



Transformation in the Era of Big Data and analytics

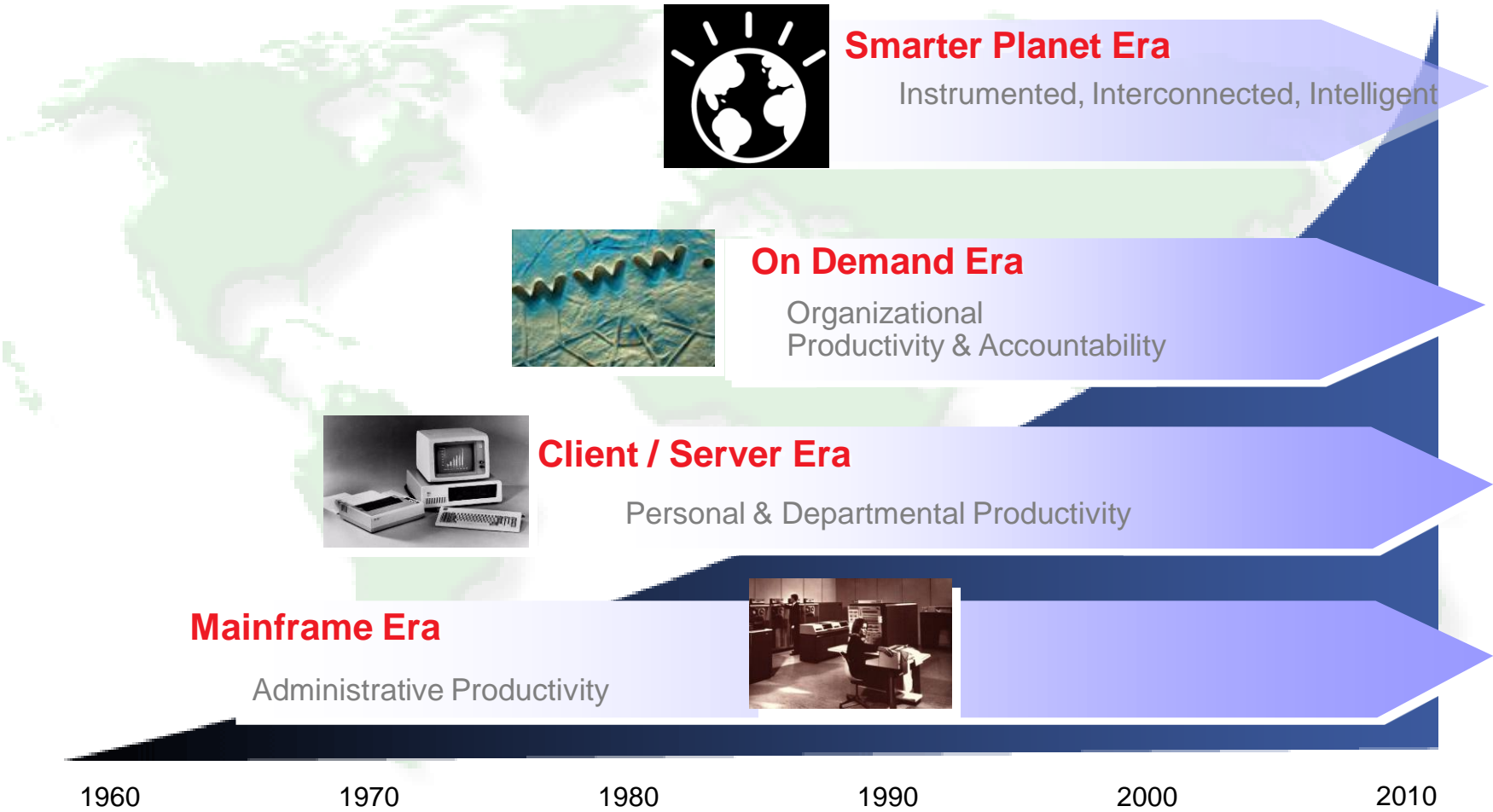
IBM InfoSphere Streams IBM BigInsight and IBM Netteza

Information On Demand

Gain Insight. Optimize Results.



Eras of Information Technology Evolution



Source: U.S. Department of Commerce



The World is Changing and Becoming More...



INSTRUMENTED



INTERCONNECTED



INTELLIGENT



The resulting explosion of information creates a need for a new kind of intelligence

...to help build a Smarter Planet

The Social Layer in a Instrumented Interconnected World



12+ TBs
of tweet data
every day

30 billion RFID tags
today
(1.3B in 2005)

4.6 billion
camera
phones
world wide

? **TBs** of
data every day

**100s of
millions of
GPS
enabled
devices** sold
annually



25+ TBs of
log data every day

76 million smart meters in
2009...
200M by 2014

2+ billion
people on
the Web by
end 2011

http





Extract Intent, Life Events, Micro Segmentation Attributes

The screenshot shows the Facebook interface with a search bar and navigation links. Below the navigation bar, there are four rows of profile pictures and their corresponding intent labels:

- Profile picture of a woman wearing a yellow hat with a blue and white pattern. Label: **Name, Birthday, Family**
- Profile picture of a black and white dog. Label: **Not Relevant - Noise**
- Profile picture of a family of four people. Label: **Monetizable Intent**
- Profile picture of a red Angry Bird character. Label: **Not Relevant - Noise**

