Name: A.Ahmed Riyas Kaan
Reg no:822221104002
year:3
Title:Big data analysis with IBM cloud databases
college: University college of engineering, Thirukkuvalai

Overview - Big Data & Analytics

Agenda

- **What is Big Data?**
  - Concepts
  - Characteristics
- **Business Motivation**
  - Big Data Challenges
  - How Big Data Impacts Every Aspect of Your Business
  - A Big Data Journey
- **IBM Big Data Platform**
  - InfoSphere Data Explorer
  - InfoSphere BigInsights
  - IBM PureData Systems, InfoSphere Warehouse
  - InfoSphere Streams
- **Big Data Use Cases**
- **Get Started**

What is Big Data?

- **All kinds of data**
  - Large volumes
  - Valuable insight, but difficult to extract
  - May be extremely time sensitive
- **Big Data is a Hot Topic Because Technology Makes it Possible to Analyze ALL Available Data**

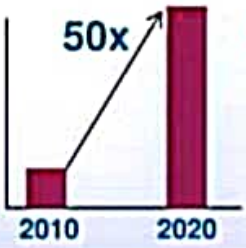
*"**Big data** technologies describe a new generation of technologies and architectures, designed to **economically extract value** from very large **volumes** of a wide **variety** of data, by enabling high **velocity** capture, discovery and/or analysis."*

Source: Matt Eastwood, IDC

Characteristics of Big Data

- V^4 = Volume Velocity Variety Veracity

Cost efficiently processing the growing Volume



35 ZB

Responding to the increasing Velocity

30 Billion RFID sensors and counting

Collectively analyzing the broadening Variety

80% of the worlds data is unstructured

Establishing the Veracity of big data sources

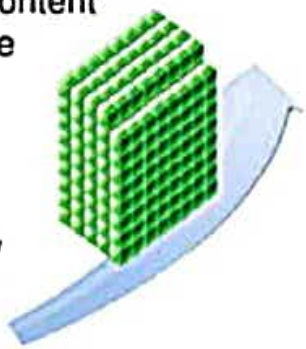
1 in 3 business leaders don't trust the information they use to make decisions

Information is at the Center of a New Wave of Opportunity...

44x

as much Data and Content Over Coming Decade

2020
35 zettabytes



Velocity

Variety

Volume

2009
800,000 petabytes

80%

Of world's data is unstructured

... And Organizations Need Deeper Insights

1 in 3 Business leaders frequently make decisions based on information they don't trust, or don't have

1 in 2 Business leaders say they don't have access to the information they need to do their jobs

83% of CIOs cited "Business intelligence and analytics" as part of their visionary plans to enhance competitiveness

60% of CEOs need to do a better job capturing and understanding information rapidly in order to make swift business decisions

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

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**

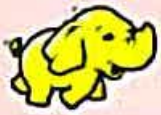
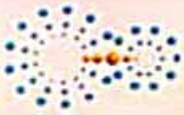
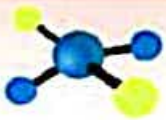
Retailer reduces time to run queries by **80%** to optimize inventory

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

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

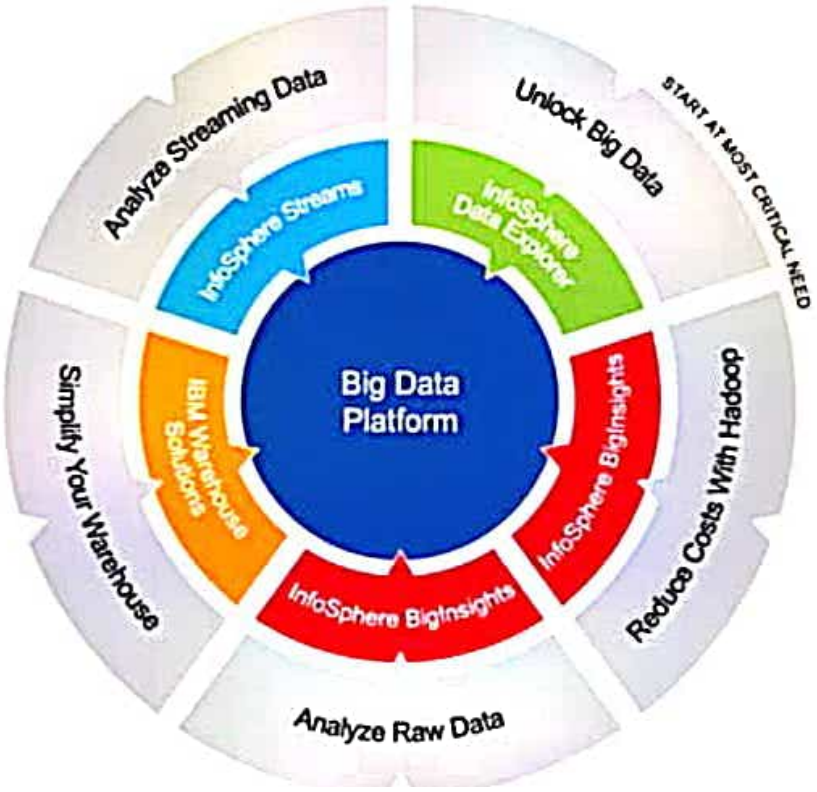
Leveraging Big Data Requires Multiple Platform Capabilities

CONSUMABLE
(What not How/Patterns/Expert Systems, +++)

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

IBM's Business-centric Big Data Platform

- **Enables you to start with a critical business needs 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

