idrogeno, lavorando con governi e investitori per creare una strada spianata alla diffusione di questa tecnologia.

Nel quadro complessivo descritto, un posto rilevante è rappresentato anche dalla holding multinazionale Stellantis N.V.

Stellantis ha sviluppato la soluzione basata su celle a combustibile a idrogeno per i veicoli commerciali leggeri (LCV) che necessitano di un'elevata autonomia, e tempi di rifornimento rapidi e zero emissioni senza comprometterne la capacità di carico.

Le architetture dei veicoli elettrici a celle a combustibile possono essere suddivise in due configurazioni: "Full-Power" e "Range-Extender". Nella configurazione Full-Power la cella a combustibile funge da principale fonte di propulsione in tutte le condizioni operative. Questo approccio richiede una cella a combustibile di grande potenza e dimensioni, insieme a una batteria di piccole dimensioni. Invece, la configurazione Range-Extender prevede una batteria di grandi dimensioni che viene supportata da una piccola cella a combustibile a bassa potenza, la cui unica funzione è quella di estendere l'autonomia del veicolo fornendo energia alla batteria. Tuttavia, in quest'ultima configurazione, la cella a combustibile non è in grado di generare energia sufficiente per alimentare il veicolo quando la batteria è completamente scarica.

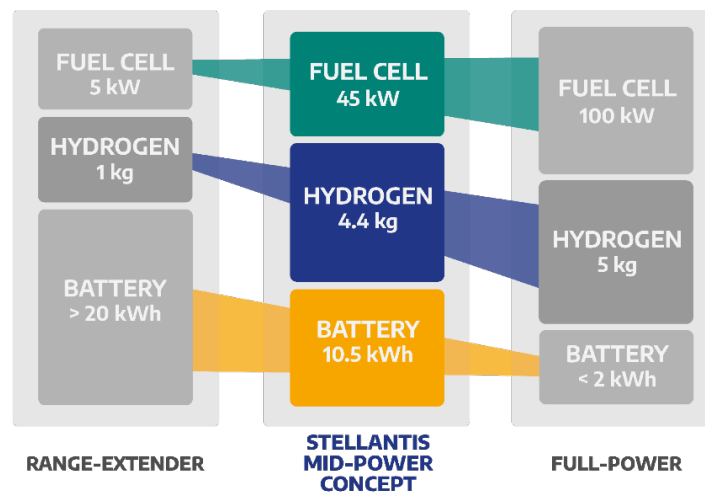


Figura 17 <https://www.stellantis.com/it/tecnologie/tecnologia-a-celle-a-combustibile-a-idrogeno>

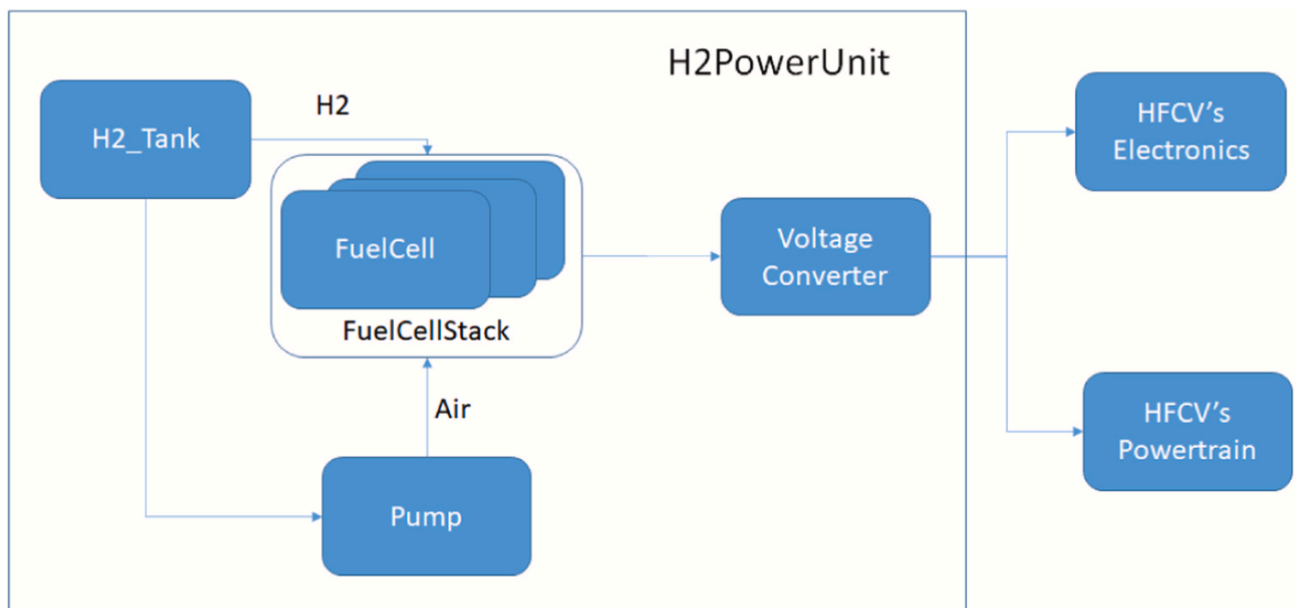
Tra le due soluzioni estreme, per rispondere alle esigenze dei clienti, la holding multinazionale ha progettato un'innovativa architettura mid-power che si pone nell'intermezzo tra le due. Si contraddistingue in quanto riesce a garantire un'autonomia record nel segmento, in dettaglio fino a 400 km per i furgoni di medie dimensioni. Invece, nei furgoni di grandi dimensioni, consente di raggiungere un'autonomia di 500 km con tempi di rifornimento ridotti a soli 5 minuti. Tutti i componenti del sistema di propulsione a celle a combustibile sono collocati al di fuori del vano di carico al fine di preservare la capacità di carico. Inoltre, il sistema può essere facilmente integrato con la piattaforma elettrica a batteria per LCV, ottenendo sinergia tra le versioni elettriche a batteria e quelle a idrogeno. Stellantis ha iniziato la produzione di veicoli a celle a combustibile a idrogeno, inclusi furgoni di medie e grandi dimensioni, presso i suoi stabilimenti di Hordain (Francia) e Gliwice (Polonia). La gamma di Stellantis comprenderà otto modelli di furgoni a idrogeno: Citroën ë-Jumpy e ë-Jumper, Fiat Professional E-Scudo e E-Ducato, Opel/Vauxhall Vivaro e Movano, nonché Peugeot E-Expert e E-Boxer.

