Beyond Watson: The Business Implications of Big Data

Shankar Venkataraman

IBM Program Director, STSM, Big Data

August 10, 2011

The World is Changing and Becoming More…

INSTRUMENTED

INTERCONNECTED

INTELLIGENT

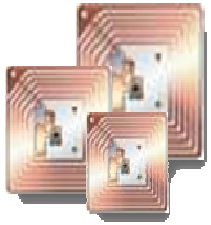
The resulting explosion of information creates a need for a new kind of intelligence

…to help build a Smarter Planet

There is an Explosion in Data and Real World Events

2 Billion **Internet users by 2011**

1.3 Billion RFID tags in 2005
30 Billion **RFID tags by 2010**



4.6 Billon
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

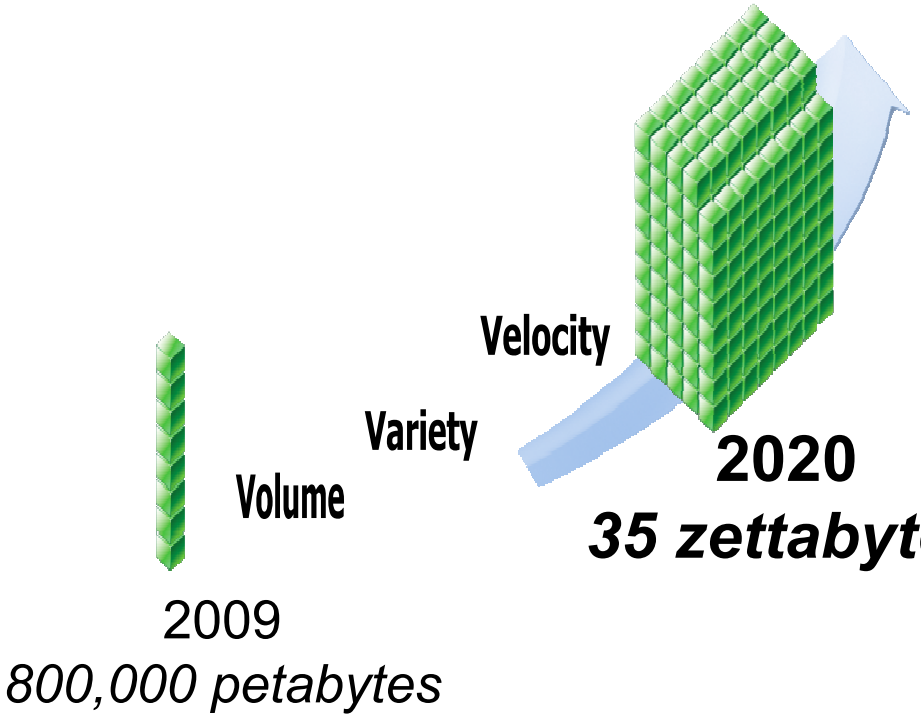
Facebook process *10 terabytes* of data every day

