IBM SolutionsConnect

Smarter Choices for Improved IT Economics

Big Data Solutions
Overview

Frank Ketelaars
Technical Leader BigData Europe

Agenda

1. **Big Data Customer Examples**
2. The IBM Big Data Platform
3. Analyzing Data at Rest
4. Analyzing Data in Motion (at Warp speed)
5. Accelerators
6. Questions and Answers

What we hear from customers . . . .

Lots of potentially valuable data is dormant or discarded due to size/performance considerations

Large volume of unstructured or semi-structured data is not worth integrating fully (e.g. Tweets, logs, . . .)

Not clear what should be analyzed (exploratory, iterative)

Information distributed across multiple systems and/or Internet

Some information has a short useful lifespan

Volumes can be extremely high

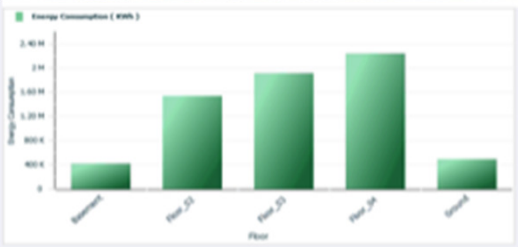
Analysis needed in the context of existing information (not stand alone)

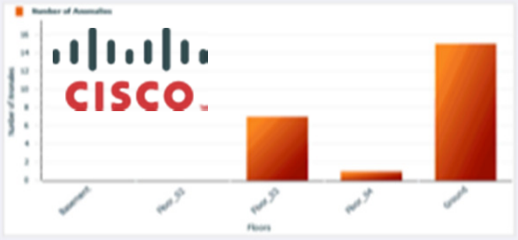
Wind Turbine Manufacturer optimizes capital investments based on 2.5 Petabytes of information.

- Model the weather to optimize placement of turbines, maximizing power generation and longevity.
- Reduce time required to identify placement of turbine from weeks to hours.
- Incorporate 2.5 PB of structured and semi-structured information flows. Data volume expected to grow to 6 PB.

Cisco turns to IBM big data for intelligent infrastructure management

- Optimize building energy consumption with centralized monitoring and control of building monitoring system
- Automates preventive and corrective maintenance of building corrective systems
- Uses Streams, InfoSphere BigInsights and Cognos
  - Log Analytics
  - Energy Bill Forecasting
  - Energy consumption optimization
  - Detection of anomalous usage
  - Presence-aware energy mgt.
  - Policy enforcement





KTH – Royal Institute of Technology analyzes real-time data streams to identify traffic patterns

Need

- Gather real-time traffic data from a variety of sources; integrate and analyze data to better manage traffic

Benefits

- Uses diverse data -- including GPS locations, weather conditions, speeds and flows from sensors on motorways, incidents and roadwork
- Analyzing large volumes of streaming data in real time is leading to smarter, more efficient and environmentally friendly traffic in urban areas

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.

