Name:D.Ashwini
Year:3
Reg.no:822221104006
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**

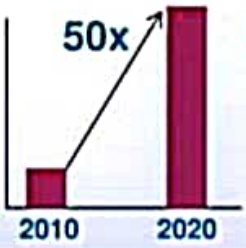
*"**Big data** technologies describe a new generation of technologies and architectures, designed to **economically extract value** from very large **volumes** of a wide **variety** of data, by enabling high **velocity** capture, discovery and/or analysis."*

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

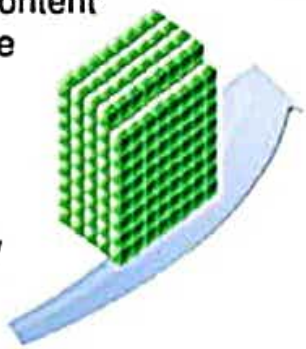
1 in 3 business leaders don't trust the information they use to make decisions

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**

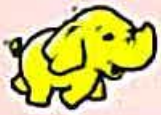
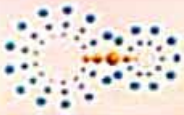
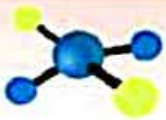
Retailer reduces time to run queries by **80%** to optimize inventory

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

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

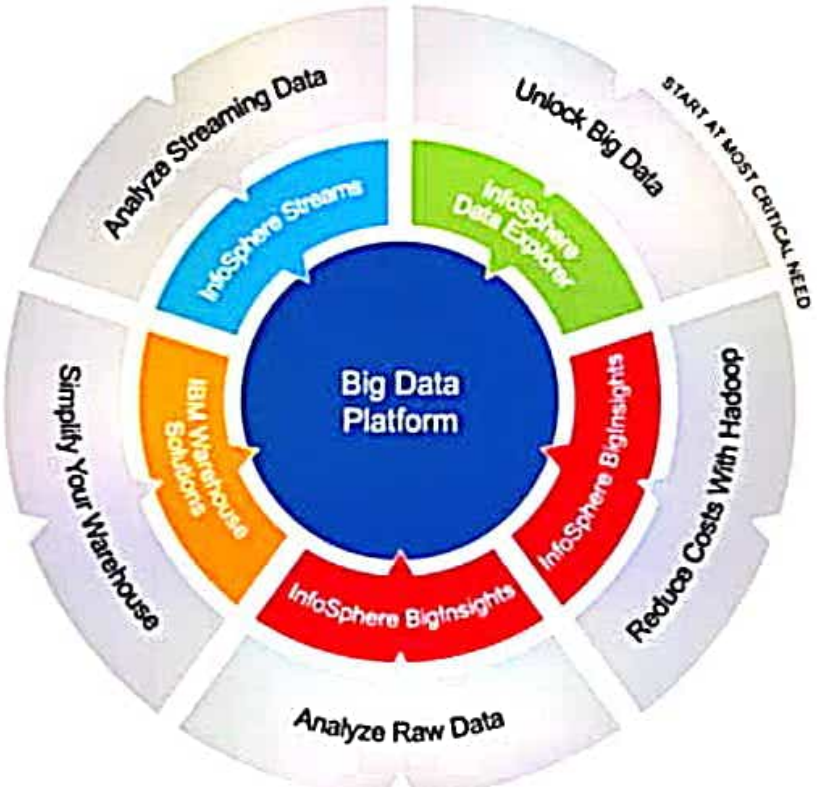
Leveraging Big Data Requires Multiple Platform Capabilities

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**

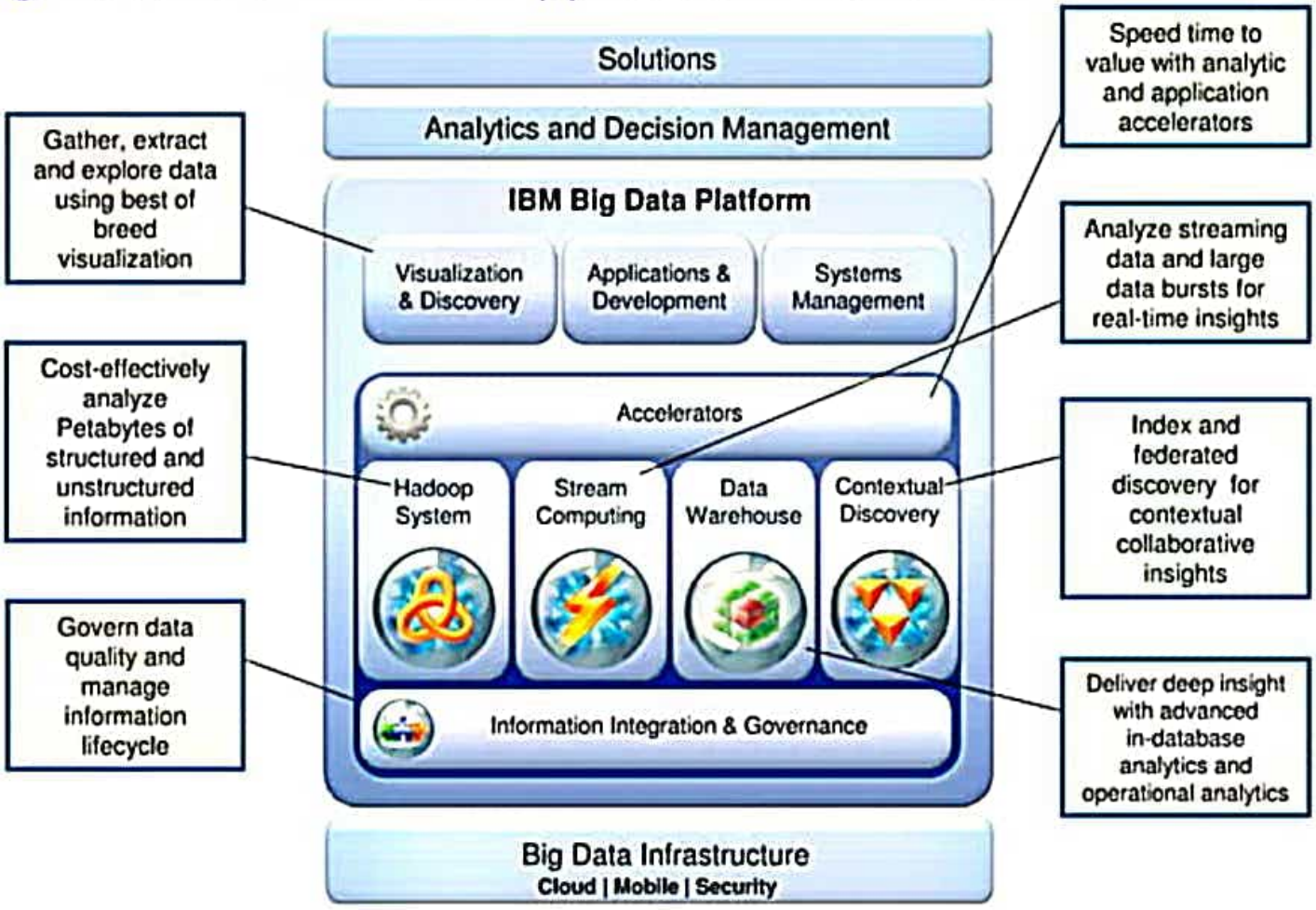
Project 1: Big Data Foundation
-Data Warehousing, Data Quality, Customer Data Hub
-Single view of the customer

