

HIGH PERFORMANCE (BIG) DATA /ANALYTICS di CINECA



BIG DATA meets HPC

- Aumento di **Volume**, **Varietà** e **Velocità** dei dati disponibili; necessità di **Veridicità**
 - Evoluzione di nuove **tecnologie** di raccolta, storage ed elaborazione
 - Maggiore **capacità analitica (anche predittiva) e interpretativa** di cui le imprese possono dotarsi
- ➔ determinano nuove opportunità per la creazione di **Valore** aggiunto per industria, amministrazione pubblica e mondo accademico

High Performance Data Analysis

- Il settore dell' HPDA può essere definito come l'insieme delle simulazioni e analisi *data-intensive* che coinvolgono operazioni con volumi di dati e complessità algoritmica sufficiente per richiedere tecnologie e approcci HPC.
- Le principali società commerciali in una varietà di segmenti di mercato si stanno rivolgendo a tecnologie HPC per i grandi carichi di lavoro di analisi dei dati che la tecnologia IT aziendale da sola non può affrontare in modo efficace. I sistemi HPC sono in grado di gestire le query più complesse, più variabili e requisiti di consegna più rapidi.

SCS/CINECA per HP-DA

- Infrastruttura
- Piattaforma Software
- Expertise

INFRASTRUTTURA

- CINECA si è dotato nel settembre 2014 di un apposita infrastruttura HPC per i Big Data (PICO) fornita da IBM attraverso procedimento di gara
- PICO si integra con i supercalcolatori già presenti in Cineca (Fermi, PLX) e ha caratteristiche specifiche per il calcolo data-intensive:
 - 70 nodi IBM NeXtScale (progettate per portare la potenza HPC in ogni data center)
 - Processore Intel Ivy Bridge 20 core per nodo
 - 128 GB RAM per nodo
 - 1480 core in totale
- Architettura di storage multi tier per applicazioni Big Data
 - Disco a stato solido locale al nodo per accelerare le performance di accesso a DB locali ai nodi
 - 16 PB di storage in linea (4PB disponibili e 12 PB su tape) con migrazione automatica su tape per dati di grandi dimensioni, dati raw o dati non strutturati
 - 48 PB di storage con compressione
- Interconnessione → 45 Gigabits

PIATTAFORMA SOFTWARE

IBM Infosphere BigInsights è la piattaforma installata su PICO che permette di affrontare tutte le sfide dei Big Data. La piattaforma unisce alle tecnologie tradizionali, perfette per attività strutturate e ripetibili, nuove tecnologie complementari che operano con rapidità e flessibilità, e che sono ideali per analisi non strutturate, rilevamento ed esplorazione di dati ad hoc.

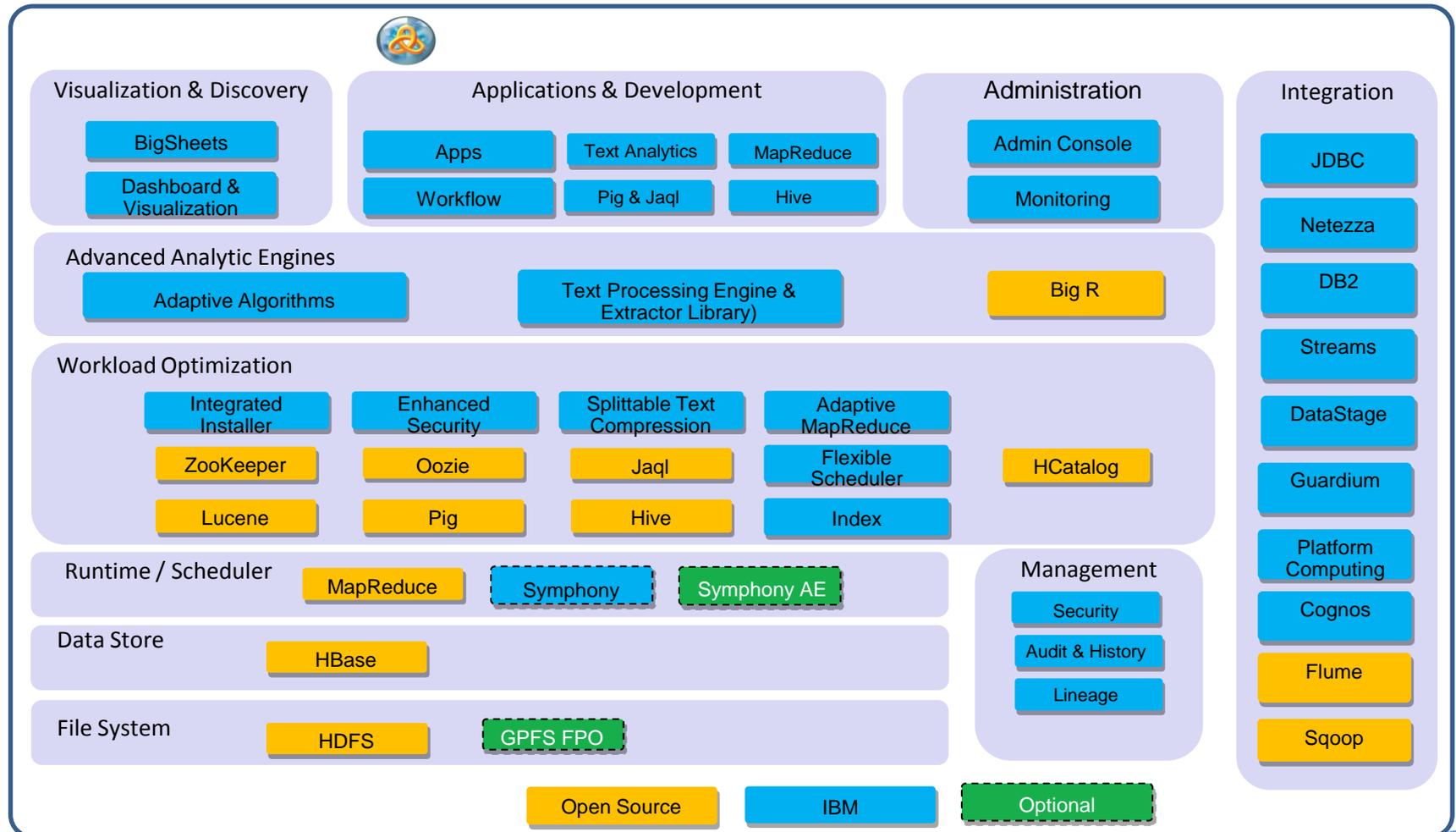
Le funzionalità principali sono:

