



BIG DATA & ANALYTICS

Industry 4.0

Manutenzione predittiva degli impianti



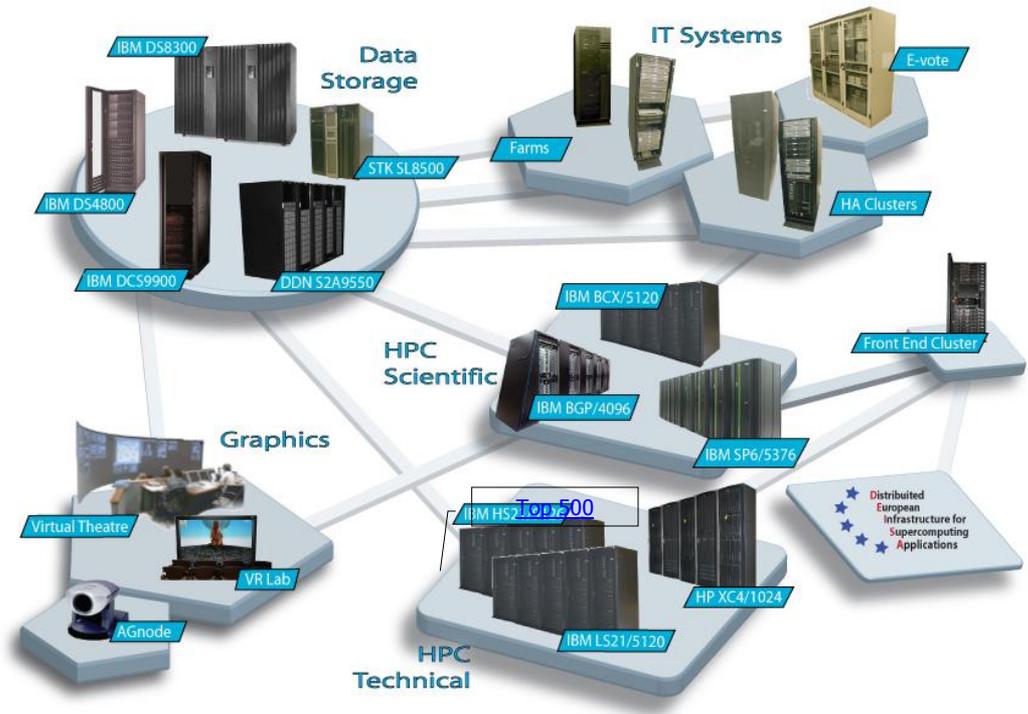
CINECA

Il Cineca è un Consorzio Interuniversitario senza scopo di lucro al servizio del sistema accademico nazionale istituito nel 1969. La missione è “promuovere l’utilizzo dei più avanzati sistemi di **elaborazione dell’informazione** a favore della ricerca scientifica e tecnologica, pubblica e privata”, e al trasferimento tecnologico alle imprese e alla Pubblica Amministrazione.

