BIG DATA ANALYTICS

Introduction to Big Data

Introduction to Big Data: *Outline*

- Introduction
- Big Data Enabling Technologies
- Hadoop Stack for Big Data

Introduction to Big Data

What is Big Data? *Definitions*

According to Bernard Marr

- Bernard Marr defines Big Data as the digital trace that we are generating in this digital era.
- This digital trace is made up of all the data that is captured when we use digital technology.

The basic idea behind the phrase Big Data is that everything we do is increasingly leaving a digital trace (or data), which we can use and analyze to become smarter. The driving forces in this brave new world are access to ever increasing volumes of data and our ever increasing technological capability to mine that data for commercial insights.

Introduction to Big Data

What is Big Data? *Definitions*

The research from Gartner

Big Data is high-volume, high-velocity and/or high-variety information assets that demand cost effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation.

Ernst and Young

Big Data refers to the dynamic, large and disparate volumes of data being created by people, tools and machines. It requires new, innovative, and scalable technology to collect, host and analytically process the vast amount of data gathered in order to derive real-time business insights that relate to consumers, risk, profit, performance, productivity management and enhanced shareholder value.

Introduction to Big Data

What is Big Data? *Definitions*

Lisa Arthur, a Forbes contributor

Big Data is a collection of data from traditional and digital sources inside and outside a company that represent a source of ongoing discovery and analysis.

What is Big Data?

Big Data is data whose scale, distribution, diversity and/or timeliness require the use of new technical architectures and analytics to enable insights that unlock new sources of business value.

- Requires new data architectures
- New tools
- New analytical methods
- Integrating multiple skills into new role of data scientist.

Introduction to Big Data

What is Big Data?

- Big Data refers to extremely large and complex datasets that are challenging to process using traditional database management tools or standard data processing methods.
- Key challenges include:
  - Capture: Collecting data
  - Curation: Organizing and maintaining data
  - Storage: Saving large volumes of data
  - Search: Retrieving specific information
  - Sharing: Distributing data efficiently
  - Transfer: Moving data between systems
  - Analysis: Extracting insights from data
  - Visualization: Presenting data in an understandable way

Introduction to Big Data

What is Big Data?

- The rise of Big Data is driven by the value gained from analyzing massive, interconnected datasets.
- These insights help:
  - Identify business trends
  - Improve research quality
  - Prevent diseases
  - Link legal references
  - Fight crime
  - Monitor real-time traffic conditions

Introduction to Big Data

Types of Big Data

Structured Data

- Structured data is highly organized and stored in fixed formats, such as tables or relational databases.
- It is easy to search, process, and analyze using traditional database tools.
- Businesses commonly use structured data for financial records, inventory management, and customer databases.
- Examples: Sales records, spreadsheets, customer databases.

Introduction to Big Data

Types of Big Data

Semi-Structured Data

- Semi-structured data combines elements of both structured and unstructured data.
- While it doesn't follow a strict format like structured data, it does contain tags or markers that organize the information to some extent.
- This type of data is increasingly common in web-based and cloud applications.
- Examples: JSON (JavaScript Object Notation) files, XML documents, CSV files.

Introduction to Big Data

Types of Big Data

Unstructured Data

- Unstructured data lacks any predefined format or structure, making it challenging to analyze with traditional tools.
- It accounts for the majority of data generated today, including multimedia content and social media posts.
- Advanced analytics and machine learning tools are often required to process unstructured data.
- Examples: Videos, social media posts, audio files, emails.

Introduction to Big Data

Types of Big Data

Quasi-Structured Data

- Quasi-structured data refers to data that does not follow a consistent format but still contains some identifiable patterns or fields.
- It is often generated by systems like web servers and clickstream logs, where the data has structure but is irregular.
- Examples: Web server logs, clickstream data, network logs.

Introduction to Big Data

Facts and Figures: (As of December 2024)

- Walmart processes 2.5 petabytes of unstructured data from 1 million customers every hour.
- Facebook: As of 2024, Facebook's data warehouse stores over 300 petabytes, with an incoming daily rate of about 600 terabytes.
- Modern Aircraft generate between 5 to 8 terabytes of data per flight, depending on the duration and complexity of the systems involved.
- More than 5 billion people worldwide are using mobile phones for activities like calling, texting, tweeting, and browsing.
- Decoding the human genome has seen significant advancements, with sequencing now achievable in a matter of hours, compared to 10 years previously.

