Information Management

IBM

IBM Big Data Platform

Turning big data into smarter decisions

Ayhan Önder

October 30, 2012

"Data is the new Oil"

In its raw form, oil has little value. Once processed and refined, it helps power the world.

Forbes.com

"Big Data has arrived at Seton Health Care Family, fortunately accompanied by an analytics tool that will help deal with the complexity of more than two million patient contacts a year…"

The New York Times

"At the World Economic Forum last month in Davos, Switzerland, Big Data was a marquee topic. A report by the forum, "Big Data, Big Impact," declared data a new class of economic asset, like currency or gold.

FT FINANCIAL TIMES World business newspaper

"Increasingly, businesses are applying analytics to social media such as Facebook and Twitter, as well as to product review websites, to try to "understand where customers are, what makes them tick and what they want", says Deepak Advani, who heads IBM's predictive analytics group."

THE WALL STREET JOURNAL.

"Companies are being inundated with data—from information on customer-buying habits to supply-chain efficiency. But many managers struggle to make sense of the numbers."

Forbes.com

"…now Watson is being put to work digesting millions of pages of research, incorporating the best clinical practices and monitoring the outcomes to assist physicians in treating cancer patients."

Los Angeles Times

The Oscar Senti-meter — a tool developed by the L.A. Times, IBM and the USC Annenberg Innovation Lab — analyzes opinions about the Academy Awards race shared in millions of public messages on Twitter."

CNBC

"Data is the new oil."
Clive Humby

Imagine the Possibilities of Harnessing *Your* Data Resources

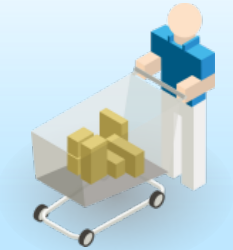
Big data challenges exist in every organization today