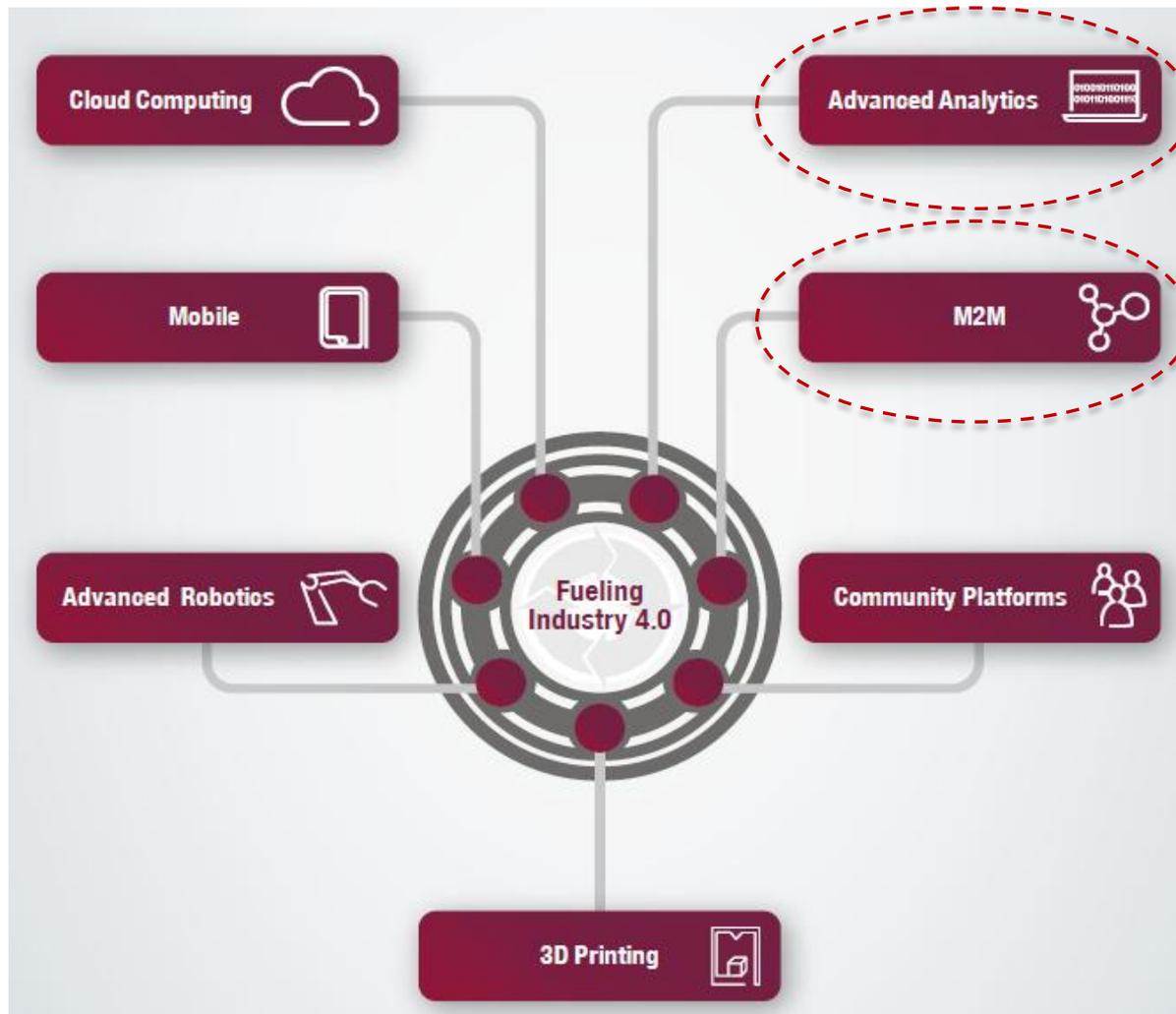
Fanno parte del Consorzio:

- MIUR
- 70 università
- 4 Enti di Ricerca

Circa 1.000 dipendenti con sedi a Bologna, Milano e Roma



Abilitatori tecnologici Industry 4.0



Fonte: Capgemini Consulting

Big Data

- Aumento di Volume, Varietà e Velocità dei dati disponibili; necessità di Veridicità
- Evoluzione di nuove tecnologie di raccolta, storage ed elaborazione dati
- Maggiore capacità di analisi dati (analytics) che necessitano di:
 - ✓ analisi più complesse (es. machine learning)
 - ✓ elaborazioni/query su grandi volumi di dati
 - ✓ elaborazioni/query sui flussi dati real-time



→ determinano nuove opportunità per la creazione di **Valore** aggiunto per le aziende e le organizzazioni.

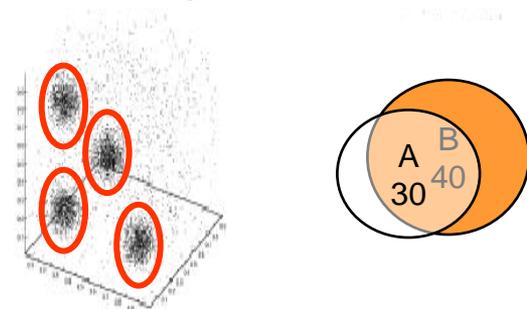
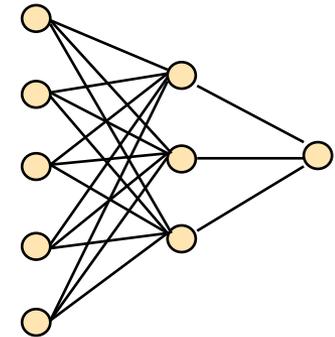
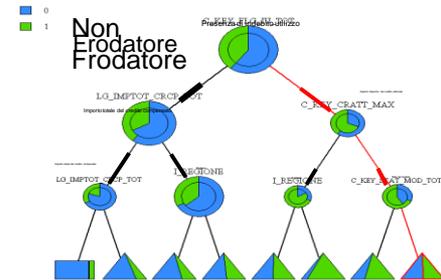
Definizione di Analytics

Analytics sono applicazioni informatiche che usano metodi matematici e statistici su sistemi computazionali altamente scalabili per estrarre valore dai dati, come trovare schemi ricorrenti (patterns), raggruppamenti (clusters) e relazioni nei dati (rules) per predire futuri comportamenti o scenari.

