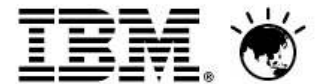**Pulse**

IBM SolutionsConnect 2013

Big Data and Analytics for Service Management
Kieran Hagan – IBM Information Management

khagan@au1.ibm.com

12/06/2013

IBM

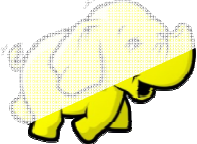# Agenda

- So what is this Big Data stuff all about anyway?

- Patterns that are taking hold

- Customer application

- A new infrastructure

- Never before possible
- Large volumes
- Unstructured Data
- Valuable insight, but difficult to extract
- It's all about Hadoop

These definitions are wrong

IBM

Resolving information management and infrastructure challenges outside traditional approaches to deliver **new business insight**

Fit For Purpose

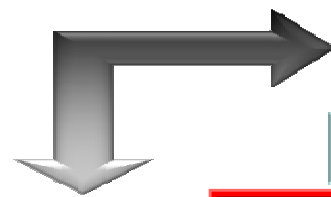**Data Virtualisation and Movement Platform** –
high speed data movement infrastructure with pre-built native connections to Data layers, Content Management, Big Data, and other types of data repositories.
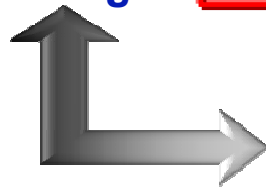
Information Catalog –
contains the metadata describing data domains, security and governance and policies

Data Fabric and Optimisation Layer –
Semantic Layer that optimises data location and storage

BI | Predictive Analytics | Prescriptive Analytics|

Analytics and Simulations

Data Virtualisation and Movement

Information Catalog

| Hadoop System | Stream Computing | Data Warehouse | Specialised Appliances | Future Platforms |

Data Fabric and Optimisation
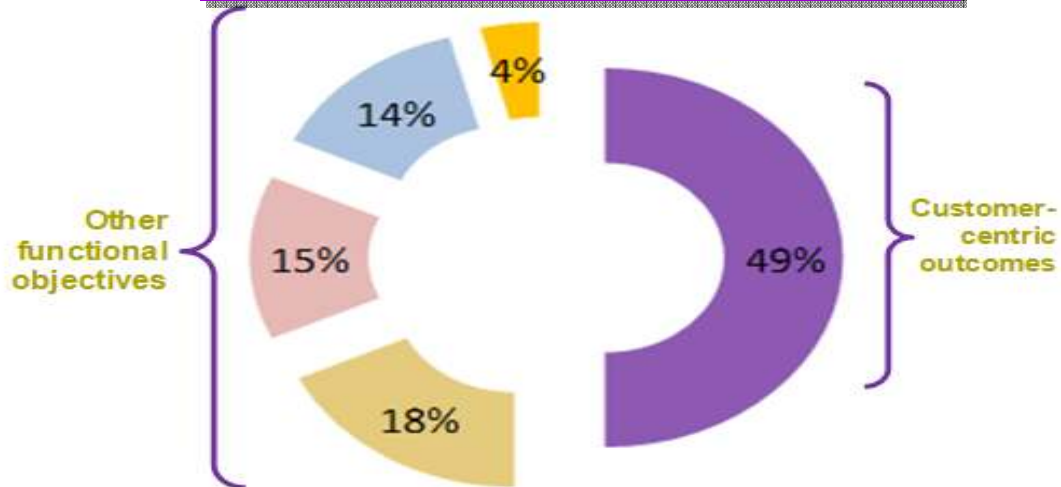
…Which enables new insight

- The idea of the Super-set
- Move from Sampling to "Absolute Knowledge"
- Combining Structured and Unstructured Analytics
- Importance of Augmented Decision Making
- Stream Computing

Where the demand is…

Big data objectives

Other functional objectives

- Customer-centric outcomes: 49%
- 14%
- 4%
- 15%
- 18%

Legend:
- Customer-centric outcomes
- New business model
- Operational optimisation
- Employee collaboration
- Risk / financial management

Top functional objectives identified by organisations with active big data pilots or implementations. Responses have been weighted and aggregated.