Nel 2023, Symbio, una joint venture paritaria tra Forvia, Michelin e Stellantis, ha aperto la sua prima gigafactory in Francia. Questo rappresenta il più grande impianto di produzione di celle a combustibile integrate in Europa, permettendo a Stellantis di ampliare ulteriormente la propria offerta di veicoli a idrogeno.

Quindi, ad oggi, l'idrogeno viene visto come un vettore energetico fondamentale per la decarbonizzazione, soprattutto dove le batterie non rappresentano la soluzione ottimale.

Dopo questa introduzione sugli *Hydrogen Fuel Cell Vehicle* e il relativo quadro attuale di mercato, l'elaborato riporta in seguito, un caso di studio applicativo sulla tipologia di veicoli HFCV in campo di manutenzione circolare e Big Data Driven. (Fonte Turner, C., Okorie, O., Emmanouilidis, C., & Oyekan, J. (2022). *Circular production and maintenance of automotive parts: An Internet of Things (IoT) data framework and practice review. Computers in Industry, 136, 103593.*)

Il caso di studio sull'Hydrogen Fuel Cell Electric Vehicle viene esaminato in termini della sua unità di potenza: H2PowerUnit. L'H2PowerUnit (Figura 18) fornisce la fonte di movimento per l'HFCV ed è composta dalle seguenti sotto-unità: *H2\_Tank*, che contiene la fonte di carburante a idrogeno; *VoltageConverter*, che trasforma la corrente dalla FuelCellStack alla tensione richiesta; *Pump*, utilizzata per pressurizzare le celle a combustibile nella FuelCellStack; *FuelCell*, che fornisce la corrente elettrica in uscita basata sulle reazioni chimiche della fonte di carburante a idrogeno; **FuelCellStack**, composta da 1 a 10 celle a combustibile individuali. In questo studio di caso, i flussi di dati provenienti dall'entità H2PowerUnit forniscono la maggior parte dei parametri: con l'integrazione di tecnologie **IoT** e **5G**, i sensori montati sull'H2PowerUnit sono ora in grado di trasmettere dati in tempo reale a server remoti, migliorando le capacità di monitoraggio e manutenzione predittiva del sistema. Esiste comunque la possibilità che informazioni provenienti da applicazioni e altri repository di documenti elettronici forniscano parametri aggiuntivi.