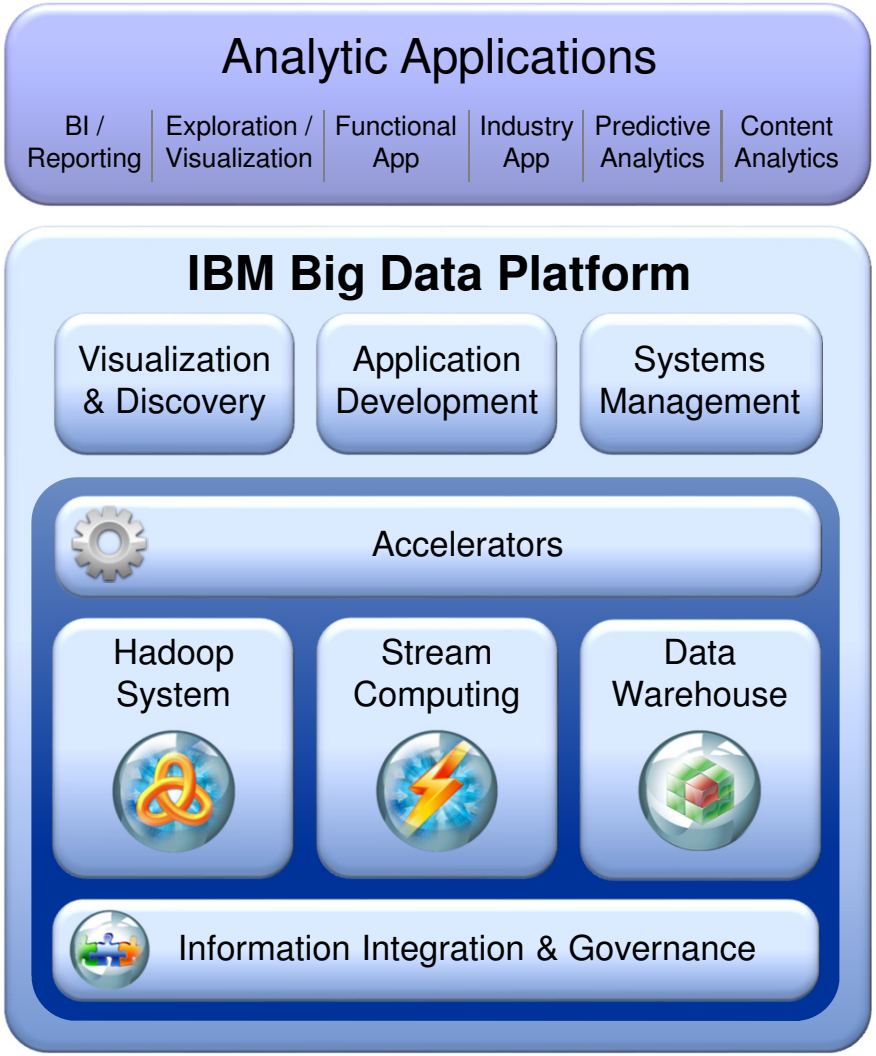
Agenda

1. Big Data Customer Examples
2. **The IBM Big Data Platform**
3. Analyzing Data at Rest
4. Analyzing Data in Motion (at Warp speed)
5. Accelerators
6. Questions and Answers

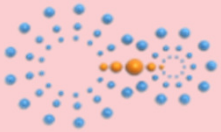
Moving the Analytics Closer to the Data

New analytic applications drive the requirements for a big data platform

- Integrate and manage the full variety, velocity and volume of data
- Apply advanced analytics to information in its native form
- Visualize all available data for ad-hoc analysis
- Development environment for building new analytic applications
- Workload optimization and scheduling
- Security and Governance



The Big Data Platform Manifesto

Understand and Navigate Federated Big Data Sources		Federated Discovery and Navigation
Manage and Store Huge Volume of any Data		Hadoop File System MapReduce
Structure and Control Data		Data Warehousing
Manage Streaming Data		Stream Computing
Analyze Unstructured Data		Text Analytics Engine
Integrate and Govern all Data Sources		Integration, Data Quality, Security, ILM, MDM

Agenda

1. Big Data Customer Examples
2. The IBM Big Data Platform
3. **Analyzing Data at Rest**
4. Analyzing Data in Motion (at Warp speed)
5. Accelerators
6. Questions and Answers

Big Data Platform - Hadoop System

- Manages a wide variety and huge volume of data
- Augments open source Hadoop with enterprise capabilities
  - Performance Optimization
  - Development tooling
  - Enterprise integration
  - Analytic Accelerators
  - Application and industry accelerators
  - Visualization
  - Security

Merging traditional and 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

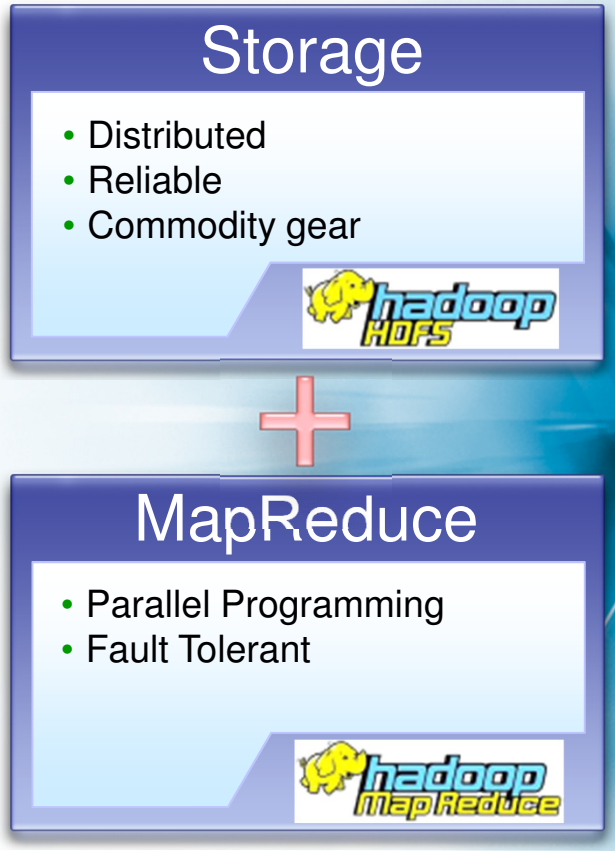
Delivers a platform to enable creative discovery