Introduction to Big Data

Facts and Figures: (As of December 2024) cont'd.

- The largest AT&T database handles vast amounts of data, including customer call records, with databases containing up to 312 terabytes of data and 1.9 trillion rows. (AT&T stands for American Telephone and Telegraph. It is one of the largest telecommunications companies in the world, providing services in telephony, broadband, and television.)
- Global Data Generation: The world generated approximately 120 zettabytes of data in 2023, expected to reach 181 zettabytes by the end of 2025, with approximately 2.5 quintillion bytes of data created daily.
- Internet of Things (IoT): The number of IoT devices worldwide is projected to reach 17.08 billion in 2024, fueling exponential data growth.

Introduction to Big Data

Facts and Figures: (As of December 2024) cont'd.

- Big Data Market Growth: The global big data and analytics market has grown to $348.21 billion in 2024, with continued expansion expected.
- Business Investment in Big Data and AI: Approximately 97.2% of businesses are actively investing in big data and artificial intelligence to enhance decision-making and operational efficiency.
- Data Science Employment: The demand for data science professionals continues to rise, with an expected 28% increase in data science jobs by 2026.

Introduction to Big Data

An Insight into Data Measurement in Big Data Analytics:

In the digital world, data is measured in terms of bytes and its larger units (kilobytes, megabytes, etc.). To give a relatable context, these units can be compared to physical objects like grains or containers of rice, helping us grasp their scale. Here's how it translates:

- Byte (B):
  - Analogy: One grain of rice.
  - Explanation: A byte is the basic unit of digital information. It typically represents a single character, such as a letter or a digit. In Big Data, it's the foundation of data storage.

Introduction to Big Data

An Insight into Data Measurement in Big Data Analytics:

- Kilobyte (KB - 10^3 or 1,000 Bytes):
  - Analogy: One cup of rice.
  - Explanation: A kilobyte is a small unit of data storage, often used to measure the size of simple text files. For example, a single-page Word document might be a few kilobytes.
- Megabyte (MB - 10^6 or 1,000,000 Bytes):
  - Analogy: 8 bags of rice.
  - Practical Use: Represents data used on personal computers or desktops. For example, a high-quality photo or a minute of audio is typically a few megabytes.

Introduction to Big Data

An Insight into Data Measurement in Big Data Analytics:

- Gigabyte (GB - 10^9 or 1,000,000,000 Bytes):
  - Analogy: 3 semi-trucks of rice.
  - Practical Use: Often used to measure storage on phones, laptops, or data consumption. For instance, a two-hour HD movie might require 4-6 GB of data.
- Terabyte (TB - 10^{12} or 1,000,000,000,000 Bytes):
  - Analogy: 2 container ships of rice.
  - Practical Use: Terabytes are used in enterprise environments or Internet-scale data. A single company's database of customer transactions, such as Walmart, may span several terabytes.

Introduction to Big Data

An Insight into Data Measurement in Big Data Analytics:

- Petabyte (PB - 10^{15} or 1,000,000,000,000,000 Bytes):
  - Analogy: Enough rice to blanket half of Jaipur.
  - Practical Use: Associated with Big Data, petabytes represent massive-scale data. For example, Facebook processes over 600 terabytes of new data daily, and its total storage exceeds 300 petabytes.
- Exabyte (EB - 10^{18} or 1,000,000,000,000,000,000 Bytes):
  - Analogy: Enough rice to blanket the West Coast of the United States or one-quarter of India.
  - Practical Use: Exabytes are used to measure global-scale data, such as total mobile data traffic in a year. By 2024, the global Internet data usage is expected to exceed several exabytes daily.