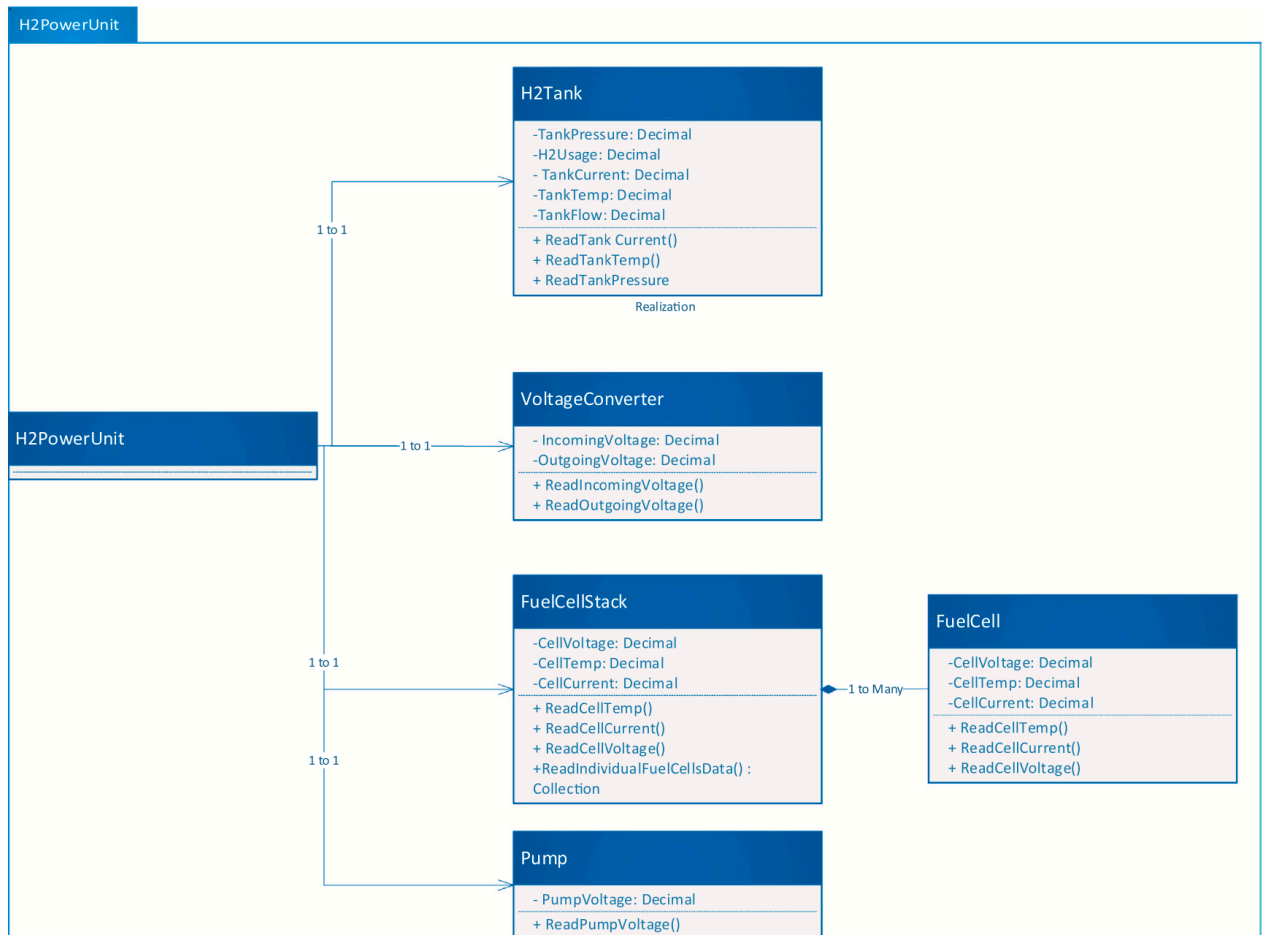


. Schematic of the case study HFCV power unit (H2PowerUnit).

Figura 18– Fonte Turner, C., Okorie, O., Emmanouilidis, C., & Oyekan, J. (2022). *Circular production and maintenance of automotive parts: An Internet of Things (IoT) data framework and practice review. Computers in Industry, 136, 103593.*



Quindi, l'architettura di un'unità di alimentazione a idrogeno per veicoli, rappresentata in Figura 19 dal diagramma Unified Modelling Language UML, è potenziata attraverso l'integrazione di sensori IoT consentendo un monitoraggio continuo e in tempo reale dei vari parametri operativi del sistema, migliorando la sicurezza, l'efficienza e la manutenzione predittiva.



. UML (Unified Modelling Language) diagram for hydrogen car power unit.

Figura 19– Fonte Turner, C., Okorie, O., Emmanouilidis, C., & Oyekan, J. (2022). Circular production and maintenance of automotive parts: An Internet of Things (IoT) data framework and practice review. Computers in Industry, 136, 103593.

### *H2PowerUnit con sensori IoT*

Unità centrale che gestisce il flusso di potenza nel veicolo a idrogeno. È monitorata attraverso sensori IoT collegati ai diversi componenti che raccolgono dati da ciascuna parte del sistema e li trasmettono a una piattaforma di controllo centralizzata, dove i dati sono analizzati in tempo reale.

- H2Tank

Elemento critico dell'unità in quanto immagazzina l'idrogeno a pressione elevata. In dettaglio si monitorano i seguenti parametri con lo scopo primario di prevenire malfunzionamenti:

Pressione del serbatoio: sensori di pressione IoT possono monitorare la pressione interna in tempo reale e inviare avvisi nel caso di variazioni anomale.

Temperatura del serbatoio: sensori termici collegati al serbatoio invierebbero informazioni sulla temperatura dell'idrogeno, fondamentale per evitare situazioni di rischio.

Flusso dell'idrogeno: sensori di flusso IoT potrebbero tracciare l'uso dell'idrogeno, ottimizzando la distribuzione e il consumo.

Sensori di corrente: rilevano il flusso di corrente nel serbatoio, indicando eventuali anomalie nella conversione o trasmissione dell'energia.

