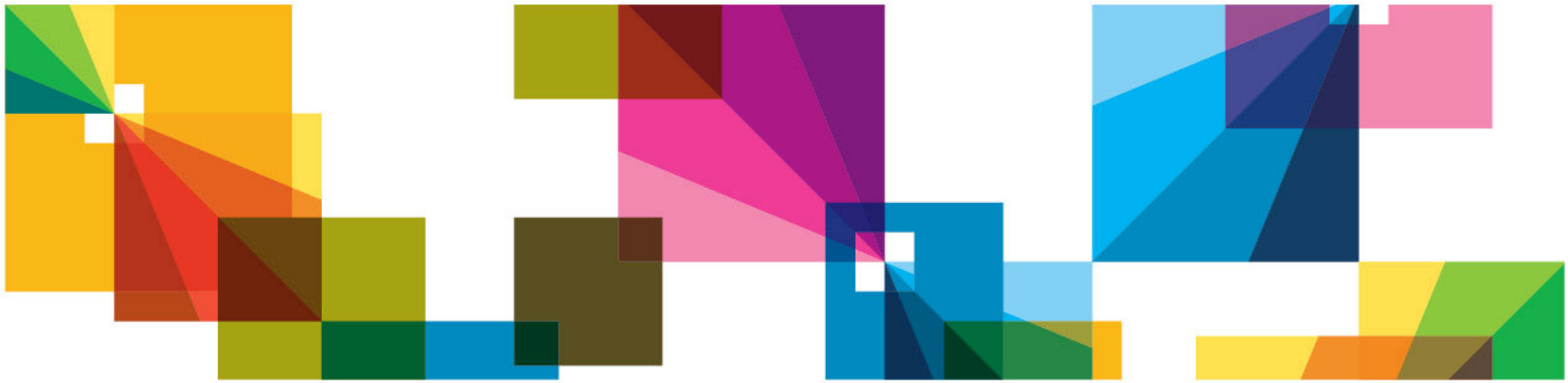
Smarter Analytics Live

IBM

Il Sole 24 ORE

Francesca Valentini

Data Warehousing all'avanguardia per le 3"V" dei Big Data: Velocity, Variety e Volume



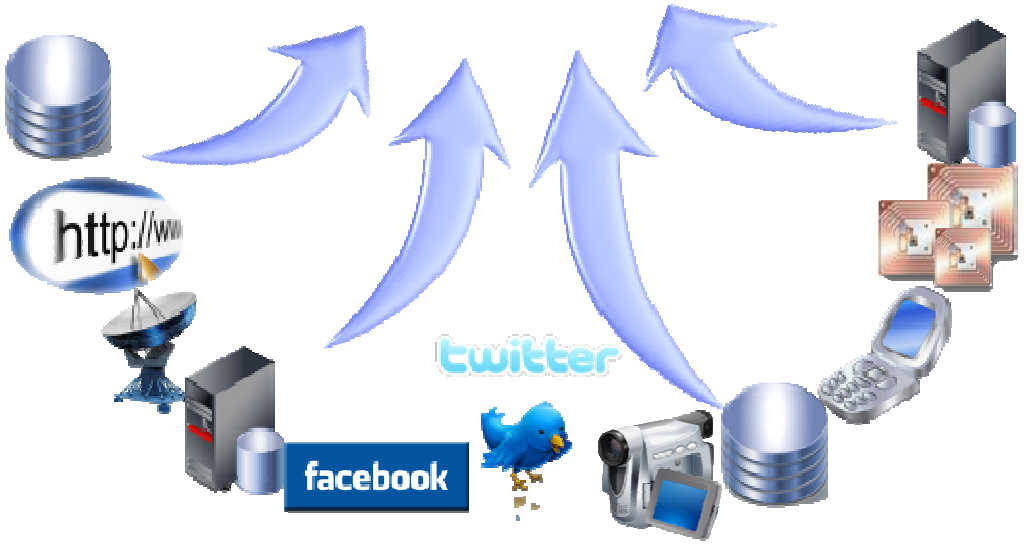
What is “BIG DATA”?

All kinds of data

Large volumes

Valuable insight, but difficult to extract

Often extremely time sensitive



Where is this data coming from?

12 TB of tweets being created each day.

Every day, the New York Stock Exchange captures 1 TB of trade information.

