

IBM Smarter Business and Technology Series

Lotusphere and InformationOnDemand Come to You

Get Social. Do Business. Gain Insight. Optimize Results.



From Information to Insight: The Big Value of Big Data

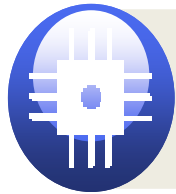
John Christopher Isaac

Senior IT Specialist, Information Management Software





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



There is an Explosion in Data and Real World Events

1.3 Billion RFID tags in 2005

30 Billion RFID tags by 2010



2 Billion Internet users by 2011



4.6 Billion Mobile Phones World Wide

Capital market data volumes grew 1,750%, 2003-06



twitter



Twitter process 7 terabytes of data every day

World Data Centre for Climate

- 220 Terabytes of Web data
- 9 Petabytes of additional data



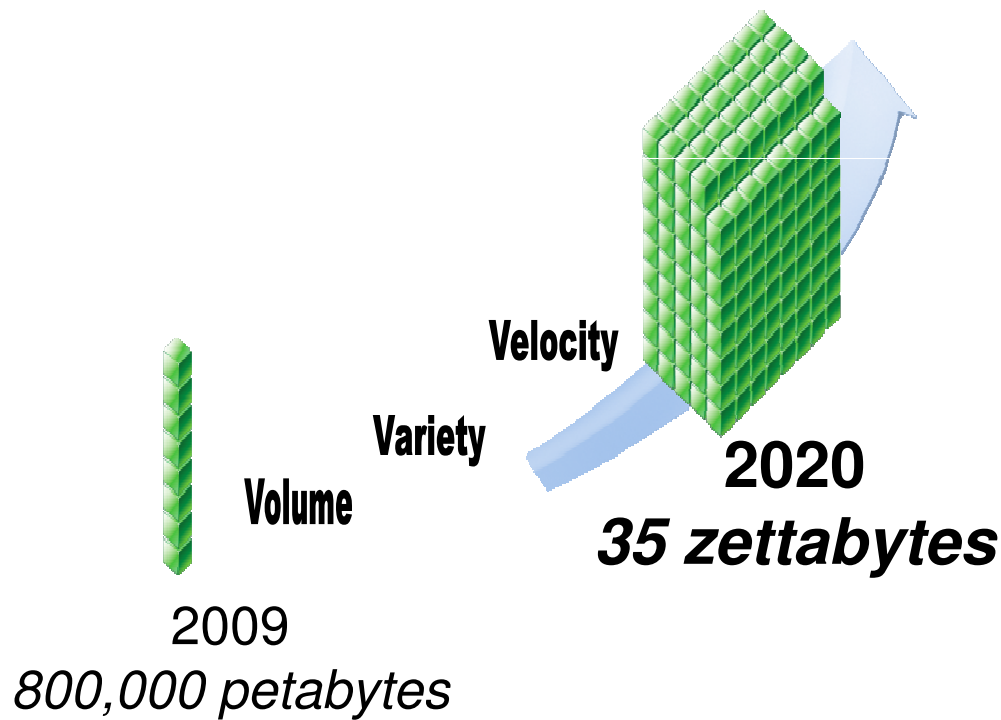
Facebook process 10 terabytes of data every day



Information is Exploding...