1. **L'analisi descrittiva** si occupa di descrivere ciò che accade e di presentare informazioni, servendosi di strumenti di Business Intelligence e di visualizzazione. (Analisi OLAP)
 2. **L'analisi predittiva** cerca di prevedere le performance e/o il comportamento degli utenti, dell'organizzazione o degli impianti in scenari futuri. (Data Mining)
 3. **L'analisi prescrittiva** si spinge oltre la previsione di risultati futuri, fornendo raccomandazioni in maniera automatica su quale soluzione scegliere tra un ventaglio di possibilità. (Cognitive Computing)
- Quali sono i clienti top ten? -> Analisi descrittiva (Reporting base)
 - Quali prodotti stanno comprando i top ten clienti nell'ultimo semestre nelle diverse aree geografiche? -> Analisi descrittiva (Reporting multidimensionale (OLAP))
 - Quali sono i profili dei miei clienti? -> Analisi dei cluster
 - Qual è la probabilità che un certo cliente mi abbandoni? -> Analisi predittiva (Machine Learning)
 - Quale sarà il comportamento di acquisto dei nostri clienti su una nuova linea di prodotti? -> Analisi prescrittiva (Cognitive Computing)

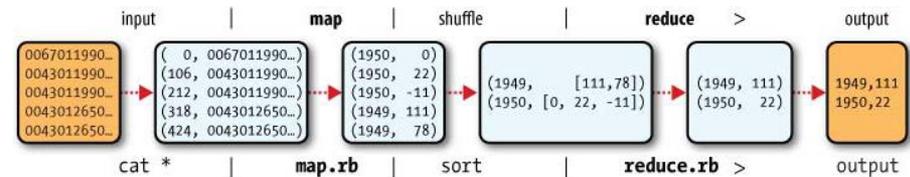
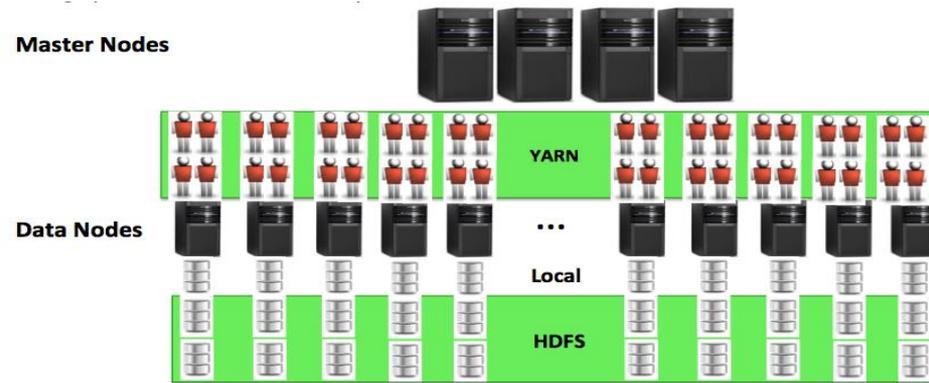
Tecniche di Data Mining

STATISTICA INFERENZIALE
Stima di grandezze statistiche
Verifica di ipotesi
Regressione Lineare
Regressione Logistica
Correlazione
Tabella di contigenza e chi quadrato
Granger Casual Modeling
CLASSIFICAZIONE/PREVISIONE SUPERVISIONATA
Decision tree -> regression tree
Decision tree -> classification tree
Naive bayes
Reti bayesiane
Random forest
Reti neurali
CLASSIFICAZIONE ED ALTRE TECNICHE NON SUPERVISIONATE
Clustering
Regole di Associazione
Sequential Pattern