5 Billion mobile phones in use in 2010. Only 12% were smartphones.

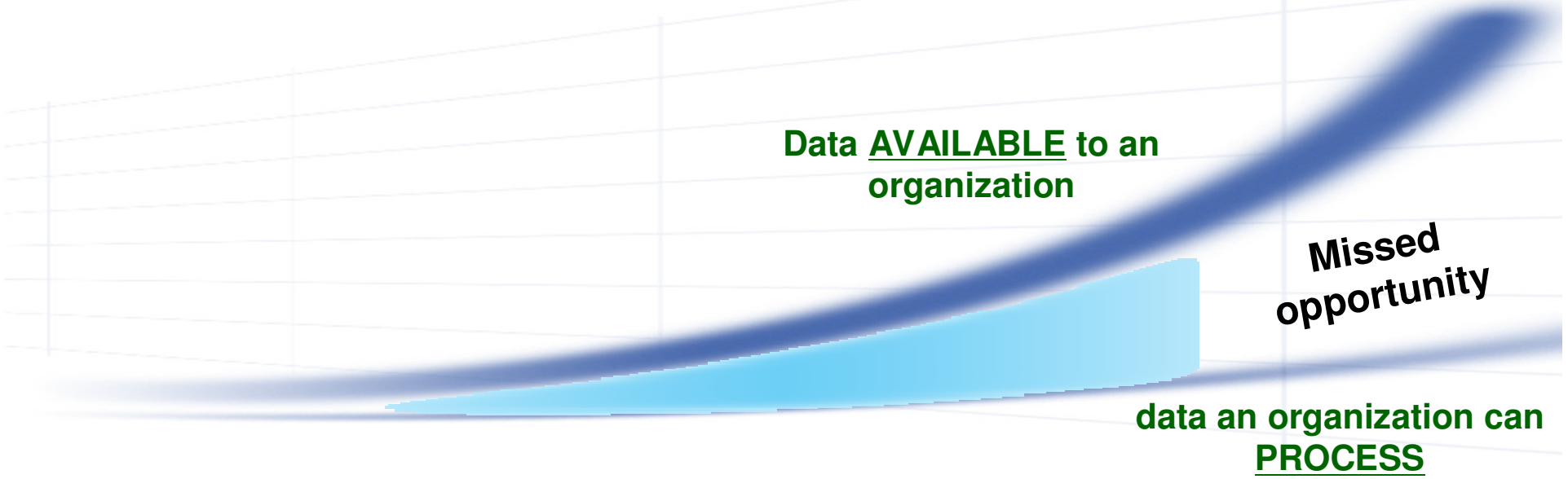
Every second of HD video generates > 2,000 times as many bytes as required to store a single page of text.

More than 30M networked sensor, growing at a rate >30% per year.

What is your business doing with it?

Source: McKinsey & Company, May 2011

Why is Big Data important?

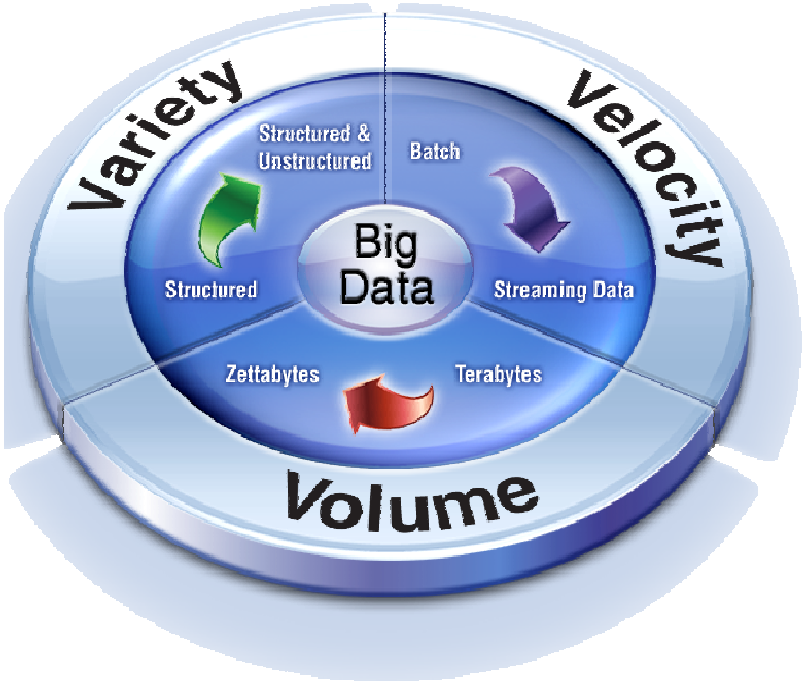


Organizations are able to process less and less of the available data.