Big data sources

| Source	%
Transactions	88%
Log data	73%
Events	59%
Emails	57%
Social media	43%
Sensors	42%
External feeds	42%
RFID scans or POS data	41%
Free-form text	41%
Geospatial	40%
Audio	38%
Still images/videos	34%
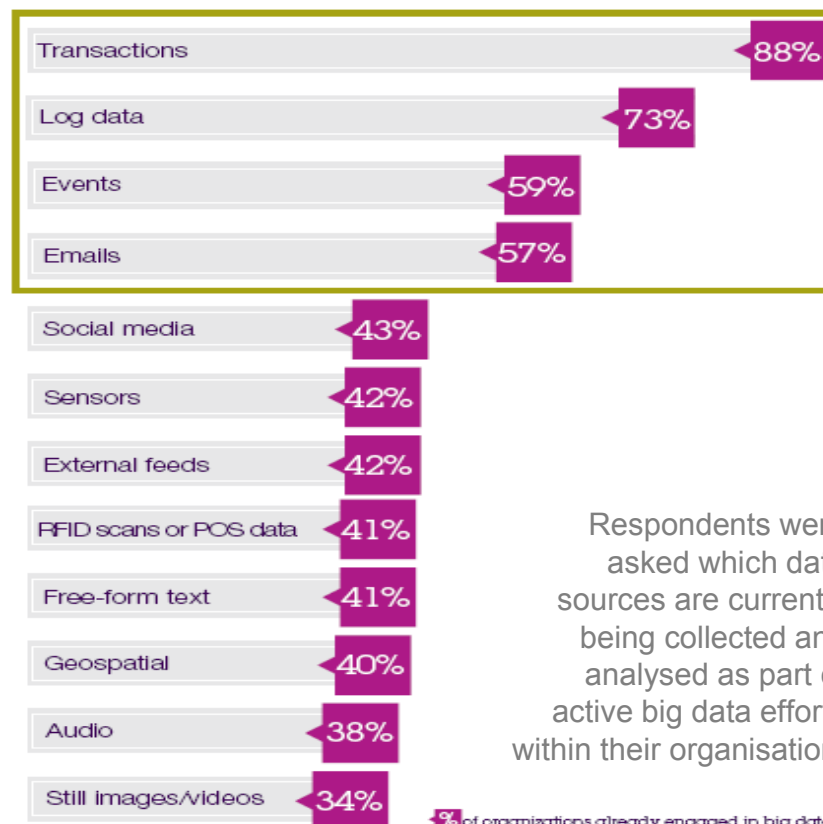
Respondents were asked which data sources are currently being collected and analysed as part of active big data efforts within their organisation.

% of organizations already engaged in big data

Pulse

IBM

Log Analysis: Problem Characteristics

Several thousand log files collected daily, data collected over several years
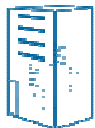
Infrastructure (Servers, Networks, Storage), Middleware (App Server, Web Server, Database Server, Messaging Server), Apps
Value in collocating and co-analysing the above data

Millions of files, petabytes of data in total, terabytes produced per day.

The relationships between logs (links shown below) have to be discovered
Large percentage of storage in an enterprise is for log data

Analysis of log data has many challenges
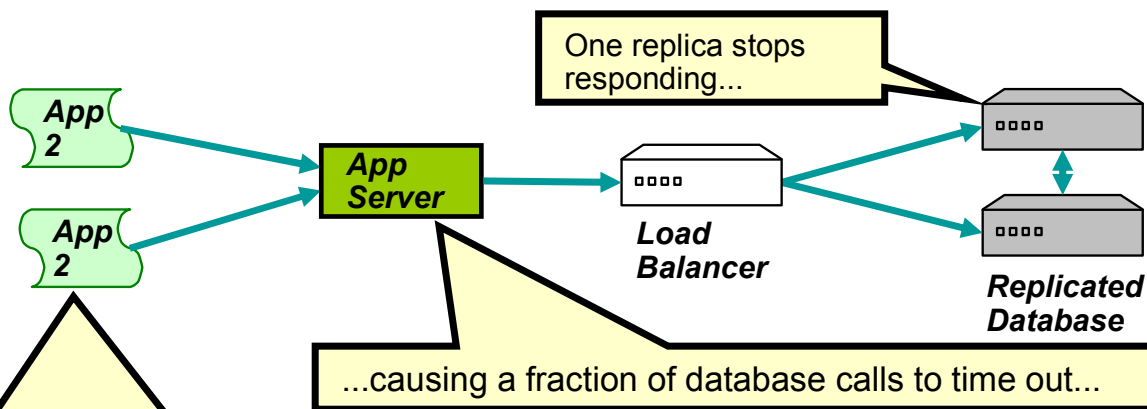
- Collection and parsing of data
- Interpretation of logs
- SMEs flooded with common bugs
- Lack of a joined up view.
- Reactive rather than proactive

App 2

App 2

App Server

Load Balancer

Replicated Database

One replica stops responding...