A Big Data Journey:
Anticipating and Improving Customer Interactions

- **Financial and tax preparation software and services**
- **$4.15B rev 2012**

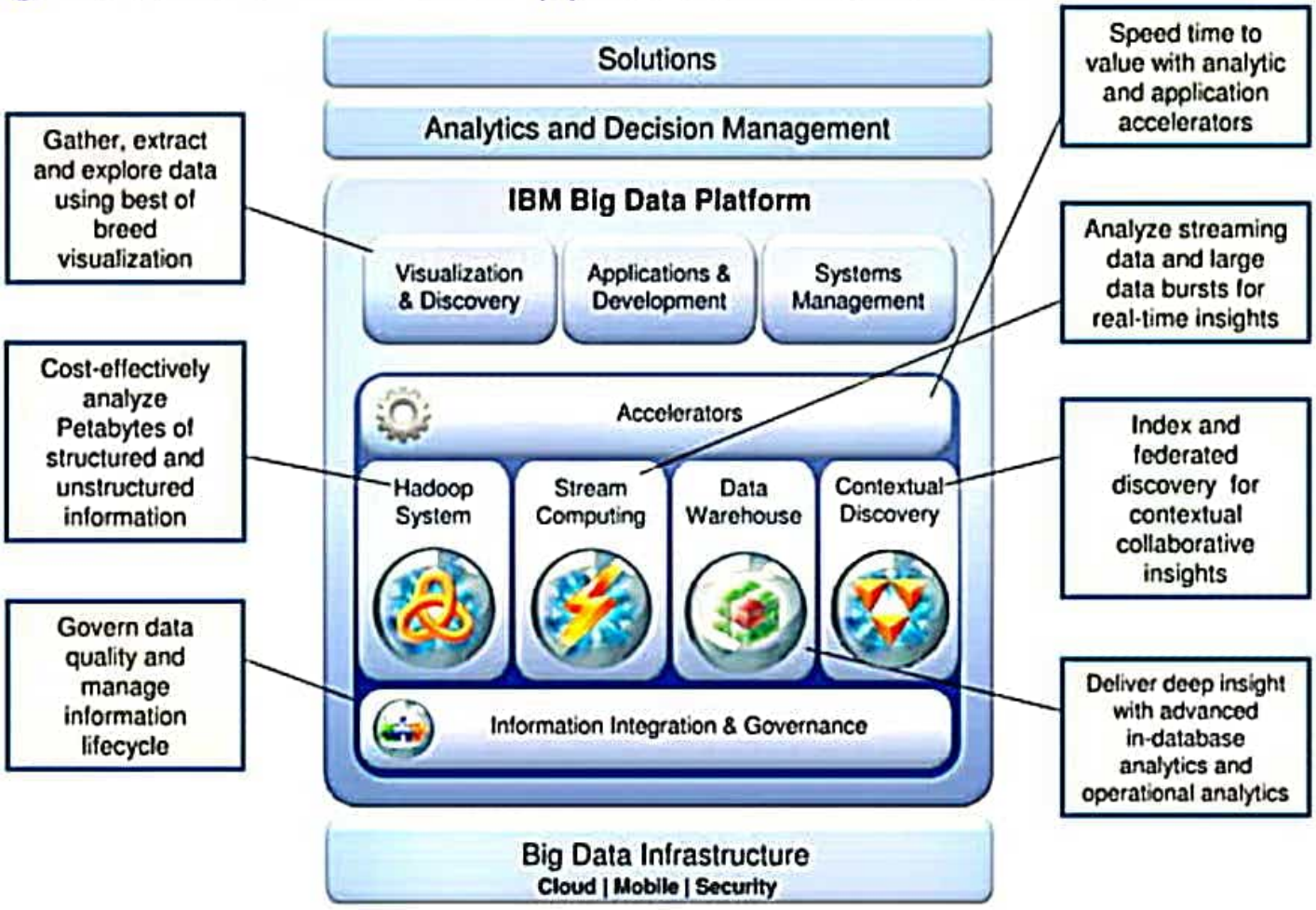
Project 1: Big Data Foundation
-Data Warehousing, Data Quality, Customer Data Hub
-Single view of the customer

Project 2: Analytics
-Customer behavior and segmentation analysis
-Reduced customer churn 10%
-$10M new revenue in 12months

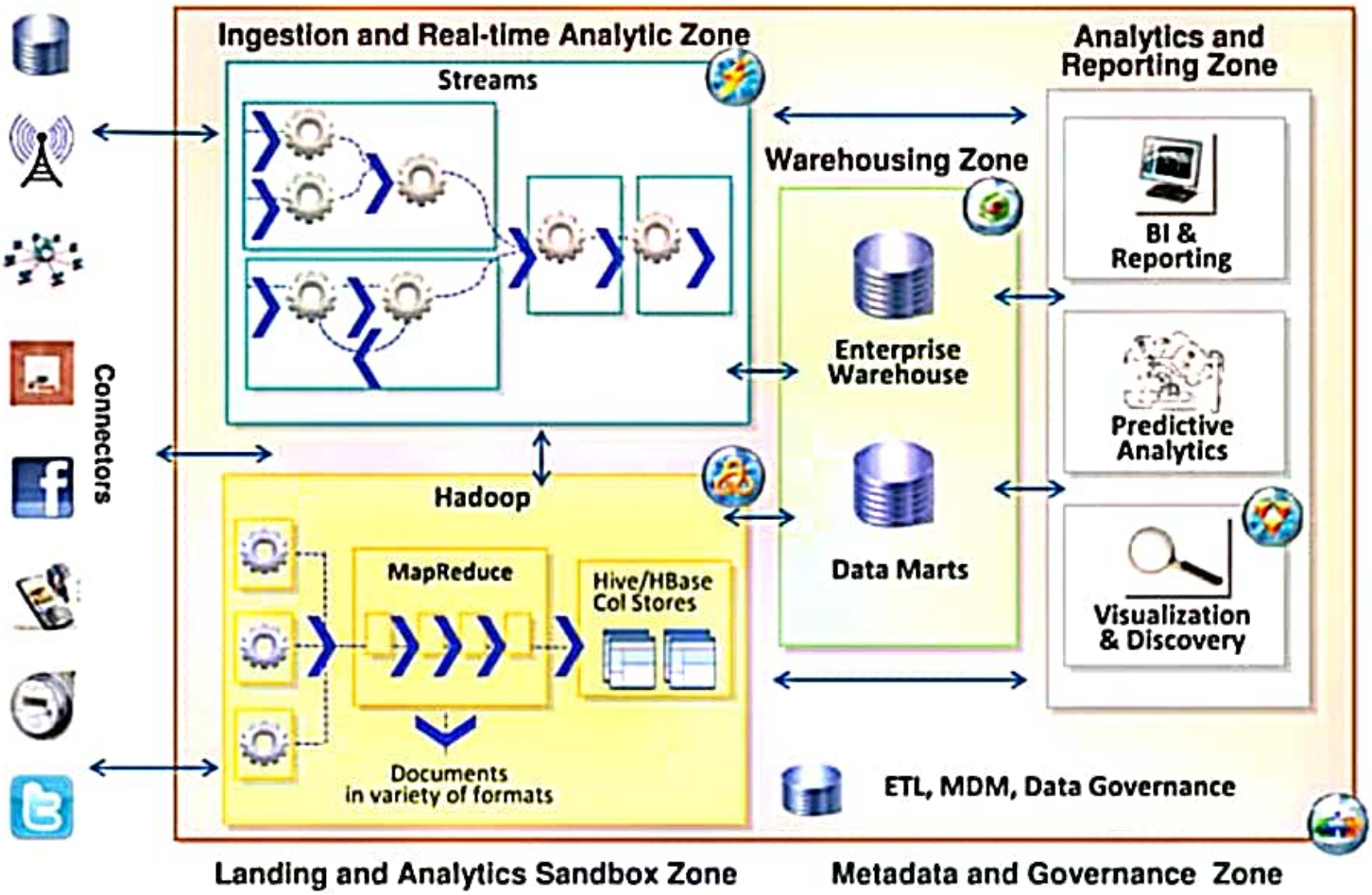
Project 3: Unstructured Data Analytics
-Social media analysis, Log Analysis, Text Analytics
-Augment customer profiles with new data sources
-Data warehouse cost optimization
-Data Exploration