Introduction to Big Data

An Insight into Data Measurement in Big Data Analytics:

- Zettabyte (ZB - 10^{21} or 1,000,000,000,000,000,000,000 Bytes):
  - Analogy: Enough rice to fill the Pacific Ocean.
  - Practical Use: Represents the total digital data in the world. As of 2023, the total amount of data created globally was approximately 120 zettabytes and is projected to reach 181 zettabytes by 2025. It is considered the Future of data measurement.
- Yottabyte (YB - 10^{24} or 1,000,000,000,000,000,000,000,000 Bytes):
  - Analogy: An earth-sized bowl of rice.
  - Practical Use: Though no current systems can manage a yottabyte, it is a theoretical concept for future data scales. Governments and research organizations may someday handle yottabytes of data.

Introduction to Big Data

An Insight into Data Measurement in Big Data Analytics:

- Brontobyte (10^{27} or 1,000 Yottabytes):
  - Analogy: Astronomical size-impossible to visualize.
  - Practical Use: Still a theoretical unit. Brontobytes are used in discussions about the distant future of Big Data Analytics, where astronomical quantities of information might be processed, such as simulations of the universe.

This analogy using rice helps bridge the gap between abstract digital concepts and tangible physical comparisons, making it easier to understand the massive scale of data handled in Big Data Analytics. As technology evolves, these units will play a critical role in describing the vast amounts of information we generate, analyze, and store daily.

Introduction to Big Data

What's making so much data?

- Sources: Social media interactions, online shopping, personal content creation.
- Machines: IoT devices, smart home appliances, sensors, and industrial machines.
- Organizations: Business operations, customer transactions, and marketing campaigns.
- Ubiquitous computing: Computing power embedded in everyday devices has become pervasive.
- More people carrying data-generating devices:
  - With over 7 billion mobile phone users worldwide (2024 estimate), devices like smartphones with apps such as Facebook, Instagram, GPS, and Cameras have become key contributors to data generation.

Introduction to Big Data

What's making so much data?

- Data on the Internet (2024 Live Stats):
  - 5.4 billion Internet users globally.
  - 2.4 billion websites registered online.
  - 350 billion emails sent daily.
  - 9.2 billion Google searches conducted every day.
  - 10 million blog posts written daily.
  - 1.2 billion tweets sent every day.
  - 9.7 billion videos viewed daily on platforms like YouTube.
  - 150 million photos uploaded daily.
  - 200 million Facebook likes daily.

Introduction to Big Data

Source of Data Generation

Source	Examples	Data Generated Per Day	Examples with Data Insights	Remarks
Social Media Platforms	Facebook, Instagram, Twitter, LinkedIn	~4.7 billion posts, likes, and shares	Facebook processes 500 terabytes of data daily, including photos, videos, and interactions.	Social media users generate high-velocity, unstructured data.
IoT Devices	Smartwatches, Home Automation, Sensors	~80 zettabytes annually (2024 projection)	A single Boeing 787 generates 40 terabytes of data per hour during flight.	IoT generates continuous real-time streaming data.
Email Communications	Gmail, Outlook, Yahoo Mail	~361 billion emails sent daily	Gmail alone accounts for over 300 billion emails sent daily worldwide, often including text and attachments.	Rich in textual data used for analytics in cybersecurity, spam filtering.
Video Streaming	YouTube, Netflix, Twitch	~1 million hours of video streamed daily	YouTube users upload 500 hours of video every minute, contributing heavily to multimedia data.	Contributes to > 60% of internet traffic, driving multimedia analytics.
E-commerce Platforms	Amazon, Flipkart, eBay	~2.5 PB (petabytes) from transactions globally	Amazon records 1 million purchase transactions per second, generating approximately 1 petabyte/day.	Tracks user behavior, purchases, and browsing patterns.
Financial Transactions	PayPal, Credit Cards, Online Banking	~2.9 billion transactions globally	Visa processes over 65,000 transactions per second at peak times.	Critical for fraud detection and risk management.
Telecommunications	Smartphones, 4G/5G Networks	~95 exabytes of mobile data consumed globally	Over 2.5 quintillion bytes of data are created daily from mobile devices alone.	Includes call records, internet usage, and messaging services.
Energy Sector	Smart Grids, Power Plants	~100 PB daily	The U.S. energy grid generates 40 terabytes of data daily from sensors and monitoring systems.	Data is used for optimization and predictive maintenance of energy systems.
Healthcare	MRI Scans, Wearable Devices	~1 TB per patient annually in developed countries	Each connected hospital bed can produce 50 MB per hour through IoT sensors.	IoT in healthcare and electronic medical records generate massive datasets.