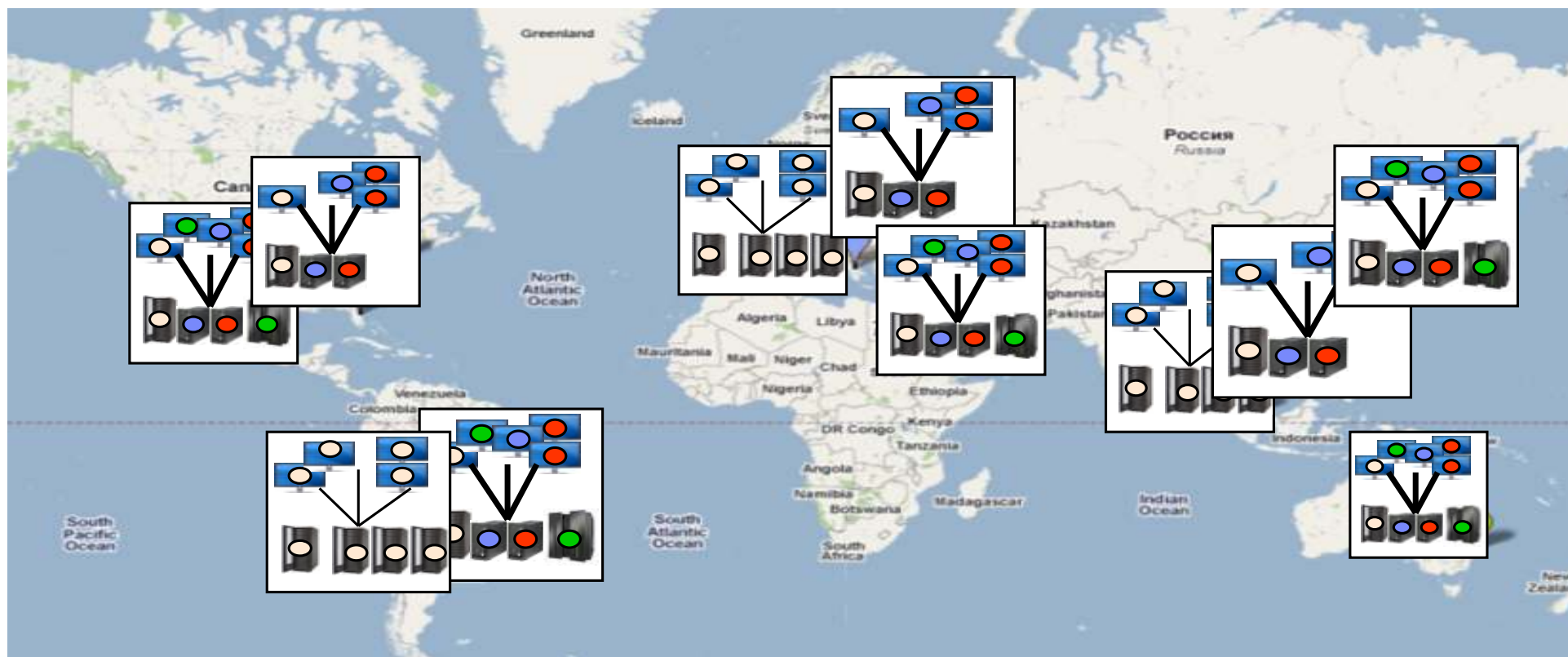...causing a fraction of database calls to time out...

...which leads to intermittent failures in the application.

IBM.