- Voltage Converter

È responsabile della trasformazione della tensione in ingresso e in uscita dal sistema:

Tensione in ingresso e in uscita: sensori di tensione IoT rilevano eventuali discrepanze o inefficienze nella conversione.

Monitoraggio delle variazioni di tensione: i dati possono essere usati per adattare dinamicamente il funzionamento del sistema in base alle necessità del veicolo e alle condizioni operative.

- Fuel Cell Stack

Serie di celle a combustibile che generano energia elettrica dalla reazione tra l'idrogeno e l'ossigeno. Si rileva:

Tensione delle celle: sensori rilevano la tensione generata da ciascuna cella, fornendo dati su come ogni singola cella sta contribuendo al sistema.

Temperatura delle celle: sensori di temperatura monitorano il surriscaldamento o il raffreddamento delle celle, ottimizzando così il funzionamento e prevenendo guasti.

Corrente delle celle: i sensori di corrente monitorano il flusso elettrico per prevenire squilibri tra le celle.

Questi dati possono essere raccolti e analizzati per identificare celle malfunzionanti o ridurre il consumo energetico, migliorando l'efficienza complessiva.

- Fuel Cell

La singola cella nella pila contiene sensori che forniscono informazioni dettagliate su:

Tensione, temperatura e corrente di una singola cella, permettendo un monitoraggio granulare e quindi l'identificazione immediata di anomalie o malfunzionamenti.

Più in generale, i dati raccolti a livello di singola cella vengono aggregati per avere una visione complessiva delle prestazioni della pila.

- Pump

Regola il flusso di idrogeno all'interno del sistema e di altri fluidi di raffreddamento, é equipaggiata con sensori IoT per monitorare:

Tensione della pompa: sensori IoT verificano la potenza necessaria alla pompa e ne ottimizzano il funzionamento.

Pressione del fluido: sensori di pressione controllano il corretto flusso dell'idrogeno per la reazione elettrochimica o di altri fluidi necessari per il raffreddamento.

Riepilogando, dal diagramma UML riportato, risultano chiare le relazioni tra i sotto-componenti di H2PowerUnit e i tipi di funzioni che richiederebbero flussi di dati forniti dai sensori. Si giunge così ad una combinazione di parametri identificati in grado di fornire un riepilogo affidabile dello stato attuale di un particolare componente automobilistico e non solo, ma anche degli stati precedentemente registrati.

Il flusso di dati che si ottiene permette di formulare specifici quesiti sulla catena di valore HFCV. La prima domanda riguarda il calo dei livelli di MPG (miles per gallon) del veicolo rilevati. Ecco che qui entrano in gioco i primi attori della catena del valore, i clienti, la cui anomalia viene visualizzata nelle MPG Dashboard. In seguito, entrano in gioco i produttori, che potrebbero voler rivedere il materiale utilizzato per produrre la Fuel Cell Stack: il materiale fondamentale è il platino che al contempo è molto costoso. Negli ultimi anni, sono stati introdotti nuovi catalizzatori a base di materiali più economici per sostituire parzialmente il platino, quali il palladio o i nanomateriali, con l'obiettivo di ridurre i costi e migliorare l'efficienza. Coloro che invece si occupano della rigenerazione del componente hanno l'intento di ritirare e riparare o rigenerare le celle a combustibile. Mentre, gli ingegneri di manutenzione analizzerebbero lo stato di salute del convertitore di tensione adattando di conseguenza le tempistiche di manutenzione. Inoltre, insieme all'equipaggio OEM, potrebbero scegliere di rivedere il sistema di produzione identificando nuovi parametri da monitorare e recuperare dal veicolo. (es. H2\_Tank, FuelCellStack).

Il rilevamento di vibrazioni ed emissioni acustiche nel veicolo invece, potrebbe innescare un'azione di manutenzione, richiedendo azioni immediate ai vari attori della catena del valore: i sensori IoT avanzati combinati con tecnologia 5G consentono il rilevamento immediato e la trasmissione in tempo reale di dati sulle vibrazioni ed emissioni acustiche, permettendo agli ingegneri di verificare i parametri relativi alla pompa, stato del serbatoio di idrogeno, delle tubazioni del sistema. Al contempo, i fornitori di componenti cercheranno di migliorare l'affidabilità della parte attraverso una riprogettazione del design basata su algoritmi di machine learning che analizzano i dati raccolti per identificare punti deboli. Mentre, nel campo della rigenerazione si potrebbe valutare la possibilità di riconfigurare le pompe difettose, prendendo in considerazione materiali più robusti o una diversa architettura, in collaborazione con i produttori OEM che nello specifico analizzeranno pompa, sottosistemi relative a pompa e serbatoio H2. Nel quadro di attori coinvolti rientrano pure gli enti governativi che si occuperanno della parte relativa alla sicurezza del sistema.

Le variazioni anomale di temperatura nel veicolo durante l'uso interesseranno i tecnici di manutenzione che individueranno probabili guasti nel convertitore di tensione, alla pompa o anomalie varie alla FuelCellStack o al Serbatoio H2. Contemporaneamente i fornitori di componenti analizzeranno i materiali usati nella FuelCellStack, e i produttori **OEM** nello specifico individueranno quali sottosistemi sono associati ai cambiamenti di temperatura.