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

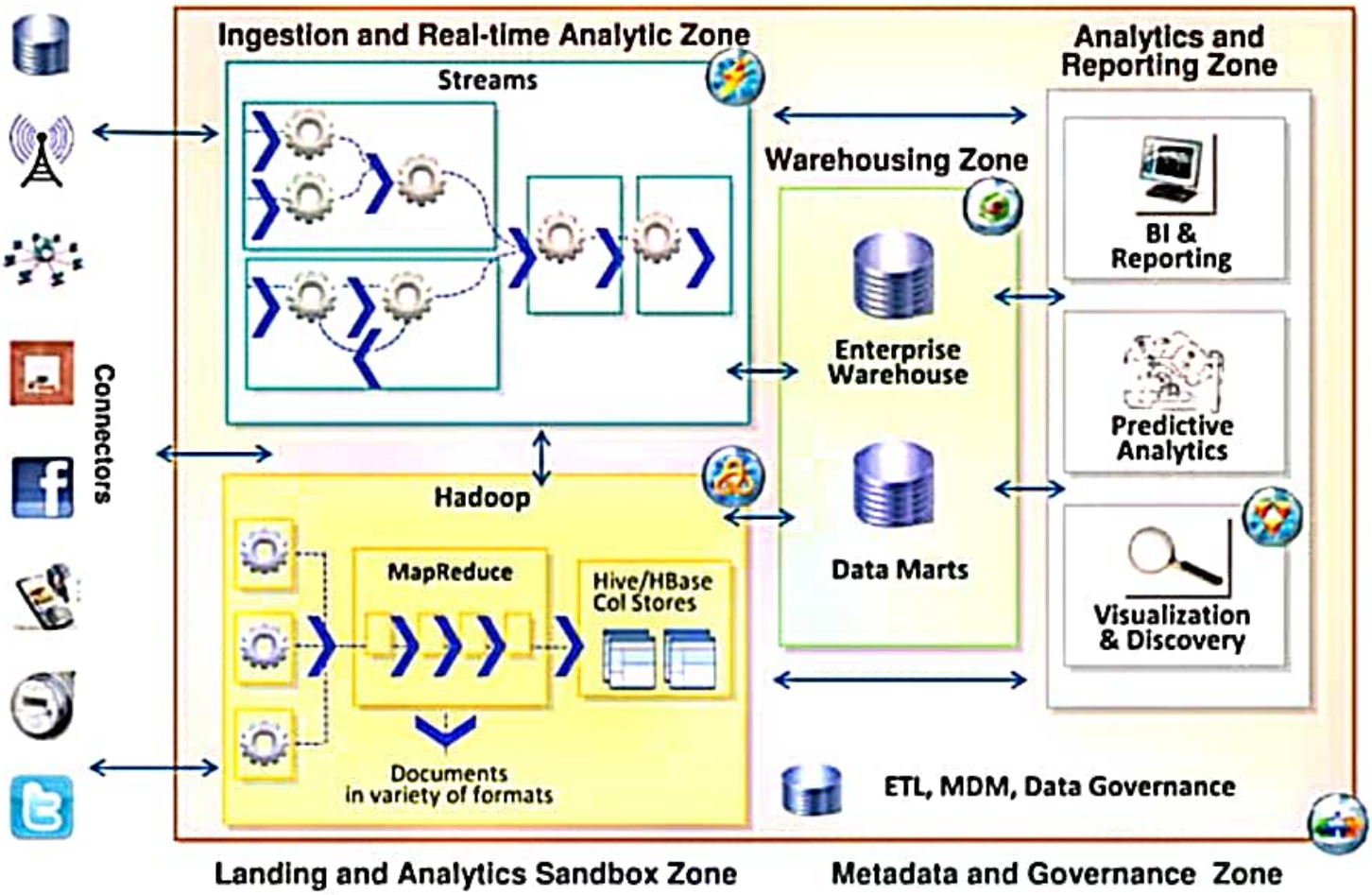
Project 3: Unstructured Data Analytics
-Social media analysis, Log Analysis, Text Analytics
-Augment customer profiles with new data sources
-Data warehouse cost optimization
-Data Exploration