Central Lab Platform – Before
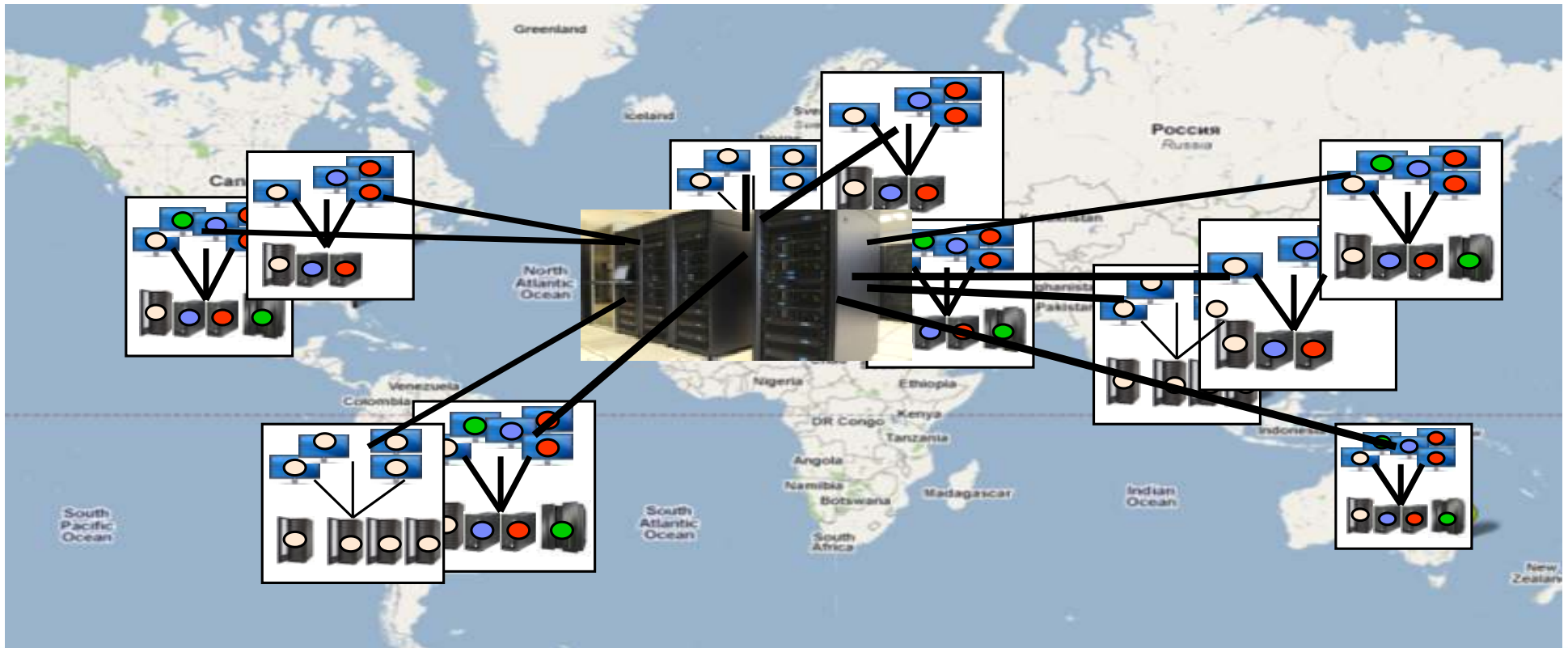
The challenges of a scattered Infrastructure are high costs and business transformation roadblocks

Scattered infrastructures can be transformed into a Centralised Cloud Platform
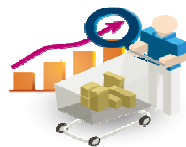
There are Many Use Cases for a Big Data Platform

Know Everything about your Customer
- Social media customer sentiment analysis
- Promotion optimisation
- Segmentation
- Customer profitability
- Click-stream analysis
- CDR processing
- Multi-channel interaction analysis
- Loyalty program analytics
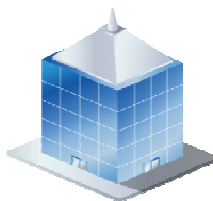- Churn prediction

Innovate New Products at Speed and Scale
- Social Media - Product/brand Sentiment analysis
- Brand strategy
- Market analysis
- RFID tracking & analysis
- Transaction analysis to create insight-based product/service offerings

Run Zero Latency Operations
- Smart Grid/meter management
- Distribution load forecasting
- Sales reporting
- Inventory & merchandising optimization
- Options trading
- ICU patient monitoring
- Disease surveillance
- Transportation network optimization
- Store performance
- Environmental analysis
- Experimental research