Government cuts acoustic analysis from hours to 70 Milliseconds

Utility avoids power failures by analyzing 10 PB of data in minutes

Hospital analyses streaming vitals to detect illness 24 hours earlier

Retailer reduces time to run queries by 80% to optimize inventory

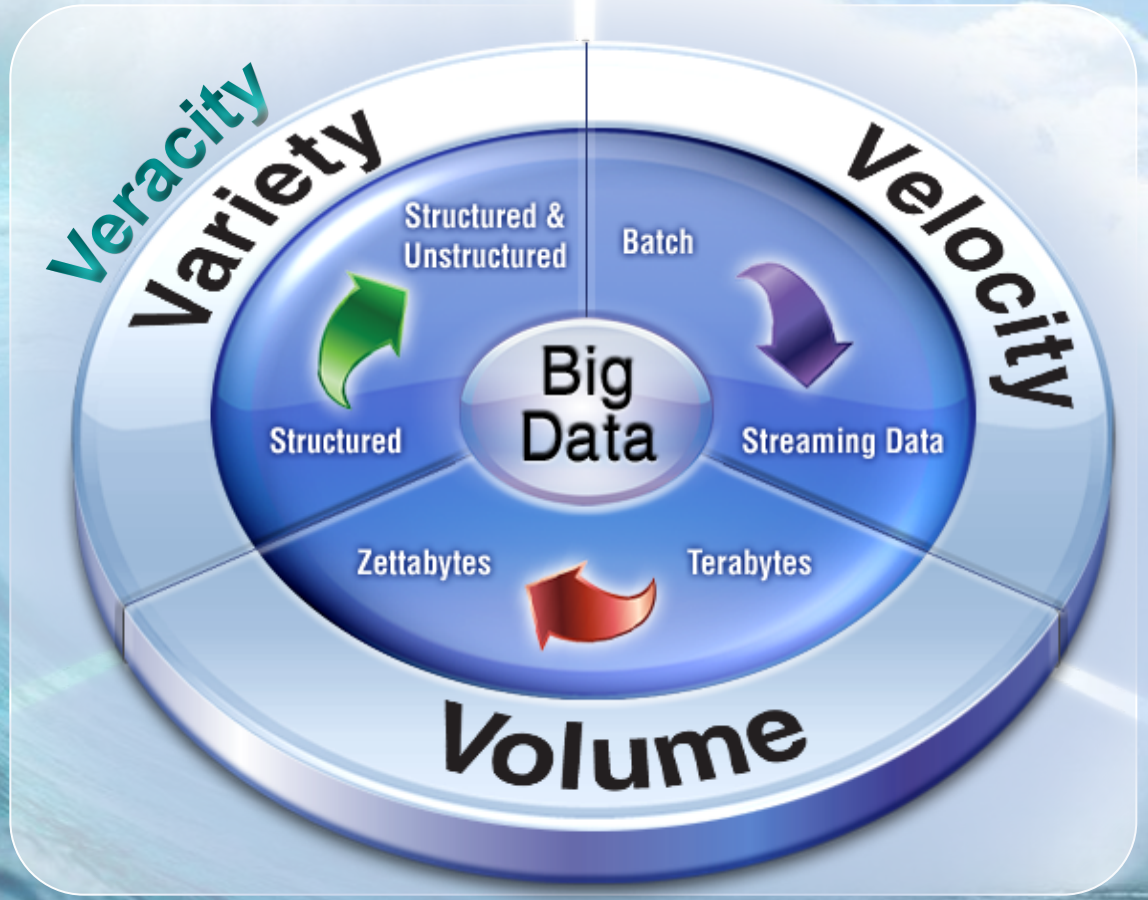


Stock Exchange cuts queries from 26 hours to 2 minutes on 2 PB

Telco analyses streaming network data to reduce hardware costs by 90%

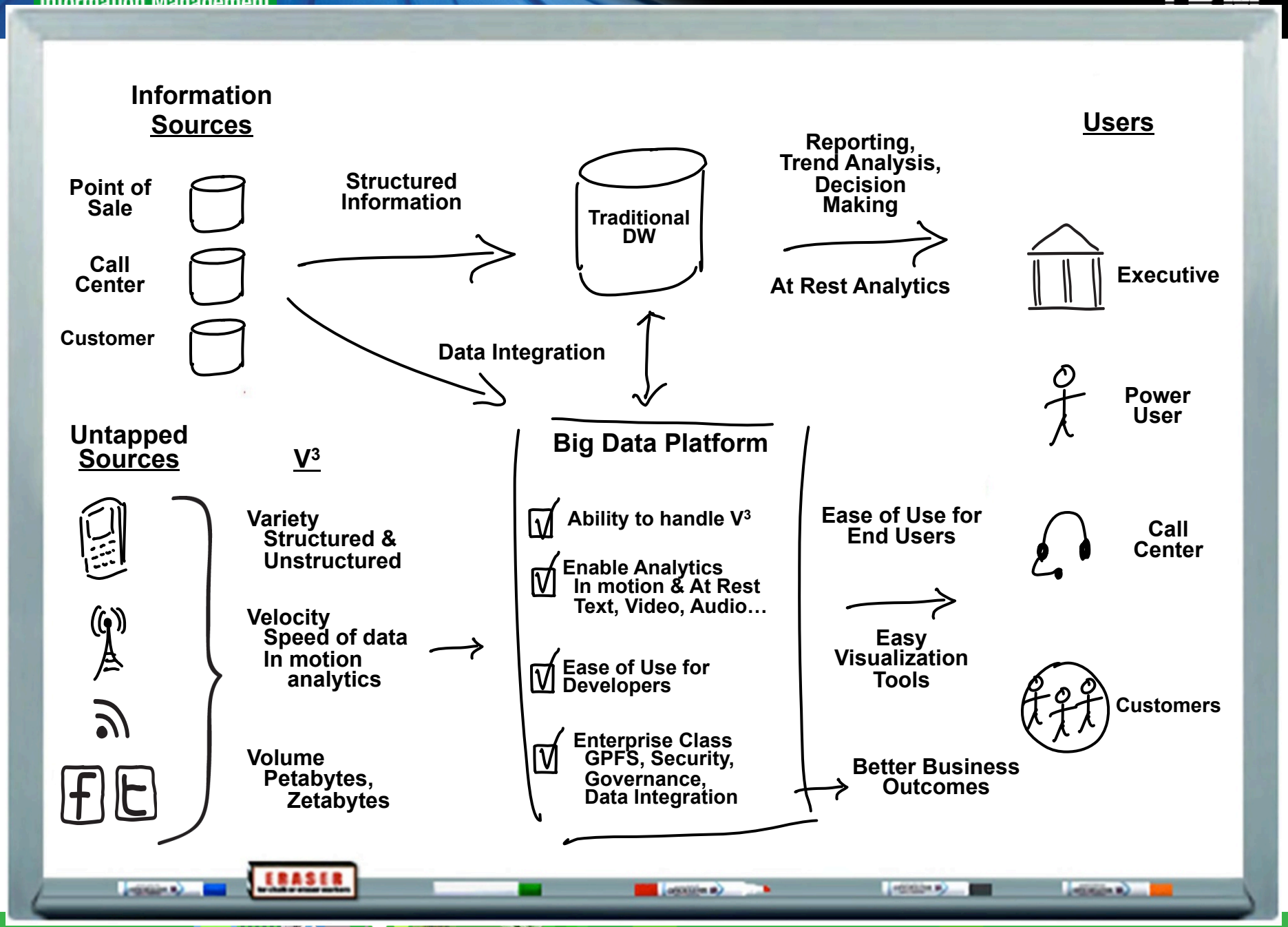
The Big Data Opportunity

Extracting insight from an immense volume, variety and velocity of data, in context, beyond what was previously possible.



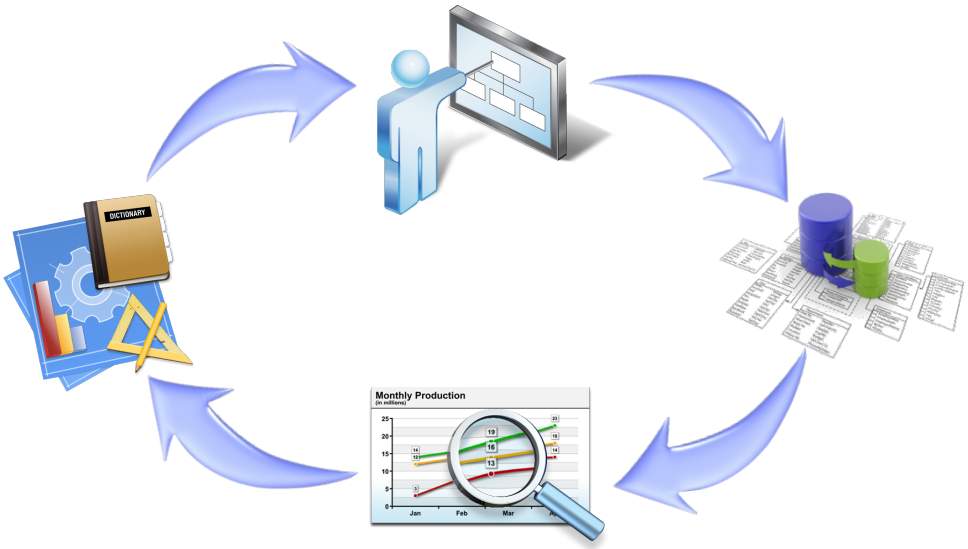
What Our Customers Tell Us . . .

- **Don't know what should be analyzed**
- **Volumes can be extremely high**
  - Potentially valuable data is dormant or discarded (size/performance)
  - Too expensive to justify integrating large volumes of unstructured data
- **Much of their data is unstructured, or in widely varying structures, which are difficult to analyze**
- **Difficult to integrate information distributed across multiple systems and the Internet**
- **Some information has a short useful lifespan**
- **Analysis needed in the context of their existing information**



Traditional Analytics:
Business Requirements Drive Solution Design

Business Defines Requirements – What Questions Should we Ask?