Enterprises are “more blind” to new opportunities.

The big data opportunity

Extracting insight from an immense volume, variety and velocity of data, in a timely and cost-effective manner.

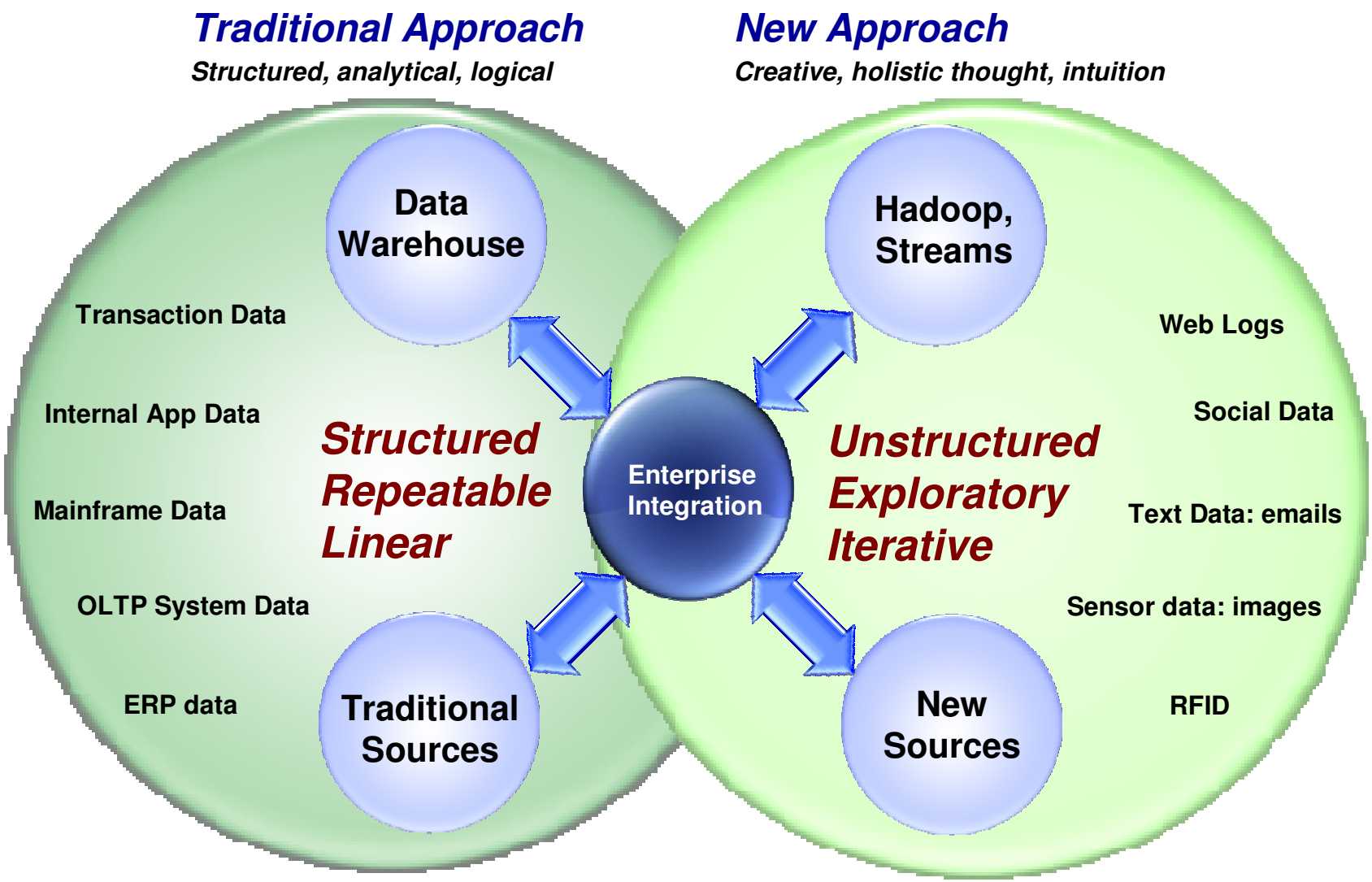


Variety: Manage the complexity of multiple relational and non-relational data types and schemas