Introduction to Big Data

An Example of Big Data at Work - Ex1: Big Data at work through crowdsourcing for real-time traffic management

Fig: Here's a visual representation of Big Data at work through crowdsourcing for real-time traffic management, featuring live traffic overlays, dynamic routing, and a green corridor for an ambulance. This showcases the practical use of Big Data in urban mobility solutions.
Image Src: AI-powered tools

Introduction to Big Data

An Example of Big Data at Work – Ex2: Big Data at work across various industries like healthcare, finance, e-commerce, transportation, and energy

Fig: Here's a visually engaging image showcasing Big Data at work across various industries like healthcare, finance, e-commerce, transportation, and energy.
Image Src: AI-powered tools

Introduction to Big Data

Challenges of Traditional RDBMS and Big Data Solutions with Real-World Examples

Problem Area	Description	Examples	Big Data Solutions	Key Examples in Action
Volume of Data	Traditional RDBMS cannot handle massive datasets effectively.	Processing petabytes of user activity logs on social media.	Distributed storage (e.g., HDFS, NoSQL databases like Cassandra, MongoDB).	Facebook processes 500 TB of data daily, including photos, videos, and interactions.
Velocity of Data	Inability to process high-speed, real-time data streams.	Analyzing live tweets to detect trends or public sentiment in real time.	Real-time data frameworks (e.g., Apache Kafka, Apache Flink).	Twitter uses Big Data tools like Kafka to process 500 million tweets per day for sentiment analysis.
Unstructured Data	RDBMS is designed for structured data, not for unstructured or semi-structured.	Handling images, videos, and IoT sensor logs.	Tools like Hadoop, Spark, and Elasticsearch for multi-format data.	YouTube processes 500 hours of video uploads per minute, leveraging Big Data for recommendations.
Slow Query Performance	Delayed retrieval and analysis make insights irrelevant by the time they're ready.	Identifying trending topics on social media takes hours instead of minutes.	In-memory computing (e.g., Apache Spark, Redis).	Netflix uses in-memory systems to generate personalized recommendations in real time.
Scalability Challenges	Traditional databases struggle to scale horizontally for distributed data.	Expanding server capacity to handle growing e-commerce transaction data.	Horizontal scaling with cloud platforms (e.g., AWS-Amazon Web Services, GCP-Google Cloud Platform, Azure).	Amazon handles 1 million transactions per second using distributed systems to manage peak loads.
Need for Real-Time Computation	Traditional RDBMS tools cannot support instantaneous insights.	Dynamic traffic routing for an ambulance using real-time road conditions.	Real-time processing systems (e.g., Spark Streaming, Flink, Storm).	Google Maps dynamically computes routes based on live traffic data crowdsourced from users.

Introduction to Big Data

Challenges of Working with Big Data

- Big data presents challenges in capturing, storing, searching, sharing, analyzing, and visualizing information.

Capturing Big Data

- Volume & Variety: Collecting large, diverse data from various sources.
- Data Quality: Ensuring data is accurate and clean.
- Real-Time Streams: Handling continuous data intake.

Storing Big Data

- Scalability: Storing growing data effectively.
- Cost Management: Balancing performance with cost efficiency.
- Data Management: Organizing data for easy access.

Introduction to Big Data

Challenges of Working with Big Data cont'd.

Searching Big Data

- Efficient Querying: Fast and accurate searches across large datasets.
- Indexing: Creating and maintaining indexes for quick retrieval.
- Semantic Search: Searching unstructured data (text, images) by meaning.

Sharing Big Data

- Privacy & Security: Protecting sensitive data and complying with regulations.
- Interoperability: Ensuring smooth data sharing between systems.
- Access Control: Managing who can access specific data.

Introduction to Big Data

Challenges of Working with Big Data cont'd.

Analyzing Big Data

- Computational Power: Needing massive resources for processing.
- Algorithm Complexity: Applying efficient analysis algorithms on large data.
- Real-Time Analytics: Handling data and insights instantly.

Visualizing Big Data

- Data Representation: Presenting complex data simply.
- Interactivity & Scalability: Creating visualizations that can handle large datasets and user interaction.
- Contextualization: Providing clear insights from data.