Project 4: Real Time Analytics
-No latency analytics
-Real time behavior prediction
-Real time customer segmentation

Big Data Platform and Application Frameworks

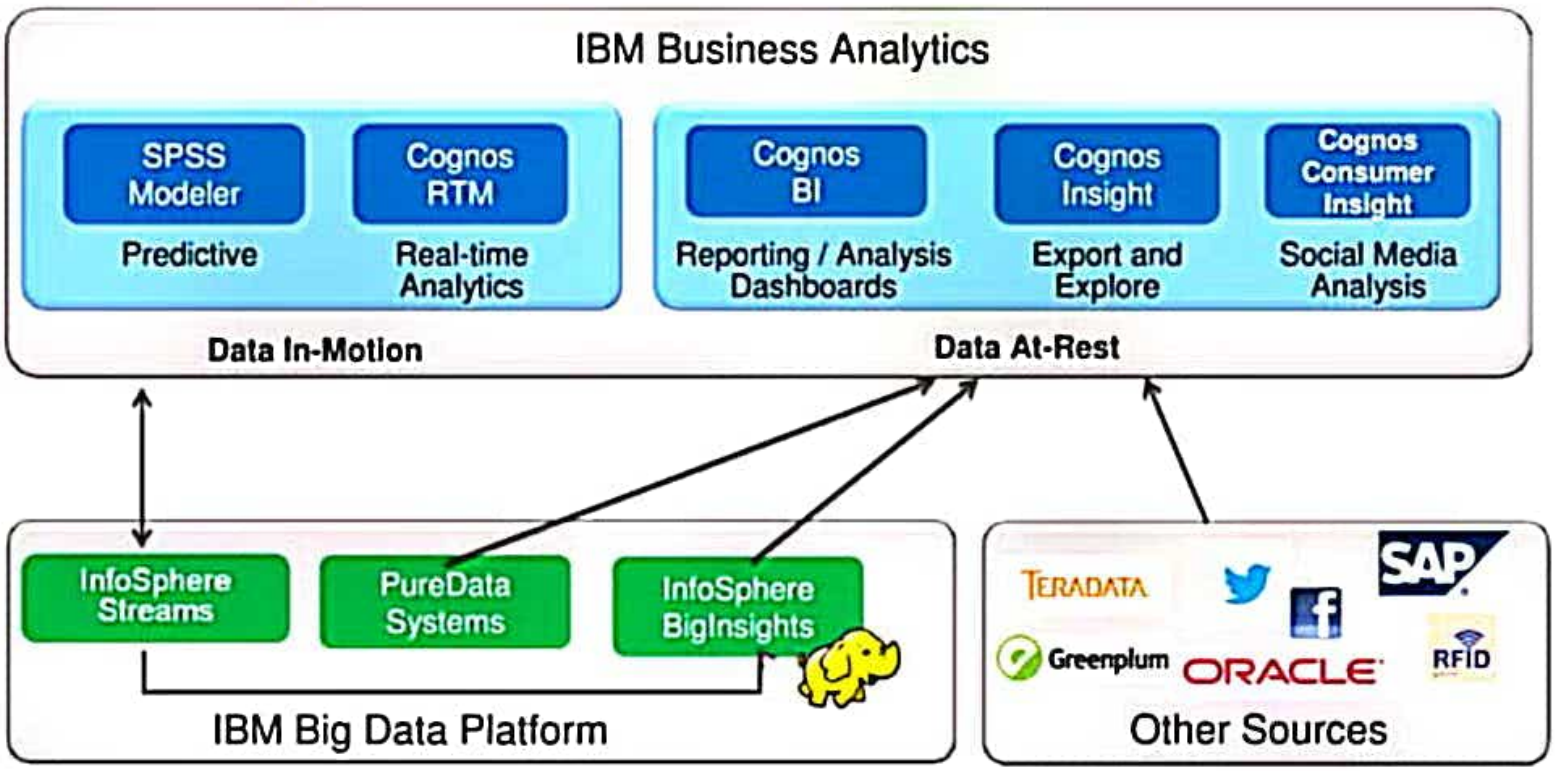


An example of the big data platform in practice



TECHNOLOGY

Example: Integrate big data sources with enterprise data



Big Data Key Use Cases:

Big Data Exploration
Find, visualize, understand all big data to improve decision making

Enhanced 360° View of the Customer
Extend existing customer views (MDM, CRM, etc) by incorporating additional internal and external information sources

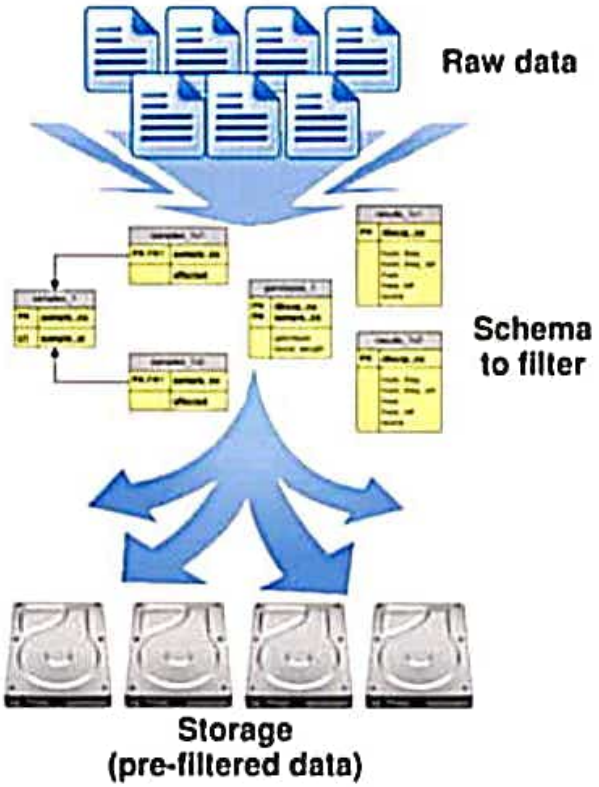
Security/Intelligence Extension
Lower risk, detect fraud and monitor cyber security in real-time

Operations Analysis
Analyze a variety of machine data for improved business results

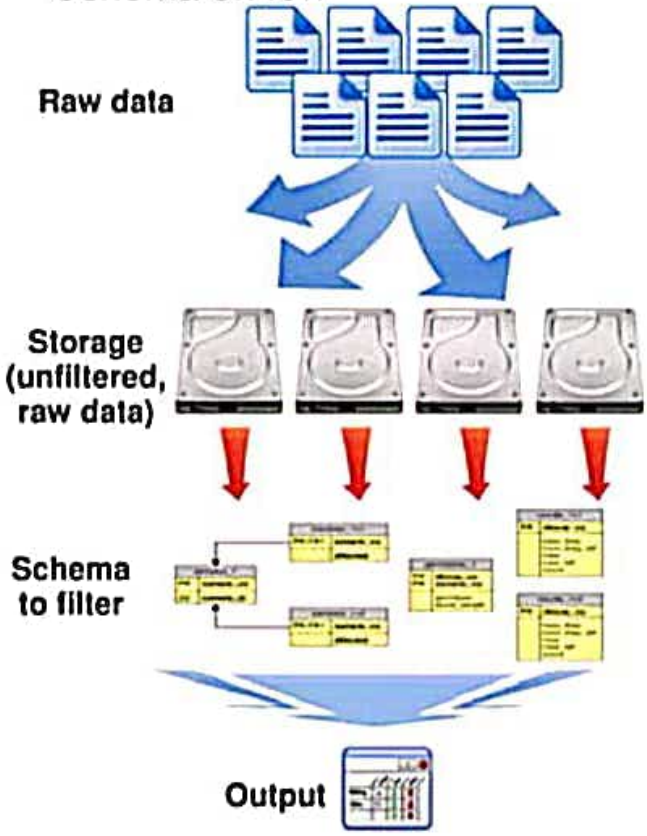
Data Warehouse Augmentation
Integrate big data and data warehouse capabilities to increase operational efficiency

Big Difference: Schema on Run

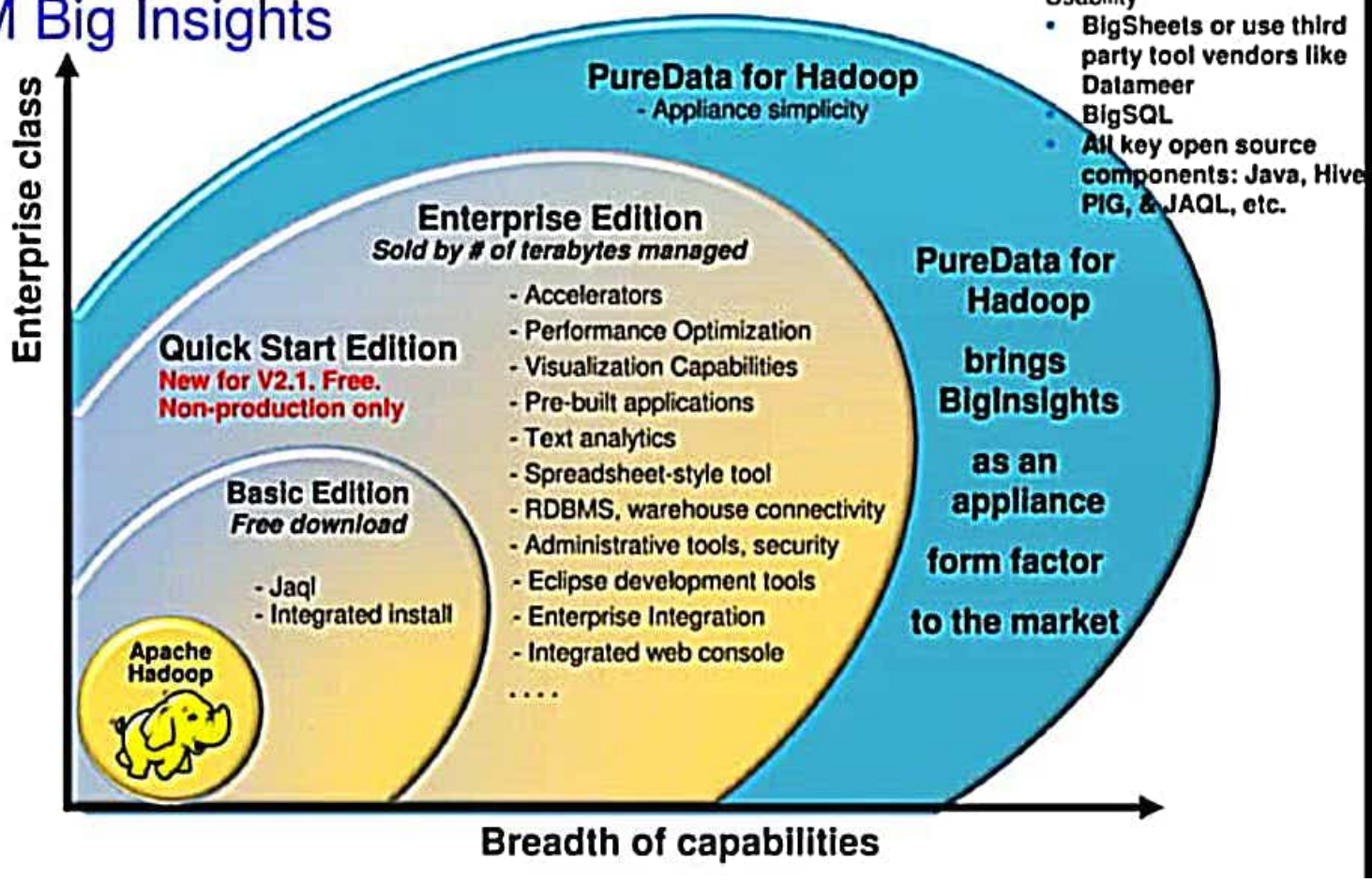
- **Regular database**
  - Schema on load