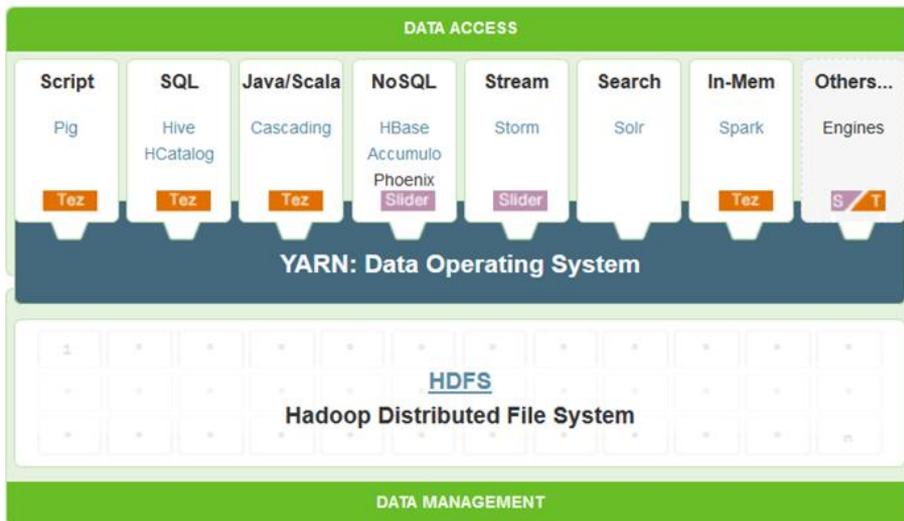
Big Data tecnologia abilitante – Hadoop e Spark

Le principali architetture di riferimento per i Big Data sono sempre più Hadoop e Spark di Apache.org. Hadoop e Spark consentono il calcolo distribuito su più nodi e in-memory. Queste architetture introducono **nuovi modelli di programmazione distribuita** particolarmente utili per il **Large Scale Machine Learning** e l'**analisi real time su streaming di dati**.



Esistono diversi moduli open source di Apache.org compatibili con le architetture Hadoop e Spark:

- ✓ Database NoSQL e in-memory (Hbase)
- ✓ SQL distribuito (Hive)
- ✓ Data Streaming e ETL (Storm, Flume, SparkStreaming)

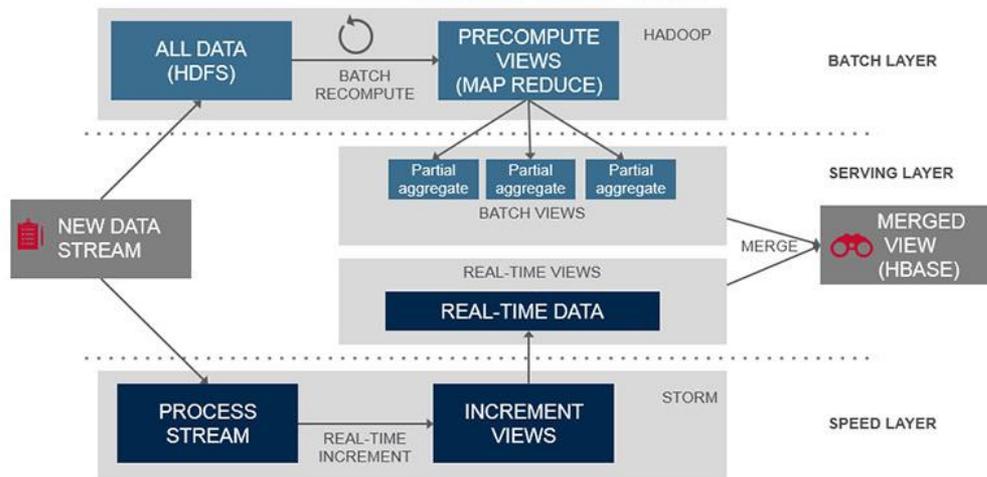


Big Data e dati dai sensori

Il flusso di dati provenienti dai sensori e dai sistemi distribuiti di raccolta devono essere raccolti in una piattaforma che li rende disponibili per le analisi in tempo reale.



Lambda Architecture



L'architettura Lambda consente di conciliare i dati provenienti dai sistemi aziendali con quelli raccolti in streaming dai device e fare analisi real-time.

Le sue componenti fondamentali sono:

- ✓ Batch layer
- ✓ Speed layer
- ✓ Serving layer

Analisi Predittiva

Rolls-Royce Aircraft Division Restructures its Business Model.