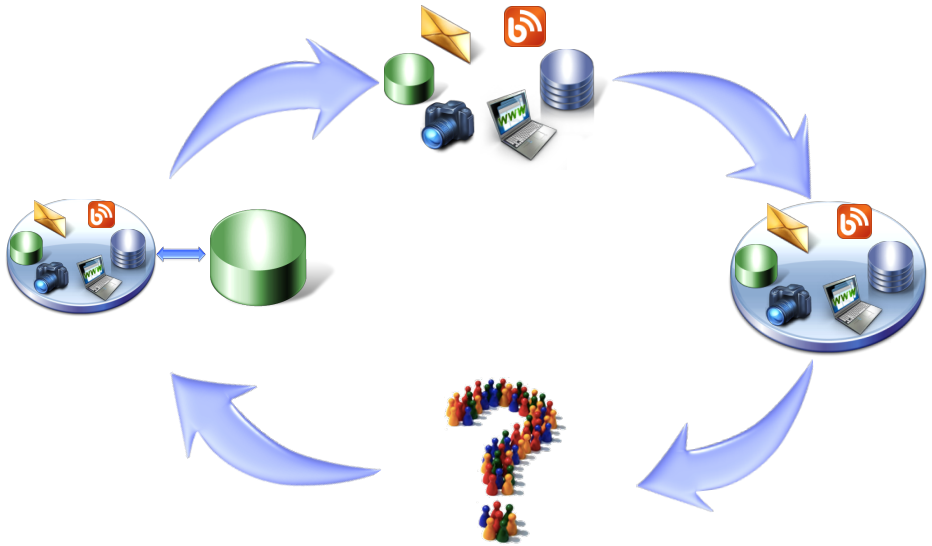
New requirements require redesign and rebuild

IT Designs a Solution with a set structure and functionality

Business executes queries to answer questions over and over

Big Data Analytics:
Information Sources Drive Creative Discovery

Business and IT Identify Information Sources Available



New insights drive integration to traditional technology

IT Delivers a Platform that enables creative exploration of all available data and content

Business determines what questions to ask by exploring the data and relationships

Merging the 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

Big Data Will Impact Every Aspect of Your Business

Know Everything about your Customer

Analyze all sources of data to know your customers as individuals, from channel interactions to social media.

Run Zero-latency Operations

Analyze all available operational data and react in real-time to optimize processes. Reduce the cost of IT with new cost-effective technologies.

Innovate New Products at Speed and Scale

Capture all sources of feedback and analyze vast amounts of market and research data to drive innovation.

Instant Awareness of Fraud and Risk

Develop better fraud/risk models by analyzing all available data, and detect fraud in real-time with streaming transaction analysis.

Exploit Instrumented Assets

Monitor assets from real-time data feeds to predict and prevent maintenance issues and develop new products and services.

In Order to Realize New Opportunities, You Need to Think Beyond Traditional Sources of Data

Transactional and Application Data

- Volume
- Structured
- Throughput

Machine Data

- Velocity
- Semi-structured
- Ingestion

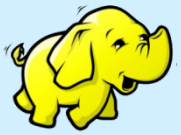
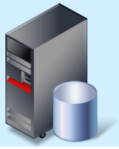
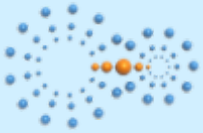
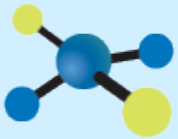
Social Data