44x as much Data and Content
Over Coming Decade

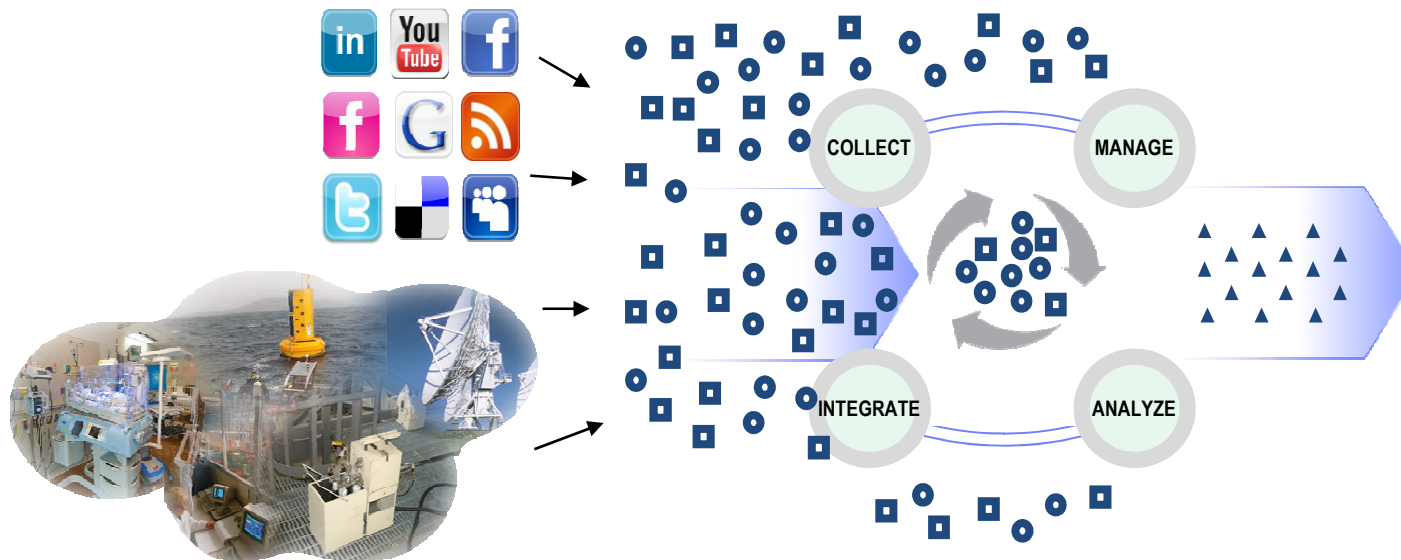
80% Of world's data
is unstructured





The BIG Data Challenge

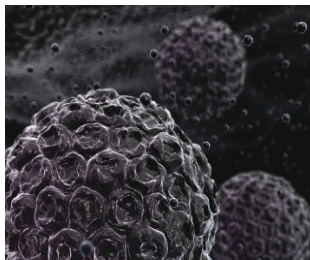
- Manage and benefit from massive and growing amounts of data
- Handle uncertainty around format variability and velocity of data
- Handle unstructured data
- Exploit **BIG Data** in a timely and cost effective fashion





Innovations

- Networking, computing and storage
- Massive Parallel Databases
- Distributed computing framework
- Real-time analytic on data in motion
- Context accumulation, sensemaking algorithms
- Advanced analytics, machine learning, text analysis, natural language
- Visualization