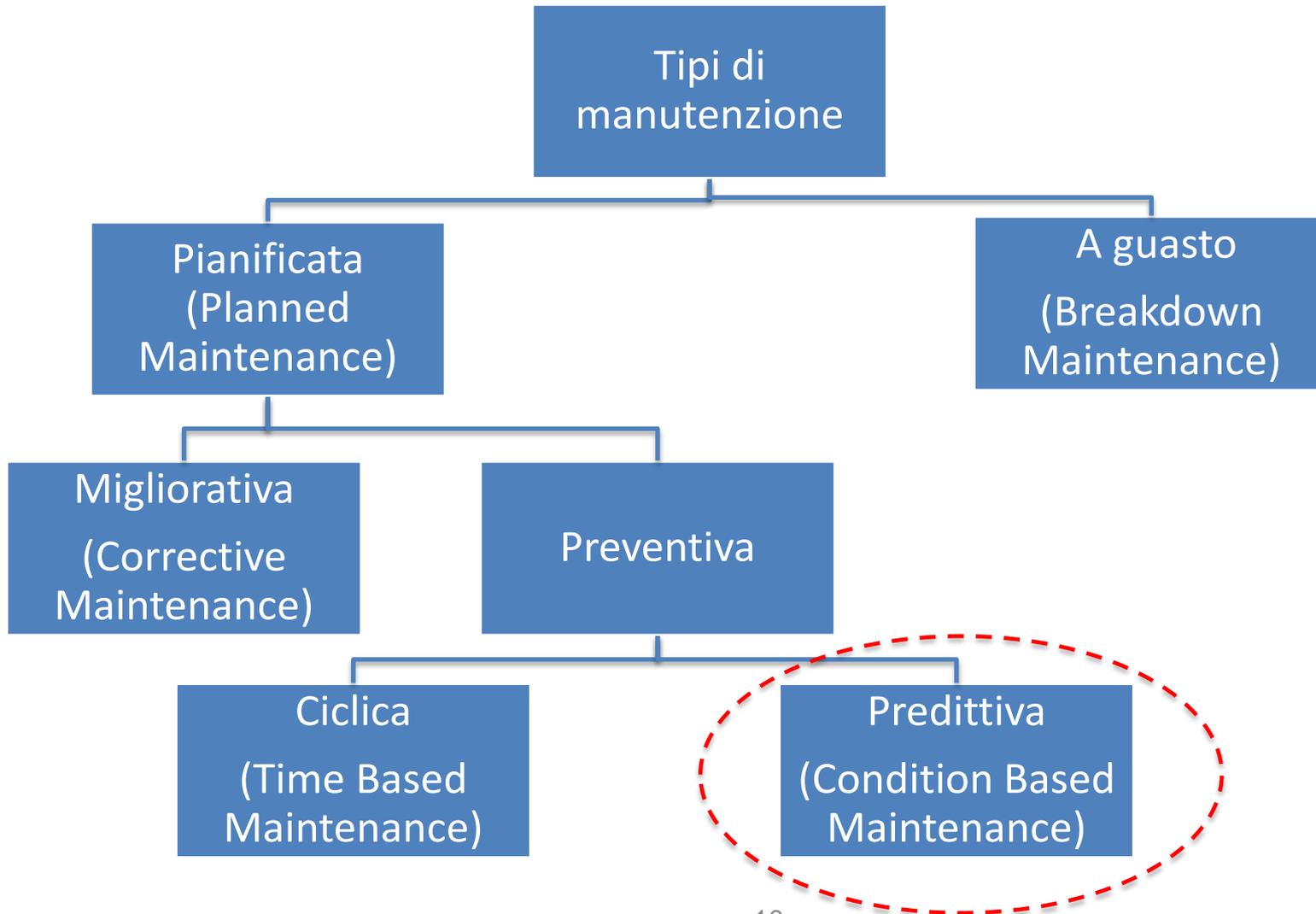
Instead of selling its engines to customers, Rolls-Royce now rents them, retaining the responsibility for repair, maintenance, and replacement. This shift was triggered by new Big Data capabilities that have allowed an unprecedented level of [predictive maintenance](#). Rolls-Royce can now [identify correlations between different part failures and different operational environments](#). This is allowing the company to [predict engine failures several days before they occur, with high accuracy and low false alarms](#).

General Electric

GE announced recently that revenues from its [Predictivity solutions](#), which make machines more productive, will exceed \$1 billion in 2014. The convergence of industrial machines, data and the Internet to lift machine productivity is termed as industrial Internet.

GE is outfitting its machines from jet engines and gas turbines to CT/PET scanners with sensors and software that monitor and collect data on health of different parts of that machine. A cloud-based Internet platform then analyzes this data through advanced analytics involving historical data points to provide machine operators and maintenance engineers with real-time information, which is used to schedule predictive maintenance checks that improve machine efficiency as well as prevent downtime to improve overall productivity. This is the essence of GE's Predictivity solutions.