- Variety
- Highly unstructured
- Veracity

Enterprise Content

- Variety
- Highly unstructured
- Volume

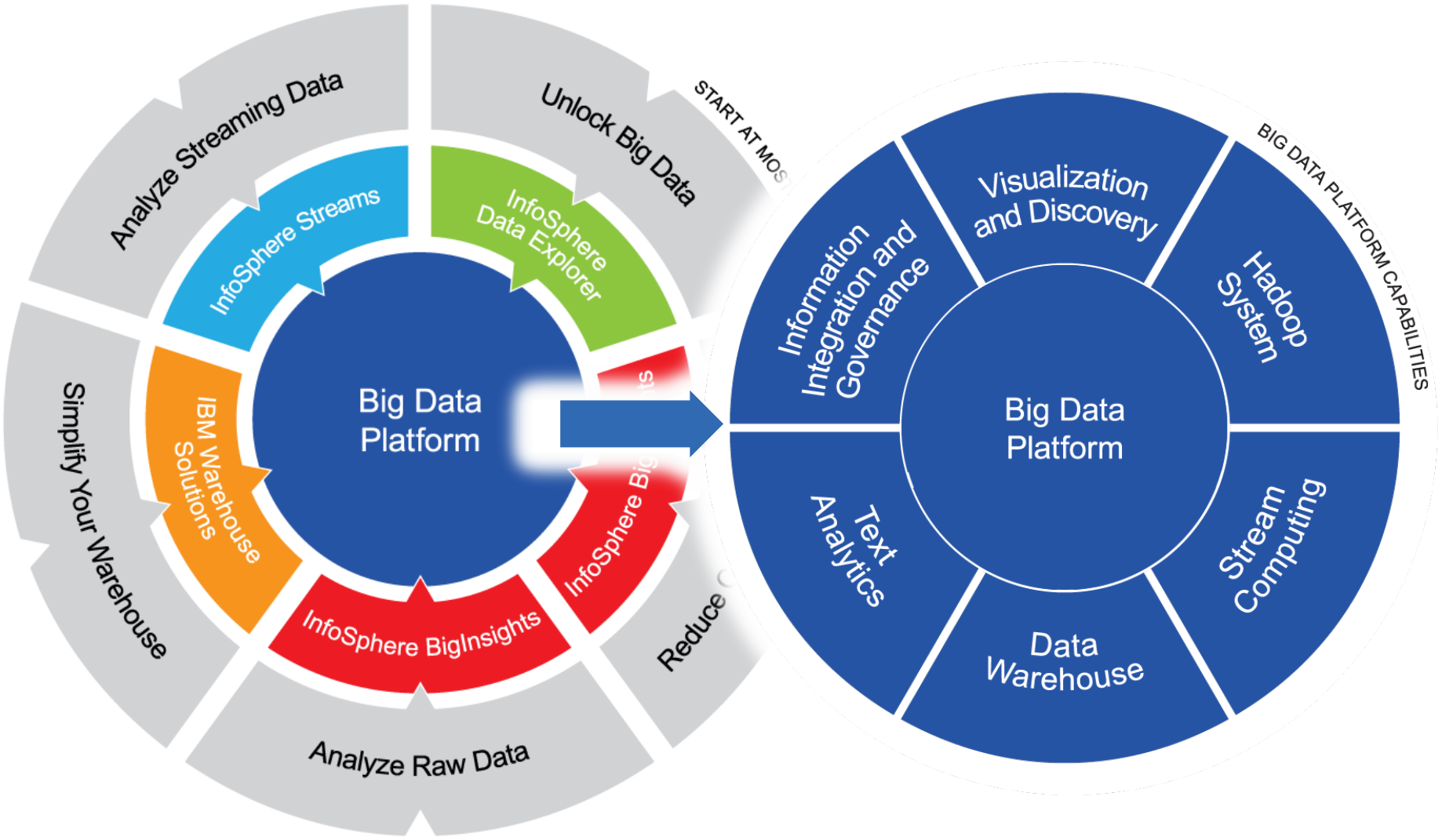
Leveraging Big Data Requires Multiple Platform Capabilities

Understand and navigate federated big data sources		Federated Discovery and Navigation
Manage & 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, Lifecycle Management, MDM

Business-centric Big Data enables you to start with a critical business pain and expand the foundation for future requirements

- "Big data" isn't just a technology—it's a business strategy for capitalizing on information resources
- Getting started is crucial
- Success at each entry point is accelerated by products within the Big Data platform
- Build the foundation for future requirements by expanding further into the big data platform

Expand with the Big Data Platform for future needs



1 – Unlock Big Data

- **Customer Need**
  - Understand existing data sources
  - Expose the data within existing content management and file systems for new uses, without copying the data to a central location
  - Search and navigate big data from federated sources
- **Value Statement**
  - Get up and running quickly and discover and retrieve relevant big data
  - Use big data sources in new information-centric applications
- **Customer examples**
  - Proctor and Gamble – Connect employees with a 360° view of big data sources
- **Get started with: InfoSphere Data Explorer**

2 – Analyze Raw Data

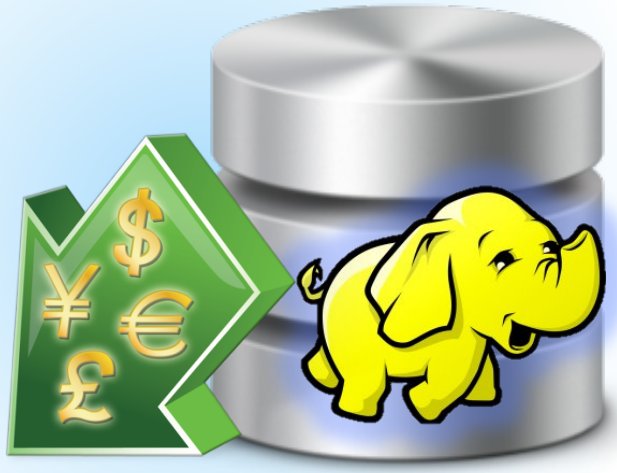
- **Customer Need**
  - Ingest data as-is into Hadoop and derive insight from it
  - Process large volumes of diverse data within Hadoop
  - Combine insights with the data warehouse
  - Low-cost ad-hoc analysis with Hadoop to test new hypothesis
- **Value Statement**
  - Gain new insights from a variety and combination of data sources
  - Overcome the prohibitively high cost of converting unstructured data sources to a structured format
  - Extend the value of the data warehouse by bringing in new types of data and driving new types of analysis
  - Experiment with analysis of different data combinations to modify the analytic models in the data warehouse
- **Customer examples**
  - Financial Services Regulatory Org – managed additional data types and integrated with their existing data warehouse
- **Get started with: InfoSphere BigInsights**