Business

Explores what questions could be asked

- Brand sentiment
- Product strategy
- Maximum asset utilization

Key points about Hadoop



Scalable

- New nodes can be added on the fly

Affordable

- Massively parallel computing on commodity servers

Flexible

- Hadoop is schema-less – can absorb any type of data

Fault Tolerant

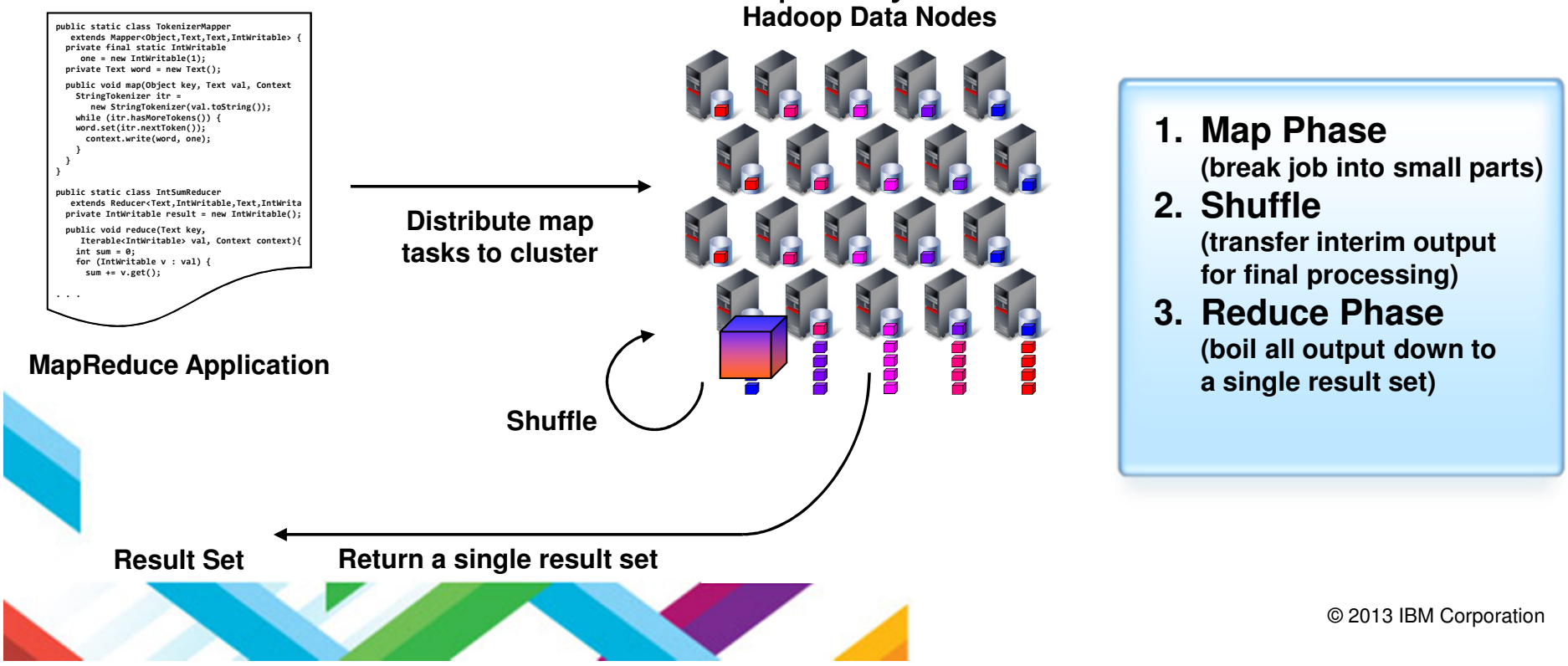
- Block Replication Factor
- Through MapReduce software framework

Hadoop Explained

Hadoop computation model

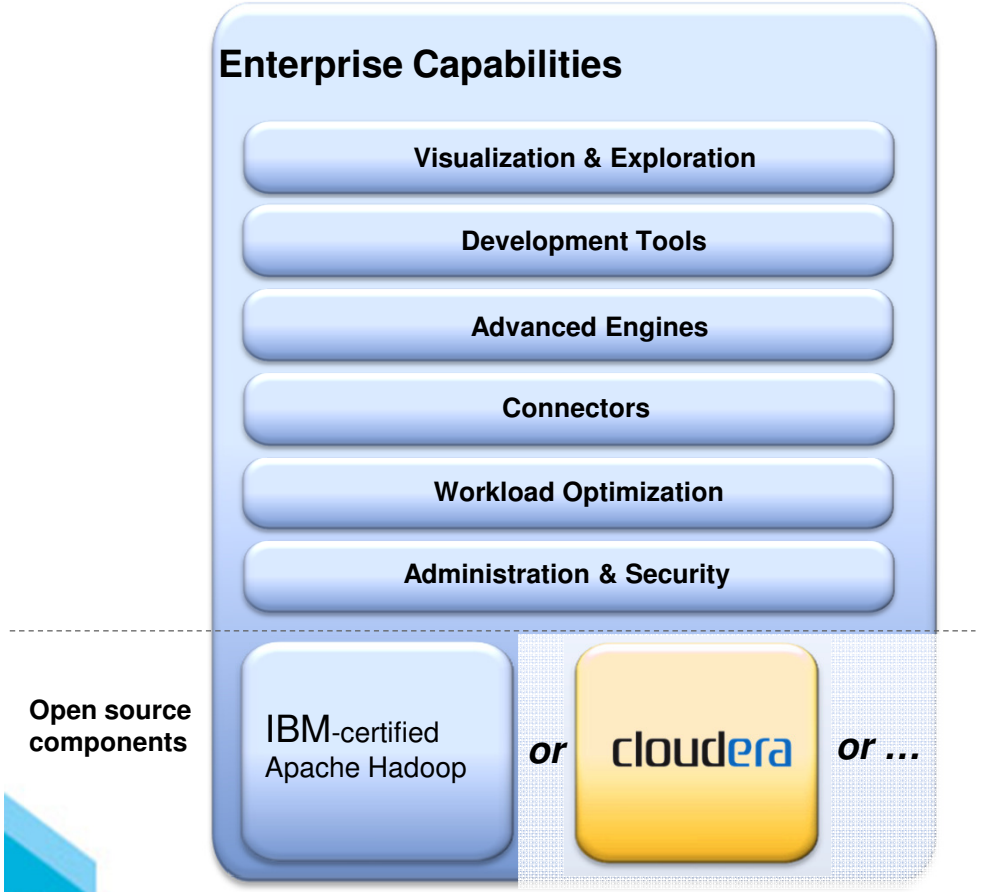
- Data stored in a distributed file system spanning many inexpensive computers
- Bring function to the data
- Distribute application to the compute resources where the data is stored