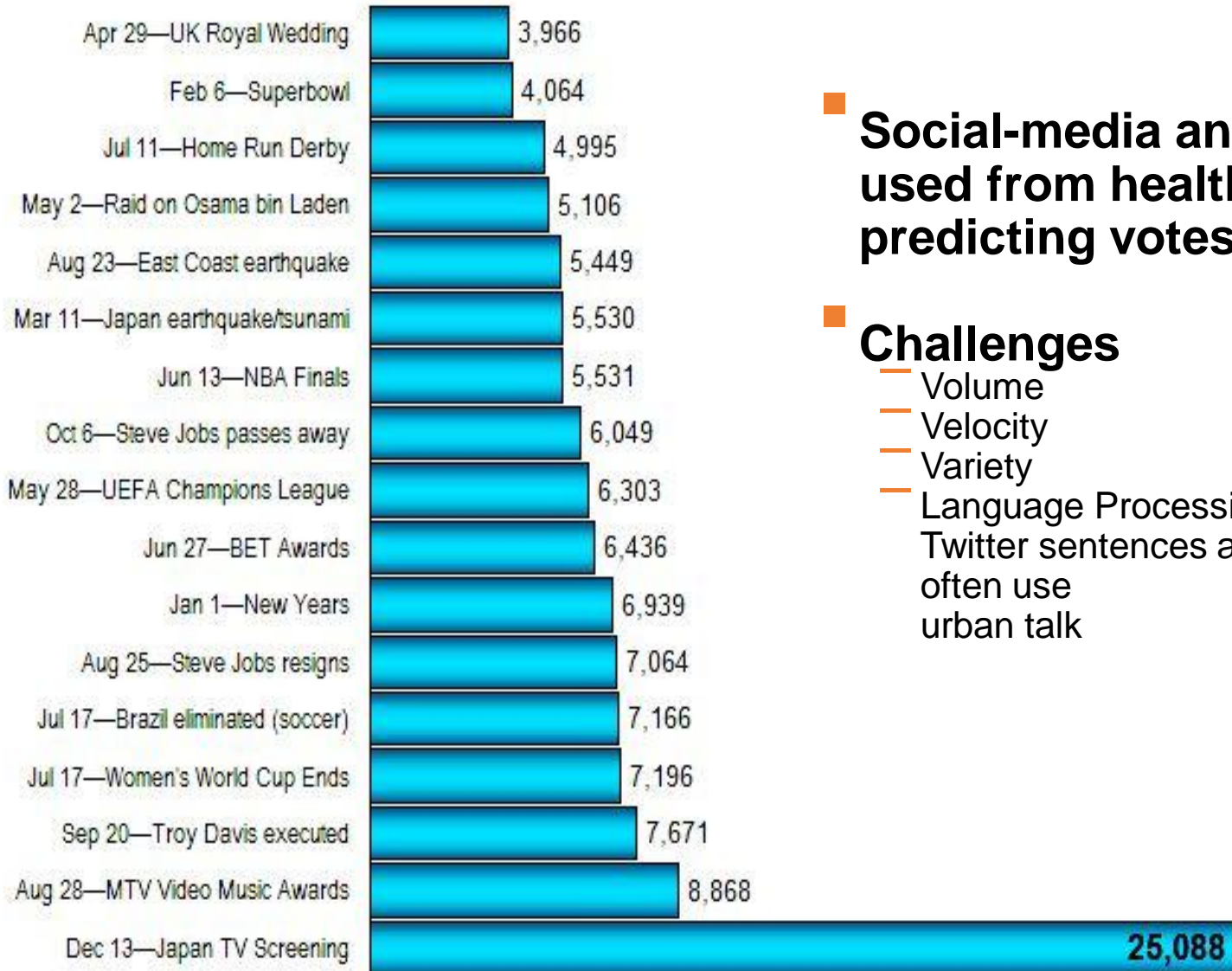
The screenshot shows the Twitter interface with a search bar and navigation links. Below the navigation bar, there are four tweets and their corresponding intent labels:

- Tweet: "What's happening?". Label: **Location**
- Tweet: "What's happening?". Label: **Wishful Thinking**
- Tweet: "What's happening?". Label: **Relocation**
- Tweet: "What's happening?". Label: **SPAMbots**





Twitter Tweets per Second Record Breakers of 2011



■ **Social-media analytics can be used from healthcare to predicting votes**

- **Challenges**
- Volume
 - Velocity
 - Variety
 - Language Processing: consider that Twitter sentences are not well formed and often use urban talk



Traditional and New (Big Data) Approaches



Traditional Approach *Structured & Repeatable Analysis*

Business Users

Determine what question to ask



IT

Structures the data to answer that question

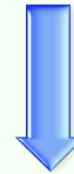


Monthly sales reports
Profitability analysis
Customer surveys

Big Data Approach *Iterative & Exploratory Analysis*

IT

Delivers a platform to enable creative discovery



Business

Explores what questions could be asked



Brand sentiment
Product strategy
Maximum asset utilization

The Social Layer in a Instrumented Interconnected World



12+ TBs
of tweet data
every day

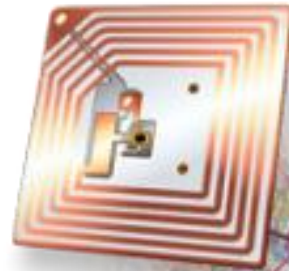
? TBs of
data every day



25+ TBs of
log data every day



30 billion RFID
tags today
(1.3B in 2005)