SHARE in Orlando 2011

Information is Exploding…

44x as much Data and Content Over Coming Decade

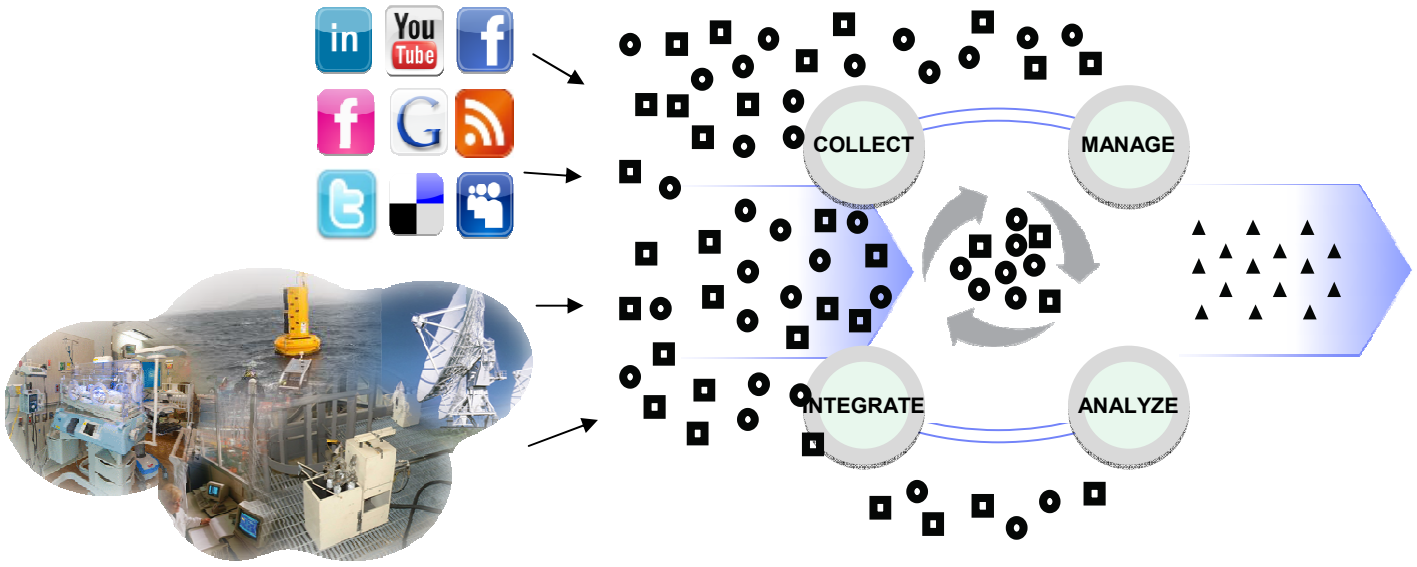
80% Of world's data is unstructured



Source: IDC, The Digital Universe Decade – Are You Ready?, May 2010

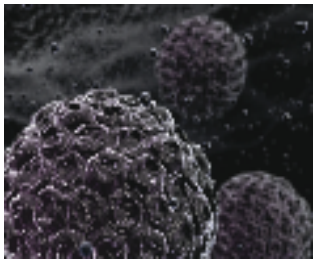
The BIG Data Challenge

- 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

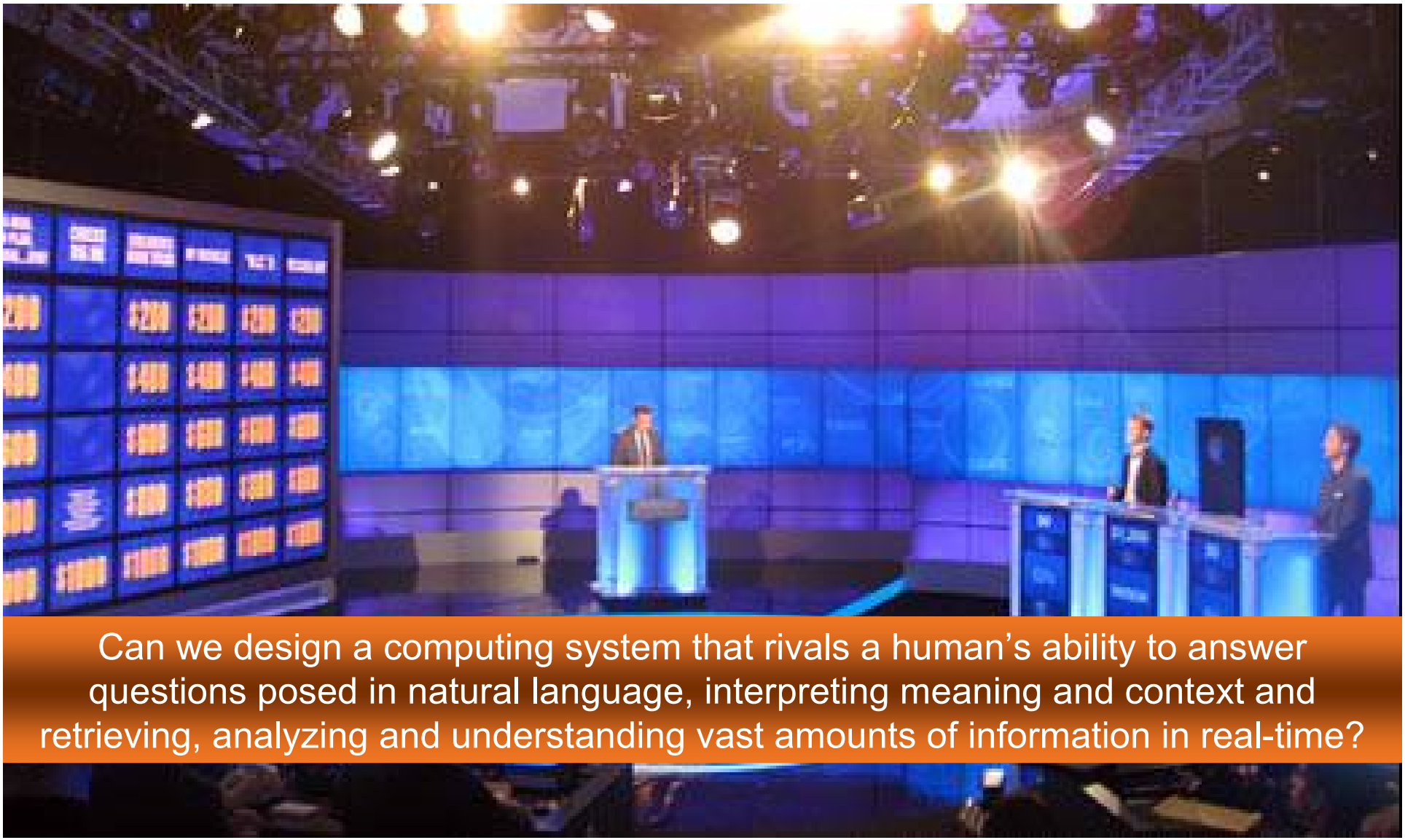
Reduce Fraud Real-time promotions

Reduce traffic & pollution

Streamline supply chain