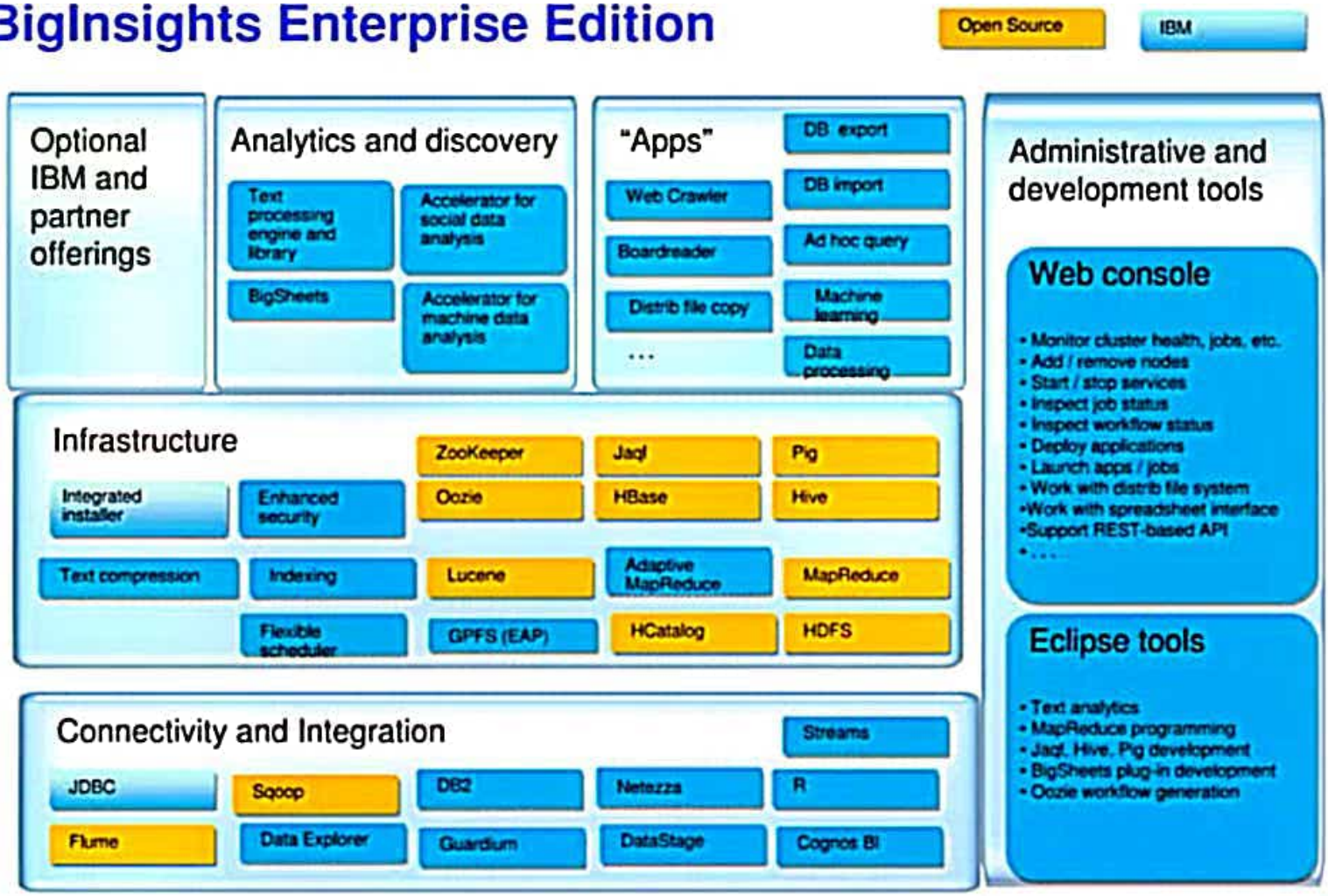
- **Big Data (Hadoop)**
  - Schema on run



From Getting Starting to Enterprise Deployment: IBM Big Insights



BigInsights Enterprise Edition



Stream Computing Represents 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