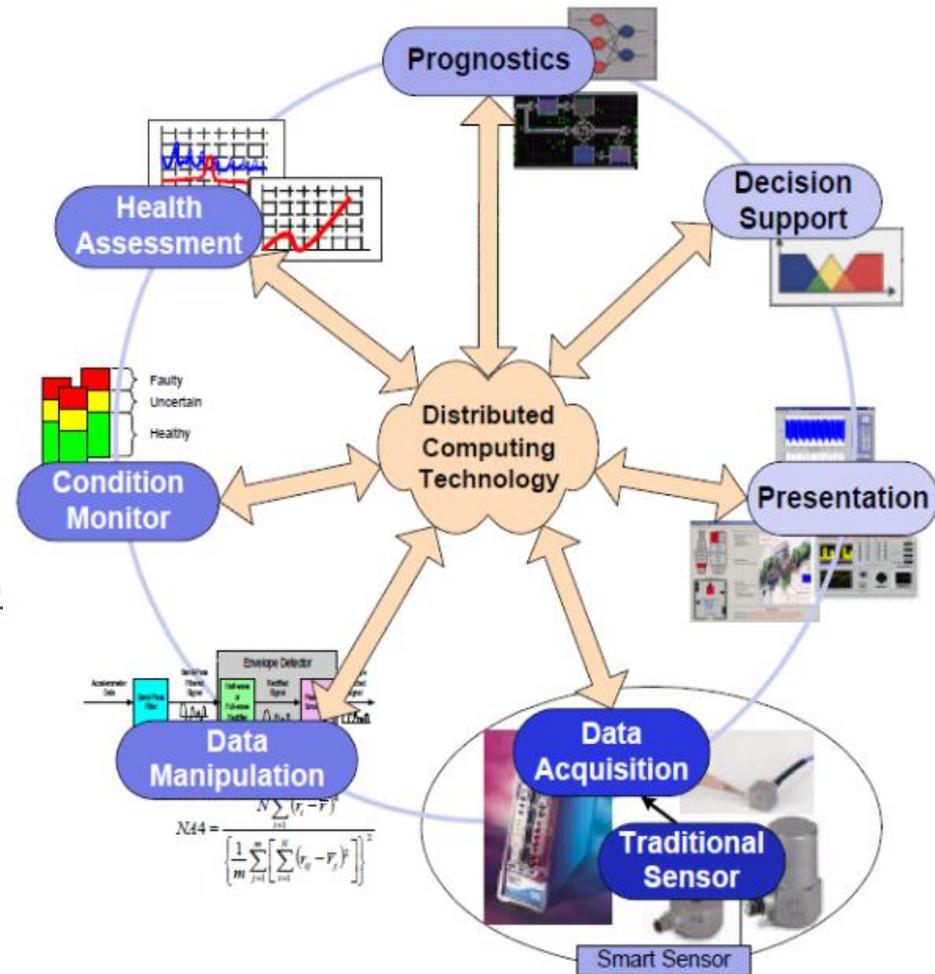
Tipi di manutenzione



Condition Based Maintenance (CBM)

Condition-based maintenance (CBM)

1. Acquisizione dati
2. Pre processing (preparazione dei dati per l'elaborazioni di data mining e di machine learning)
3. Monitoraggio dell'impianto (fornisce semplici allarmi basati sui limiti operativi)
4. Valutazione della salute dell'impianto (genera record diagnostici sulla base dell'analisi dei trend per valutare il livello di degrado)
5. **Predizione (genera previsioni sulla probabilità di guasti usando tecniche di machine learning)**
6. Supporto alle decisioni (analisi multidimensionale della Business Intelligence)
7. Reporting complessivo (cruscotto che rappresenta tutte le informazioni generate dai precedenti sistemi)



Obiettivi dell'analisi predittiva nel CBM

Gli obiettivi più comuni di analisi sono:

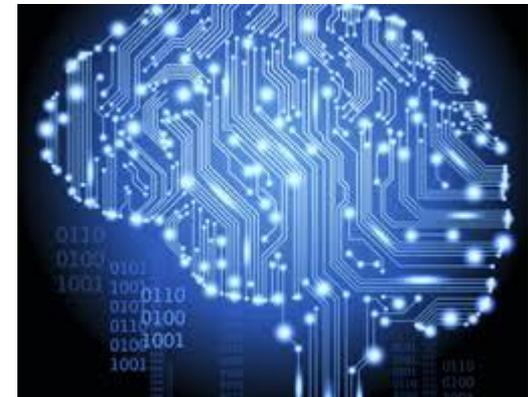
- 1) Predire il Tempo Residuo di Funzionamento (Time to Failure)
- 2) Prevedere se una risorsa si guasterà entro un determinato periodo di tempo
- 3) Prevedere se una risorsa si guasterà entro differenti finestre temporali
- 4) Individuare regole (patterns) che determinano un guasto

Analisi Predittiva nel CBM

La manutenzione predittiva nella Condition-based maintenance (CBM), si focalizza sull'individuare la probabilità di guasti **prima** che avvengano.