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

Big Data Platform and Application Frameworks

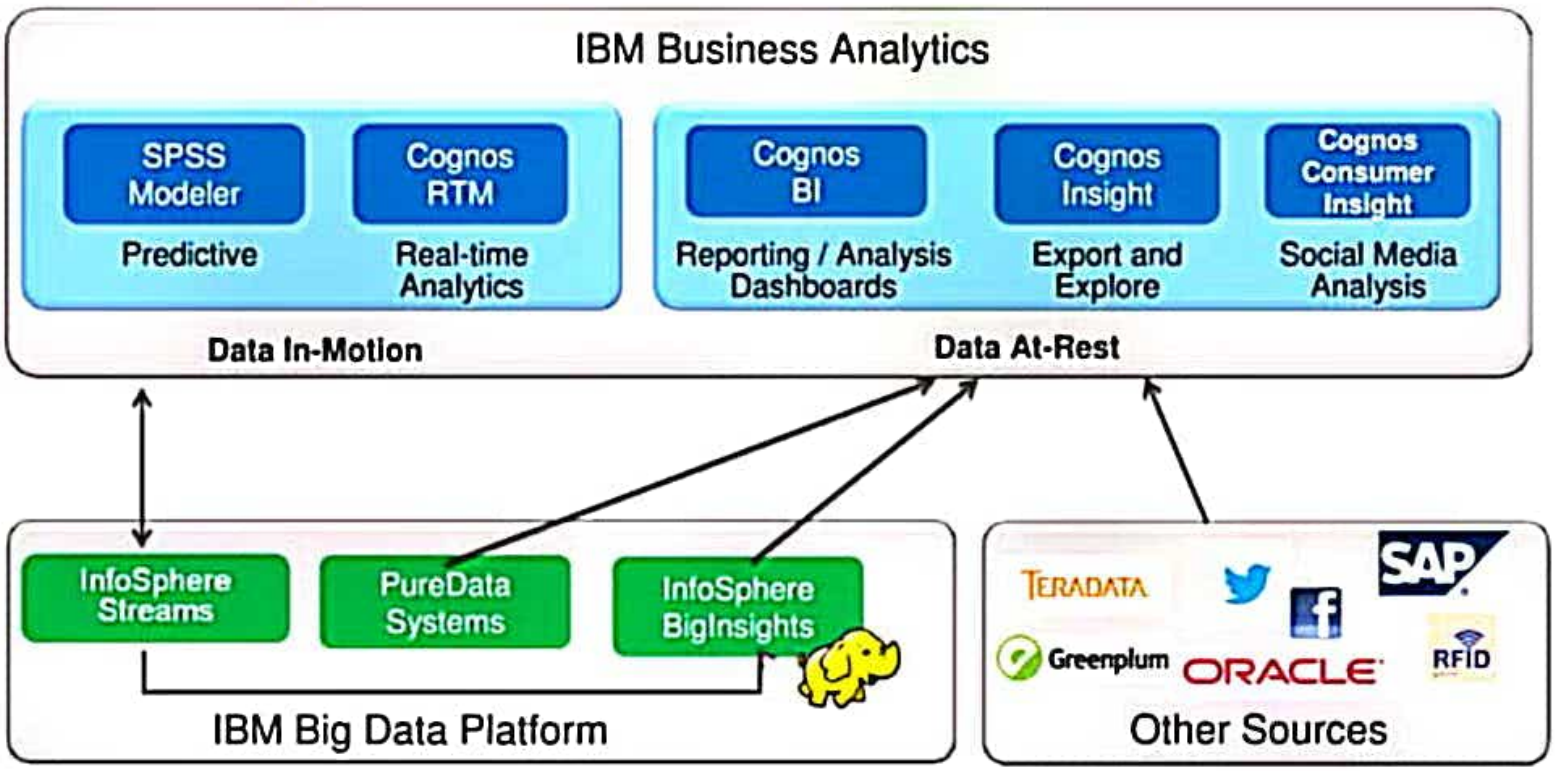


An example of the big data platform in practice



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

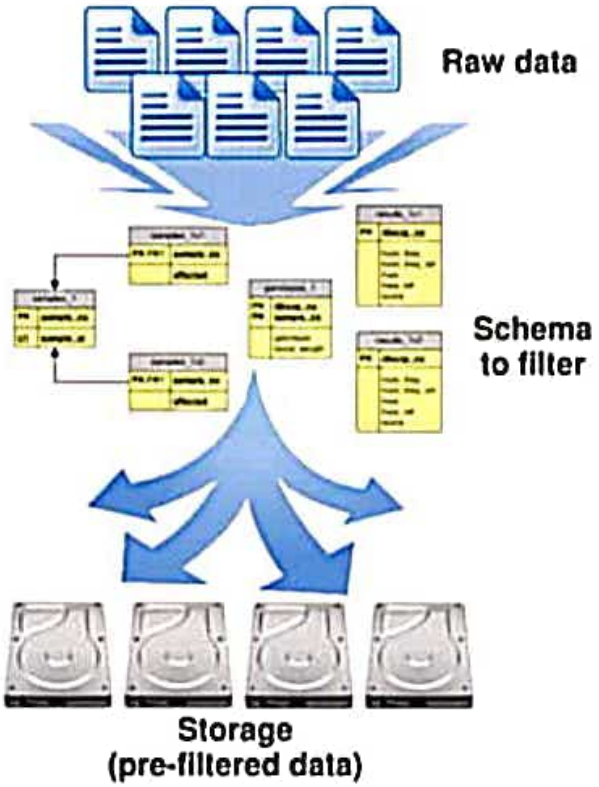
Security/Intelligence Extension
Lower risk, detect fraud and monitor cyber security in real-time

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

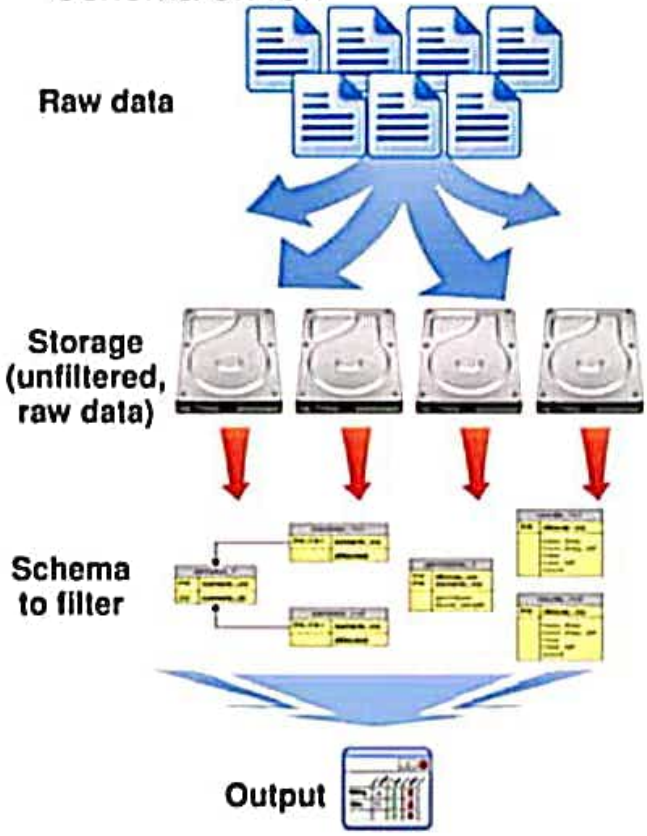
Data Warehouse Augmentation
Integrate big data and data warehouse capabilities to increase operational efficiency

Big Difference: Schema on Run

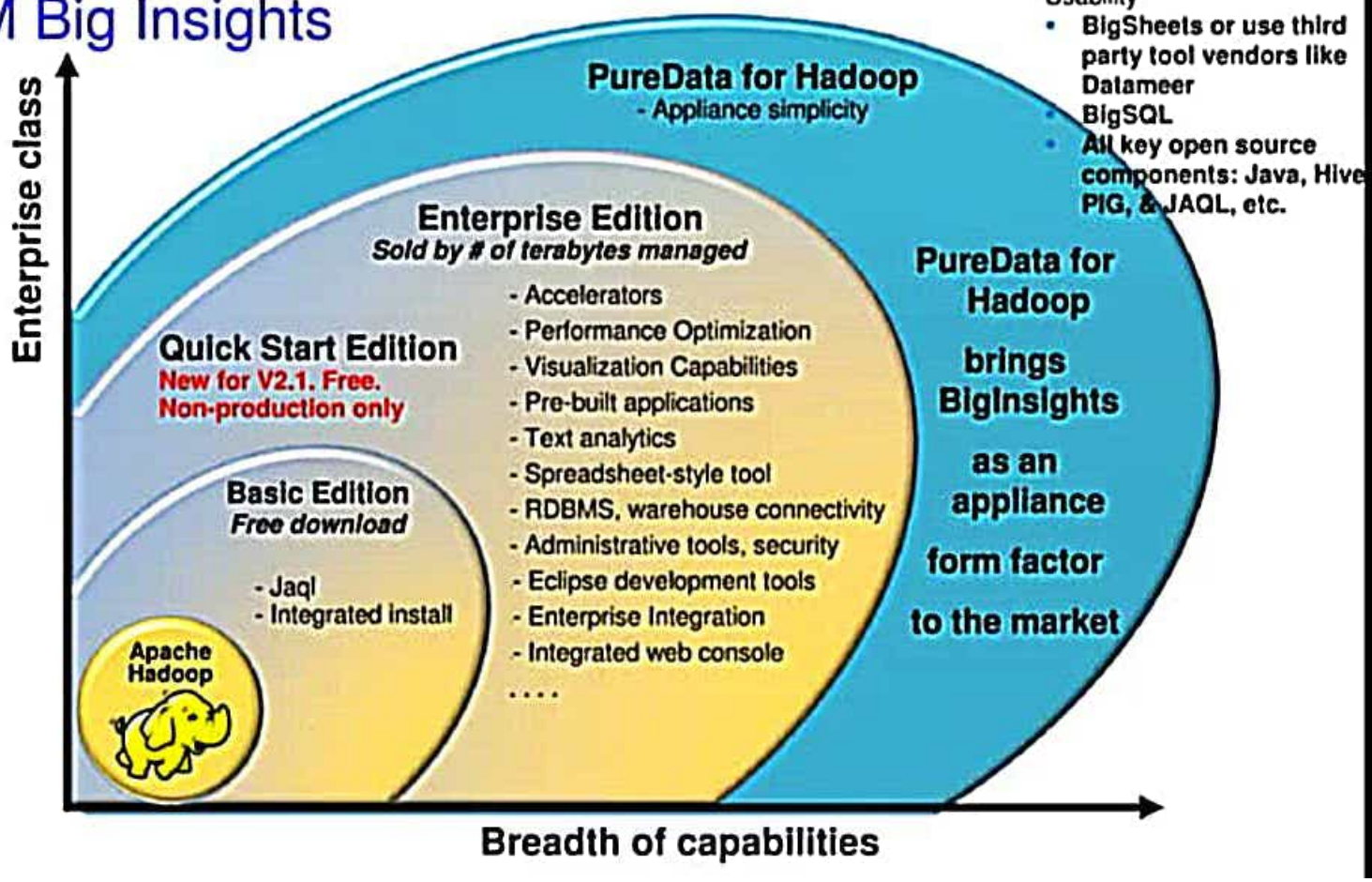
- **Regular database**
  - Schema on load