Scalable to thousands of nodes and petabytes of data



1. **Map Phase**
   (break job into small parts)
2. **Shuffle**
   (transfer interim output for final processing)
3. **Reduce Phase**
   (boil all output down to a single result set)

BigInsights: Value Beyond Open Source



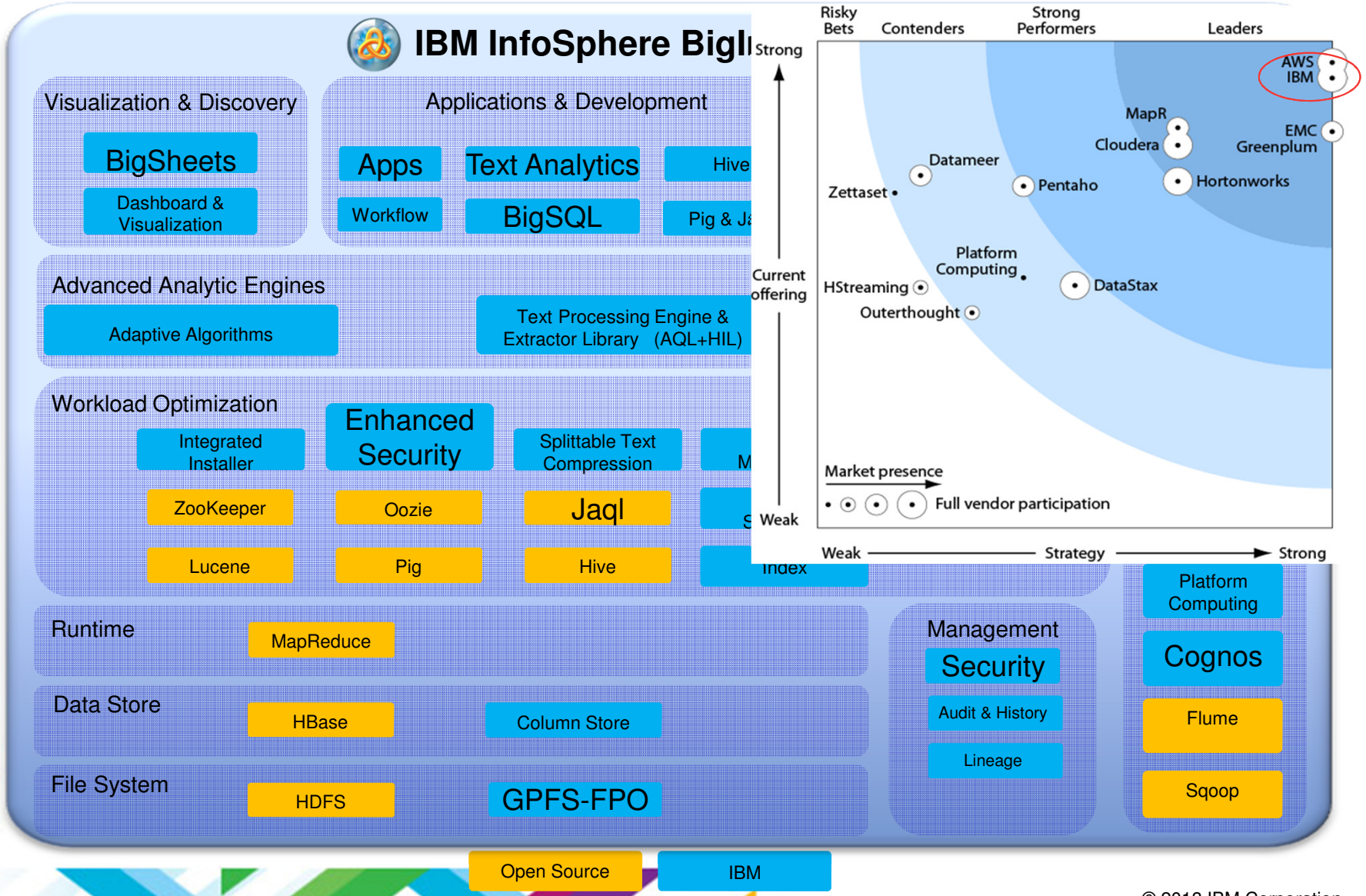
Key differentiators

- Built-in analytics
  - Text engine, annotators, Eclipse tooling
  - Interface to project R (statistical platform)
- Enterprise software integration
- Spreadsheet-style analysis
- Integrated installation of supported open source and other components
- Web Console for admin and application access
- Security
- Platform enrichment: additional security, performance features, . . .
- World-class support
- Full open source compatibility

Business benefits

- Quicker time-to-value due to IBM technology and support
- Reduced operational risk
- Enhanced business knowledge with flexible

BigInsights Enterprise Edition Components



Agenda

1. Big Data Customer Examples
2. The IBM Big Data Platform
3. Analyzing Data at Rest
4. **Analyzing Data in Motion (at Warp speed)**
5. Accelerators
6. Questions and Answers

Big Data Platform - Stream Computing

- Built to analyze data in motion
  - Multiple concurrent input streams
  - Massive scalability
- Process and analyze a variety of data
  - Structured, unstructured content, video, audio
  - Advanced analytic operators