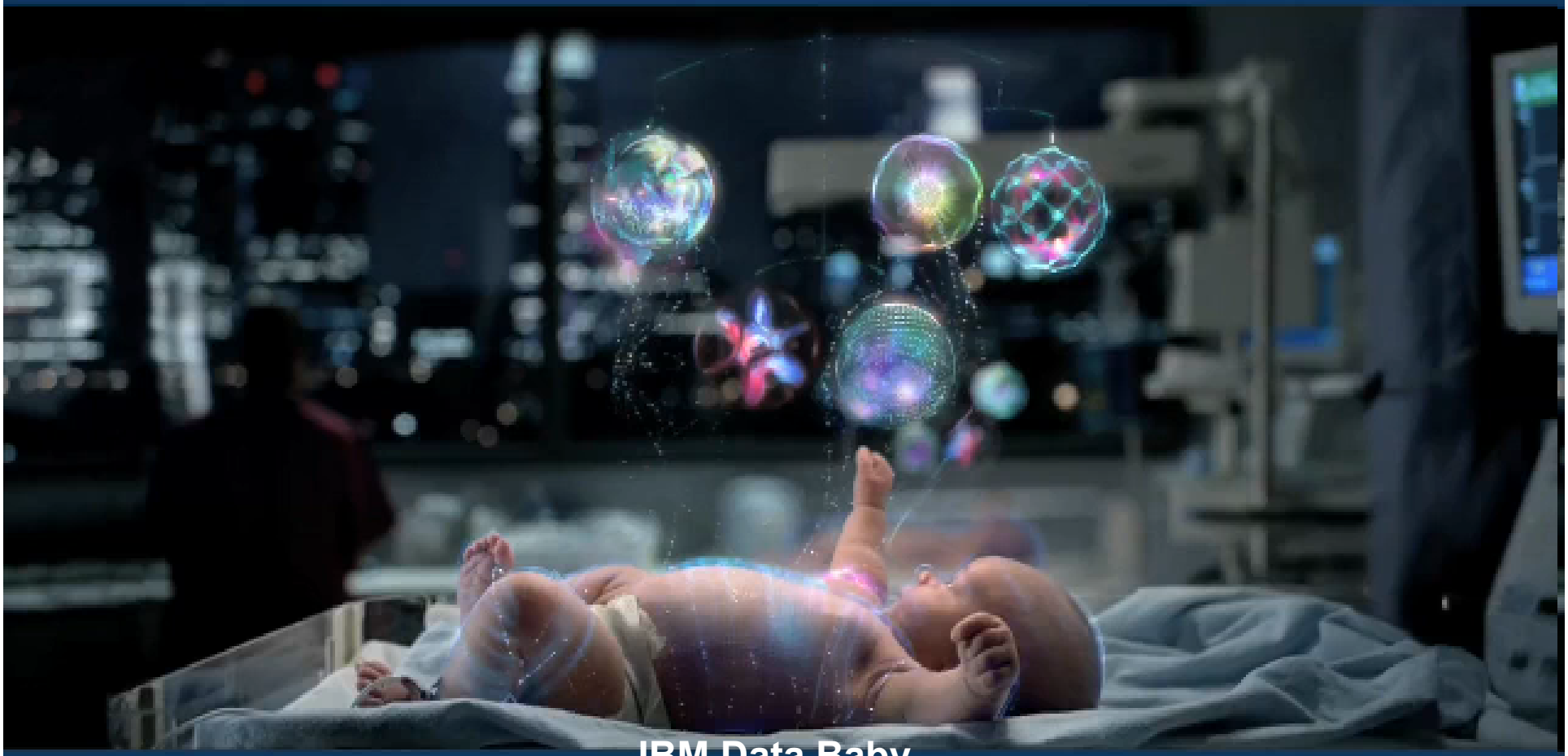
Smarter law enforcement

IBM 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?

Big Data Analytics in Smarter Hospitals



IBM Data Baby
youtube.com

Organizations Need Deeper Insights From Their Data

1 in 3

Business leaders frequently make decisions based on information they don't trust, or don't have

83%

of CIOs cited "Business intelligence and analytics" as part of their visionary plans to enhance competitiveness

1 in 2