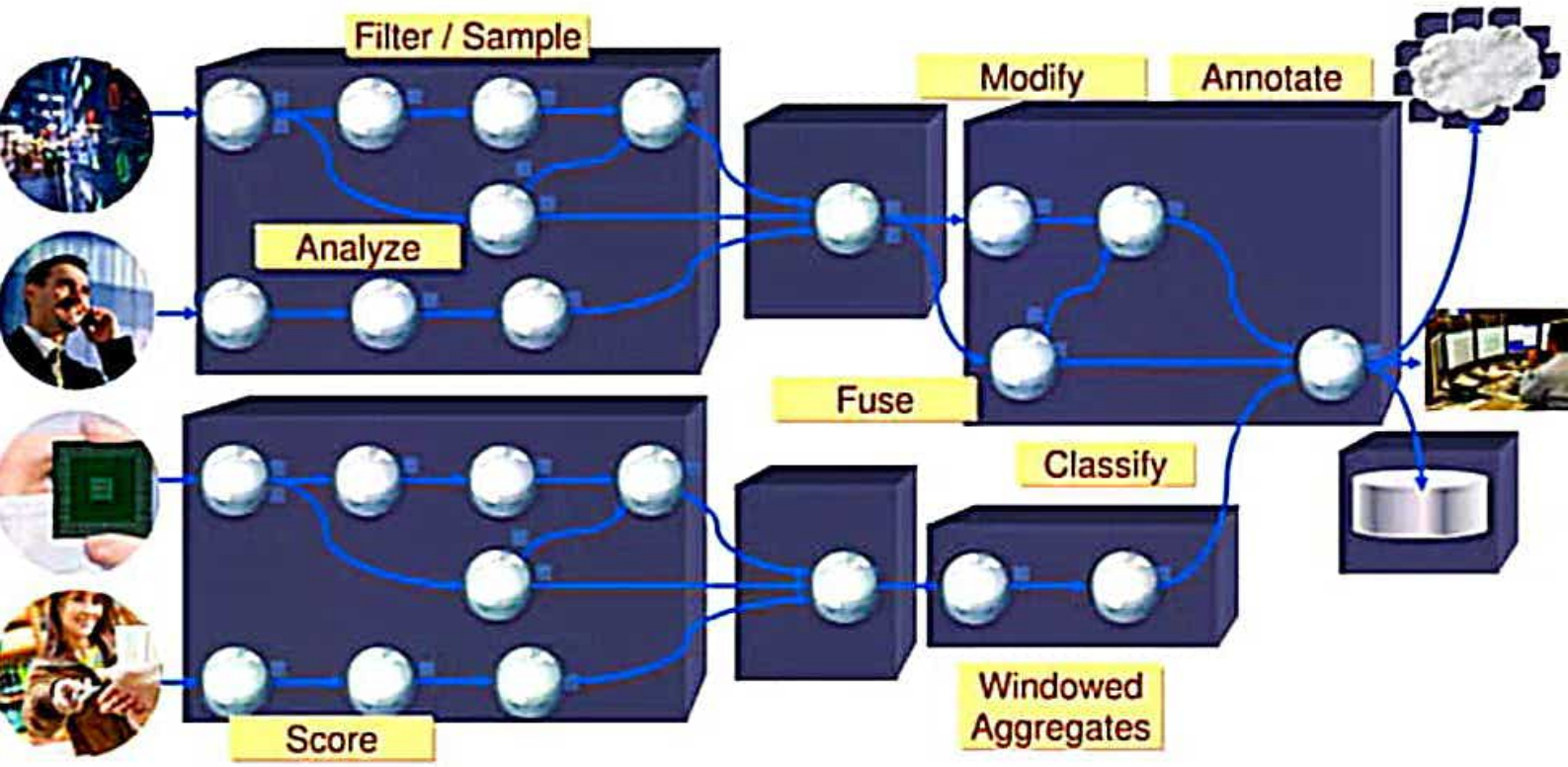
Current fact finding

Analyze data in motion – before it is stored

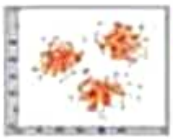
Low latency paradigm, push model

Data driven – bring data to the analytics

Big Data in real-time with InfoSphere Streams



Analytic Accelerators Designed for Velocity (and Variety)

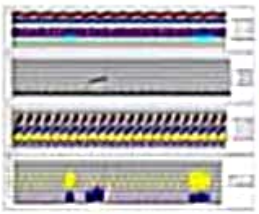


Mining in Microseconds
(included with Streams)

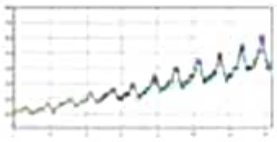
Acoustic
(IBM Research)
(Open Source)

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

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



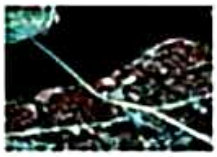
Advanced Mathematical Models
(IBM Research)



Predictive
(IBM Research)

$$\sum_{population} R(s_i, a_i)$$

Statistics
(included with Streams)