Stream Computing - A Paradigm Shift

Traditional Computing

Historical fact finding

Find and analyze information stored on disk

Batch paradigm, pull model

Query-driven: submits queries to static data

Stream Computing

Current fact finding

Analyze data in motion – before it is stored

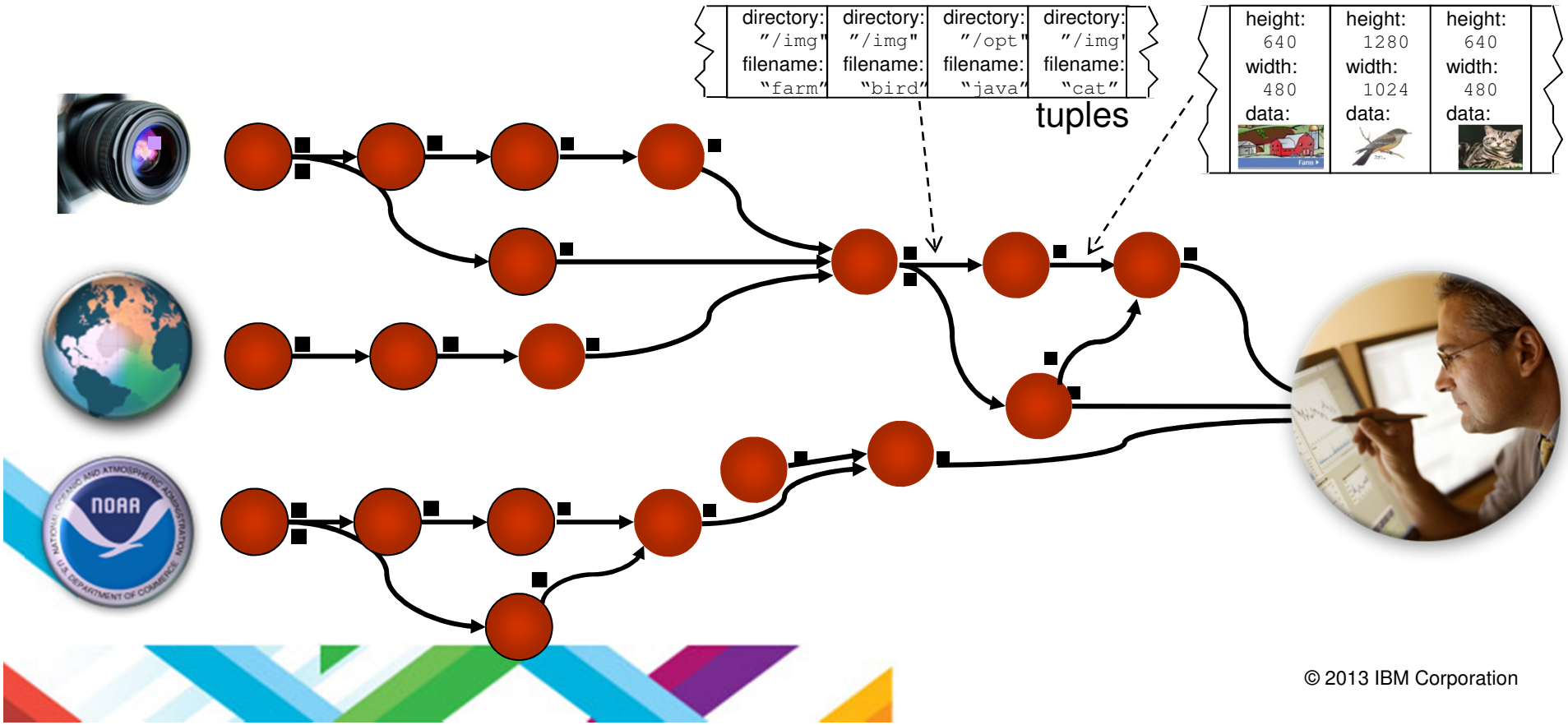
Low latency paradigm, push model

Data driven – bring data to the analytics

Stream Computing - Illustrated