Business leaders say they don't have access to the information they need to do their jobs

35%

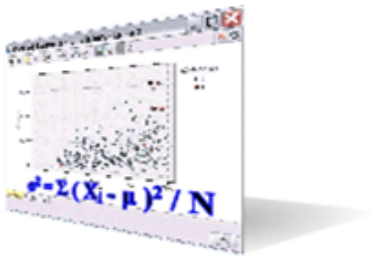
of Customers will look to replace their current warehouse with a pre-integrated Warehouse solution in the next 3 years, only 14% have today

Source (TDWI: Next Generation Data Warehouse Platforms Q4 2009)

IT Needs integrated, enterprise-grade capabilities

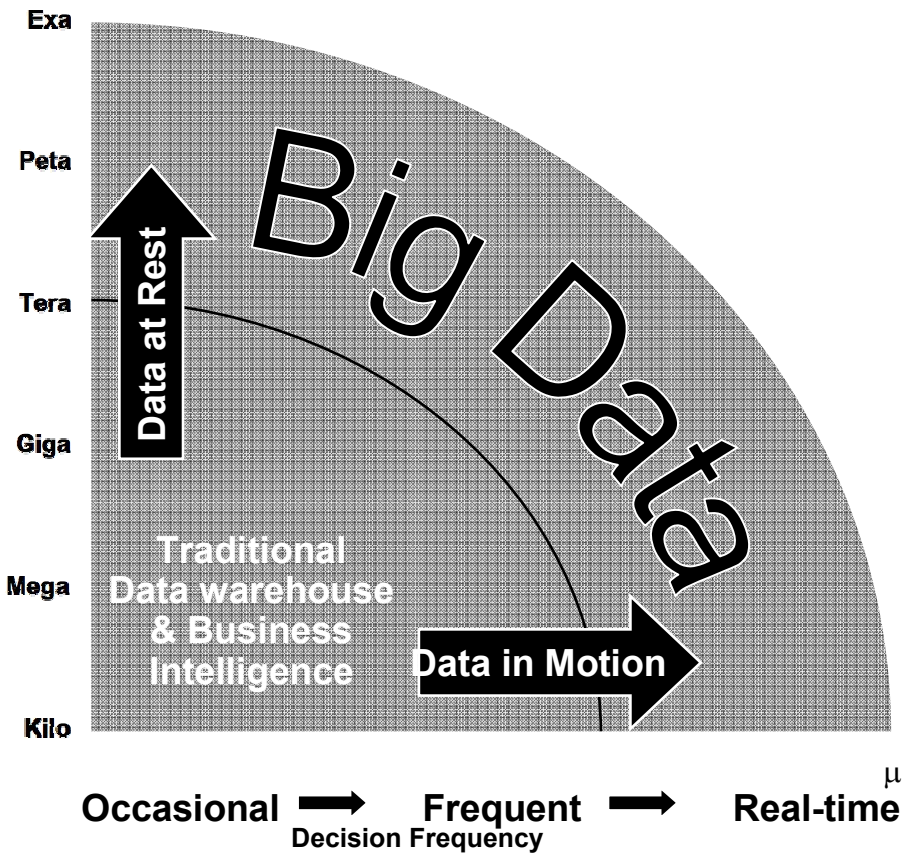
- *Extract insights from new information sources*
- *Improve response time to business needs*