Velocity: Streaming data and large volume data movement

Volume: Scale from terabytes to zettabytes

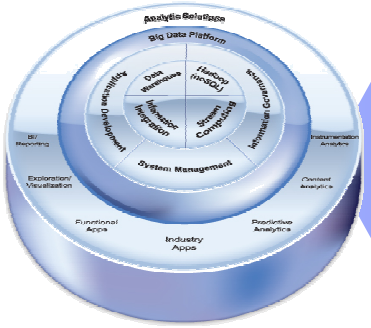
Complementary approaches for different Big Data use cases

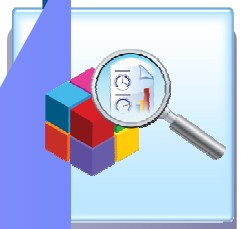


Big Data Paradigm Shift

VOLUME & VARIETY Structured & Unstructured	**REAL TIME VELOCITY** Data in Motion Analytics
STOP SAMPLING Move to Absolute Knowledge	**FIT FOR PURPOSE** New Technology, New Skills

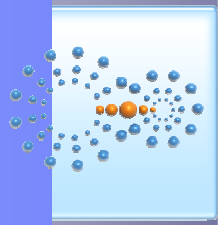
What does a Big Data platform do?





Analyze a Variety of Information

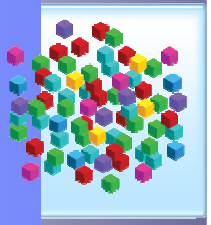
Novel analytics on a broad set of mixed information that could not be analyzed before



Analyze Information in Motion

Streaming data analysis

Large volume data bursts & ad-hoc analysis



Analyze Extreme Volumes of Information

Cost-efficiently process and analyze petabytes of information

Manage & analyze high volumes of structured, relational data

Discover & Experiment

Ad-hoc analytics, data discovery & experimentation

Manage & Plan

Enforce data structure, integrity and control to ensure consistency for repeatable queries

What can you do with big data?

Financial Services
- Fraud detection
- Risk management
- 360° View of the Customer

Utilities
- Weather impact analysis on power generation
- Transmission monitoring
- Smart grid management

Transportation
- Weather and traffic impact on logistics and fuel consumption

IT
- Transition log analysis for multiple transactional systems
- Cybersecurity

Health & Life Sciences
- Epidemic early warning system
- ICU monitoring
- Remote healthcare monitoring

Retail
- 360° View of the Customer
- Click-stream analysis
- Real-time promotions

Telecommunications
- CDR processing
- Churn prediction
- Geomapping / marketing
- Network monitoring

Law Enforcement
- Real-time multimodal surveillance
- Situational awareness
- Cyber security detection

The IBM Big Data Platform

InfoSphere Information Server
High volume data integration and transformation

InfoSphere Streams
Low Latency Analytics for streaming data

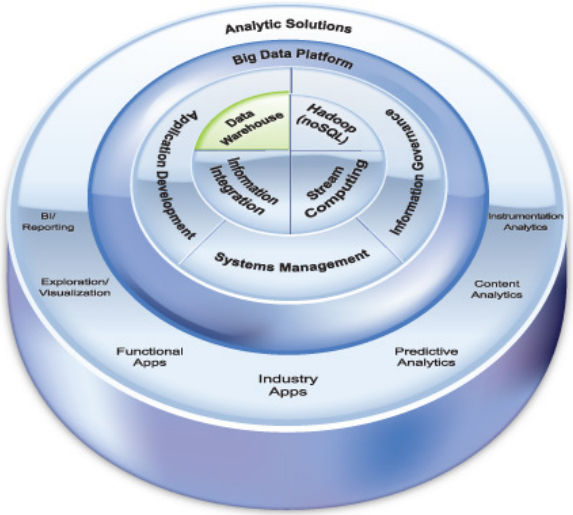
IBM InfoSphere Warehouse
Large volume structured data analytics