Introduction to Big Data

Challenges of Working with Big Data cont'd.

Overall Challenges

- Data Governance: Managing data quality, privacy, and compliance.
- Legacy System Integration: Connecting big data solutions with existing systems.
- Talent & Expertise: Finding skilled professionals for big data systems.

Introduction to Big Data

Characteristics or V's of Big Data

- Big Data is defined by several key characteristics, often referred to as the "V's." These characteristics provide a framework to understand the complexity, scale, and applications of big data.
- There is no one definition of Big Data, but there are certain elements that are common across the different definitions, such as velocity, volume, variety, veracity and value .
- These are the V's of Big Data or characteristics of Big Data.

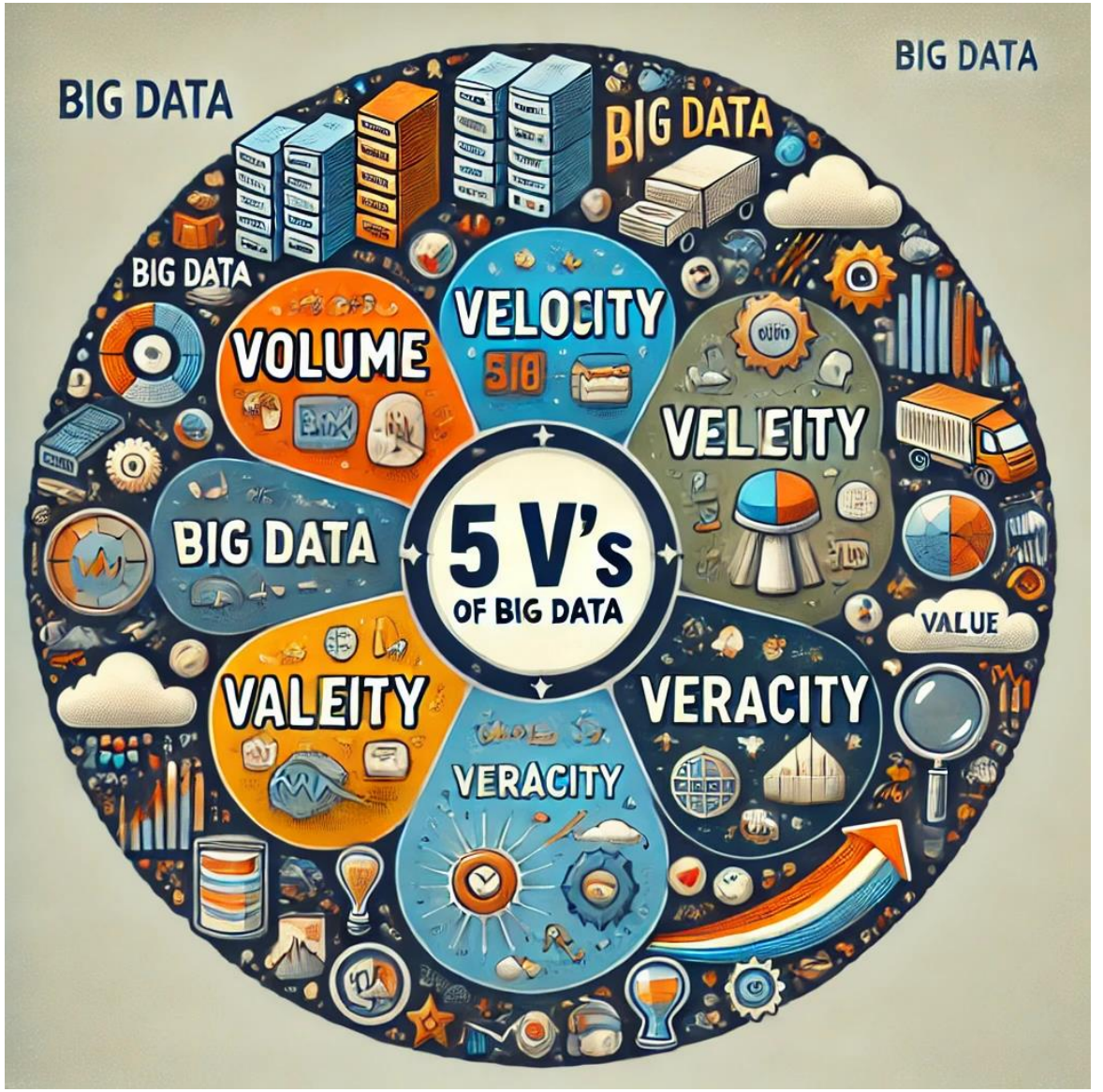


Image Src: AI-powered tools

Introduction to Big Data

Characteristics or V's of Big Data

Characteristic	Definition	Real-Time Example	Usage
Volume (Scale)	Refers to the sheer quantity of data generated and stored. It can be measured in terabytes, petabytes, or even exabytes.	Social media platforms like Facebook generate massive amounts of user-generated content, including posts, videos, and images daily.	Used for storing and processing massive datasets, e.g., in data lakes or cloud storage.
Velocity (Speed)	The speed at which data is generated, processed, and analyzed. This includes real-time data streaming.	Stock market data and high-frequency trading systems rely on real-time data updates for decision-making.	Enables real-time analytics for applications like fraud detection or live monitoring.
Variety (Complexity)	Represents the diversity of data types, including structured, unstructured, and semi-structured data.	Emails, videos, photos, sensor data, and transaction logs all represent different types of data handled by companies.	Allows integration of different data sources for holistic insights.
Veracity	The quality and trustworthiness of data. It deals with uncertainties, biases, and inconsistencies in data.	Health data collected from wearable devices can sometimes be inaccurate or incomplete, affecting analysis.	Improves decision-making by ensuring accurate and reliable datasets.
Value	The usefulness and insights derived from data. Not all collected data adds value unless analyzed effectively.	E-commerce platforms analyze purchase patterns to provide personalized recommendations to users.	Extracts actionable insights to enhance decision-making and business strategies.

Introduction to Big Data

Characteristics or V's of Big Data - Additional V's

Characteristic	Definition	Real-Time Example	Usage