4.6
billion
camera
phones
world
wide



100s of
millions
of GPS
enabled
devices sold
annually



76 million smart meters in
2009...
200M by 2014

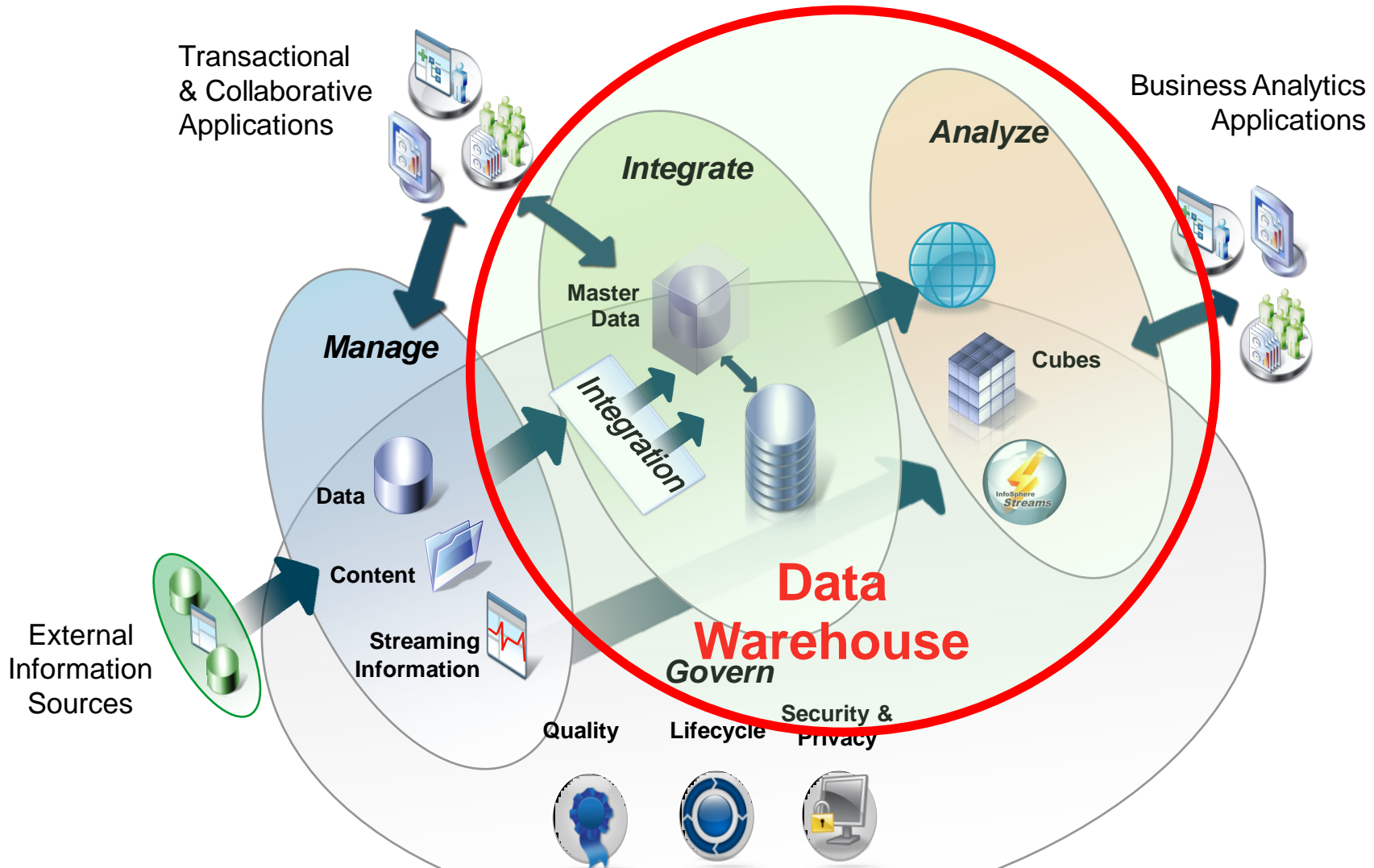


2+
billion
people on
the Web by
end 2011



Data Warehouse and Business Intelligence

A Flexible Platform for Managing, Integrating, Analyzing and Governing Information



Streams and Oceans of Information



Information streams

High speed information flowing in real-time, often transient

- Information from sensors, instruments, etc.
- Information flowing from real-time logs and activity monitors
- Streaming content like audio and video
- High speed transactions like tickers, trades, or traffic systems



Information oceans

Information stored outside conventional systems. Data may originate from the Web or different internal different systems

- Collection of what has streamed
- Information from social media, logs, click streams, emails, etc.
- Unstructured or mixed schema documents like claims, forms, desktop applications, etc.
- Structured data from disparate systems

What is BIG Data?

The processing of an **immense volume** and **variety** of data for the purpose of analytics (generating insight), at a **velocity** that was previously impossible.



Volume

- Scale from terabytes to zettabytes

Variety

- Variable structures, structured and unstructured

Velocity

- Streaming data and large volume data movement

To derive insight from previously untouched data and integrate that insight into your business operations – data warehouses, business processes, and applications.

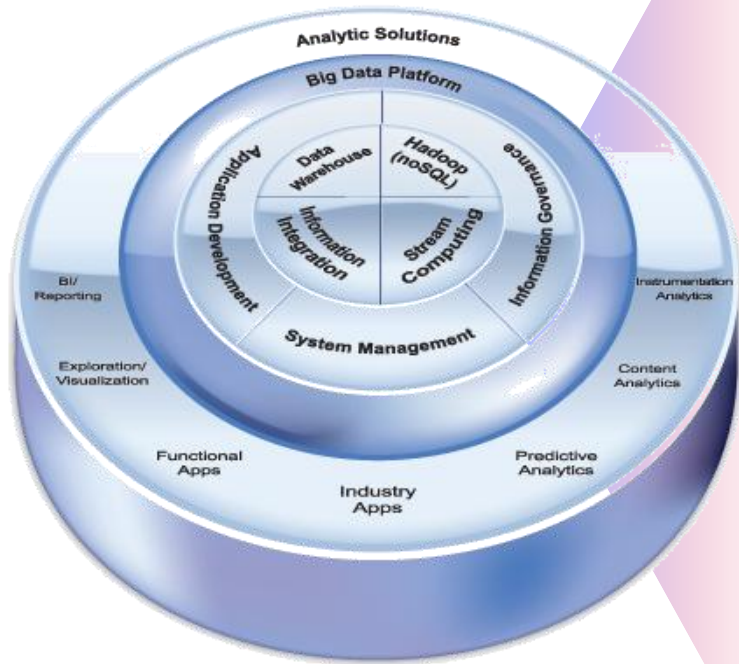
Challenges

- **Not clear** what should be analyzed (exploratory)
- **Lots of potentially valuable data** is dormant or discarded due to size/performance considerations
- Information distributed across multiple systems and/or Internet
- **Large volume of unstructured data** is not worth integrating fully (e.g. Tweets)
- Some information has a **short useful lifespan**
- Volumes can be extremely high
- Analysis needed in the **context** of their existing information (not standalone)