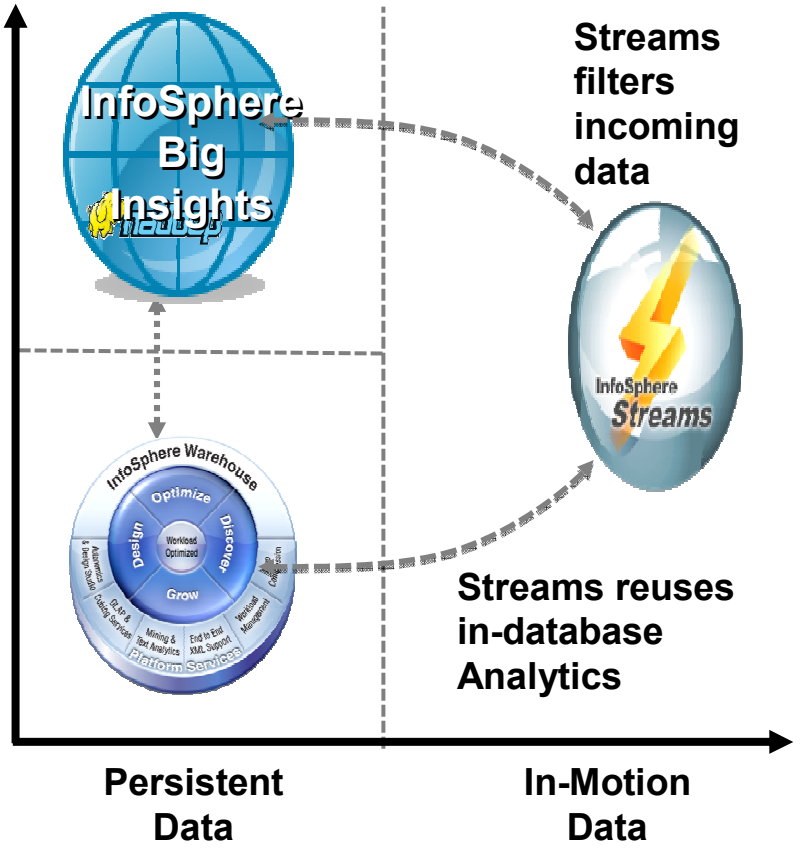
- *Run analytics on more data*
- *Integrate insights with operational systems*
- *Embed real-time process support*