Ulteriore caso di anomalia che è possibile rilevare ed analizzare riguarda l'aumento della condensazione nel tubo di alimentazione.

**Table 2**

Questions the Ontology Could be Used to Answer.

Value Chain Questions	Suppliers	Maintenance Engineers	Customers	OEM Manufacturer	Government Regulators	Re-manufacturer
1. Why has the Miles Per Gallon (MPG) measure dropped?	Quality of raw materials used to develop FuelCellStack e.g platinum	VoltageConverter,; Maintenance timings; H2_Tank; FuelCellStack.	MPG dashboard	Entire ontology and systems approach	Ethical sourcing issues Infrastructure issues Geo-political climate issues Driving up efficiency	Repair/Refurbish part under investigation
2. Why is there an increase of sound or vibration coming from the vehicle?	Reliability of parts Design and compactibility issues	Pump	Acoustic emissions	Pump	Reliability of parts	Pump reconditioning
3. Why is there an increase in condensation from tail pipe or in the vehicle?	Materials used for the H2_Tank and pipes	H2_Tank and pipe	Water condensations	sub-systems related to pump and H2_Tank	Safety of systems	H2-Tank reconditioning
4. Why is there an increase or drop in temperature in the vehicle?	Materials used in the FuelCellStack	FuelCellStack; VoltageConverter; Pump faults; H2_Tank.	Thermostat warning	sub-systems associated with temperature changes	Safety of systems	Repair/Refurbish

In generale, si segue il seguente flusso basato sui big data driven rilevati:

*I. Domande sulla catena del valore*

- Problemi che potrebbero verificarsi nei sistemi di celle a combustibile, come la riduzione di MPC, aumento di rumori o vibrazioni, aumento della condensazione o variazioni di temperatura.

*II. Coinvolgimento dei vari attori*

- Fornitori: si occupano principalmente della qualità delle materie prime e della compatibilità dei pezzi.
- Ingegneri di manutenzione: lavorano su componenti specifici come convertitori di tensione, serbatoi H2, pompe e celle a combustibile.
- Clienti: notano problemi come emissioni acustiche o variazioni termiche sui cruscotti.
- Produttori OEM: si concentrano sui sottosistemi e sulle relative problematiche rilevate.
- Enti di regolazione governativa: si preoccupano di questioni come la sicurezza e le implicazioni etiche.
- Specialista in prodotti ricondizionati: si occupano di riparazioni e sostituzioni di componenti difettosi.

Il fine ultimo della seguente analisi, nell'ottica di una perfetta economia circolare nel campo della rigenerazione, sarebbe quello di sfruttare i dati raccolti in tempo reale e la relativa ontologia sviluppata, ad esempio quella appena descritta, per alimentare modelli predittivi che riescano a stabilire il momento ottimale nel ciclo di vita di un componente per rigenerarlo o ripararlo, al fine di massimizzarne la vita utile. È da tenere presente però, che quando si considera la progettazione dei processi di manutenzione, l'intuizione e la conoscenza acquisita nel tempo giocano un ruolo cruciale, soprattutto se si parla di attività legate a prodotti automobilistici complessi. Al contempo, basandosi solo sulla conoscenza acquisita nel corso degli anni, le pratiche di manutenzione si evolvono lentamente nonostante vi siano costantemente nuovi processi che emergono velocemente con l'introduzione di nuovi modelli di veicoli. Tuttavia, grazie all'adozione crescente di tecnologie avanzate come l'intelligenza artificiale e i sistemi IoT, è possibile migliorare l'approccio tradizionale con processi di manutenzione più dinamici e adattivi. La riprogettazione dinamica dei processi basata su sensori, e osservazioni in tempo reale raccolte dai tecnici di manutenzione, rappresenta ora uno strumento fondamentale non solo per i tecnici esperti ma anche per i nuovi lavoratori che dovranno apprendere le pratiche di manutenzione. Questo approccio facilita una curva di apprendimento più rapida e aumenta l'efficienza operativa.

Applicando il concetto di manutenzione intelligenza al caso di studio descritto, è possibile individuare un numero aggiuntivo di parametri contestuali alle failures di specifiche parti dell'unità di potenza decritta: sensore di perdita di idrogeno, indicatore di messa a terra, livello di condensa dell'aria sulla valvola, misurazione nel serbatoio di idrogeno condizioni di guida reali, indicatore di guasto, storia delle parti, modello statistico delle condizioni di guida. Tali parametri contestuali aggiuntivi, se forniti ai vari attori della catena del valore, delineeranno delle attività di manutenzione predittiva al fine di supportare la gestione circolare del fine vita dei veicoli.

Expansion of HFCV power unit parameter set for intelligent maintenance purposes.