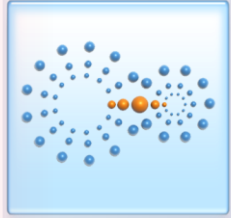


What Does a Big Data Platform Do?



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 and ad-hoc analysis



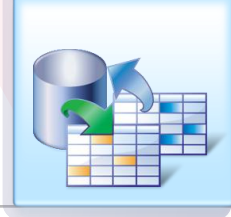
Analyze Extreme Volumes of Information

Cost-efficiently process and analyze PBs of information
Manage & analyze high volumes of structured, relational data



Discover and Experiment

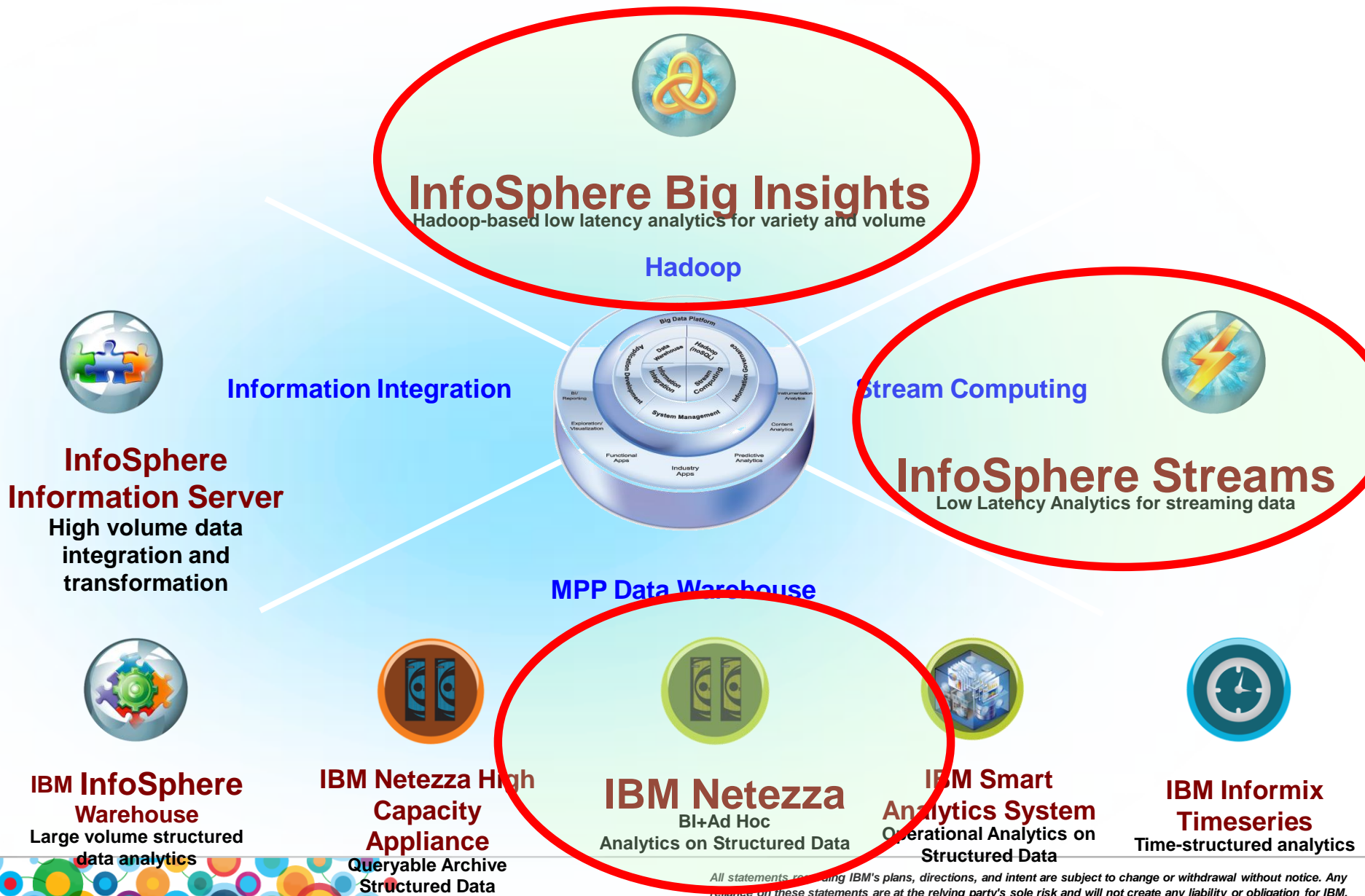
Ad-hoc analytics, data discovery and experimentation



Manage and Plan

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

The IBM Big Data Platform



All statements regarding IBM's plans, directions, and intent are subject to change or withdrawal without notice. Any reliance on these statements are at the relying party's sole risk and will not create any liability or obligation for IBM.