Il cuore del Sistema CBM è lo strato 5 (Prognostics) il quale si basa sull'analisi predittiva applicando tecniche di **machine learning** e utilizzando tecnologie di **big data** sui dati del deterioramento e dei guasti forniti dallo strato 3 (condition monitoring) e sui dati dell'analisi dei trend forniti dello strato 4 (health assessment).

L'applicazione di machine learning per predire situazioni di probabili guasti si basa sul *costruire un modello usando dati storici e **addestrarlo** con casi noti, per essere in grado di identificare o classificare situazioni di potenziali guasti e non. Per poi usare questo modello per predire situazioni di guasti nel corso del funzionamento dell'impianto.* Il modello dovrà essere **validato** usando dati reali di test prima di applicarlo. La validazione fornisce una indicazione (matrice di confusione) sull'attendibilità del modello individuando i veri positivi, i veri negativi, falsi positivi e i falsi negativi.



Analisi Predittiva nel CBM - Esempio

Obiettivi

- 1) Prevedere per un Gruppo di Continuità (UPS) se si verificherà un evento grave (livello 3) entro 4 minuti.



Variabili per ogni evento osservato

- ✓ Type, Timestamp, Severity, 18 variabili (battery voltage, input voltage, input current, Output voltage, output frequency, percentage of load ...) a 2 e 4 minuti dal verificarsi di un evento.

Tecniche di Machine Learning utilizzate

- ✓ Binary classification
 - Decision Tree (c5.0)

Type	Timestamp	Severity	D2BatteryStatus	D2BatteryVoltage	...	D2ResidualCharge	D2ResidualTime	D4BatteryStatus	D4BatteryVoltage	...	D4ResidualCharge	D4ResidualTime	classificazione
126	10/06/2015 15:13	3	2	226,7	...	88	65	2	244,8	...	100	0	1
63	10/06/2015 15:15	3	2	244,8	...	100	999	2	248,44	...	94,2	68,8	1
126	10/06/2015 15:18	3	2	244,8	...	98,5	48	2	244,8	...	100	999	1
63	10/06/2015 15:24	3	2	244,8	...	100	87	2	244,8	...	100	0	1
...
126	26/09/2015 22:34	3	2	244,8	...	100	999	2	244,8	...	100	999	1
0	27/09/2015 21:12	-1	2	244,8	...	100	999	2	244,8	...	100	999	0
126	27/09/2015 21:16	3	2	244,8	...	100	999	2	244,8	...	100	999	1
0	28/09/2015 04:45	-1	2	244,8	...	100	999	2	244,8	...	100	999	0
126	28/09/2015 04:49	3	2	244,8	...	100	999	2	244,8	...	100	999	1

Analisi Predittiva nel CBM – Esempio

Risultati

Previsione di eventi gravi considerando le variabili a 2 e 4 minuti dal verificarsi di un evento.