Activity	Power Unit entities	Parameters required	Additional Context Parameters Needed	Circularity Implications
Safely remove hydrogen fuel cell stack	FuelCellStack FuelCell	WithCurrent (hasVoltage exactly 1 xsd:decimal) WithTemp (hasTemperature exactly 1 xsd:decimal)	Hydrogen leak sensor Electrical grounding of Hydrogen Circuit H2PowerUnit Earthing Indicator	Regular check for leaks, grease, corrosion foreign objects and replace if necessary. Cells and stack can be recycled once replaced
Safely remove Hydrogen Fuel Tank	FuelTank	WithCurrent (hasVoltage exactly 1 xsd:decimal) WithPressure (hasPressure exactly 1 xsd:decimal) WithTemp (hasTemperature exactly 1 xsd:decimal)	Air Condensation level on Valve measurement in Hydrogen Tank	Replace if leak is detected. Fuel tank can be recycled or remanufactured if faulty.
Diagnose Fuel Cell Fault Warning	FuelCellStack FuelCell	WithCurrent (hasVoltage exactly 1 xsd:decimal) WithTemp (hasTemperature exactly 1 xsd:decimal)	Fault Indicator Part history Model of Average Driving conditions	Fault warning system should be regularly tested and replaced if bad. Replace and recycle

In sintesi, la tabella si concentra su come eseguire la manutenzione intelligente di un veicolo HFCV, basandosi sostanzialmente sul monitoraggio di parametri critici, e all'approccio sostenibile tramite il riciclo e la riparazione.

Dallo studio riportato, risulta evidente come alla base della manutenzione digitale sia importante considerare l'intero ciclo di vita dei componenti. Tale caso di studio può costituire l'elemento centrale in un generatore automatizzato di processi di manutenzione: auto-circular simulator. Quest'ultimo, parte integrante del quadro per la manutenzione circolare abilitata dall'IoT (Figura 16), è in grado di analizzare dati in tempo reale e offline provenienti dalle attività di manutenzione e dai veicoli in uso, grazie all'esistenza di uno strumento di comunicazione dinamico che raccoglierà dati sulle cause dei problemi di manutenzione nei veicoli, generando processi basati su dati dei sensori e codici di guasto. Ciò permetterà di monitorare il degrado e l'affidabilità dei componenti, fornendo dati per gestire il trattamento a fine vita in modo più efficace. Lo strumento genererà vari output, come schemi di veicoli, configurazioni ottimali delle catene di approvvigionamento e diagrammi di flusso per le attività di manutenzione.

### 2.5.2. STELLANTIS CIRCULAR ECONOMY: “SUSTAINera”

Nel seguente elaborato, tra le tematiche applicative dei big data driven affrontate, ho deciso di soffermarmi e riportare l'analisi di uno studio volto a costruire il quadro per la manutenzione circolare abilitata dall'IoT, in quanto ad oggi, in campo automotive, l'economia circolare è diventata una strategia imprescindibile. Ne è l'esempio la holding multinazionale Stellantis N.V. che ha deciso di puntare sulla circolarità, aprendo il primo Hub in Europa dedicato all'economia circolare dell'automotive. In particolare, Stellantis ha promosso il suo modello di Economia Circolare con una business unit dedicata, che si basa sul raggiungimento di due obiettivi: azzerare le emissioni di carbonio entro il 2038 e definire un nuovo modello di business “cradle to cradle” che genererà oltre 2 miliardi di euro di ricavi entro il 2030 nell'ambito del piano strategico Dare Forward 2030. In dettaglio, quest'ultima mira a quadruplicare i ricavi derivanti dall'estensione della vita utile di componenti e servizi e ad aumentare di 10 volte, rispetto al 2021, i ricavi ottenuti dal riciclo entro il 2030.

L'Azienda ha sviluppato un piano completo con un approccio a 360 gradi basato sulla strategia delle 4R: Reman (rigenerazione), Repair (riparazione), Reuse (riutilizzo) e Recycle (riciclo).

- Reman: i componenti usati, usurati o difettosi vengono smontati, puliti e rigenerati seguendo le specifiche OEM. Attualmente, sono disponibili quasi 12.000 componenti per 40 linee di prodotti, comprese batterie per veicoli elettrici.
- Repair: i componenti danneggiati vengono riparati e reinstallati nei veicoli dei clienti. I centri di e-repair, presenti in 21 sedi in tutto il mondo, si occupano principalmente della riparazione delle batterie dei veicoli elettrici.
- Reuse: circa 4,5 milioni di componenti multimarca in buone condizioni vengono recuperati da veicoli a fine vita e rivenduti in 155 Paesi attraverso la piattaforma di e-commerce B-Parts.
- Recycle: gli scarti di produzione e i veicoli a fine ciclo vita vengono reintrodotti nel processo produttivo. In soli sei mesi, la business unit dedicata ha riciclato un milione di componenti.

Con l'espansione delle attività legate all'economia circolare, la business unit ha lanciato l'identità SUSTAINera per ricambi e accessori. La Holding multinazionale ha deciso di investire in tale campo poiché, a seguito di un'analisi del ciclo di vita dei prodotti più venduti in ogni famiglia di segmento, si è confermato una riduzione fino all'80% dei materiali e fino al 50% del consumo energetico rispetto ai ricambi nuovi equivalenti.