IBM InfoSphere Big Insights

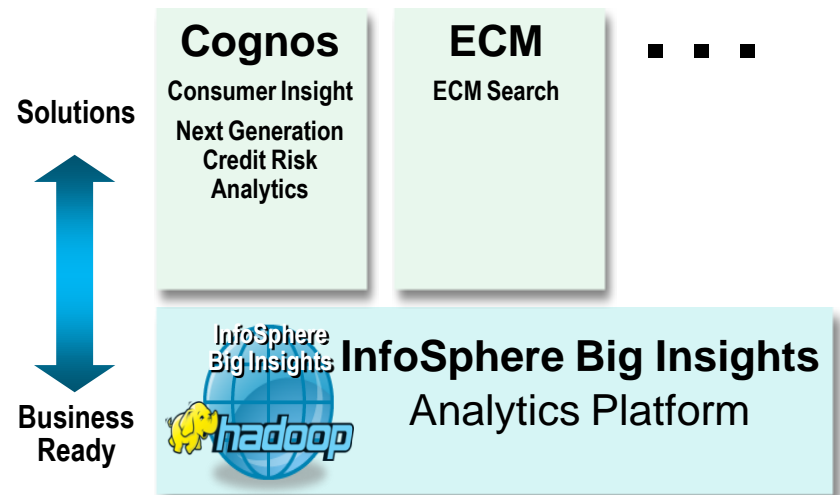
A Business Ready Platform to Collect, Store and Analyze BIG Data

Enables solutions for *Big Data*, cost effectively

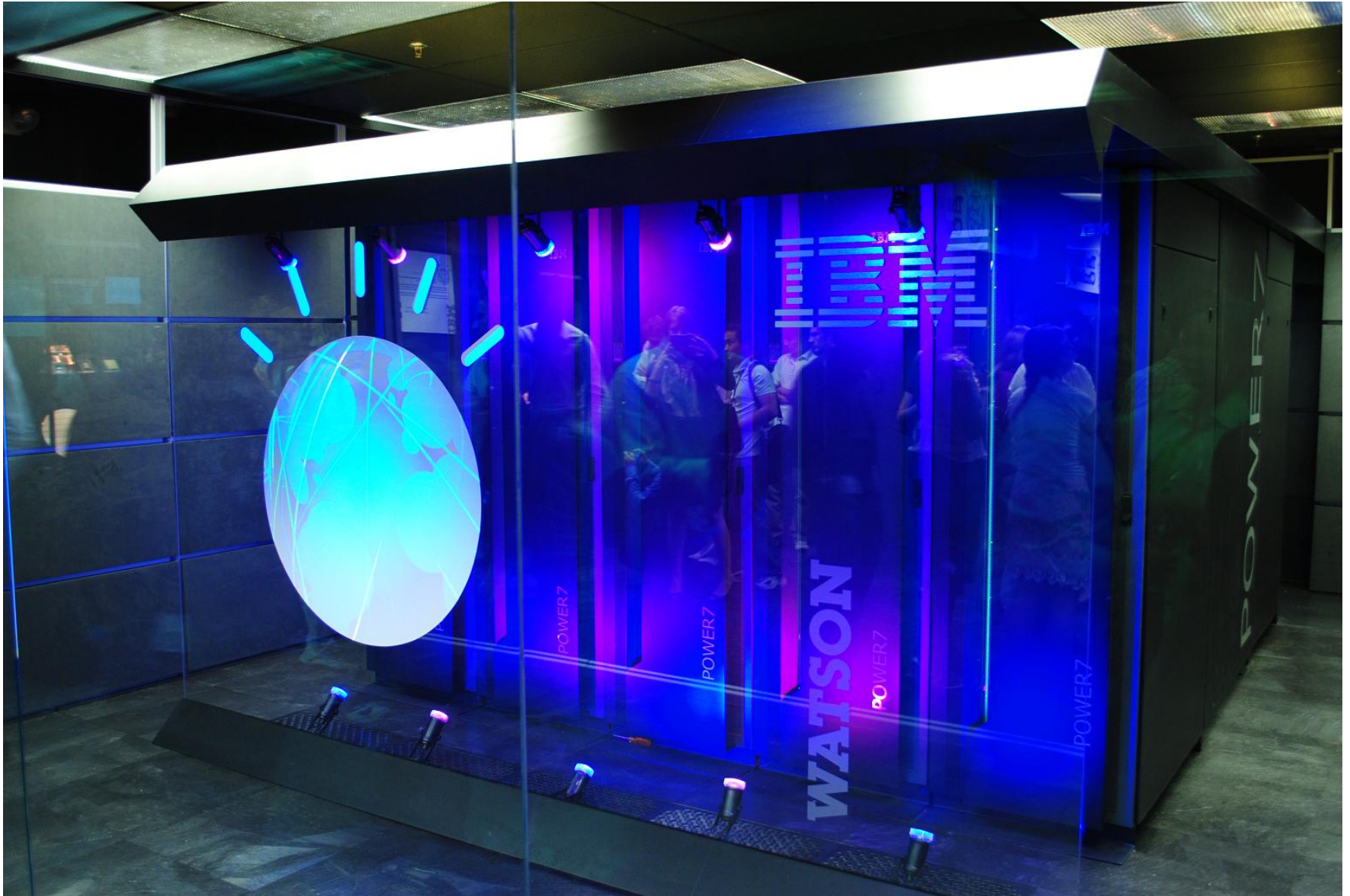
- Linearly scalable on commodity hardware, terabytes to petabytes
- Supports unstructured and structured information
- Based on Hadoop Open Source framework, fully supported by IBM

Business Ready

- *InfoSphere BigSheets* – General purpose solution environment to gather, explore and visualize
- *Advanced Analytics* – Integrates with Cognos, SPSS, LanguageWare



IBM's Watson : The Future of Computing



All statements regarding IBM's plans, directions, and intent are subject to change or withdrawal without notice. Any reliance on these statements are at the relying party's sole risk and will not create any liability or obligation for IBM.