A framework for Big Data in motion

- Reads data from the data stream (source operators)
- Processes the data immediately – sub-second analysis after events occur
- Can write processed data to a variety of destinations (sink operators)



Massively Scalable Stream Analytics

Linear Scalability

- Clustered deployments – unlimited scalability

Automated Deployment

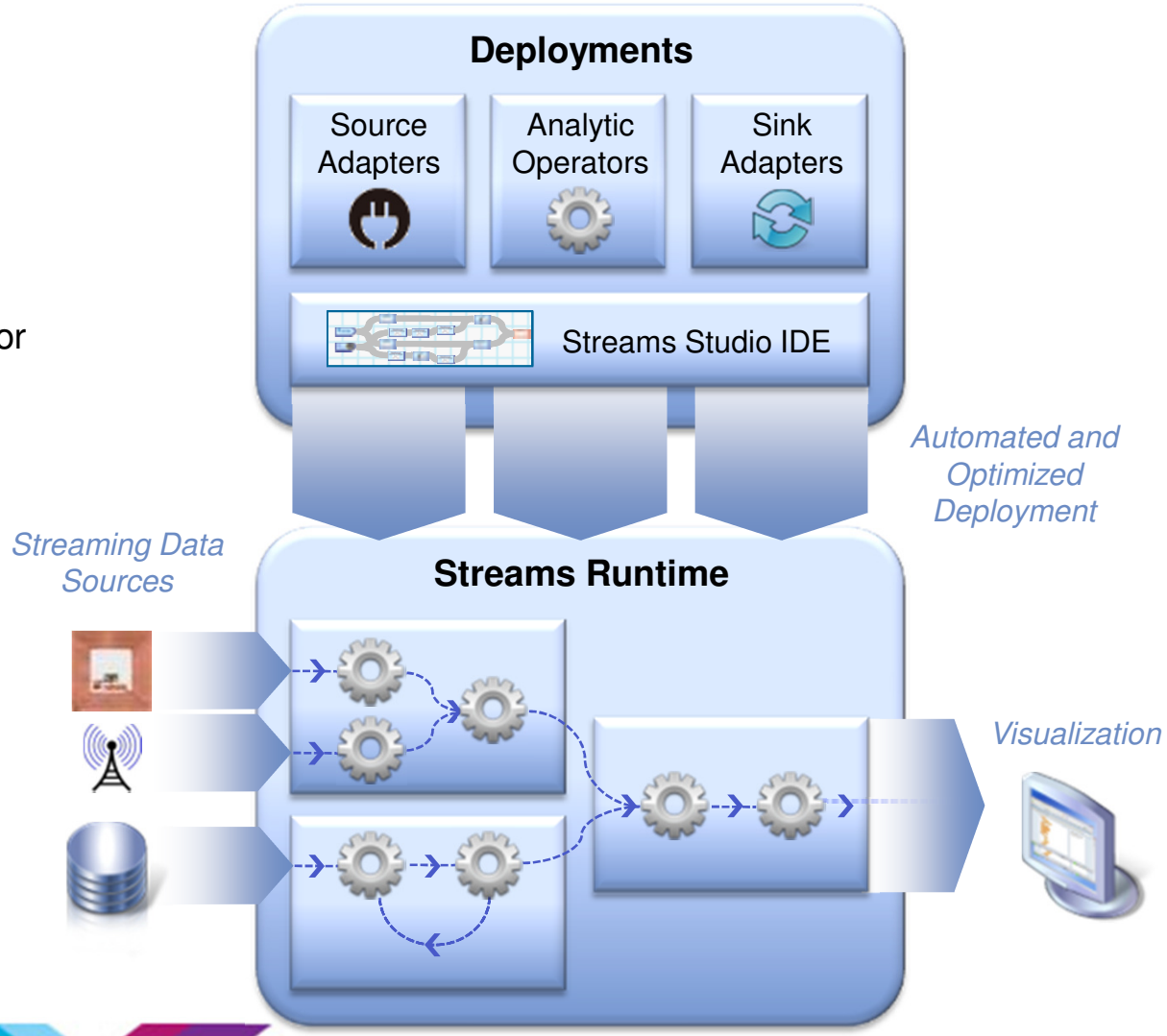
- Automatically optimize operator deployment across nodes