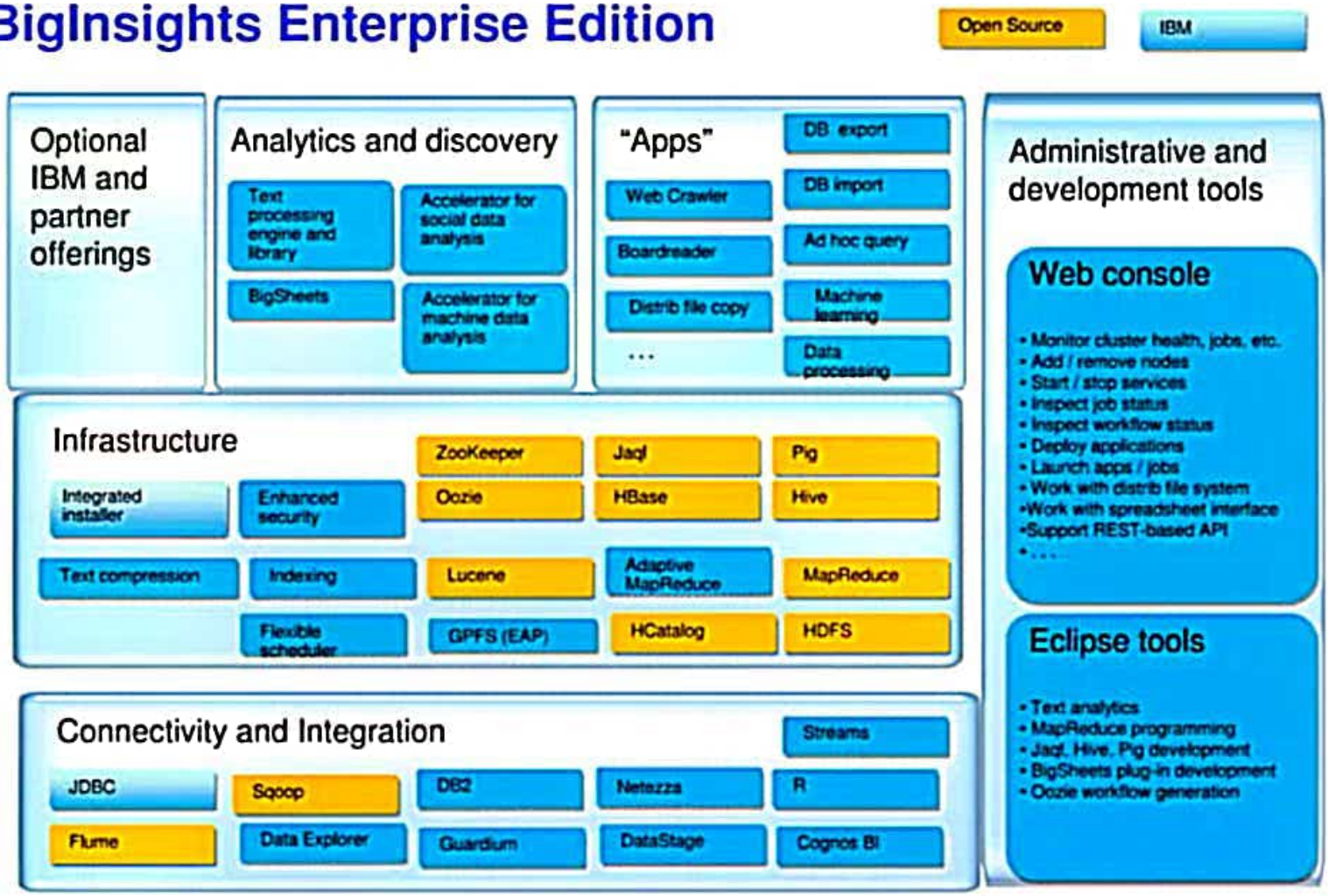
- **Big Data (Hadoop)**
  - Schema on run