Exploit Instrumented Assets
- Network analytics
- Asset management and predictive issue resolution
- Website analytics
- IT log analysis

Instant Awareness of Risk and Fraud
- Multimodal surveillance
- Cyber security
- Fraud modeling & detection
- Risk modeling & management
- Regulatory reporting

Pulse

IBM

Acxiom reduces data latency with IBM in order to drive a competitive advantage

Need

- Accommodate business and data growth volumes to maintain the richest and most accurate marketing database and keep information up to data 24x7

Benefits

- Provides the capability to process 8 to 9 billion records (that is 6 to 7 TB of data) per month

- Captures over 2,000 unique elements on each U.S. household

- Reduced data latency achieved with 3-minute bulk data loads

Vestas optimises capital investments based on **3 Petabytes** of information.

Capabilities Utilised:

InfoSphere BigInsights

InfoSphere Warehouse

- Model the weather to optimise placement of turbines, maximising power generation and longevity.

- Reduce time required to identify placement of turbine from weeks to hours.

- Incorporate 3 PB of structured and semi-structured information flows.

- Data volume expected to grow to 6 PB.

Vestas

Asian Telco reduces billing costs and improves customer satisfaction

Capabilities:

Stream Computing
Analytic Accelerators

Real-time mediation and analysis of

5B CDRs per day

Data processing time reduced from

12 hrs to 1 min

Hardware cost reduced to 1/8th

Proactively address issues
(e.g. dropped calls) impacting customer satisfaction.

14

TerraEchos uses streaming data technology to support covert intelligence and surveillance sensor systems
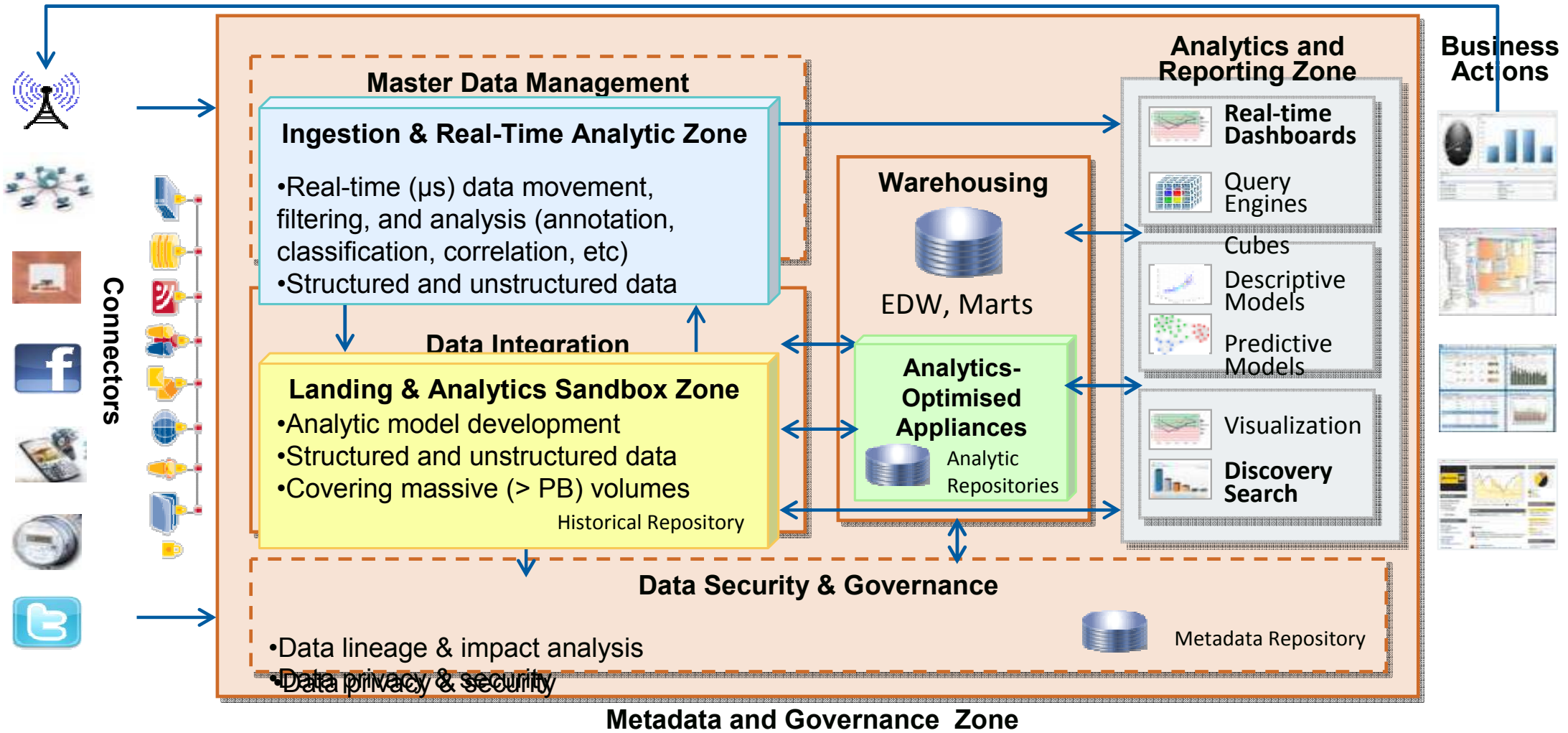
Need