Figura 20- Fonte <https://www.stellantis.com/it/news/eventi-recenti/apertura-ufficiale-del-primo-hub-di-economia-circolare-sustainera>

In Figura 20, è mostrata la struttura e l'organizzazione del primo Hub Europeo di Economia Circolare SUSTAINera, situato a Torino.

Si individua come primo step il settore del Reman, ovvero la rigenerazione dei componenti usati. Qui arrivano motori endotermici, cambi e batterie elettriche dalle officine certificate, che vengono smontati, puliti e rigenerati. I componenti rigenerati vengono rivenduti attraverso il circuito dell'after-sale, mentre i materiali non recuperabili finiscono in fonderia per essere riciclati.

Segue l'area Sorting, dove i pezzi provenienti dalle officine vengono selezionati e indirizzati verso i flussi di rigenerazione o di riciclo. Secondo le stime di Stellantis, entro il 2025 circa 2,5 milioni di componenti usati alimenteranno queste attività con l'obiettivo di trattare 8 milioni di pezzi entro il 2030.

L'hub include anche un'area di smontaggio dei veicoli con una capacità iniziale di 10.000 auto all'anno per la rivendita di parti, e un'area di ricondizionamento, dove vengono riparati fino a 66 veicoli al giorno per un totale di 15.000 all'anno. Al momento, si lavora su veicoli con un chilometraggio massimo di 30.000 km, tutti certificati Stellantis e pronti per il mercato tramite la piattaforma Spoticar.

Anche se non fisicamente presente ed individuabile nell'organizzazione del seguente Hub, un aspetto centrale dell'approccio circolare di Stellantis è l'ecodesign su cui la multinazionale lavora attraverso un costante flusso di feedback. In particolare, durante lo smontaggio e la rigenerazione dei veicoli, è possibile notare aree migliorabili nella progettazione che devono essere segnalate al reparto design.

# CONCLUSIONE

Questa tesi si propone di contribuire alla conoscenza della gestione dei big data in campo Automotive, con l'intento di supportare lo sviluppo di competenze strategiche e operative per comprendere e gestire le sfide e le opportunità connesse a questa tecnologia, al fine di supportare le decisioni strategiche e migliorare i processi organizzativi.

Nello specifico applicativo Automotive, dall'analisi condotta nel seguente elaborato, in campo Big Data Driven ed Economia Circolare, emerge come il quadro per la manutenzione circolare abilitata dall'IoT illustrato in Figura 16, potrebbe essere messo in atto per guidare e gestire i flussi di attività legate alla business unit circolare SUSTAINera grazie ad una collaborazione tra SUSTAINera e MOBILSIGHT, la Stellantis's data business (§ 2.2). In particolare, sarebbe ottimale l'applicazione dello specifico caso di studio riportato nel seguente elaborato (§ 2.5.1.), riguardo *Hydrogen Fuel Cell Vehicle* HFCV, sugli otto modelli di veicoli commerciali a idrogeno già in commercio: *Citroën ë-Jumpy* e *ë-Jumper*, *Fiat Professional E-Scudo* e *E-Ducato*, *Opel/Vauxhall Vivaro* e *Movano*, nonché *Peugeot E-Expert* e *E-Boxer*. L'obiettivo iniziale è costruire una sinergia tra le due business unit, che fino a questo momento non è stata ancora realizzata. Al fine di ottenere una gestione ottimale dell'attività delineata, si richiede la presenza di un ulteriore Hub di Economia Circolare, come quello SUSTAINera, localizzato nei pressi di uno degli stabilimenti produttivi dei veicoli commerciali ad idrogeno.



# BIBLIOGRAFIA

[file:///C:/Users/Alessia/Downloads/PIIT\\_Landing\\_page\\_C2V\\_approfondimento\\_V1.4%20\(1\).pdf](file:///C:/Users/Alessia/Downloads/PIIT_Landing_page_C2V_approfondimento_V1.4%20(1).pdf)

Fonte Kumar Rahul. *Data Life Cycle Management in Big Data Analytics*. *Procedia Computer Science* 173 (2020) 364–371

Fonte Uthayasankar Sivarajah. *A study on big data analytics and innovation: From technological and business cycle perspectives*. *Technological Forecasting & Social Change* 202 (2024) 123328

*Appunti del corso Business Intelligence per Big Data*, Eliana Pastor, Politecnico di Torino, 2024

Annu Roy (Autore) Rishiraj Deb (Autore), Gaurav Arora (Autore), *Self-Service Analytics with Power BI: Learn how to build an end-to-end analytics solution in Power BI (English Edition)*, BPB Publications, 2023.

Sinan Ozdemir (Autore), Paolo Poli (Traduttore), *Data science. Guida ai principi e alle tecniche base della scienza dei dati*, s.l, Apogeo, 2017.