- **Analitica basata su Hadoop:** elabora e analizza qualsiasi tipo di dati presenti nei cluster di server di largo consumo.
- **Stream computing:** esegue un'analisi continua di volumi ingenti di dati in streaming, con tempi di risposta inferiori al millisecondo.
- **Data Warehousing:** offre importanti informazioni operative, con analytics in-database avanzata.
- **Governance e integrazione delle informazioni:** consente di comprendere, ripulire, trasformare, gestire e offrire informazioni affidabili alle tue iniziative business più importanti.

Servizi di supporto alla piattaforma:

- **Visualizzazione rilevamento:** aiuta gli utenti finali a esplorare dataset di grandi dimensioni e complessi.
- **Sviluppo delle applicazioni:** semplifica il processo di sviluppo di applicazioni big data.
- **Gestione di sistemi:** monitora e gestisce sistemi big data per prestazioni sicure e ottimizzate.
- **Acceleratori:** velocizza il time to value tramite moduli analitici e specifici del settore d'industria.

IBM InfoSphere BigInsights



EXPERTISE

- Progettazione e gestione di infrastrutture di High Performance Computing calcolo e dati per accademia e industria. Sviluppo HPC SW per accademia e industria.
- Esperienza pluriennale di CINECA in BA e BI, data, text and multimedia mining, concept mapping, in diversi domini.
- Il team Knowledge Discovery & Management raccoglie varie professionalità impegnate nella estrazione, codifica, strutturazione e gestione della conoscenza.

Rete di contatti e precedenti esperienze CINECA su BA /BI

- Settori Istituzionali CINECA:
 - EDUCATION (università)
 - SANITA' (ospedali, asl)
 - PA (amministrazione locali/smart cities)
- Altri settori:
 - Bancario e assicurativo
 - GDO
 - Energetico

INFINITE APPLICAZIONI di BD Analytics

 Banking <ul style="list-style-type: none">• Optimize Offers and Cross Sell• Contact Center Efficiency and Problem Resolution• Payment Fraud Detection & Investigation• Counterparty Credit Risk Management	 Insurance <ul style="list-style-type: none">• Claims Fraud• Next Best Action• Catastrophe Modeling	 Telco <ul style="list-style-type: none">• Pro-active Call Center• Network Analytics• Location Based Services• IT/Network Infrastructure Transformation• Smarter Campaigns	 Energy & Utilities <ul style="list-style-type: none">• Smart Meter Analytics• Distribution Load Forecasting/Scheduling• Condition Based Maintenance• Create & Target Customer Offerings	 Media & Entertainment <ul style="list-style-type: none">• Business process transformation• Audience & Marketing Optimization• Multi-Channel Enablement• Digital commerce optimization
 Retail <ul style="list-style-type: none">• Actionable Customer Insight• Merchandise Optimization Playbook• Dynamic Pricing	 Travel & Transport <ul style="list-style-type: none">• Customer Analytics & Loyalty Marketing• Capacity & Pricing Optimization• Predictive Maintenance Analytics	 Consumer Products <ul style="list-style-type: none">• Optimized Promotions Effectiveness• Micro-Market Campaign Management• Real Time Demand Forecast	 Government <ul style="list-style-type: none">• Threat Prediction and Prevention• Detect and Prevent Improper Payments• Single View	 Healthcare <ul style="list-style-type: none">• Measure & Act on Population Health• Engage Consumers in their Healthcare• Increase visibility into drug safety and effectiveness
 Automotive <ul style="list-style-type: none">• Data Warehouse Optimization• Predictive Asset Optimization (PAO)• Actionable Customer Intelligence	 Chemical & Petroleum <ul style="list-style-type: none">• EDW Smart Consolidation & Augmentation• Operational Surveillance, Analysis & Optimization• Engineering & Operational Data Exploration & Mining	 Aerospace & Defense <ul style="list-style-type: none">• Uniform Information Access Platform• Data Warehouse Optimization• Predictive Asset Optimization (PAO)	 Electronics / Industrial Products <ul style="list-style-type: none">• Channel Driven Customer Analytics (CDCA)• Predictive Asset Monitoring & Optimization (PAMO)	 Life Sciences <ul style="list-style-type: none">• Increase visibility into drug safety and effectiveness