Variability	Inconsistent and dynamic nature of data.	Social media trends changing rapidly.	Adapts to changes in data patterns over time for dynamic models.
Visualization	Presenting data insights in a comprehensible format.	Dashboards used for business analytics.	Simplifies interpretation of complex datasets for stakeholders.

Introduction to Big Data

Examples of the V's in Action

Characteristic	Example in Action
Volume	A single Boeing 737 generates 240 terabytes of flight data during its operation.
Velocity	Payment systems like Visa process thousands of transactions per second globally.
Variety	Smart city sensors collect diverse data such as traffic flow, air quality, and noise levels.
Veracity	Financial data from multiple sources is cross-verified to ensure accuracy in risk assessment.
Value	Predictive maintenance in manufacturing industries reduces downtime by analyzing machine sensor data.

Introduction to Big Data

Unlocking the Power of Big Data (Harnessing Big Data)

Traditional Approach

- Operational databases used to store customer details and other information.
- Led to the rise of relational databases for managing transactional data.

OLTP (Online Transaction Processing)

- Focus on transaction processing using relational databases.
- Used for managing day-to-day operations and storing transactional data.

OLAP (Online Analytical Processing)

- Focus on data warehousing.
- Consolidates data from multiple databases for complex analysis and decision-making.

Introduction to Big Data

Unlocking the Power of Big Data (Harnessing Big Data)

RTAP (Real-Time Analytical Processing)

- Stream data: Real-time data is continuously generated and processed.
- Stream computation: Applied to analyze data in motion and provide real-time insights.
- Improves business response time and decision-making.

Summary

- OLTP: Transaction processing with relational databases.
- OLAP: Data warehousing for analytical processing.
- RTAP: Real-time analytics for big data insights.

Introduction to Big Data

Big Data Analytics

- Big Data Analytics refers to the process of examining large and varied datasets-often referred to as big data-to uncover hidden patterns, correlations, trends, and insights.
- It involves using advanced analytics techniques like machine learning, predictive modeling, and statistical analysis to make data-driven decisions and predictions in real time.
- Big data is more real-time in nature than traditional Data Warehouse (DW) applications.
- Traditional Data Warehouse (DW) architectures (e.g. Exadata, Teradata) are not well-suited for big data applications.
- Shared nothing, massively parallel processing, scale-out architectures are well suited for big data applications.