- Deployed security surveillance system to detect, classify, locate, and track potential threats at highly sensitive national lab
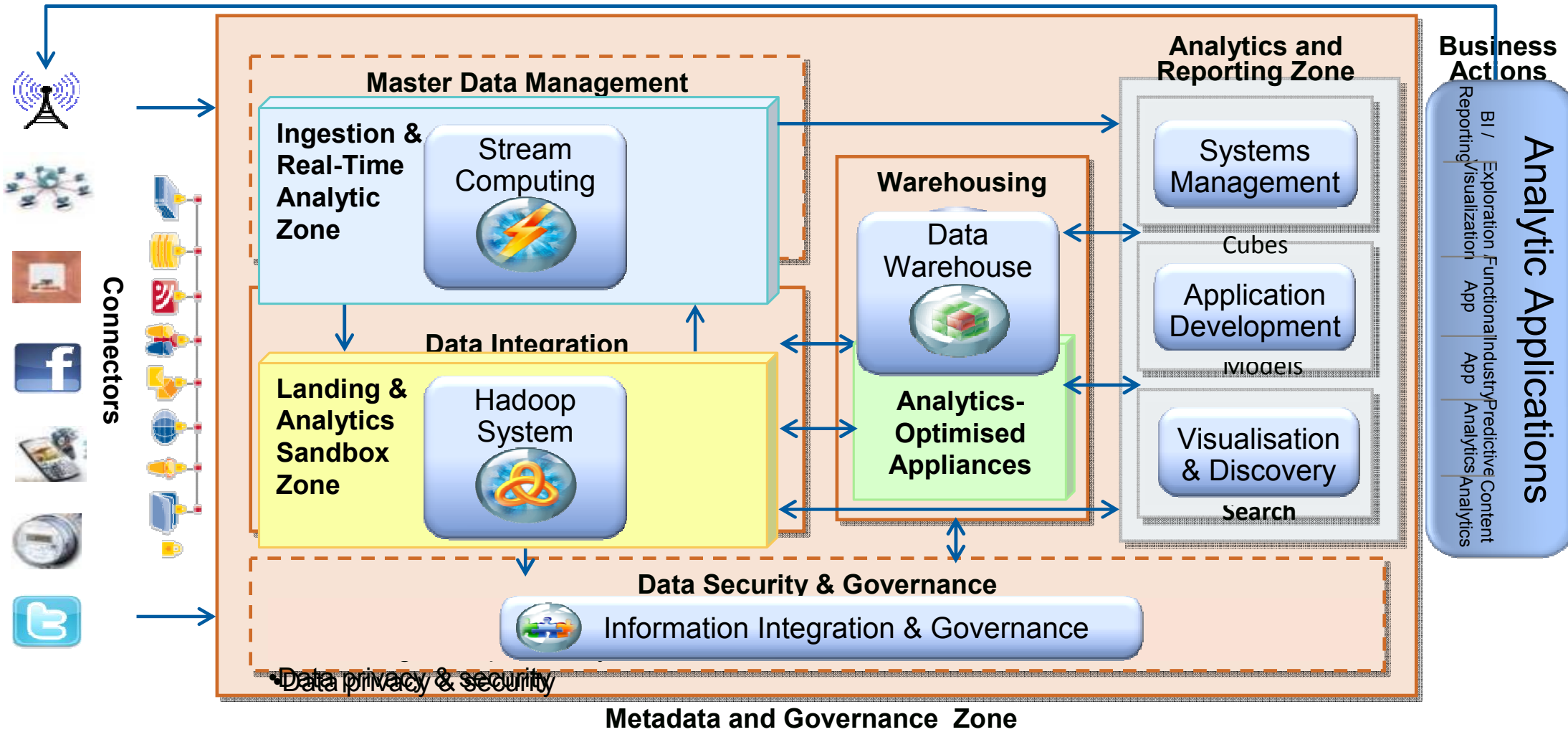
Benefits