Disease prevention



Reducing customer churn



Reduce Fraud
Real-time promotions



Reduce traffic & pollution



Streamline supply chain

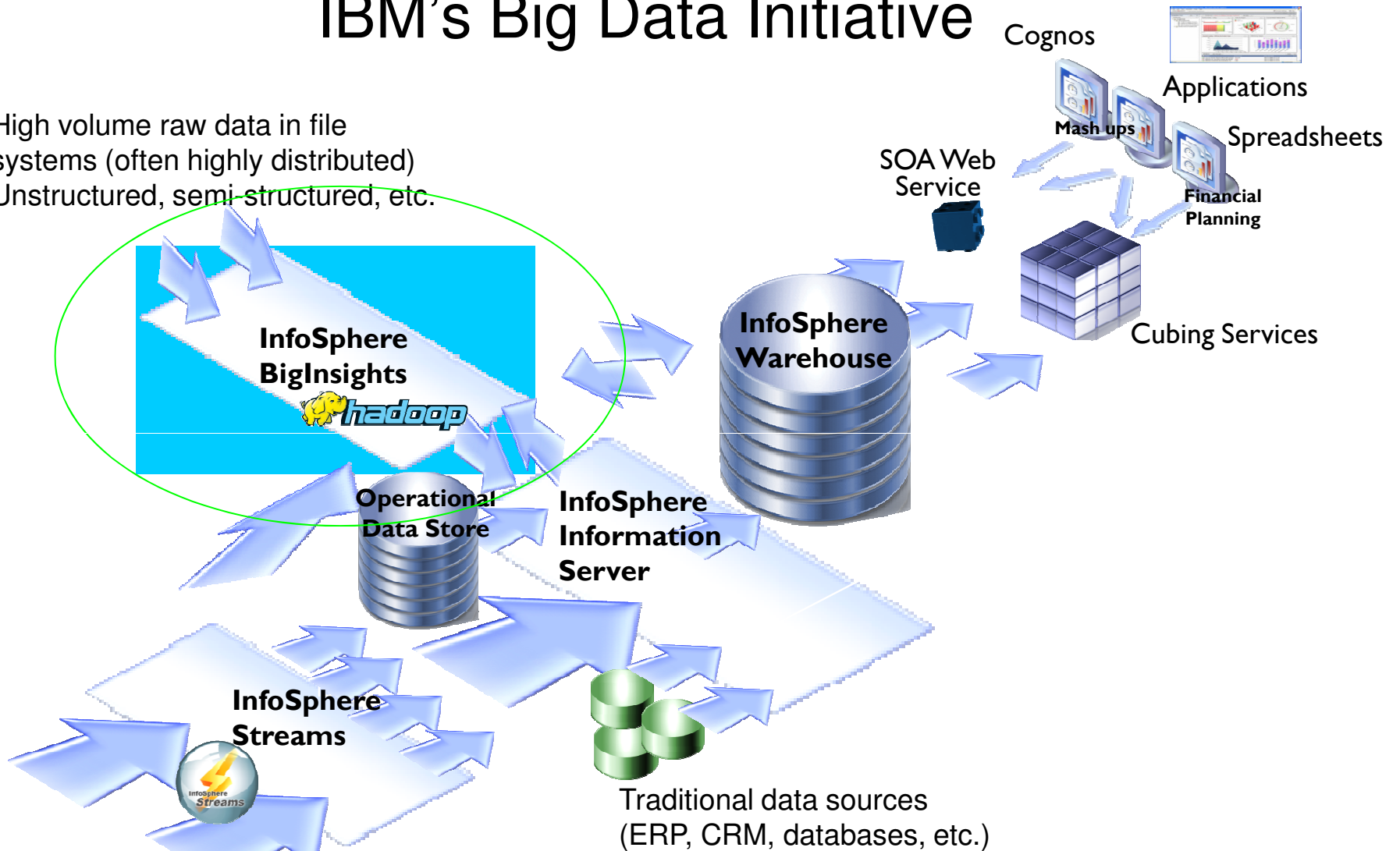


Smarter law enforcement



IBM's Big Data Initiative

High volume raw data in file systems (often highly distributed)
Unstructured, semi-structured, etc.