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

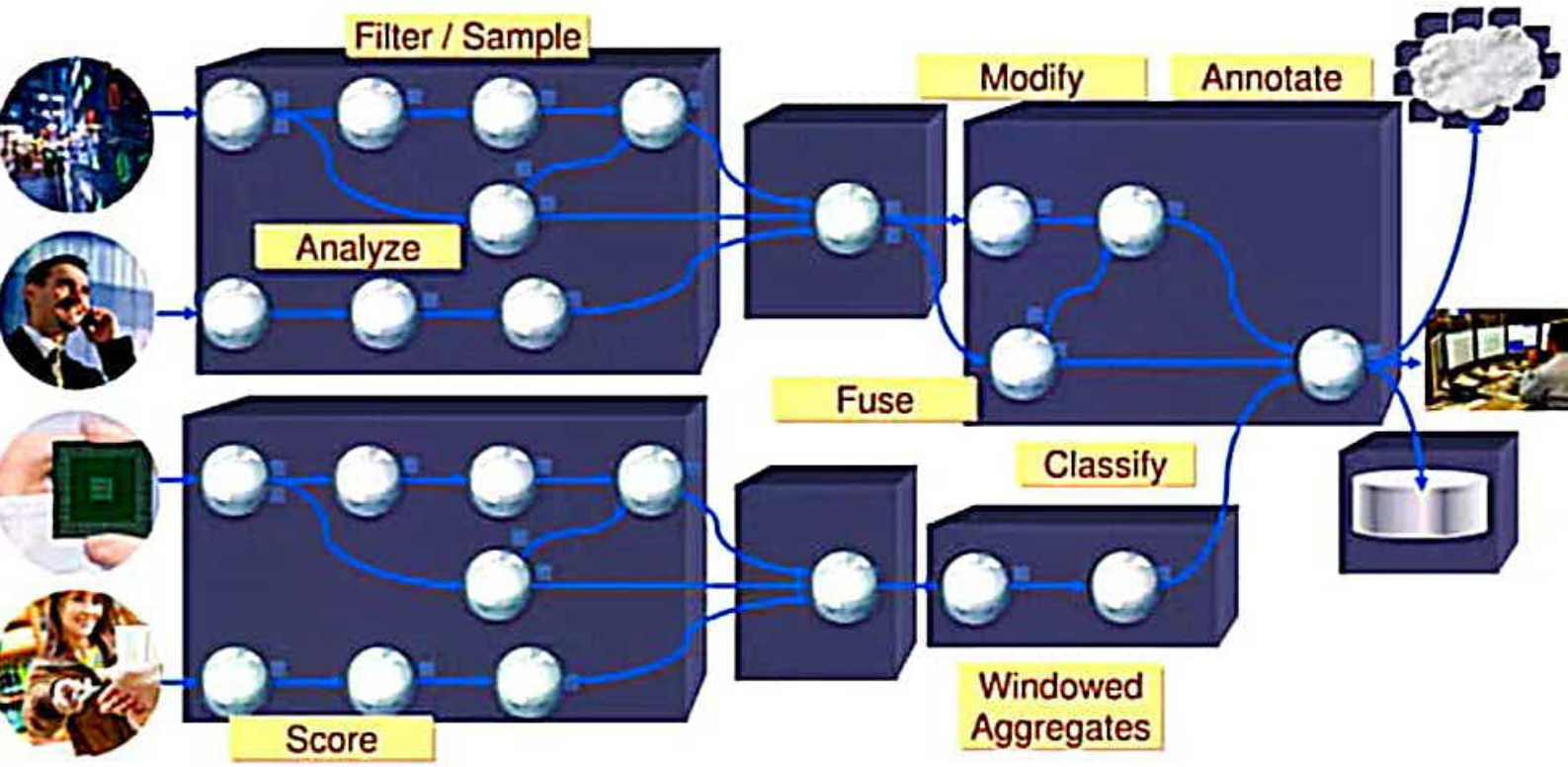
Current fact finding

Analyze data in motion – before it is stored

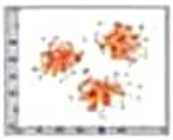
Low latency paradigm, push model

Data driven – bring data to the analytics

Big Data in real-time with InfoSphere Streams



Analytic Accelerators Designed for Velocity (and Variety)

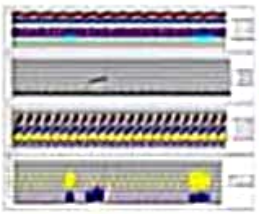


Mining in Microseconds
(included with Streams)

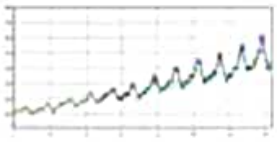
Acoustic
(IBM Research)
(Open Source)

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

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



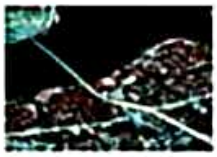
Advanced Mathematical Models
(IBM Research)



Predictive
(IBM Research)

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

Statistics
(included with Streams)

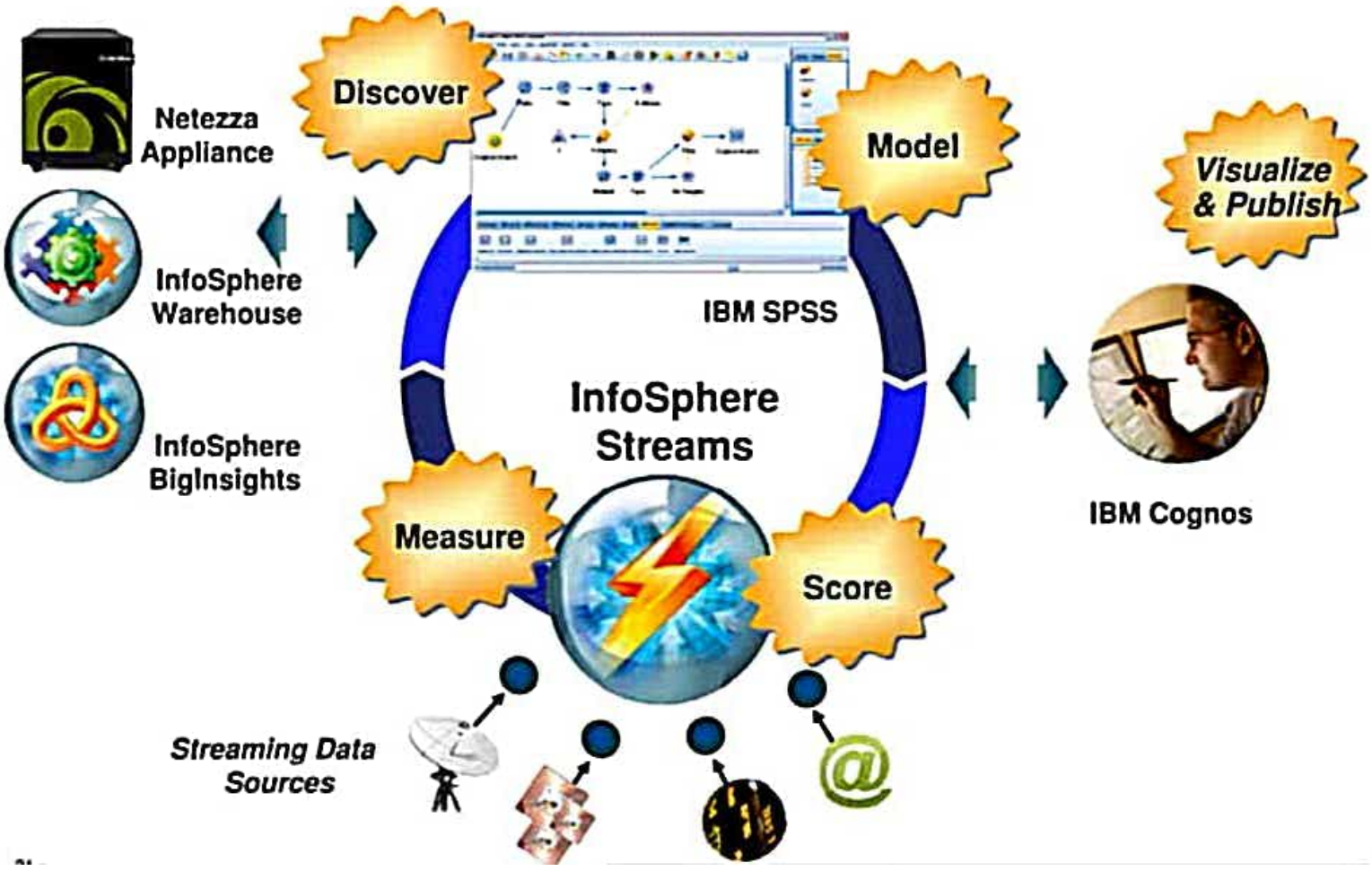


Geospatial
(IBM Research)