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

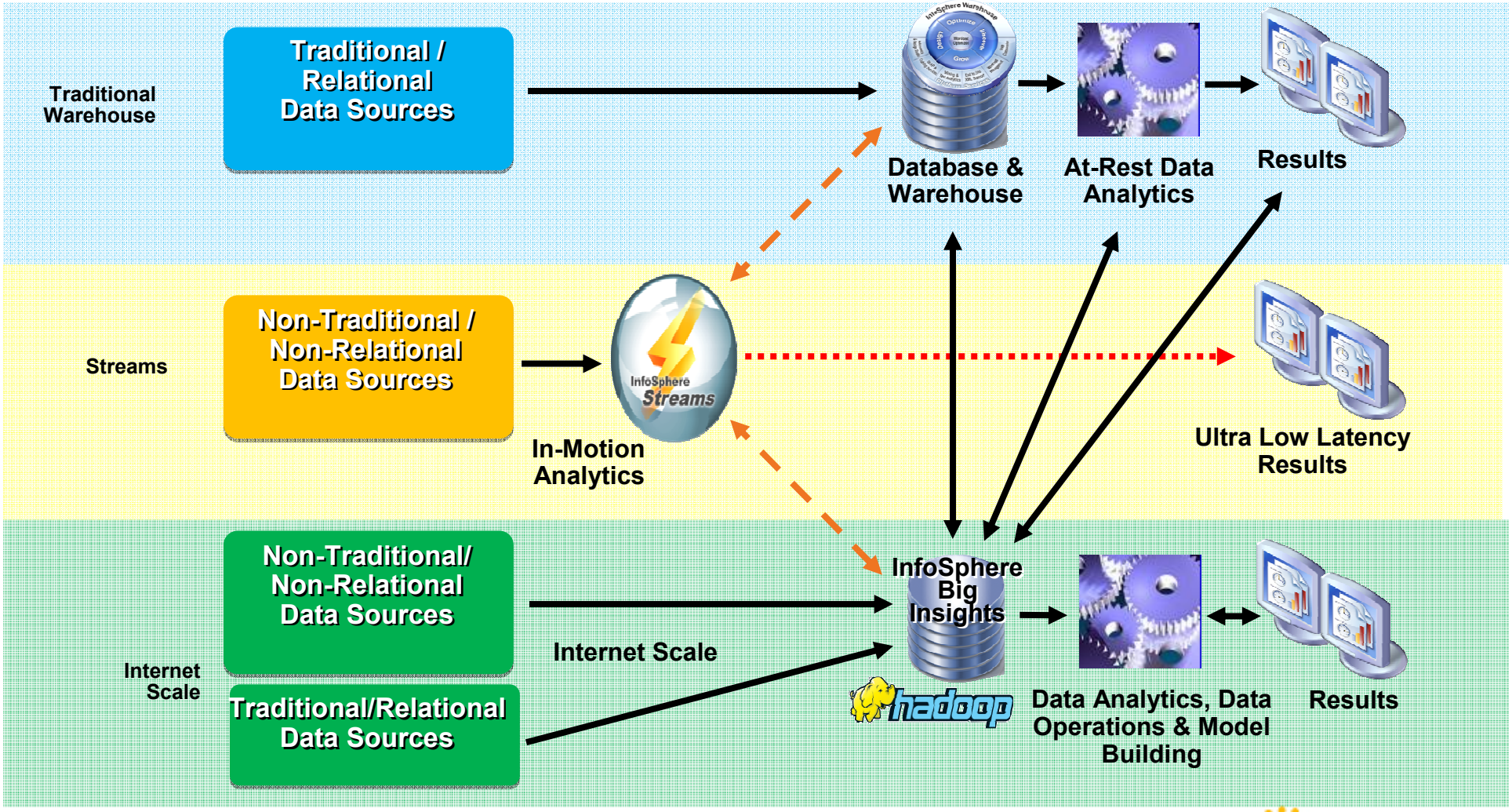
“Big Data” brings new opportunities





Source: Global Technology Outlook 2011

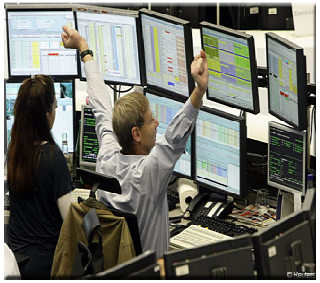
The BIG Data Ecosystem: Interoperability is Key



Applications for Big Data Analytics are Endless

Neonatal Care

Trading Advantage



Environment

Law Enforcement

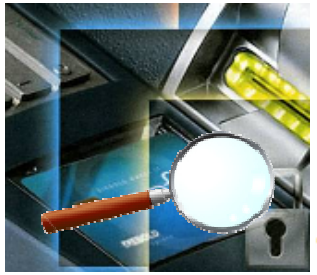
Customer Retention

Telecom

Manufacturing

Traffic Control

Fraud Prevention



SHARE in Orlando 2011

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.

Build Faster Real-time Trading Systems

Scenario

- Identify and execute trades
- Process over 5M events per second with average latency of 150 microseconds

Requirement

- Consuming, analyzing and acting on market data while maintaining sub-millisecond response time under extreme data loads
- Incorporate content feeds, news text, audio, video, to establish greater context for better decisions

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

Real-time Predictive Analytics at Hospitals

Scenario

- Early detection of potentially life threatening conditions at ICUs to lower patient morbidity and better long term outcomes
- Enable physicians to verify new clinical hypotheses

Requirement

- Real-time analytics and correlations on physiological data streams such as blood pressure, temperature, EKG, Blood oxygen saturation, etc.