- Reduces time to capture and analyse 275MB of acoustic data from hours to one-fourteenth of a second

- Enables analysis of real-time data from different types of sensors and 1,024 individual channels to support extended perimeter security

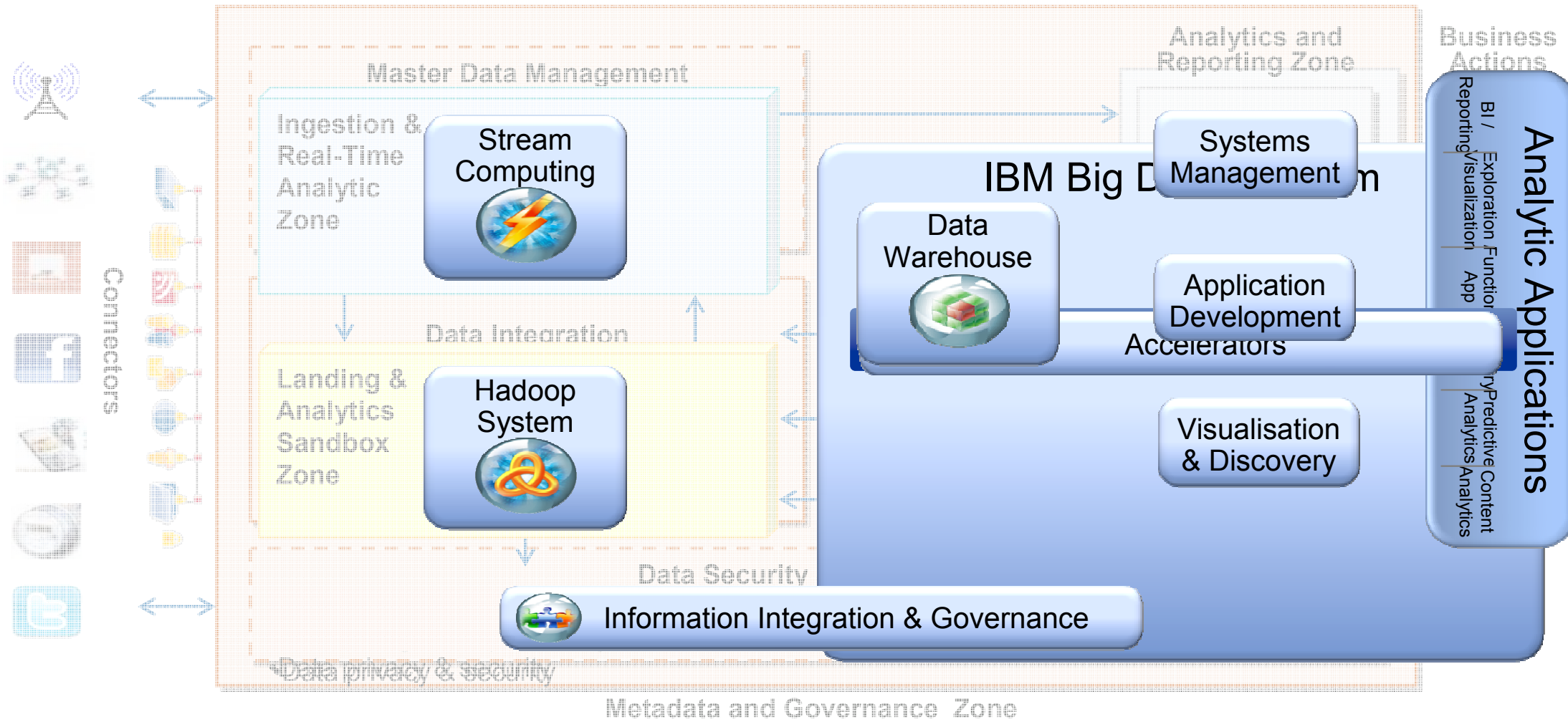- Enables a faster and more intelligent response to any threat