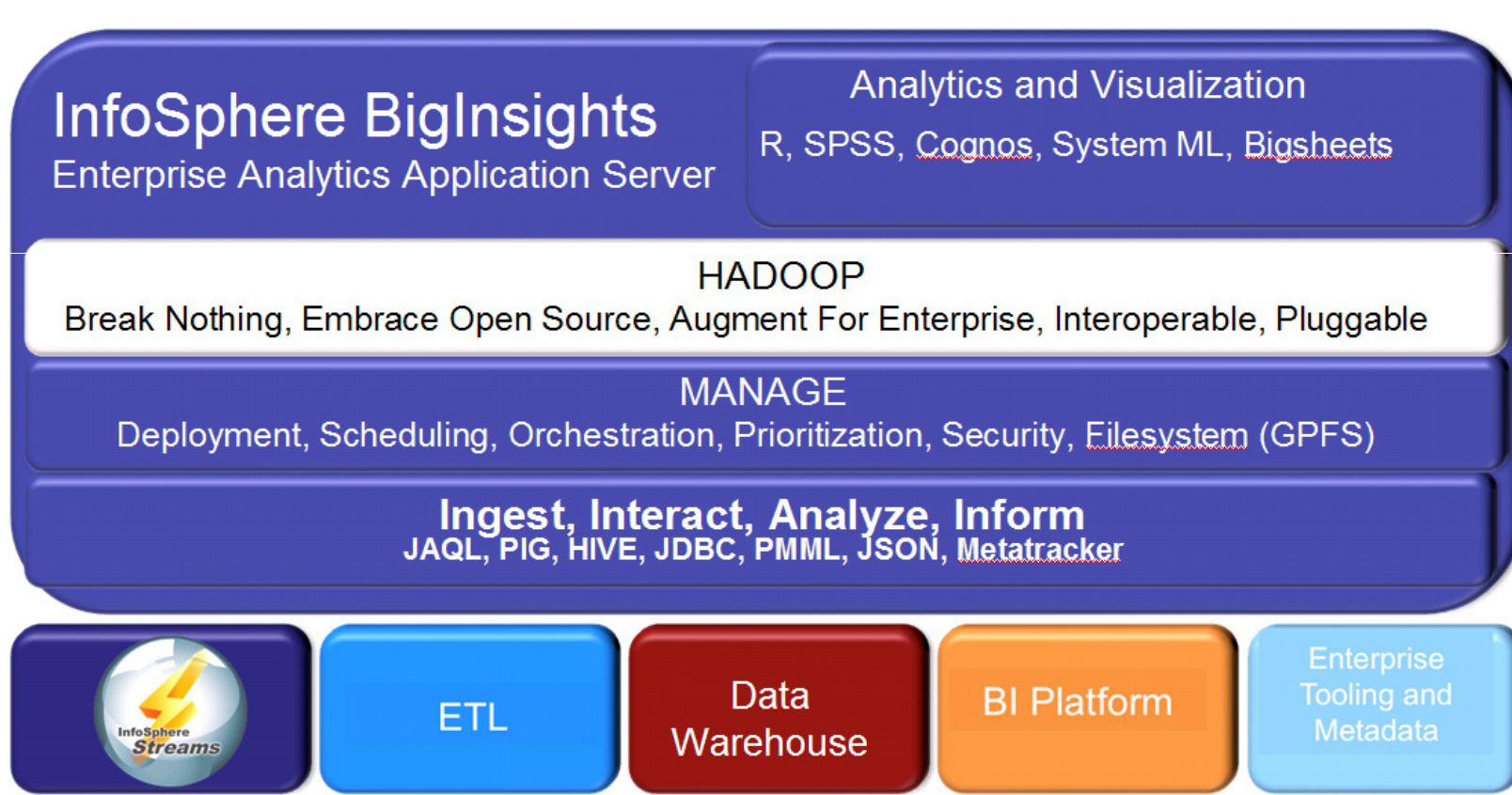
Real-time streaming Data (structured and unstructured)

- Event detection and capture of Real time Data



What is Big Insights?

- A suite of open source and IBM software for managing and analyzing large volumes of raw data





What is Hadoop?

- Apache Hadoop = free, open source framework for data-intensive applications
 - Inspired by Google technologies (MapReduce, GFS)
 - Well-suited to batch-oriented, read-intensive applications
 - Originally built to address scalability problems of Nutch, an open source Web search technology
- Enables applications to work with thousands of nodes and petabytes of data in a highly parallel, cost effective manner
 - CPU + disks of commodity box = Hadoop “node”
 - Boxes can be combined into clusters
 - New nodes can be added as needed with
 - Data formats
 - How data is loaded
 - How jobs are written





What Hadoop Is Not

- Not a replacement for your OLTP database or your data warehouse
- Not replacement for your ETL strategy
- Not a real-time complex event processor



Two Key Aspects of Hadoop

- MapReduce framework
 - How Hadoop understands and assigns work to the nodes (machines)
- Hadoop Distributed File System = HDFS
 - Where Hadoop stores data
 - A file system that spans all the nodes in a Hadoop cluster
 - It links together the file systems on many local nodes to make them into one big file system



MapReduce Explained

- "Map" step:
 - Input split into pieces
 - Worker nodes process individual pieces in parallel (under global control of the Job Tracker node)
 - Each worker node stores its result in its local file system where a reducer is able to access it
- "Reduce" step:
 - Data is aggregated ('reduced' from the map steps) by worker nodes (under control of the Job Tracker)
 - Multiple reduce tasks can parallelize the aggregation



Insights from Big Data using BigSheets

A New Class of Applications

What is it?

An insight engine for enabling ad-hoc business insights for business users - at web scale

How does it work?

- 1 – Point BigSheets to data sources of interests
- 2 – Transform data into a form that can be analyzed
- 3 – “What if tooling” - browser-based visual front end - spreadsheet metaphor to create worksheets for exploring/visualizing the big data

What's different?