Performance Optimization

- Parallel & pipeline operations
- Efficient multi-threading

Analytics on Streaming Data

- Analytic accelerators for a variety of data types
- Optimized for real-time performance

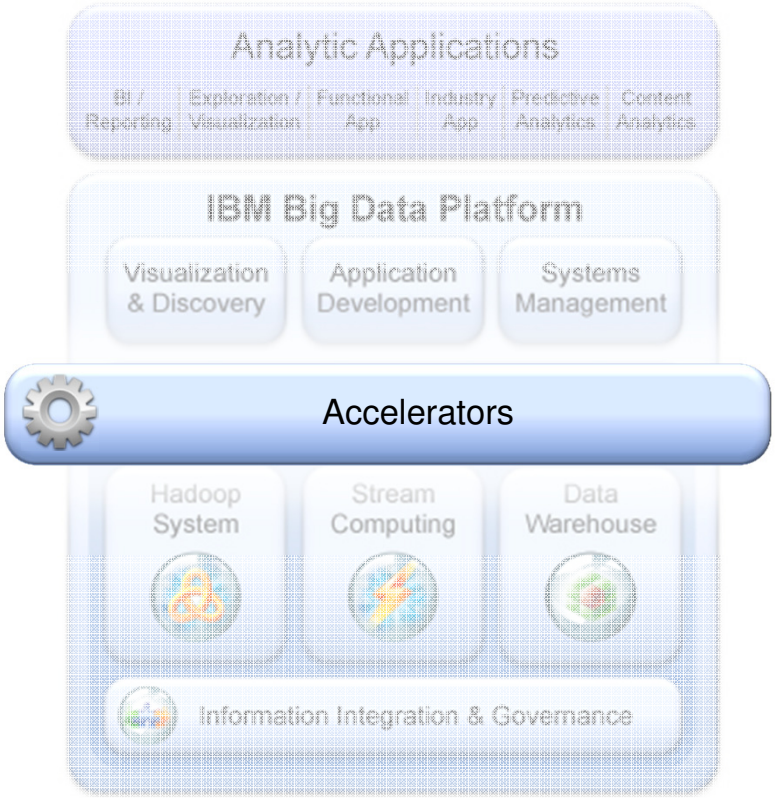


Agenda

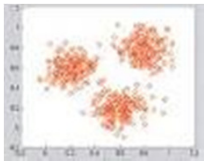
1. Big Data Customer Examples
2. The IBM Big Data Platform
3. Analyzing Data at Rest
4. Analyzing Data in Motion (at Warp speed)
5. **Accelerators**
6. Questions and Answers

Big Data Platform - Accelerators

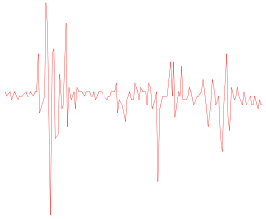
- Analytic accelerators
  - Analytics, operators, rule sets
- Industry and Horizontal Application Accelerators
  - Analytics
  - Models
  - Visualization / user interfaces
  - Adapters



Analytic Accelerators – Designed for Variety



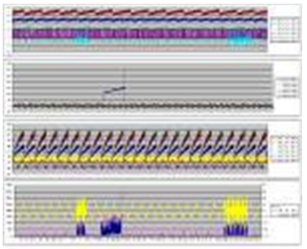
Mining in Microseconds
(included with Streams)



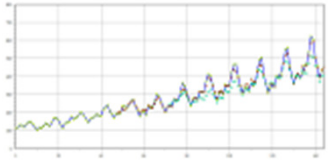
Acoustic
(IBM Research)
(Open Source)

Text
(listen, verb),
(radio, noun)

Simple & Advanced Text
(included with Streams)
(IBM Research)
(Open Source UIMA)



Advanced Mathematical Models
(IBM Research)



Predictive
(IBM Research)

$$\sum_{population} R(s_t, a_t)$$

Statistics
(included with Streams)

Geospatial
(IBM Research)

Image & Video
(Open Source)

Advanced Text Analytics

How it works

- **Parses unstructured text and detects meaning with annotators**
- **Understands the context in which the text is analyzed**
- **Hundreds of pre-built annotators for names, addresses, phone numbers, and others**
  - Parts of speech support for English, Spanish, French, German, Portuguese, Dutch, Japanese, Chinese
- **Distills structured info from unstructured text**
  - Sentiment analysis
  - Consumer behavior
  - Illegal or suspicious activities

…

Benefits

- **More precise and correct answers**
  - 2x vs. marketplace alternatives
- **50% faster than manual method**
  - Used to build world-class text analysis applications
- **Run faster text analysis**
  - 10x or more vs. marketplace alternatives

Application Accelerators

IBM Accelerator for Social Data Analytics

- B2C businesses
- Sample applications: Customer acquisition / retention, Customer Segmentation or Micro Segmentation, Marketing Campaign Optimization, Lead generation, Brand Management or Surveillance