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

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

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

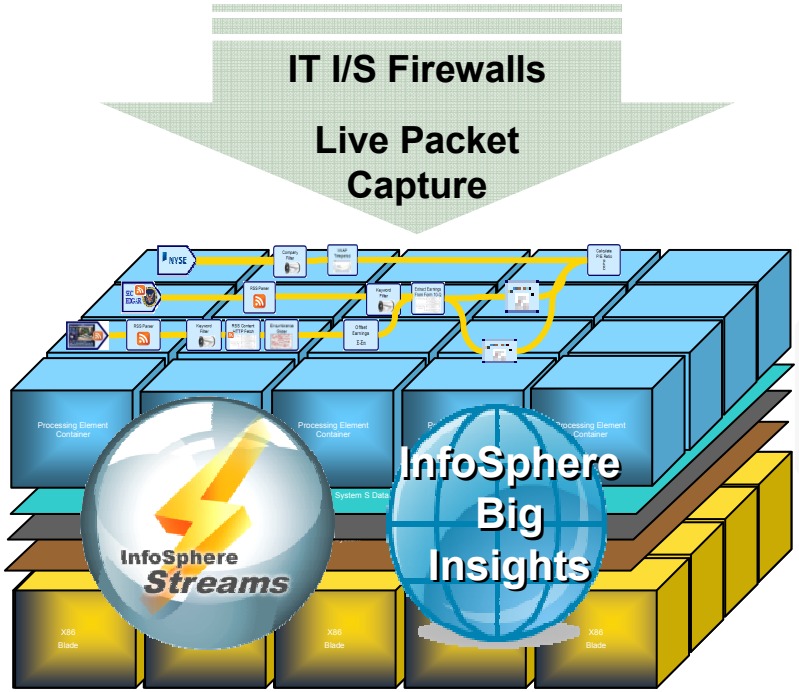
- Weblog and click-stream analysis
- Integrated view between behavior data and transaction histories

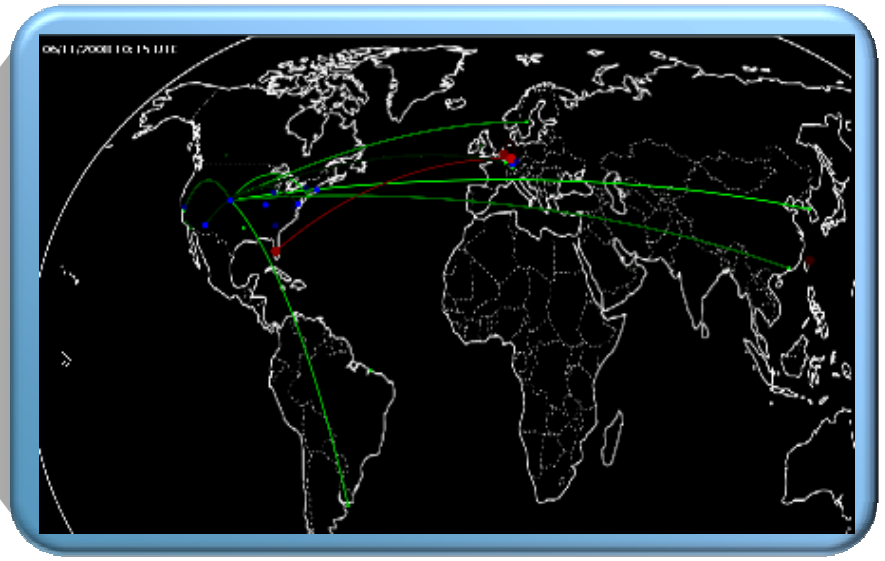
Law Enforcement and Security – Federal Government

- 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

Early detection of Cyber Security Breach and Attack





- ❑ DNS / DHCP / Netflow sources
- ❑ Botnet Behavior modeling
- ❑ External C&C Feeds (live DB queries)