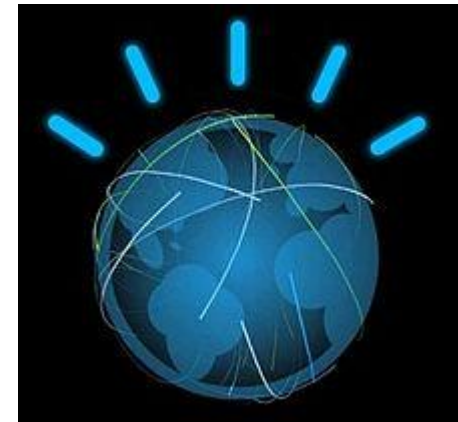


IBM's Watson : The Future of Computing

IBM Power 750 servers

No. of Server	90
No. of Racks	10
CPU Speed	3.5 GHz
No. of Cores per Processors	8
No. of Threads per Cores	4
Total no. of cores	2880 cores
Memory	16 TB

Computing power 80 TeraFLOPs
49th in the Top 50 Supercomputers list





IBM's Watson : The Future of Computing

Software

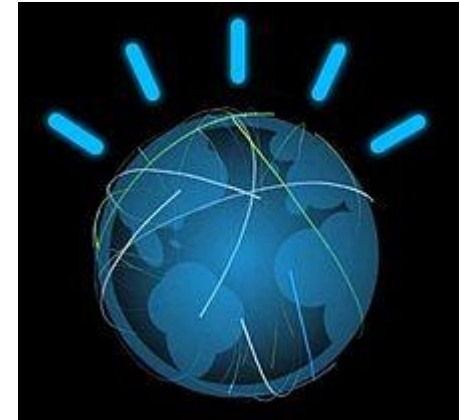
Java and C++

IBM Big Insights (Hadoop framework)

UIMA (Unstructured Information Management Architecture) framework

IBM's DeepQA software

SUSE Linux Enterprise Server 11



IBM





IBM's Watson : The Future of Computing

The sources of information

Encyclopedias

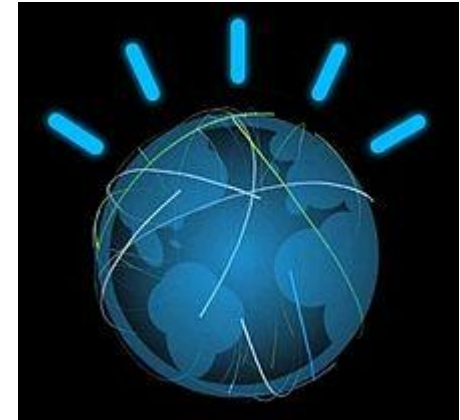
Dictionaries, Thesauri, Newswire Articles, and Literary works.

Wikipedia, DBPedia, WordNet, and Yago,

.....

No. of Pages	200 million pages
Size of Information	4 TB.

No connection to the Internet during the game.