- Unlocking insights embedded in unstructured data
- Analyzing data previously unavailable to analyze





Applications for Big Data Analytics are Endless

Neonatal Care



Trading Advantage



Environment



Law Enforcement



Customer Retention



Telecom



Manufacturing



Traffic Control



Fraud Prevention



IBM Smarter Business and Technology Series

Lotusphere and InformationOnDemand Come to You



IBM Watson Demonstrated Power of Big Data Analytics



Can we design a computing system that rivals a human's ability to answer questions posed in natural language, interpreting meaning and context and retrieving, analyzing and understanding vast amounts of information in real-time?

IBM Smarter Business and Technology Series

Lotusphere and InformationOnDemand Come to You



Big Data Analytics in Smarter Hospitals



IBM Data Baby
youtube.com



Enhancing Fraud Detection for Banks and Credit Card Companies

Scenario

- Build up-to-date models from transactional to feed real-time risk-scoring systems for fraud detection

Requirement

- Analyze volumes of data with response times that are not possible today
- Apply analytic models to individual client, not just client segment.





Transaction Analysis for Banking Industry

Scenario

- Analyze transaction issues from federated systems and applications to provide up-to-date account status with less turnaround time

Requirement

- Collect, aggregate, and analyze log data from various application systems
- Handle logs in different formats and correlating errors across applications
- Reduce response time to less than 2 minutes





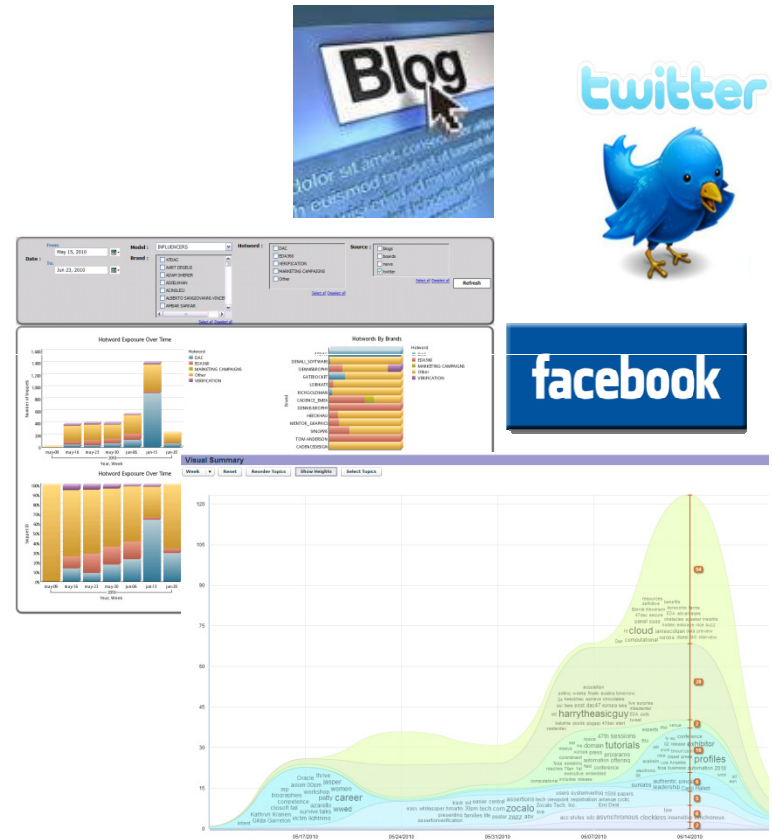
Sentiment Analysis for Products, Services and Brands

Scenario

- Monitor data from various sources such as blogs, boards, news feeds, tweets, and social medias for information pertinent to brand and products, as well as competitors

Requirement

- Extract and aggregate relevant topics, relationships, discover patterns and reveal up-and-coming topics and trends





Advanced Pharmaceutical and Medical Supply Chain Management

Scenario

- Sensors data to track and trace across supply chain to improve visibility
- Achieve compliance with ePedigree government regulations, combat deadly threat of counterfeit drugs

Requirement

- Saleable infrastructure to handle input from real-time sensors, including equipments to manage temperature sensitive pharmaceuticals