	Reali	
Previsti	0	1
0	52	33
1	18	83

Matrice di confusione

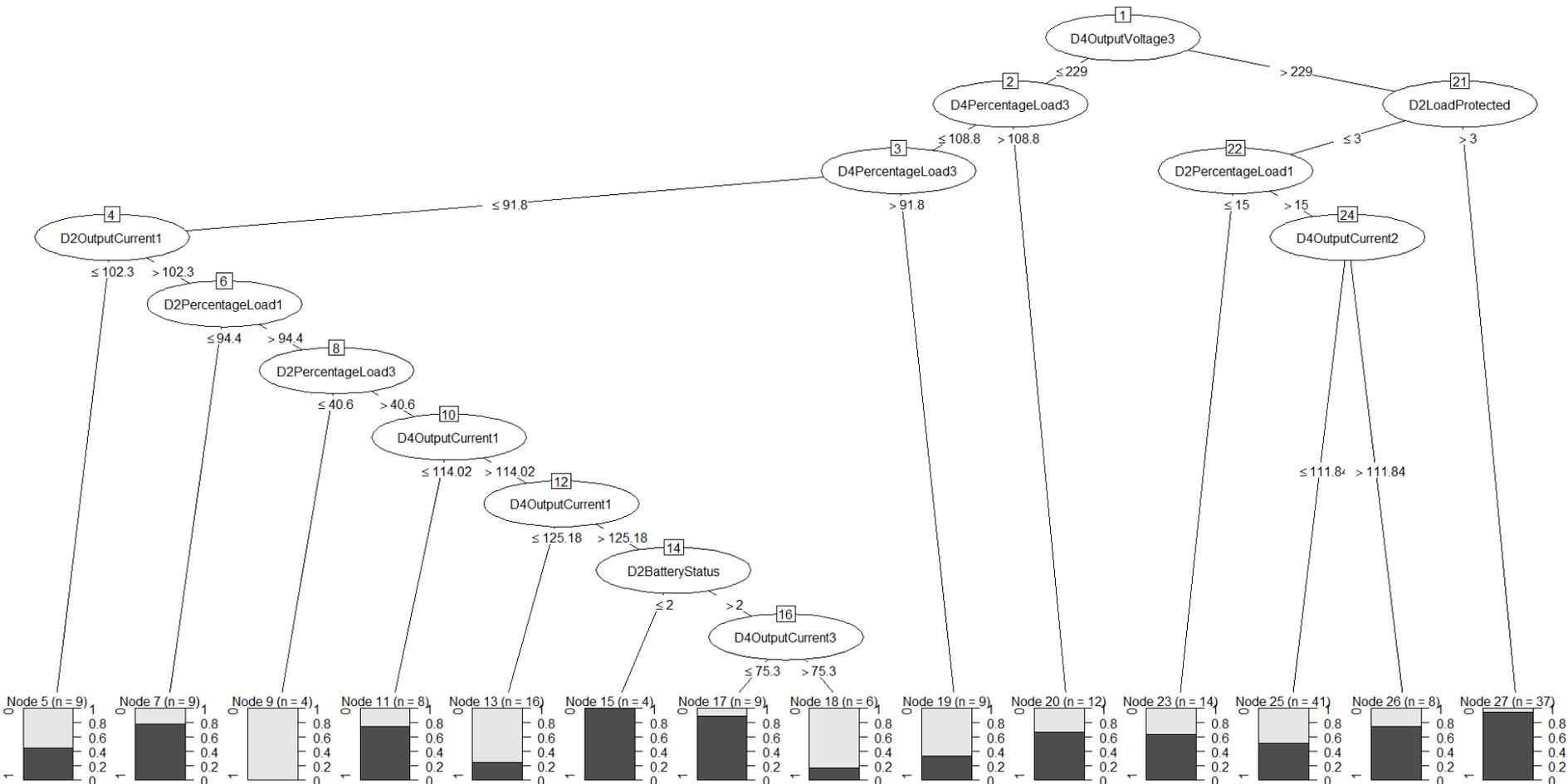
Accuratezza	0,7258
Sensibilità	0,7155
Specificità	0,7429

Accuratezza= Percentuale di osservazioni correttamente classificate

Sensibilità= Percentuale di osservazioni di classe 1 (evento grave) correttamente classificate

Specificità= Percentuale di osservazioni di classe 0 (nessun evento grave) correttamente classificate

Analisi Predittiva - Risultati



Il Laboratorio Big Data & Analytics

Il Laboratorio di Big Data & Analytics è una iniziativa di CINECA, nel campo della High Performance Analytics per promuovere la sua diffusione e aiutare i decisori aziendali e i professionisti ICT a comprendere le strategie, le potenzialità e le tecnologie dei Big Data e delle tecniche di Data Mining.

PIATTAFORME SOFTWARE:

- IBM Big Insights
- Hortonworks Data Platform

ARCHITETTURE:

- Data Streaming Analysis
- Large Scale Machine Learning

TECNOLOGIE:

- Hadoop (HDFS, MapReduce), YARN
- Spark SQL, Hive e HBase
- Storm, Spark Streaming
- Kafka & MQTT
- Spark R e Distributed R
- Spark MLLIB, H2O

INFRASTRUTTURA:

HPC IBM NeXtScale server appositamente progettata per i casi di calcolo “data-intensive”:

- 70 nodi IBM NeXtScale con interconnessione a 5 GB/sec
- Intel Ivy Bridge 20 core per nodo, 1480 core in totale
- 128 GB RAM per nodo
- 40 TB SSD locale al nodo, 16 PB di storage in linea