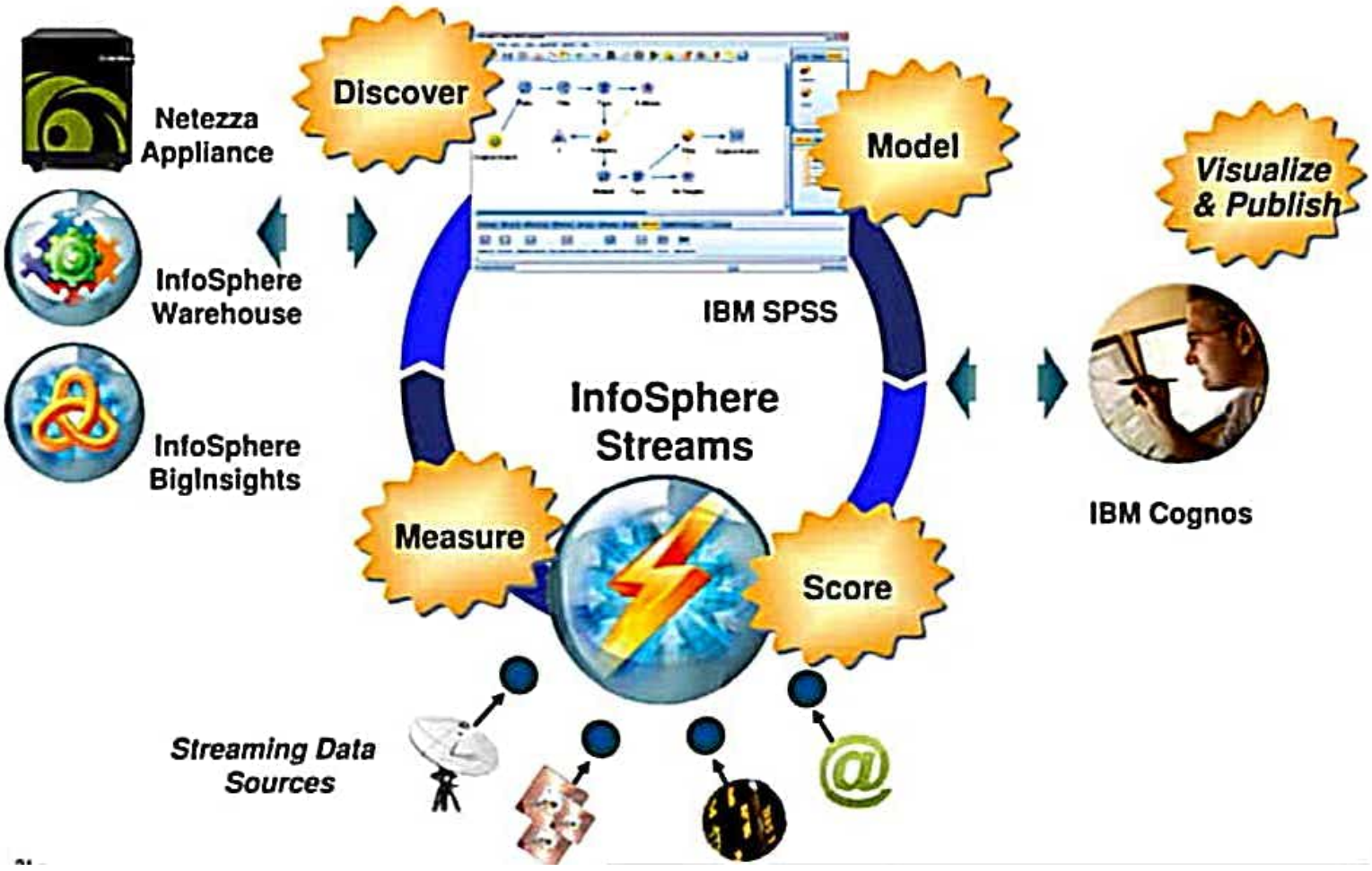


Geospatial
(IBM Research)

Image & Video
(Open Source)

OpenCV

Putting it all together ...end-to-end big data solution



#1 challenge customers face in Big Data: Unlocking the value of information through a single interface

InfoSphere Data Explorer

Providing unified, real-time access and fusion of big data unlocks greater insight and ROI

- Analyzes structured & unstructured data—in place
- Unique positional indexing
- Unlimited scalability
- Advanced data asset navigation
- Pattern clustering

- Virtual documents Contextual intelligence
- Text analytics
- Secure data integration
- Query transformation
- Easy-to-deploy big data applications
- User-friendly customisable interface

Improve customer service & reduce call times

Create unified view of ALL information for real-time monitoring

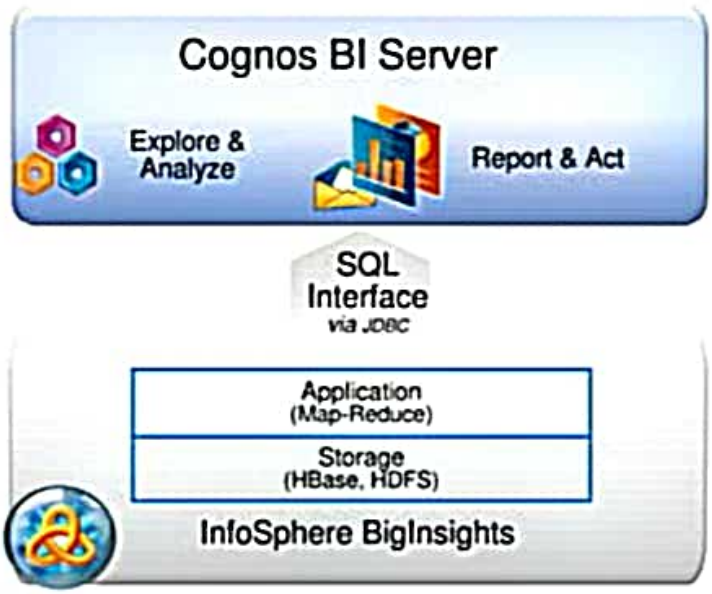
Increase productivity & leverage past work increasing speed to market

Analyze customer analytics & data to unlock true customer value

Identify areas of information risk & ensure data compliance

Cognos Business Intelligence optimized for Big SQL

- **Big SQL enables the Cognos BI server to delegate many types of analytical computations to BigInsights MapReduce processing instead of computing them locally at a performance cost like it would do with Hive**
- **Faster response times due to increased opportunity for query processing to occur closer to the data**
- **Not hindered by the latency and other limitations of querying Hadoop via Hive**



Performance – Cognos BI + DB2 BLU

18X
Faster cube load*

14X
Faster DB Query*

38x
Average Acceleration

Of database queries for reporting[2]

Cognos BI
+
DB2 BLU
+
Power

2. Based on internal tests.

IBM PureData Systems

Meeting Big Data Challenges – Fast and Easy!