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

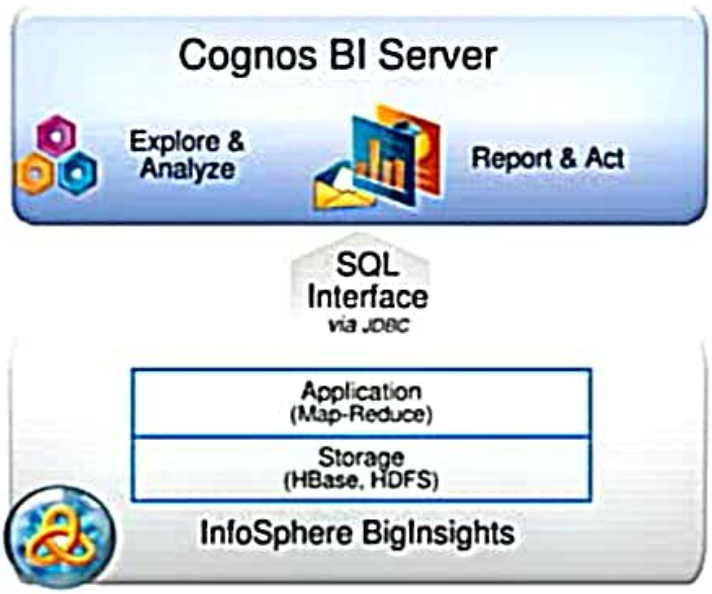
Increase productivity & leverage past work increasing speed to market

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.

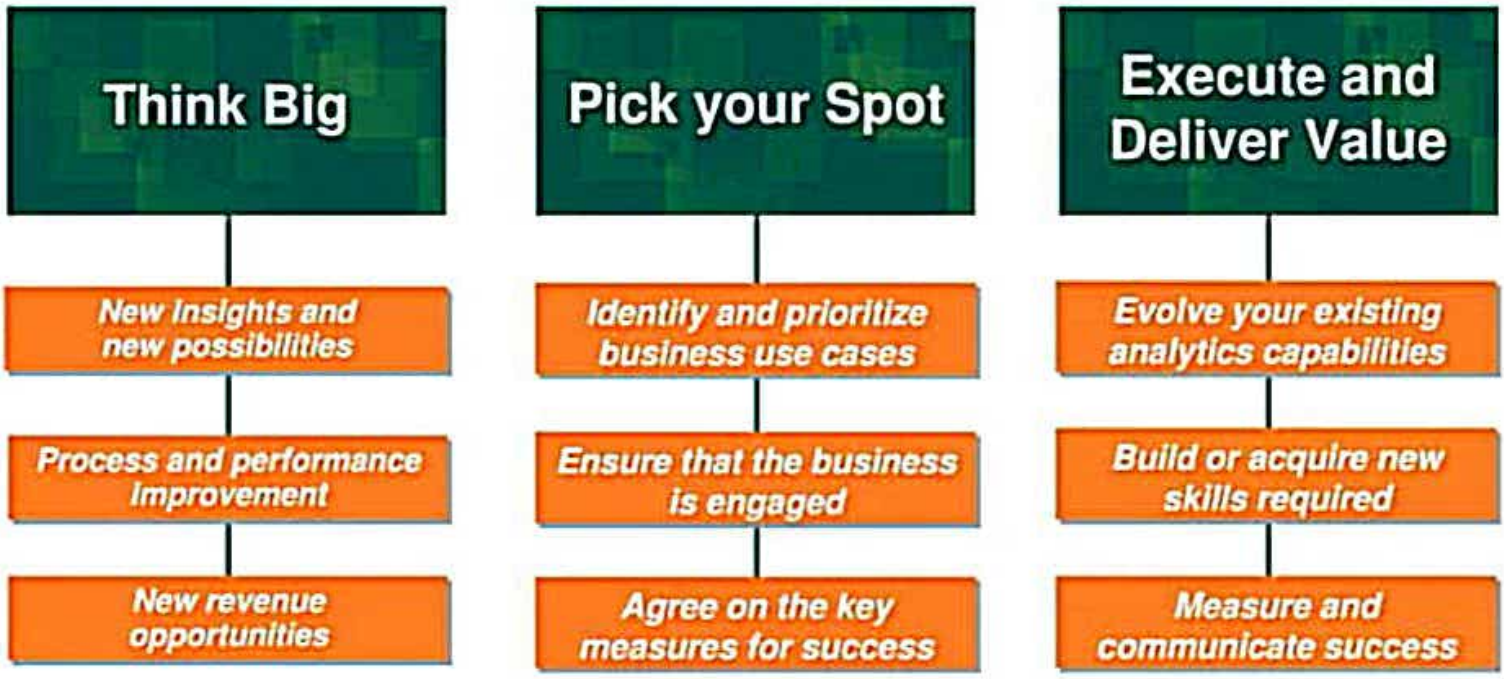
Clients Achieve Breakthrough Outcomes With IBM's Big Data Platform

	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

Incorporating Advanced Machine Learning Algorithms in Big Data Analysis

Unlock the power of big data with advanced machine learning algorithms for predictive analysis and anomaly detection.

The Rise of Big Data

Explosion of Data

The proliferation of digital technology has generated an unprecedented volume of data.