IBM Netezza Capacity Appliance
Queryable Archive Structured Data

IBM Smart Analytics System
Operational Analytics on Structured Data

IBM Informix Timeseries
Time-structured analytics

IBM Big Data Platform – Data Warehouse



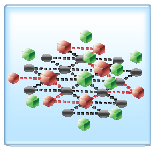
MPP Data Warehouse

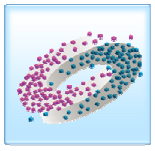
High performance analytics on a large volume of data structured for control, consistency, and integrity

Integration to share data and insights among data warehouse and other big data systems

Common application development tools & languages for data warehouse and other big data systems to facilitate skills reuse

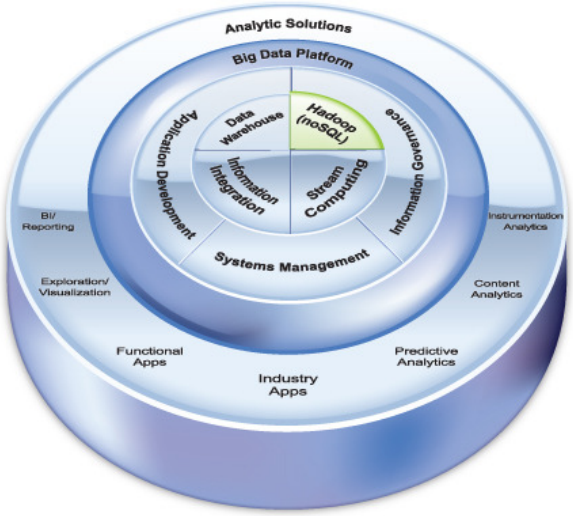
Use Cases





- **Deep Analytics** - business intelligence, predictive and advanced analytics on large volumes of data
- **Operational Analytics** – operational system access (up to 1000s/sec) to insights from large volumes of data
- **Combined structured and unstructured analysis –** combine data warehouse data with unstructured information in Hadoop for additional insights
- **Combine high velocity and historical analysis** – feed real-time analysis to data warehouse for historical analysis; closed feedback loop
  – data warehouse analysis informs tuning/changes to streams analytics

IBM Big Data Platform - BigInsights



BigInsights

A set of capabilities to cost-effectively analyze a wide variety and large volume of information to gain insights that were not previously possible. Ability to analyze data in its native format, without imposing a schema/structure, to enable fast ad-hoc ananlysis.

IBM's solution builds on the Apache Hadoop project and incorporates latest technology from IBM research.

Use Cases

Analyze a Variety of Information

Novel analytics on a broad set of mixed information that could not be analyzed before

Analyze Extreme Volumes of Information

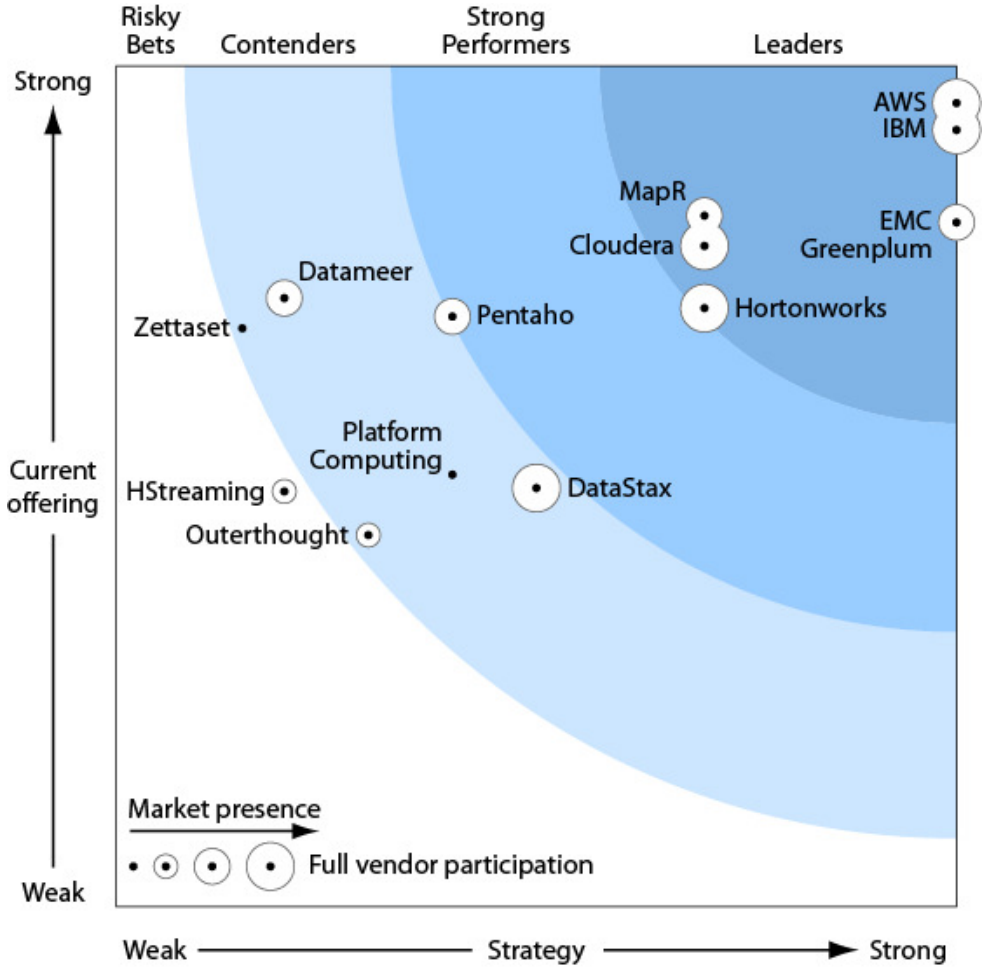
Cost-efficiently process and analyze petabytes of information