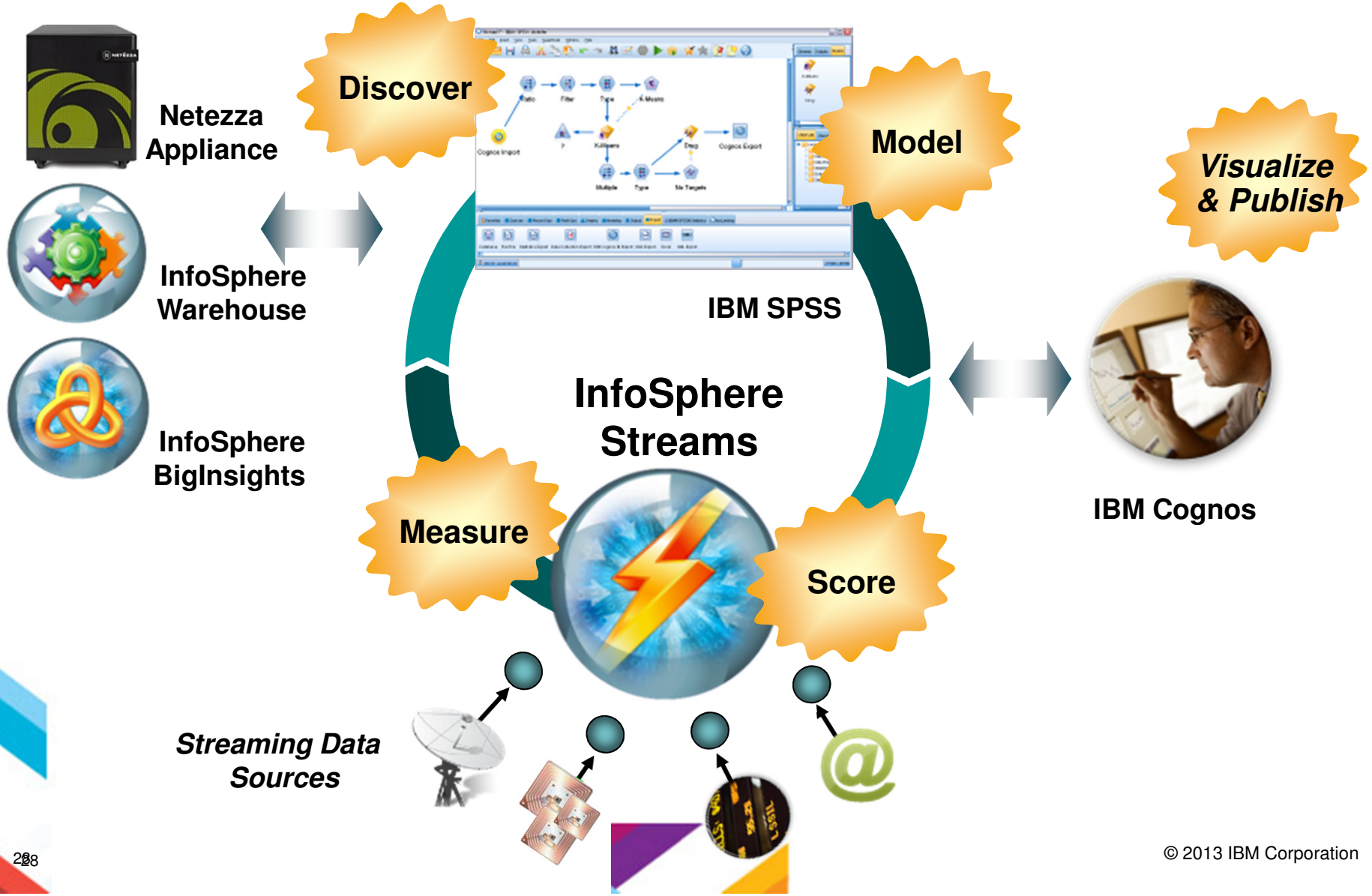
IBM Accelerator for Machine Data Analytics

- Cross-industry: manufacturing, oil & gas, energy and utility, healthcare, travel and transportation, CPG, Retail, etc.
- Operational efficiency monitoring, security incident investigation. proactive maintenance, troubleshooting, outage prevention, efficiency tracking, etc.

IBM Accelerator for Telco Event Data Analytics

- Telcos
- Campaign management, real-time promotion, fraud detection, service assurance and network monitoring,

Putting it all together



धन्यवाद
Hindi

多謝
Traditional Chinese

ขอบคุณ
Thai

Спасибо
Russian

Thank You
English

Danke
German

شكرا
Arabic

Merci
French

Obrigado
Brazilian Portuguese

Gracias!
Spanish

多谢
Simplified Chinese

Bedankt
Nederlands

நன்றி
Tamil

ありがとうございました
Japanese

감사합니다