- ✓ Botnet nodes / Malware
- ✓ IP/MAC identifying suspects

Remediation Infrastructure / Ticketing

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

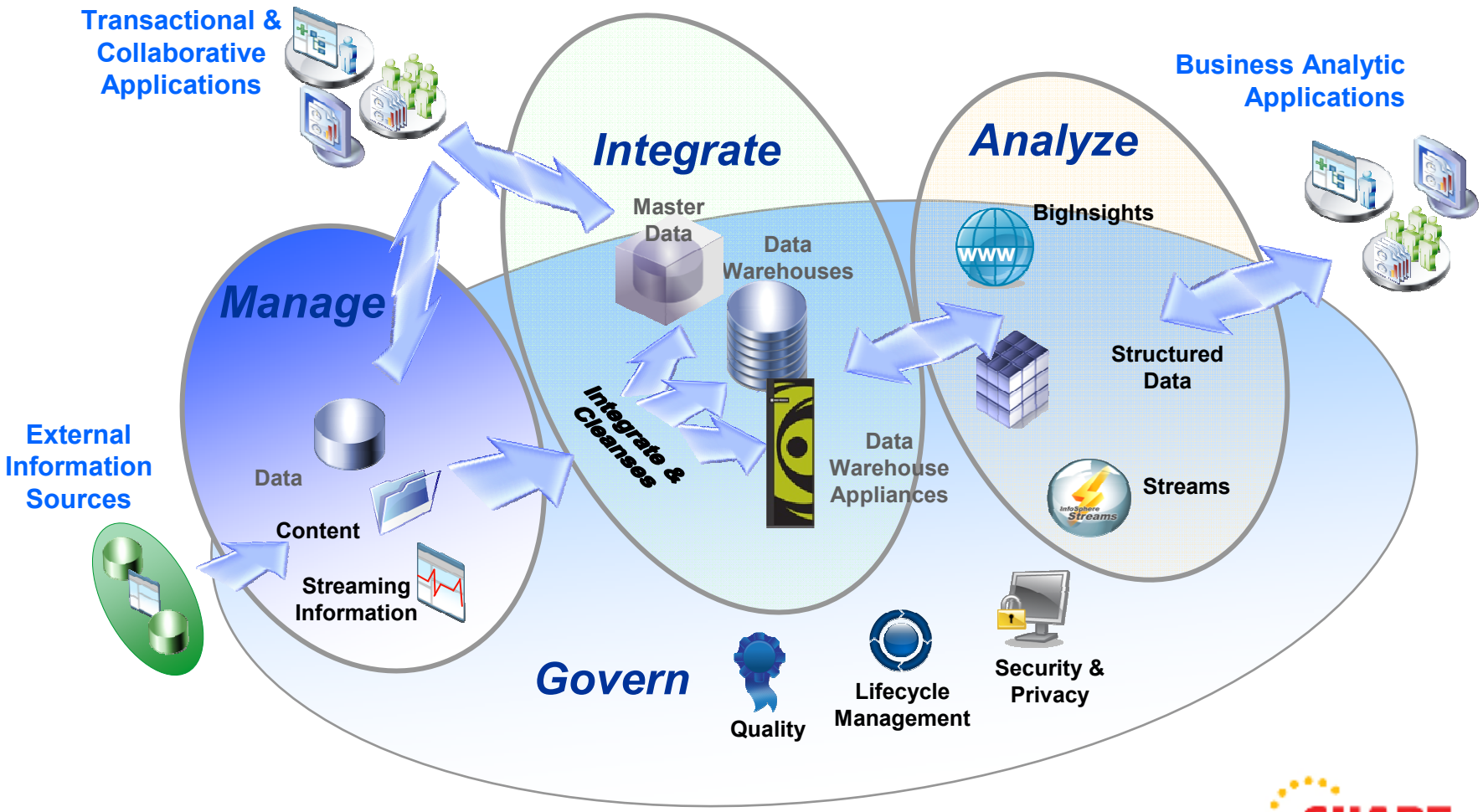
Operational Efficiency

Cross-sell/Up-Sell, Reduced Activation Time

Corporate Visibility

Marketing Campaign & Service Analytics

“BIG Data” is Integrated Part of IBM Middleware



IBM is Uniquely Positioned to Handle "BIG Data" Analysis

- ***Scale to petabytes and thousands of users*** for core data analysis with linear processor scalability
- ***Deep integration with Cognos and SPSS***
- Run third-party analytic models from the data warehouse to allow ***highly scalable, efficient analytics processing***
- ***Integrated analysis and analytic model consistency*** without having to load everything into the warehouse

SHARE in Orlando 2011