IBM

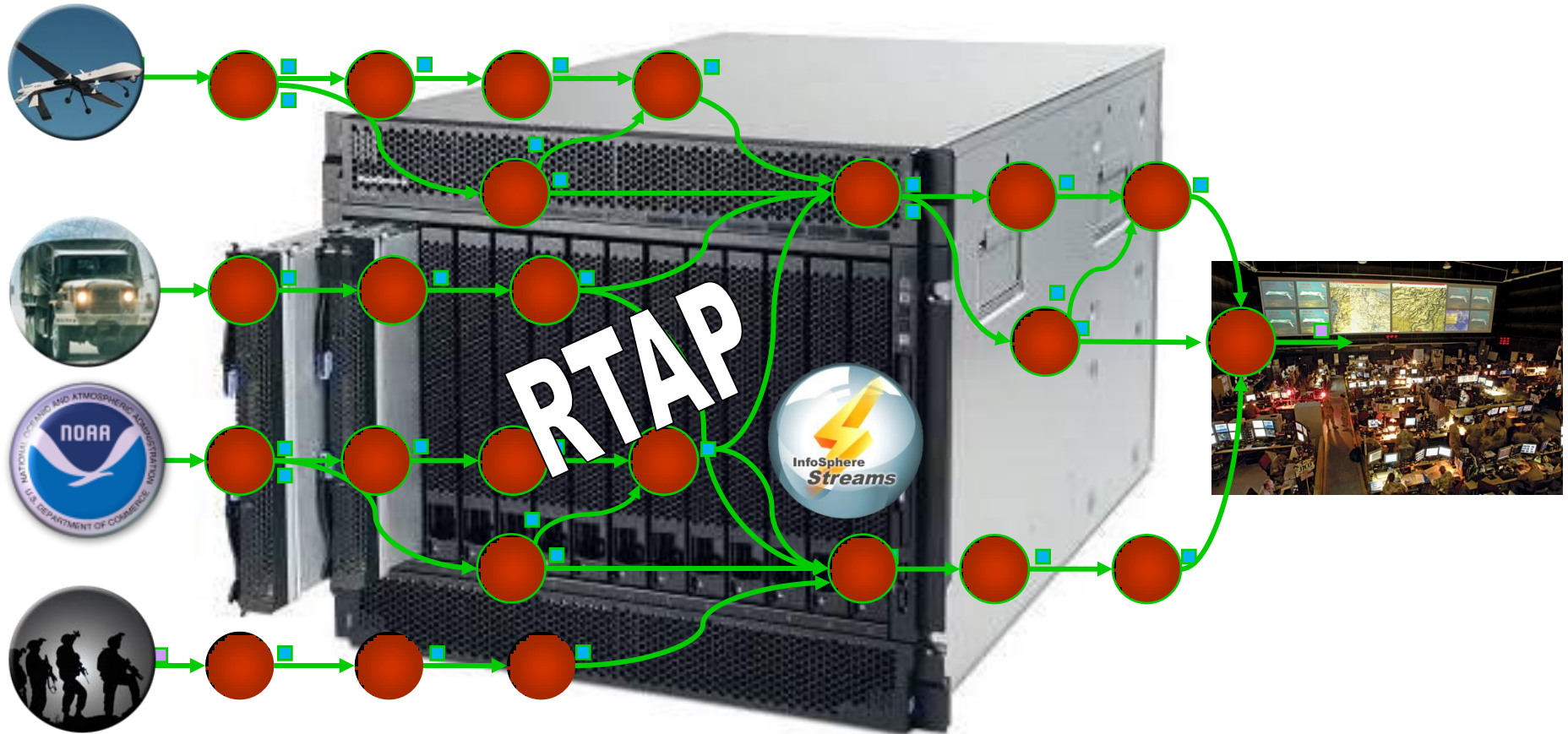


InfoSphere Stream: In-Motion analytics

In-Motion analytics for High throughput and Ultra-low latencies

Continuous Ingestion

Continuous Queries /Analytics on data in motion



IBM Netezza Data Warehouse & Analytic Appliance

Simple Appliance for Deep Analytics

- Appliance simplicity
- Integrated database, server and storage
- 10-100x faster than traditional systems
- Purpose-built engine for analytics
- Petascale data capacity
- Standard interfaces

Netezza.
Up and running in 24 hours, not 24 days.

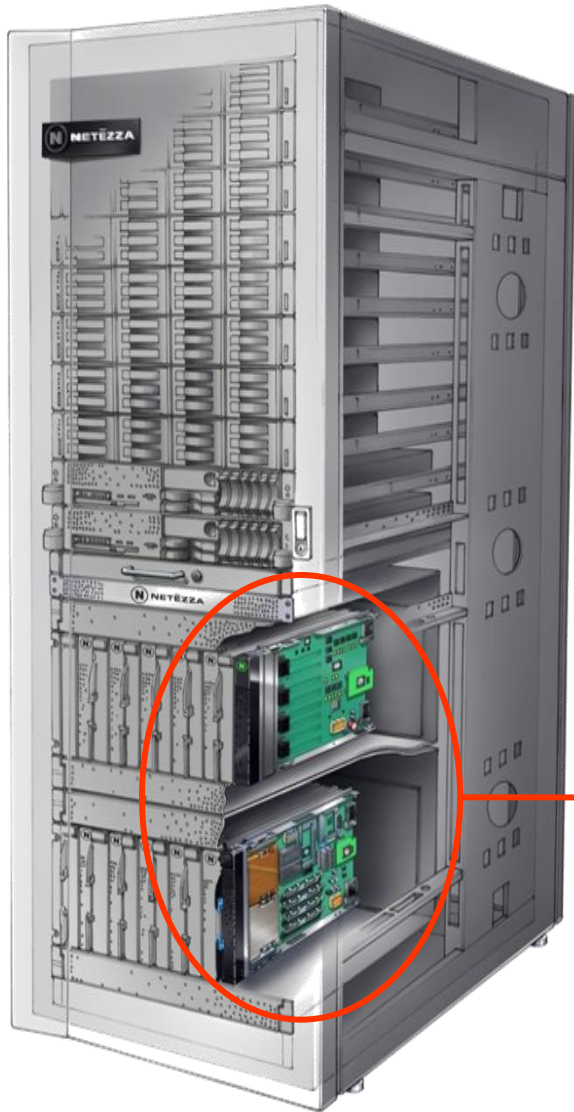
Get set up in hours instead of days, and start counting returns in minutes instead of hours. All with the IBM Netezza data warehouse appliance for high-performance analytics. It gives you analytics reports at supersonic speeds. At a fraction of the cost of Oracle Exadata. Get real, actionable business results fast.

ibm.com/facts

COST comparison based on publicly available information as of 2/20/2011 for an Oracle Exadata X2-2M full rack and a Netezza 250T1. The cost to acquire Netezza appliances may include additional services and software as well as other variables. Oracle Exadata hardware, IBM and IBM logo, Netezza, Simple, Fast and the green eye are trademarks registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. © Licensed to IBM under the IBM International Business Machines Corporation logo.