PureData
System for Operational Analytics

For apps like Real-time Fraud Detection...
Operational data warehouse services optimized to balance high performance analytics and real-time operational throughput

---

PureData
System for Analytics

For apps like Customer Analysis...
Data warehouse services optimized for high-speed, peta-scale analytics and simplicity

---

PureData
System for Hadoop

For Exploratory Analysis & Queryable Archive
Hadoop data services optimized for big data analytics and online archive with appliance simplicity

---

PureData
System for Transactions

For apps like E-commerce...
Database cluster services optimized for transactional throughput and scalability

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

Every Industry can Leverage Big Data and Analytics.

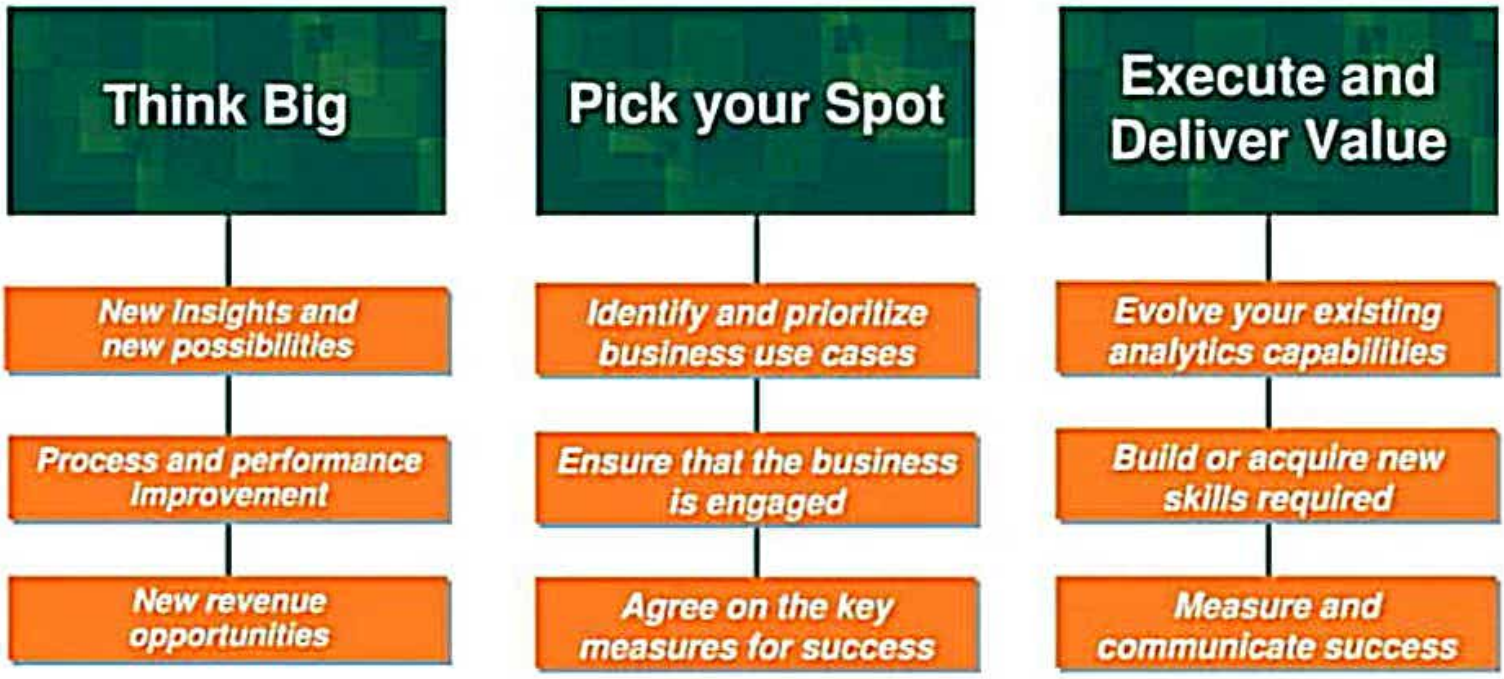
Clients Achieve Breakthrough Outcomes With IBM's Big Data Platform

	Imperative	Primary Capability	Business Value
Aircraft Manufacturer	Secure single point of access to all enterprise data	InfoSphere Data Explorer	Provide single point of access to disparate data sources
Vestas	Run Zero Latency Operations	InfoSphere BigInsights	Reduce maintenance costs and differentiate by optimal turbine placement
ufone	Know Everything about your Customers	InfoSphere Streams	Analyzed call records to drive real-time promotions & reduce churn
T-Mobile	Exploit Instrumented Assets	PureData for Analytics	Increased network availability by identifying and fixing holes
NYSE Euronext	Instant Awareness of Risk and Fraud	PureData for Analytics	Analysis time on 2 PB of data cut from 26 hours to 2 minutes

A Catalyst for ISV and Partner Innovation

Traditional Approach	Transformational Outcomes
Managing rising cost of care	PREMIER: Combining data from hundreds of hospitals to improve results across the healthcare continuum
Historical analysis of subscriber data	ROAMWARE: 2 million events analyzed per minute, delivering real-time insight to mobile operators
Customer segmentation based on loyalty data	agility: Capturing information from all interactions to improve customer lifetime value
Anti-corruption and bribery compliance program	EY: Use Big Data analytics to prioritize and isolate areas of risk or rogue activity
Manual supply chain integration	Manhattan: Provide visibility, analysis and reporting across the entire supply chain (planning -> execution)
Treat-first, seek-payment-later and write off bad debt	MEDEANALYTICS: Measure and predict patient payment behavior, reduce risk from bad debt and boost collection rates
Random parking meter patrols & search for open spots	STREETLINE: Analyzing parking systems to maximize revenue & improve the parking experience in cities

Get started!


Think Big
New insights and new possibilities
Process and performance improvement
New revenue opportunities
Pick your Spot
Identify and prioritize business use cases
Ensure that the business is engaged
Agree on the key measures for success
Execute and Deliver Value
Evolve your existing analytics capabilities
Build or acquire new skills required
Measure and communicate success

Thank You