TerraEchos

How organisations are evolving their architecture

Connectors

Master Data Management

Ingestion & Real-Time Analytic Zone

- Real-time (µs) data movement, filtering, and analysis (annotation, classification, correlation, etc)
- Structured and unstructured data

Data Integration

Landing & Analytics Sandbox Zone
- Analytic model development
- Structured and unstructured data
- Covering massive (> PB) volumes

Historical Repository

Warehousing

EDW, Marts

Analytics-Optimised Appliances

Analytic Repositories

Analytics and Reporting Zone

- **Real-time Dashboards**
- Query Engines
- Cubes
- Descriptive Models
- Predictive Models
- Visualization
- **Discovery Search**

Business Actions

Data Security & Governance

- Data lineage & impact analysis
- Data privacy & security

Metadata Repository

Metadata and Governance Zone

IBM provides the complete platform to support this evolution

IBM provides the complete platform to support this evolution

InfoSphere BigInsights – A Closer Look

User Interfaces

- Visualisation
- Dev Tools
- Admin Console

Accelerators

- Text Analytics
- Application Accelerators

BigInsights Engine

- Map Reduce +
- Indexing
- Workload Mgmt
- Security
- Apache Hadoop

Integration

- Databases
- Content Management
- Information Governance

More *Than Hadoop*

- Performance and workload optimisations

- Unique text analytic engines

- Spreadsheet-style visualisation for data discovery and exploration

- Built-in IDE and admin consoles

- Enterprise-class security

- High-speed connectors to integration with other systems
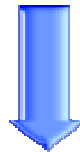
- Analytical accelerators

Merging the Traditional and Big Data Approaches

Traditional Approach
Structured and Repeatable Analysis

Business Users

Determine what question to ask

IT
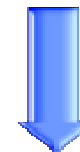
Structures the data to answer

Monthly sales reports
Profitability analysis
Customer surveys

Big Data Approach
Iterative and Exploratory Analysis

IT

Delivers a platform to enable creative discovery

Business Users

Explores what questions could be asked
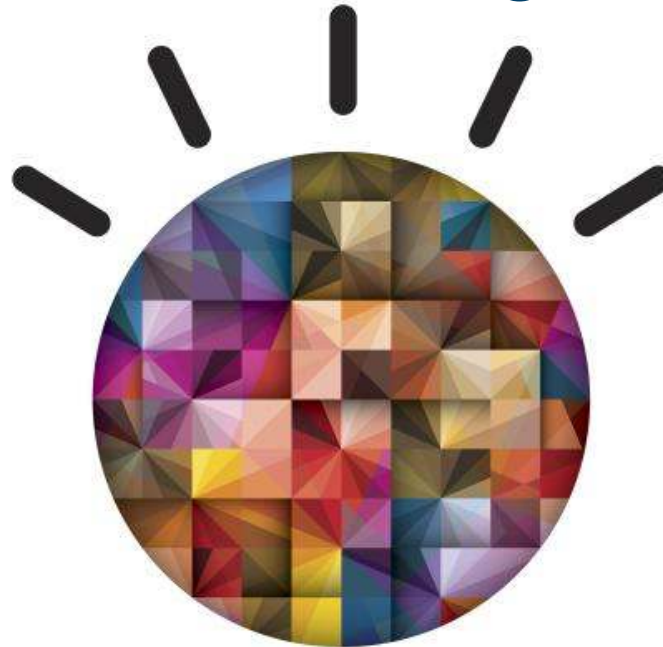
Brand sentiment
Product strategy
Maximum asset utilisation

IBM

Why Data Matters: The Age of Analytics

Visit YouTube to watch Why Data Matters – Age of Analytics

http://www.youtube.com/watch?v=f-dfWLaDBPE

For more information:
ibm.com/bigdata

#ibmbigdata

THINK
BIG