The Netezza S-Blade™



Complementary Analytics

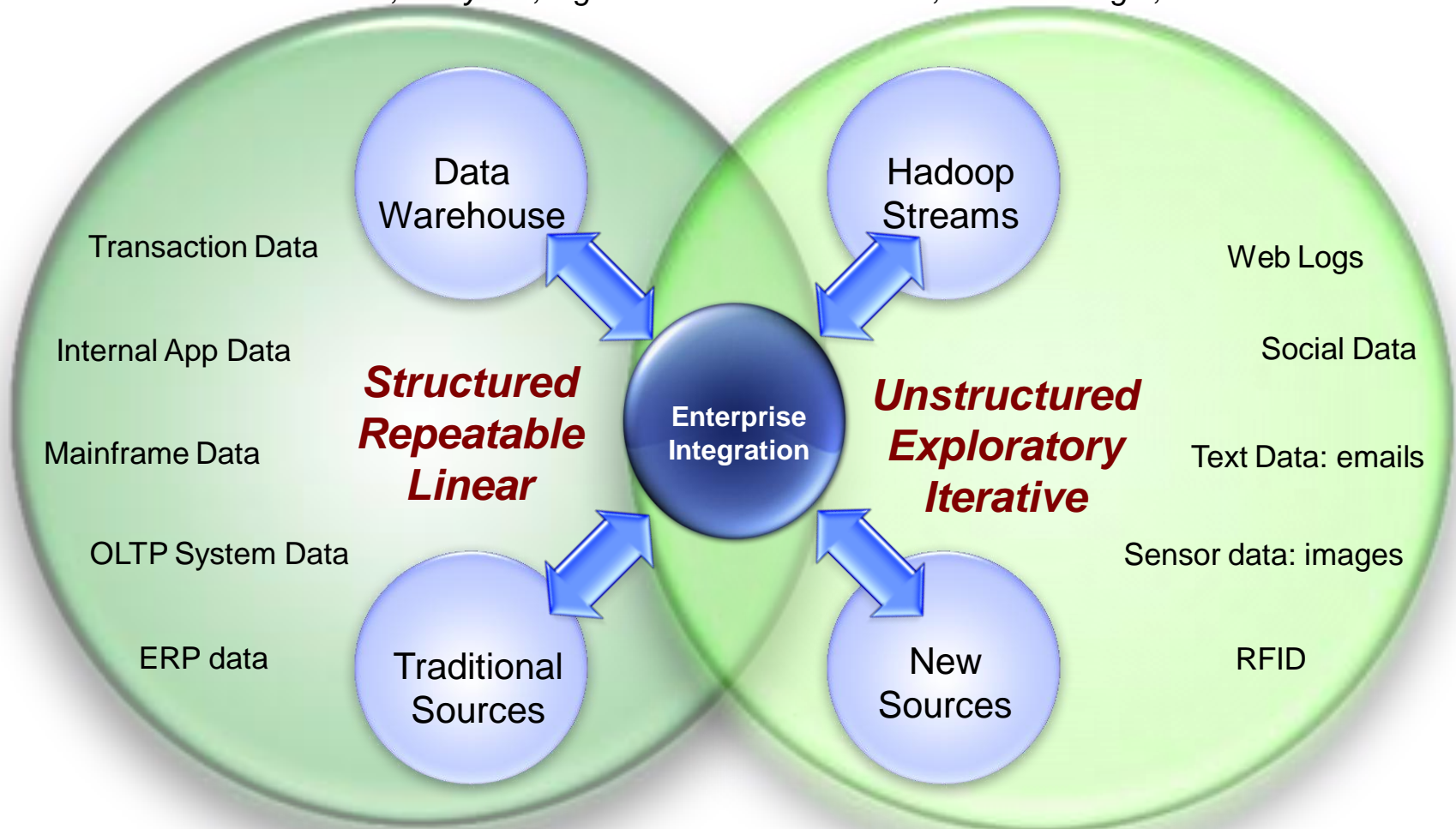


Traditional Approach

Structured, analytical, logical

New Approach

Creative, holistic thought, intuition



Catalina Marketing

- At checkout, real-time analysis of current transaction and customer's past purchasing history triggers printouts of customer specific offers and promotions
- **2.5 petabytes of data;** 300 million retail transactions added per week
 - 195 million shopper households and 400+ billion market basket records
- By doing **in-database analytics** on Netezza, move of “big data” into slower processing analytic server avoided
 - Models that previously took **half a day to process can now be scored within 60 seconds**



CATALINA MARKETING®



Banco Bilbao Vizcaya Argentaria (BBVA)



- Deployed IBM Cognos Customer Insight, *based on IBM InfoSphere BigInsights and Apache Hadoop*, to analyze Internet and Social Media sentiment (*5.8 terabytes of data*) about the bank, enabling BBVA to work on negative opinions, build on positive ones, and generally improve the bank's reputation
- BBVA can now **"listen to the voice of its clients"** and understand what they think about the bank's services, publicity campaigns, new products and customer service from a totally new viewpoint

BBVA

IBM Cognos Consumer Insight

The Hertz Corporation



- Implemented enterprise feedback management system to automate categorization of **text-based customer feedback** from thousands of web surveys, emails and text messages in real-time
- **Unstructured data** is analyzed using linguistic rules **improving categorization consistency from 43% to 85%**
- Less time reading the comments and more time making operational improvements based on data previously locked inside surveys



IBM Content Analytics

Mindshare Reveal

*Business Partner:
Mindshare Technologies*

Credit Card Company



- Analyzing semi-structured authorization logs to extract new customer insights on card usage
 - **Apache Hadoop** used to manage structured and unstructured data
 - **10 terabytes of authorization logs each month**
 - Need to keep for 7 years for regulatory compliance → **840 terabytes** stored data
- Enables timely decisions for marketing opportunities as well as fraud detection/prevention

IBM InfoSphere BigInsights



Major Telecom Company

- Analyzing system log data across all layers of the IT infrastructure to improve overall system reliability
 - *10 terabytes per day of log data growing to 75 TB/day by 2015*
- Performing real-time root-cause analysis on log data when problems occur
- Performing retrospective analysis on months of system log data to improve accuracy



IBM InfoSphere Streams

IBM InfoSphere BigInsights

Applications for BIG Data Analytics are Endless



Neonatal Care



Trading Advantage



Environment



Law Enforcement



Radio Astronomy



Telecom



Manufacturing



Traffic Control



Fraud Prevention



Imagine the Possibilities ...in a World with No Limits

Information from Everywhere



- Data & content
- Apps, web & sensors
- At rest & in motion
- Integrated & federated

Radical Flexibility



- Virtualization at every level
- Automated administration
- Easy-to-use analytics

Extreme Scalability



- “Big data” analytics
- Real-time stream processing
- Efficient parallelism
- Workload-optimized

IBM Big Data - Platform for the V³



VOLUME



Manage massive volumes of data at-rest or as it is streaming into the organization.

VARIETY



Manage data & content “As Is” – regardless of source or format.

VELOCITY



Ingest & analyze millions of records per second with ultra low latencies.





THINK

BIG

BIG