3 – Simplify your Warehouse

- **Customer Need**
  - Improve general-purpose data warehouse performance – optimize queries that take hours to run
  - Enterprise data warehouse is encumbered by too much data for too many purposes
  - Ability to ingest huge volumes of structured data and run multiple concurrent deep analytic queries
- **Value Statement**
  - Speed – 10-100x faster performance on deep analytics
  - Simplicity – minimal administration and tuning of the appliance
  - Integrated, workload-optimized systems; get up and running quickly
- **Customer examples**
  - Catalina Marketing – executing 10x the amount of predictive workloads with the same staff
- **Get started with: IBM PureData, Smart Analytics System, InfoSphere Warehouse**

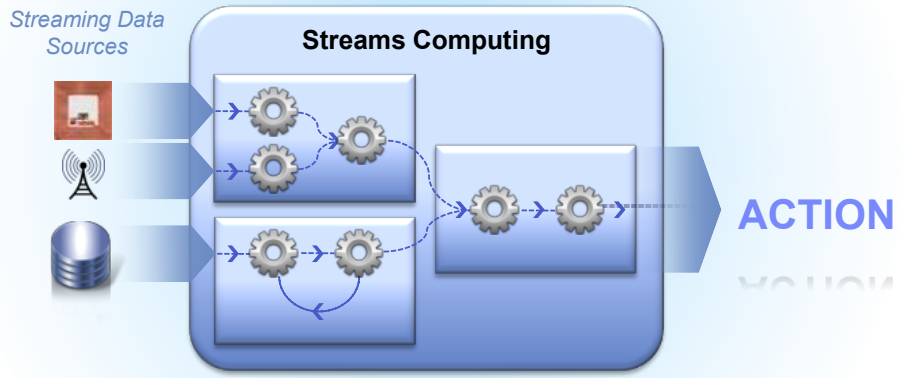
4 – Reduce costs with Hadoop

- **Customer Need**
  - Reduce the overall cost to maintain data in the warehouse – often its seldom used and kept ‘just in case’
  - Lower costs as data grows within the data warehouse
  - Reduce expensive infrastructure used for processing and transformations
- **Value Statement**
  - Support existing and new workloads on the most cost effective alternative, while preserving existing access and queries
  - Lower storage costs
  - Reduce processing costs by pushing processing onto commodity hardware and the parallel processing of Hadoop
- **Customer examples**
  - Financial Services Firm – move processing of applications and reports to Hadoop HBase while preserving existing queries
- **Get started with: IBM InfoSphere BigInsights**



5 – Analyze Streaming Data

- **Customer Need**
  - Harness and process streaming data sources
  - Select valuable data and insights to be stored for further processing
  - Quickly process and analyze perishable data, and take timely action
- **Value Statement**
  - Significantly reduced processing time and cost – process and then store what's valuable
  - React in real-time to capture opportunities before they expire
- **Customer examples**
  - Ufone – Telco Call Detail Record (CDR) analytics for customer churn prevention
- **Get started with: InfoSphere Streams**