*Management dell'economia circolare. Principi, drivers, modelli di business e misurazione*, di Natalia Marzia Gusmerotti (Autore), Marco Frey (Autore), Fabio Iraldo (Autore)

*Appunti del corso Strumenti e metodi per la sostenibilità dei sistemi edilizi e territoriali*, Caterina Mele, Politecnico di Torino, 2021

*Idrogeno. Produzioni e applicazioni*, di Davide Scullino, Sandit Libri, 2023

<https://www.bmw.com/it/innovation/come-funzionano-le-auto-a-idrogeno.html>

Turner, C., Okorie, O., Emmanouilidis, C., & Oyekan, J. (2022). *Circular production and maintenance of automotive parts: An Internet of Things (IoT) data framework and practice review*. *Computers in Industry*, 136, 103593. <https://doi.org/10.1016/j.compind.2021.103593>

Orlovska, J., Wickman, C., & Söderberg, R. (2018). *Big data usage can be a solution for user behavior evaluation: An automotive industry example*. *Procedia CIRP*, 72, 117-122. <https://doi.org/10.1016/j.procir.2018.03.102>

Mostefaoui, A., Merzoug, M. A., Haroun, A., Nassar, A., & Dessables, F. (2022). *Big data architecture for connected vehicles: Feedback and application examples from an automotive group*. *Future Generation Computer Systems*, 134, 374-387. <https://doi.org/10.1016/j.future.2022.04.020>

*Stellantis Internal Meeting*

# SITOGRAFIA

[https://it.wikipedia.org/wiki/Big\\_data](https://it.wikipedia.org/wiki/Big_data)

<https://www.oracle.com/it/big-data/what-is-big-data/#three>

<https://www.oracle.com/it/a/ocom/docs/top-22-use-cases-for-big-data.pdf>

[\[eu.industrial.panasonic.com/panasonicc2vmanutenzionepredittiva?utm\\\_source=google&utm\\\_medium=cpc&utm\\\_campaign=Syso\\\_C2V\\\_Manutenzionepredittiva&utm\\\_content=Syso&gad\\\_source=1&gclid=Cj0KCQjwIN6wBhCcARIsAKZvD5h2Vnv3XvmCxOJvVNICKQgfqlQuBgbTRrT\\\_y\\\_92Wj-uArwFSbnUH0kaAgbKEALw\\\_wcB\]\(https://information-eu.industrial.panasonic.com/panasonicc2vmanutenzionepredittiva?utm\_source=google&utm\_medium=cpc&utm\_campaign=Syso\_C2V\_Manutenzionepredittiva&utm\_content=Syso&gad\_source=1&gclid=Cj0KCQjwIN6wBhCcARIsAKZvD5h2Vnv3XvmCxOJvVNICKQgfqlQuBgbTRrT\_y\_92Wj-uArwFSbnUH0kaAgbKEALw\_wcB\)](https://information-</a></p></div><div data-bbox=)

[file:///C:/Users/Alessia/Downloads/PIIT\\_Landing\\_page\\_C2V\\_approfondimento\\_V1.4%20\(1\).pdf](file:///C:/Users/Alessia/Downloads/PIIT_Landing_page_C2V_approfondimento_V1.4%20(1).pdf)

<https://www.mckinsey.com/capabilities/operations/our-insights/manufacturing-analytics-unleashes-productivity-and-profitability>

<https://www.purestorage.com/it/knowledge/what-is-an-open-source-database.html>

<https://cloud.google.com/learn/advantages-of-cloud-computing?hl=it>

[www.elsevier.com/locate/techfore](http://www.elsevier.com/locate/techfore)

<https://www.ingenio-web.it/articoli/le-sfide-dell-idrogeno-nel-mercato-automobilistico-un-futuro-incerto/>

<https://www.magnetimarelli-parts-and-services.it/azienda/project-5-0-the-digital-wave.html#prodid=B88EF364E0D999C764B16BDD741BF5404AA1E5A181306027757DC4045664ABCC>

<https://www.mobilisights.com/en/technology.html>

<https://www.pirelli.com/tyres/it-it/auto/tecnologia/cyber-tyre>

<https://www.redhat.com/it/topics/linux/what-is-sap-hana-migration>

<http://www.alkit.se/index.php/front-page/industry/industry-projects/wice/>

<https://www.press.bmwgroup.com/italy/article/detail/T0444870IT/pionieri-dell-idrogeno:-bmw-group-e-toyota-motor-corporation-portano-la-collaborazione-ad-un-livello-superiore-per-offrire-opzioni-di-veicoli-elettrici-a-celle-a-combustibile-fcev-per-le-autovetture?language=it>

<https://www.stellantis.com/it/tecnologie/tecnologia-a-celle-a-combustibile-a-idrogeno>

<https://www.stellantis.com/it/news/eventi-recenti/apertura-ufficiale-del-primo-hub-di-economia-circolare-sustainera>

<https://www.media.stellantis.com/it-it/corporate-communications/press/l-economia-circolare-di-stellantis-ha-registrato-una-forte-crescita-nel-2023-e-proietta-la-stessa-tendenza-nel-2024>