Discovery and Experimentation

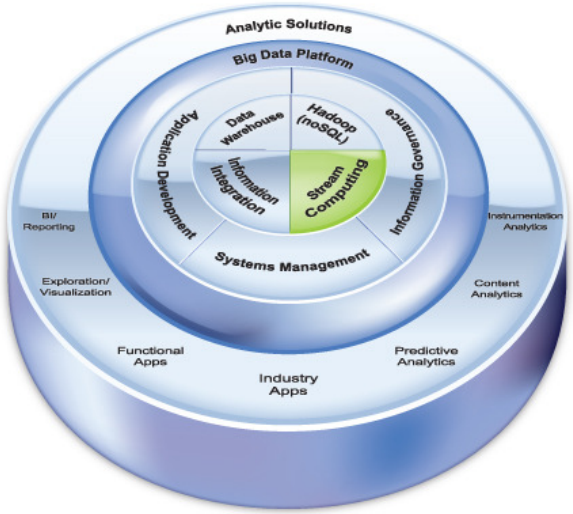
Quick and easy sandbox to explore data and determine its value

Forrester Wave™: Enterprise Hadoop Solutions, Q1 '12

Feb 2012 **"The Forrester Wave™: Enterprise Hadoop Solutions, Q1 2012"**



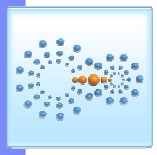
IBM Big Data Platform - Streams



Stream Computing

A platform for managing information streams and applying big data analytics to data in motion. Analyze a wide variety of data in its native format, at a massive volume and scale (terabytes per second).

Use Cases



Analyze Information in Motion

Low-latency analysis of streaming information

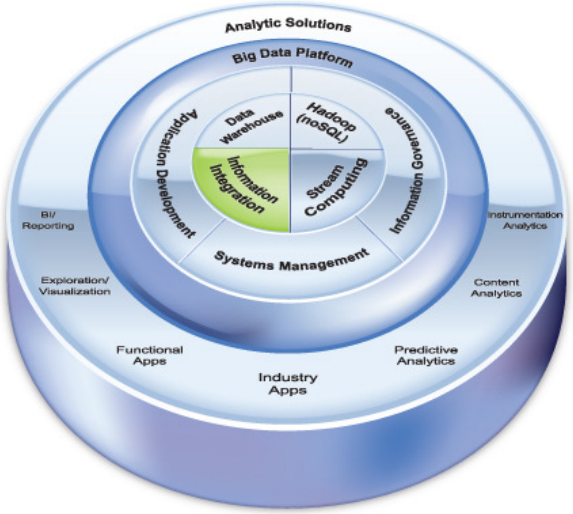
Analyze Extreme Volumes of Information in Motion

Terabytes per second, petabytes per day.