Entry Points are Accelerated by Products Within the Big Data Platform

1 – Unlock Big Data

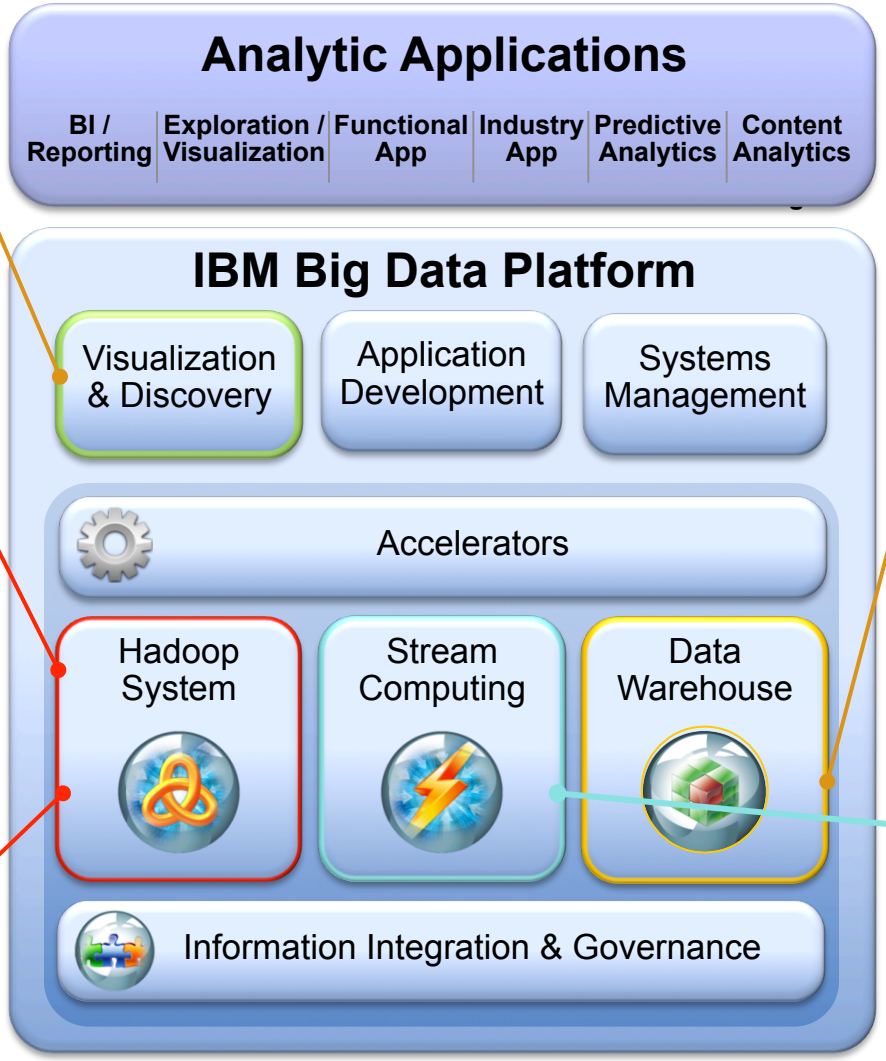
InfoSphere Data Explorer

2 – Analyze Raw Data

InfoSphere BigInsights

4 – Reduce costs with Hadoop

InfoSphere BigInsights



3 – Simplify your warehouse

IBM Warehouse Solutions

5 – Analyze Streaming Data

InfoSphere Streams

There are Many Use Cases for a Big Data Platform

Know Everything about your Customer

- Social media customer sentiment analysis
- Promotion optimization
- 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

Instant Awareness of Risk and Fraud

- Multimodal surveillance
- Cyber security
- Fraud modeling & detection
- Risk modeling & management
- Regulatory reporting

Exploit Instrumented Assets

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

Achieve Breakthrough Outcomes With Big Data Capabilities

Achieve Breakthrough Outcomes

- Know Everything About Your Customers
- Run Zero-latency Operations
- Innovate new products at Speed and Scale
- Instant Awareness of Fraud and Risk
- Exploit Instrumented Assets

With Unique Capabilities

- Visualization and Discovery
- Hadoop
- Data Warehousing
- Stream Computing
- Integration and Governance
- Text Analytics

To Analyze Any Big Data Type

- Transactional / Application Data
- Machine Data
- Social Media Data
- Content

The Platform Advantage

- The platform provides benefit as you move from an entry point to a second and third project
- Shared components and integration between systems lowers deployment costs
- Key points of leverage
  - Reuse text analytics across Streams and Hadoop
  - HDFS connectors between Streams and Information Integration
  - Common integration, meta data and governance across all engines
  - Accelerators built across multiple engines – common analytics, models, and visualization