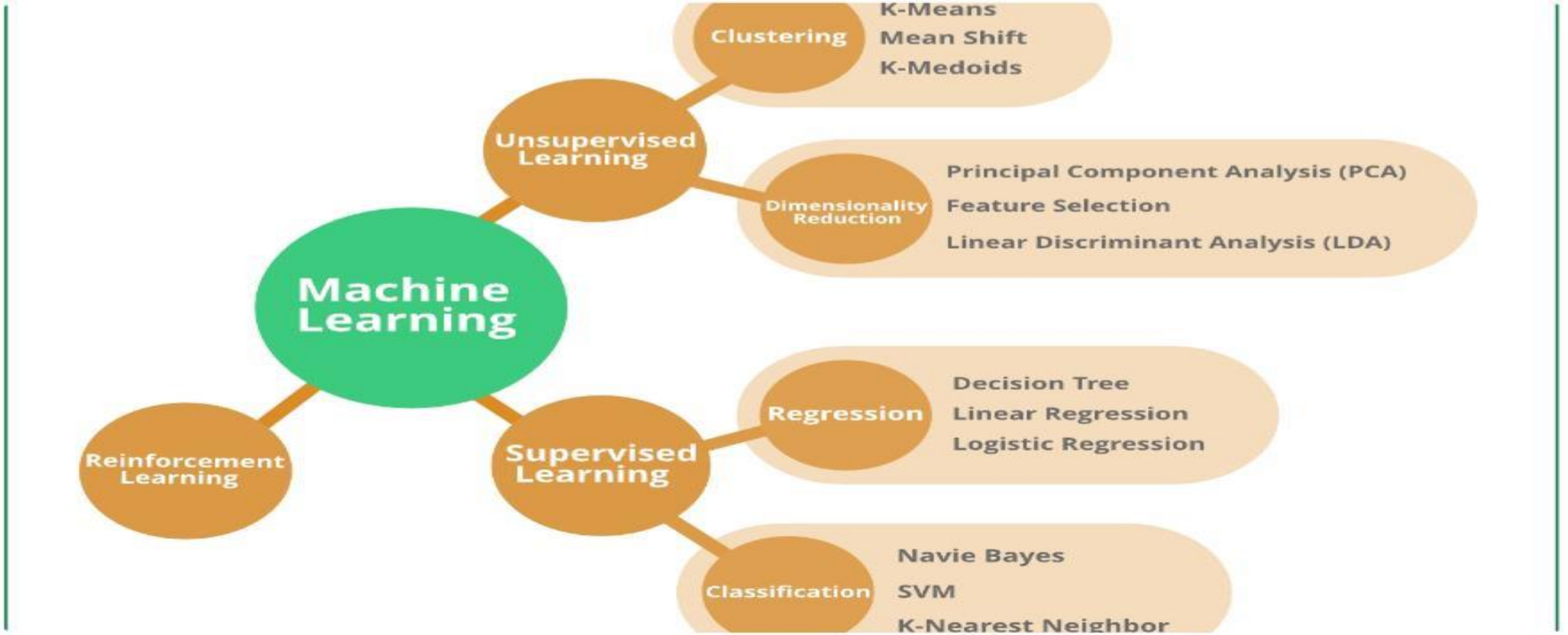
Challenges and Opportunities

Extracting meaningful insights from big data presents both challenges and exciting opportunities.

Transforming Industries

Big data analytics are revolutionizing industries such as healthcare, finance, and e-commerce.

The Power of Machine Learning

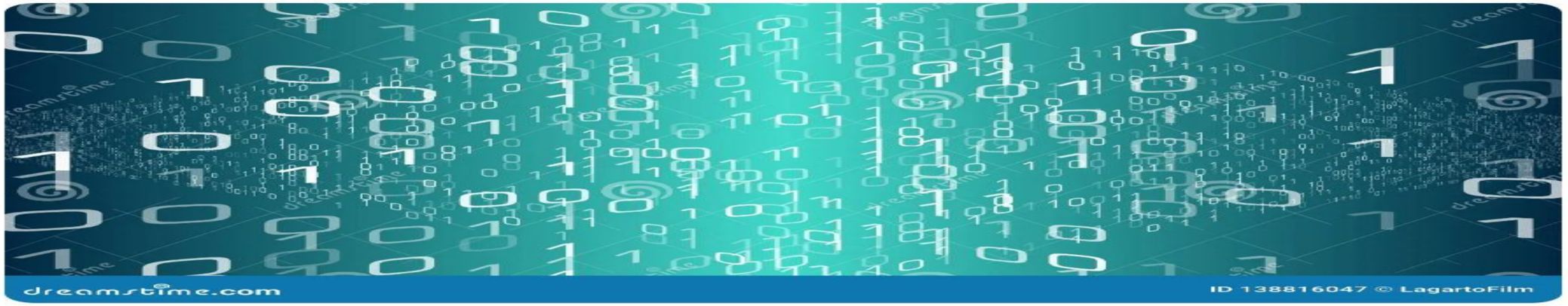


Artificial Intelligence

Machine learning algorithms enable computers to learn from data and make intelligent decisions.

Data Scientists' Toolbox

Advanced machine learning algorithms are essential tools for data scientists.



potential of big data.

Predictive Analysis: Peering into the Future

1 **Anticipating Outcomes**

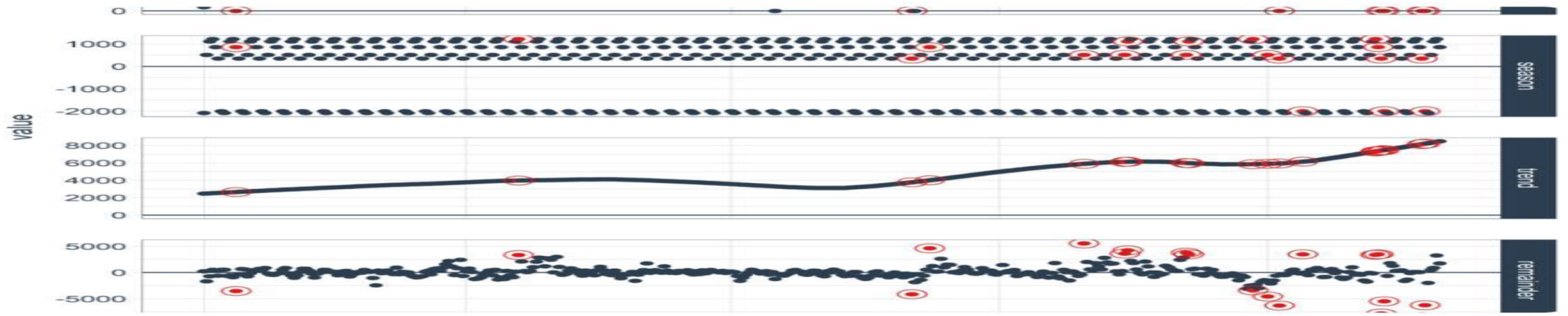
Advanced algorithms analyze historical and real-time data to make accurate predictions.

2 **Optimizing Decision-Making**

Businesses leverage predictive analysis to gain a competitive edge and enhance decision-making.

3 **From Weather to Stocks**

Anomaly Detection: Uncovering the Unusual



Safeguarding Systems

Advanced algorithms detect anomalies in big data, helping to identify potential threats and risks.

Cybersecurity Applications

Anomaly detection plays a crucial role in cybersecurity, thwarting attacks and protecting sensitive data.

Unmasking Fraud

Financial institutions utilize anomaly detection algorithms to detect fraudulent activities.

1

Healthcare

Machine learning algorithms aid in diagnosing diseases and personalizing treatment plans.

2

Retail

Big data analysis enables personalized product recommendations and targeted marketing campaigns.

3

Transportation

Algorithms optimize logistics and improve traffic management, reducing congestion and emissions.

Implementing Advanced Algorithms

Data Quality

Ensure the accuracy and completeness of data before applying advanced algorithms.

Computational Resources

Powerful hardware and distributed computing infrastructure are essential for processing big data.

Ethical Considerations

Address the ethical implications of using algorithms, such as privacy and bias concerns.

Start Building the Big Data Analysis Solution Using IBM Cloud Databases

Welcome to the world of big data analysis leveraging IBM cloud databases, where we explore the power of data-driven insights and innovation.

Benefits of Using IBM Cloud Databases for Big Data Analysis

Scalability

With IBM cloud databases, you can easily scale your data analysis infrastructure to handle large volumes of information.

Flexibility

IBM cloud databases provide a flexible environment, allowing you to adapt and tailor the solution to meet your specific needs.

Cost Efficiency

By leveraging the cloud, you can significantly reduce infrastructure costs associated with traditional on-premises solutions.

Overview of IBM Cloud Databases

1

Database as a Service (DBaaS)

IBM Cloud Databases offers managed database services, enabling you to focus on data analysis instead of database management.

2

Wide Range of Database Options

Choose from various databases, including relational, NoSQL, and time-series, to suit the specific requirements of your big data analysis.

3

Seamless Integration

Integrate IBM Cloud Databases with other IBM cloud services and third-party tools, enhancing your data analysis capabilities.

Key Features and Capabilities of IBM Cloud Databases for Big Data Analysis

High Availability

Ensure uninterrupted data access with built-in high availability features, reducing potential downtime.