Introduction to Big Data

Applications of Big Data Analytics

1. Healthcare

- Application: Predicting patient outcomes, optimizing hospital operations, and discovering new treatments.
- Example: Hospitals use big data to analyze medical records and identify patterns to predict patient outcomes, such as potential complications or diseases.

2. Retail and E-commerce

- Application: Personalizing customer experiences, improving inventory management, and enhancing supply chain operations.
- Example: Online retailers like Amazon use big data analytics to recommend products to customers based on past purchases and browsing behavior.

Introduction to Big Data

Applications of Big Data Analytics

3. Finance

- Application: Fraud detection, risk management, and customer segmentation.
- Example: Banks use big data to detect unusual transactions and flag potential fraud in real-time, analyzing patterns in transaction data.

4. Manufacturing

- Application: Predictive maintenance, process optimization, and quality control.
- Example: Companies like GE use big data to predict equipment failures and schedule maintenance before a breakdown occurs, reducing downtime.

Introduction to Big Data

Applications of Big Data Analytics

5. Transportation and Logistics

- Application: Route optimization, demand forecasting, and fleet management.
- Example: Ride-sharing companies like Uber use big data to optimize routes and pricing dynamically based on traffic conditions and demand.

6. Energy

- Application: Smart grids, energy consumption optimization, and renewable energy forecasting.
- Example: Power companies use big data to predict energy demand and optimize power distribution through smart grid technology.

Introduction to Big Data

Applications of Big Data Analytics

7. Telecommunications

- Application: Network optimization, customer churn prediction, and service personalization.
- Example: Telecom companies like Verizon use big data analytics to monitor network performance and predict service outages, improving customer satisfaction.

8. Education

- Application: Personalized learning, student performance analysis, and curriculum development.
- Example: Platforms like Coursera use big data to analyze student learning patterns and provide tailored recommendations for improving performance.

Introduction to Big Data

Applications of Big Data Analytics

9. Government

- Application: Smart cities, crime prediction, and disaster response.
- Example: Cities use big data to monitor traffic flow, optimize public transportation, and predict crime hotspots.

10. Education

- Application: Player performance analysis, injury prediction, and fan engagement.
- Example: Teams like the NBA use big data to analyze player performance, improve training regimens, and make strategic game decisions.

Introduction to Big Data

Real-Time Examples of Big Data Analytics

Amazon (Retail)

- Real-Time Application: Real-time product recommendations based on browsing behavior and past purchases.
- Impact: Increased sales and customer satisfaction through personalized experiences.

Netflix (Entertainment)

- Real-Time Application: Personalized content recommendations based on viewing history and preferences.
- Impact: Improved user engagement and retention.

Introduction to Big Data

Real-Time Examples of Big Data Analytics

Uber (Transportation)

- Real-Time Application: Dynamic pricing based on real-time demand, traffic, and location data.
- Impact: Optimized ride availability and pricing for customers and drivers.

Spotify (Music)

- Real-Time Application: Real-time playlist generation and song recommendations based on listening habits.
- Impact: Enhanced user experience and increased user retention.

Introduction to Big Data

Real-Time Examples of Big Data Analytics

Tesla (Automotive)