Analyze a Variety of Information

Analyze a variety of data in its native format – streaming audio, video, spatial, among others

IBM Big Data Platform – Information Server



Information Integration

High performance integration on a large volume of information for Data warehouses, Hadoop systems, streaming information sources, and transactional systems

Data transformation & remediation to ensure the appropriate level of quality

Use Cases

Use common infrastructure and reuse business logic for:

Extract and load data into Hadoop and traditional systems

Load data/insights from Hadoop and Streaming Analytics to the warehouse for deeper analysis

IBM Big Data Platform

Analytic Knowledge Accelerators

- Text Analytics, Sentiment Analytics
- Statistical & Predictive Analytics (ML)
- Entity Integration Analytics

Enterprise Robustness

- Visualization & Application Tools
- Cluster and Workload Management
- Security & Governance

Integration

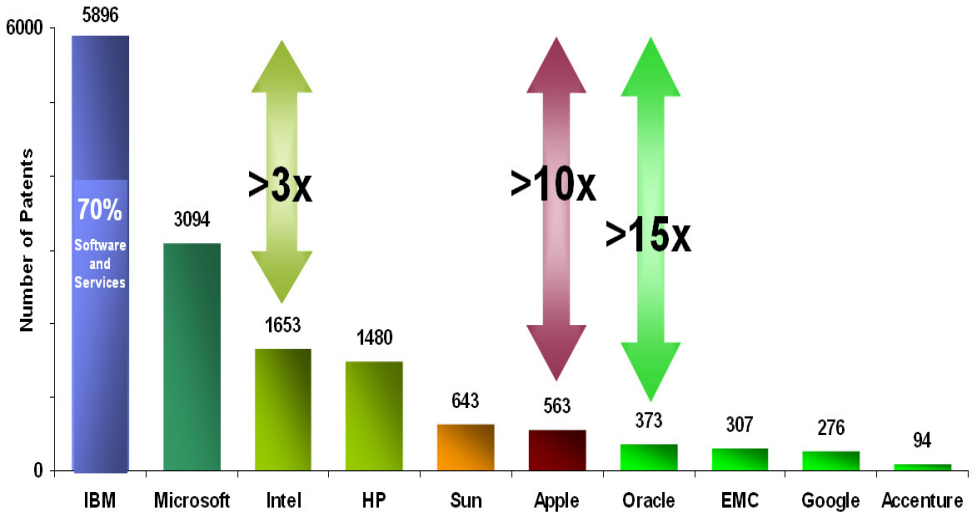
- With current sources
- With new Big Data sources

IBM's commitment to Big Data

- $100M new investment into analytics announced 2Q11
- IBM spent $14+ billion in 24 analytics acquisitions in 5 years
- IBM has the largest commercial research organization on Earth
  - 200+ mathematicians developing breakthrough analytics
- Largest patent portfolio in the industry

NYConvergence
Digital Media Technology in the Tri-State Region

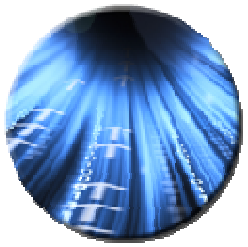
Over the last five years, the company spent $14 billion on the acquisition of two dozen data tools companies. IBM believes its future relies on helping customers manage and learn from the large amount of data available today. The company is currently working on integrating its system, **Watson**, into the health care field as a physician's assistant by feeding it medical specific domain information.



For 18 consecutive years
IBM inventors have received the most U.S. patents

5,000 patents in a single year

IBM's unique strengths in Big Data



1. *Big Data in Real-Time*

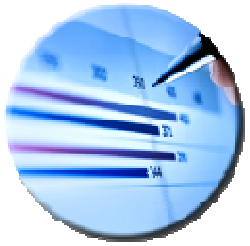
- Ingest, analyze and act on massive volumes of streaming data.
- Faster AND more cost-effective for specific use cases. (10x volume of data on the same hardware.)

2. *Fit for purpose analytics*

- Analyzes a variety of data types, in their native format – text, geospatial, time series, video, audio & more.

3. *Enterprise Class*

- Open source enhanced for reliability, performance and security.
- High performance warehouse software and appliances
- Ease of use with end users, admin and development UIs.



4. *Integration*

- Integration into your IM architecture.
- Pre-integrated analytic applications.