Data Security

Protect your data with robust security measures, including encryption, access controls, and monitoring.

Advanced Analytics

Leverage advanced analytics capabilities, such as machine learning and artificial intelligence, to extract

Steps to Start Building the Big Data Analysis Solution Using IBM Cloud Databases

1

Evaluate Data Requirements

Understand the nature of your data and define the specific requirements for your big data analysis solution.

2

Select Database Option

Choose the most suitable IBM Cloud Database based on the characteristics and needs of your data analysis project.

3

Create Database Instances

Provision and configure the necessary database instances in the IBM Cloud environment to support your big data analysis workload.

Best Practices for Implementing and Maintaining the Big Data Analysis Solution

Data Governance

Establish clear data governance policies and practices to ensure data quality, consistency, and compliance.

Ongoing Monitoring

Regularly monitor the performance and health of your big data analysis solution to identify and address any potential issues.

Continuous Learning

Stay updated with the latest advancements in big data analysis techniques and technologies to drive continuous improvement.

Continue Building the Big Data Analysis Solution

Unlock the full potential of your data by harnessing advanced analysis techniques and visualizing the results.

Exploratory Data Analysis

1

Discover Hidden Patterns

Uncover valuable insights and patterns within your data through exploratory analysis methods.

2

Data Preprocessing

Prepare your data for analysis by cleaning, transforming, and integrating multiple data sources.

3

Data Visualization

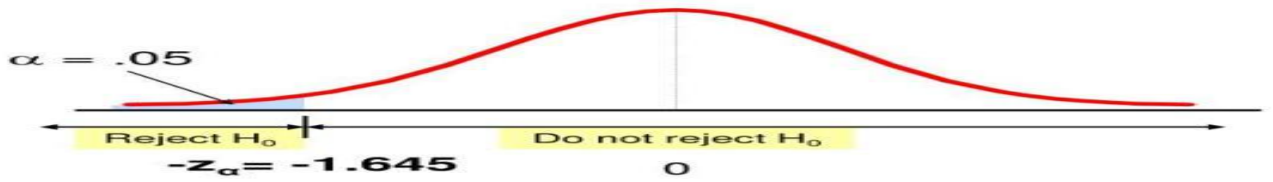
Visualize your data to gain a better understanding of its underlying structure and relationships.

Statistical Methods

Hypothesis Testing Example

(continued)

- 4. Determine the rejection region

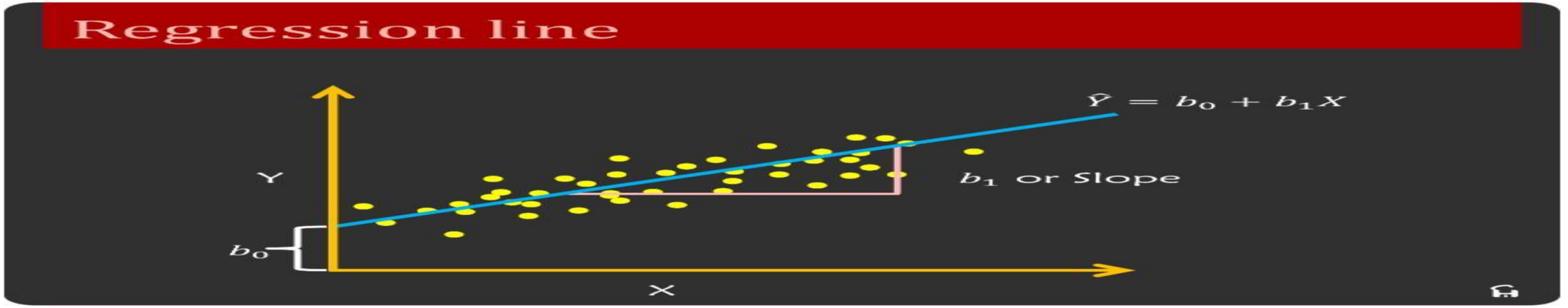


This is a one-tailed test with $\alpha = .05$.
Since **σ is known**, the cutoff value is a z value:

Reject H_0 if $z < z_\alpha = -1.645$; otherwise do not reject H_0

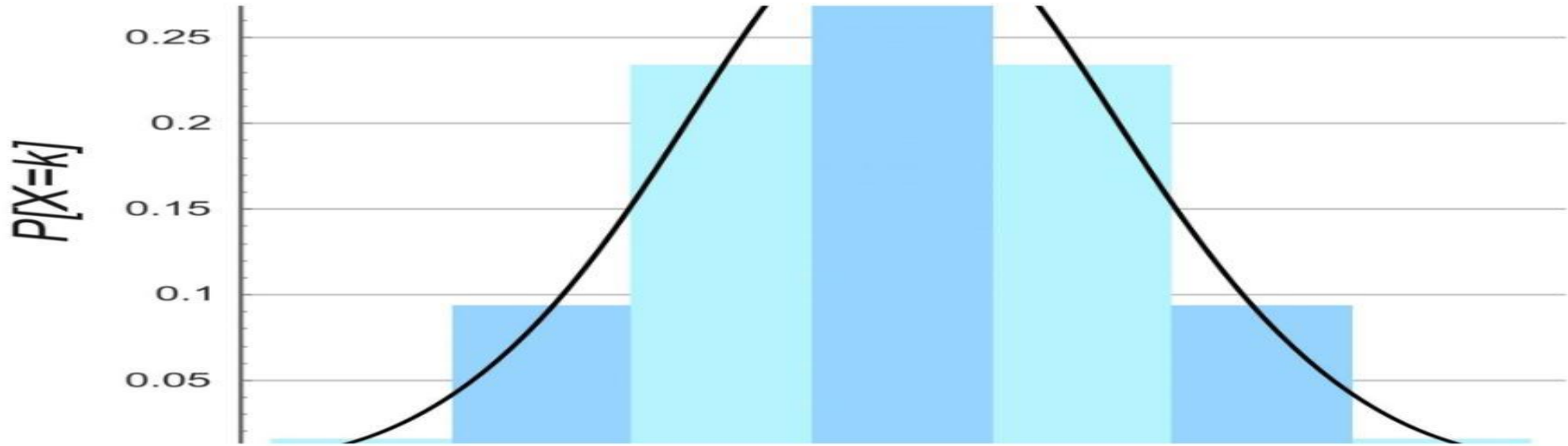
Hypothesis Testing

Make data-driven decisions by conducting hypothesis testing and drawing statistically significant conclusions.



Regression Analysis

Uncover relationships between variables and predict future outcomes using regression analysis techniques.



Data Distribution

Understand the distribution of your data and its implications for analysis and decision-making.

Machine Learning Algorithms

Supervised Learning

Train models to make accurate predictions or classifications using labeled training data.

Unsupervised Learning

Extract meaningful patterns and relationships from unlabeled data through clustering and dimensionality reduction.

Reinforcement Learning

Develop intelligent systems that learn and improve from interacting with their environment.

Data Visualization

- Effective visualizations are key to understanding complex data.
- Explore various types of visualizations, such as charts, graphs, and maps.
- Discover powerful tools like Tableau and Power BI for creating impactful visual representations.
- Learn best practices for designing clear, informative, and visually appealing data visualizations.


Thank
you!