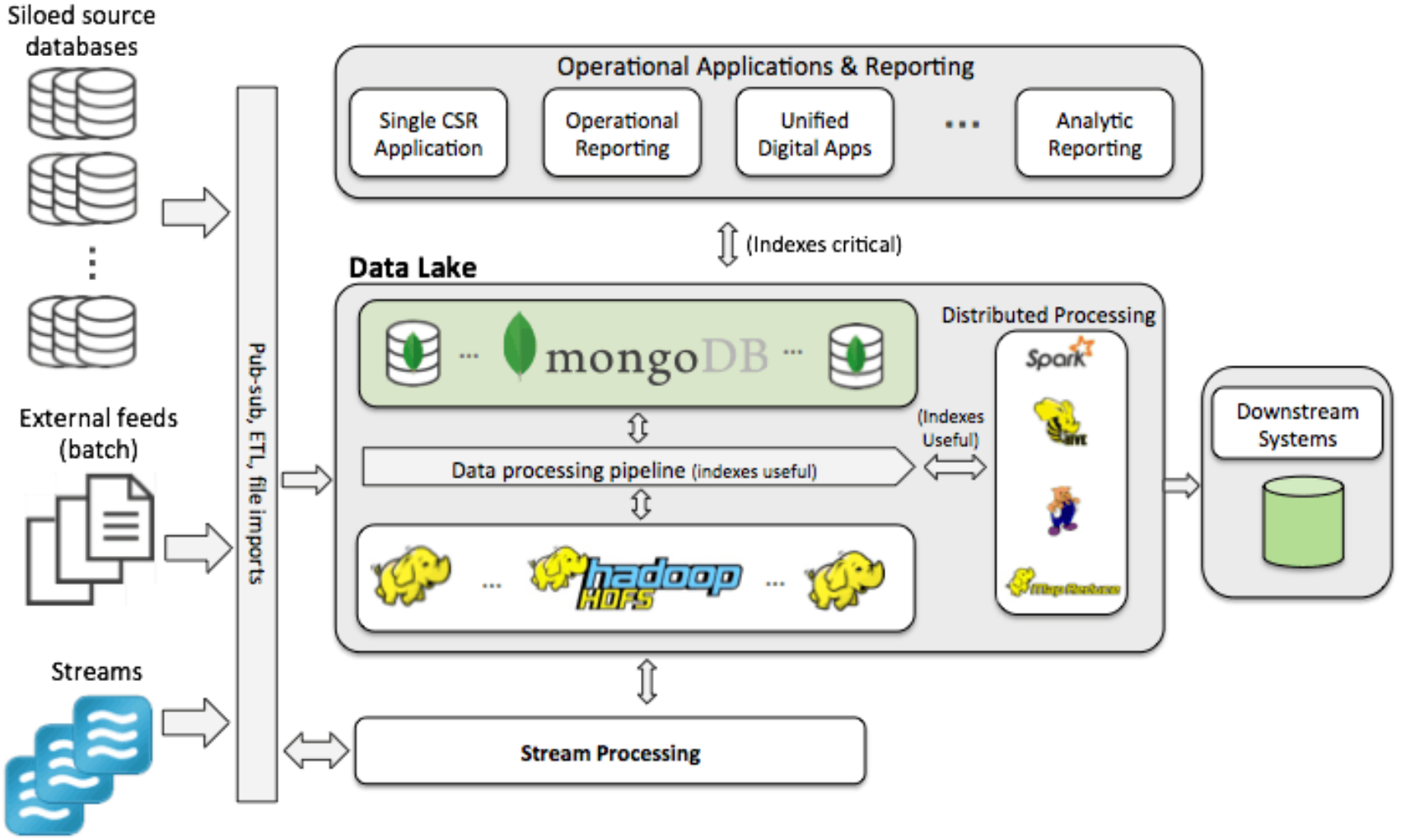
- Real-Time Application: Real-time monitoring of vehicle performance and driver behavior through sensors and IoT data.
- Impact: Predictive maintenance and continuous improvement of self-driving algorithms.

Social Media (Facebook, Twitter, etc.)

- Real-Time Application: Analyzing user data in real time for targeted advertising and content recommendations.
- Impact: Increased user engagement and ad revenue.

Introduction to Big Data

Components of Big Data Architecture



Stream icon from: https://en.wikipedia.org/wiki/File:Activity_Streams_icon.png

Introduction to Big Data

Key Components of Big Data Architecture

1. Data Sources:
   - Origins of data, such as databases, IoT devices, social media, and emails.
2. Data Storage:
   - Stores structured data in relational databases and unstructured data in NoSQL databases like MongoDB or HDFS.
3. Batch Processing:
   - Processes large datasets using tools like Hadoop.
4. Stream Processing:
   - Handles real-time data streams using tools like Apache Spark Streaming.
5. Analytical Data Store:
   - Stores processed data for querying and analysis, using tools like Hive or NoSQL stores.
6. Analytics and Reporting:
   - Extracts insights from data for business decisions.
7. Orchestration:
   - Automates workflows for data processing and movement.

Introduction to Big Data

Big Data Technology: Big Data – The Moving Parts