Customer Acquisition and Retention

Scenario

- Reconcile what business know about a customer's behavior in physical stores with web stores
- Take action based on insights to enable new levels of customer services

Requirement

- Weblog and click-stream analysis
- Integrated view between behavior data and transaction histories





Law Enforcement and Security – Federal Government

- Streams of information including video surveillance, wire taps, communications, call records, etc.
- Millions of streams per second with low density of critical data
- Identify patterns and relationships among vast information sources



"The **US Government** has been working with IBM Research **since 2003** on a **radical new approach** to data analysis that enables high speed, scalable and complex analytics of heterogeneous data streams in motion. The project **has been so successful** that US Government **will deploy additional installations** to enable other agencies to achieve greater success in various future projects" - US Government



Infrastructure Optimization for Telco Companies

Scenario

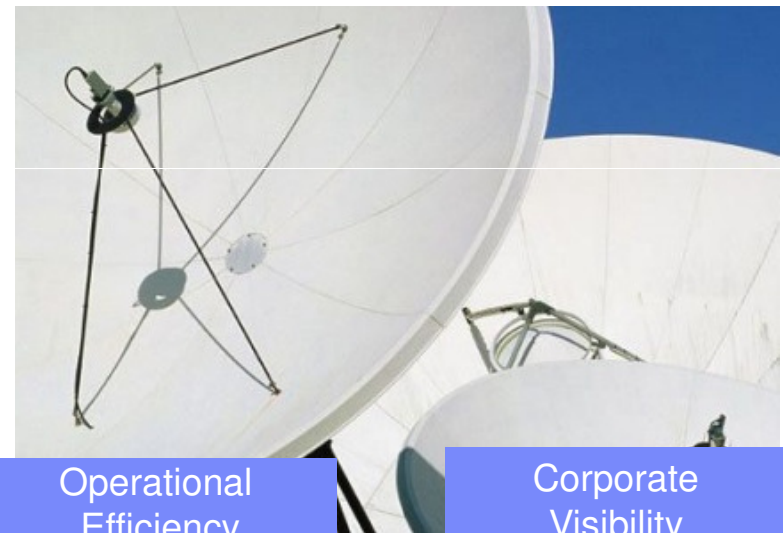
- Mediate CDRs to billing systems, eliminate delays associated de-duplications; improve speed and quality of billing process and campaign execution

Requirement

- Real-time summarization of information
- Abilities to handle billions of call records
- Integrated enterprise-wide performance management across all LOB (mobile, fixdlin, media, B2B)

Data Infrastructure Optimization

Single, real-time data feed for Fraud, BI & Revenue Assurance systems



Operational Efficiency

Cross-sell/Up-Sell, Reduced Activation Time

Corporate Visibility

Marketing Campaign & Service Analytics



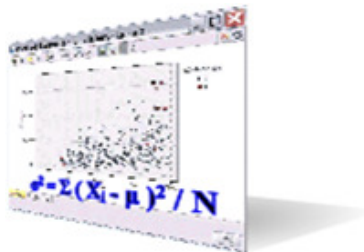
IT Needs integrated, enterprise-grade capabilities



- *Extract insights from new information sources*
- *Improve response time to business needs*



- *Run analytics on more data*
- *Integrate insights with operational systems*
- *Embed real-time process support*



- *Make analytics available to more users*
- *Integrated new insights with existing analysis, queries, reports, and predictive models*



Our Engagement Models

InfoSphere BigInsights

Workshops

- On-site workshops with customers and IBM jStart team
- *Visit ibm.com/jstart

BigInsight cloud access

- Access to BigInsights images on IBM test/development cloud
- *Contact Tom Deutsch (tdeutsch@us.ibm.com)

Strategic business partnering

- Cooperative engagements for individual customer solutions





Summary

- IBM sees great value in using Hadoop to process Big Data
- We have a strategy for Big Data & Commercial Hadoop
- BigInsights is a foundation for next generation analytics
- BigSheets allows business users to find insight in Big Data
- Today's we we are announcing a Tech Preview of BigInsights

IBM Smarter Business and Technology Series

Lotusphere and InformationOnDemand Come to You



THANK
YOU

