

Introduction to Big Data Analysis Introduction to Big Data

**Case Study Royal Manufacturing**



Company Background Royal Manufacturing is an engine manufacturing company renowned for its advanced solutions and products of aircraft engines Company utilizes engine performance data captured during flights for their analysis and performance improvement This data is analyzed after the flight with recorded readings of 100 parameters They generates 3 petabytes of data while producing 6000 turbine fan blades annually Emily Johnson CTO at Royal Manufacturing

**Case Study Royal Manufacturing**



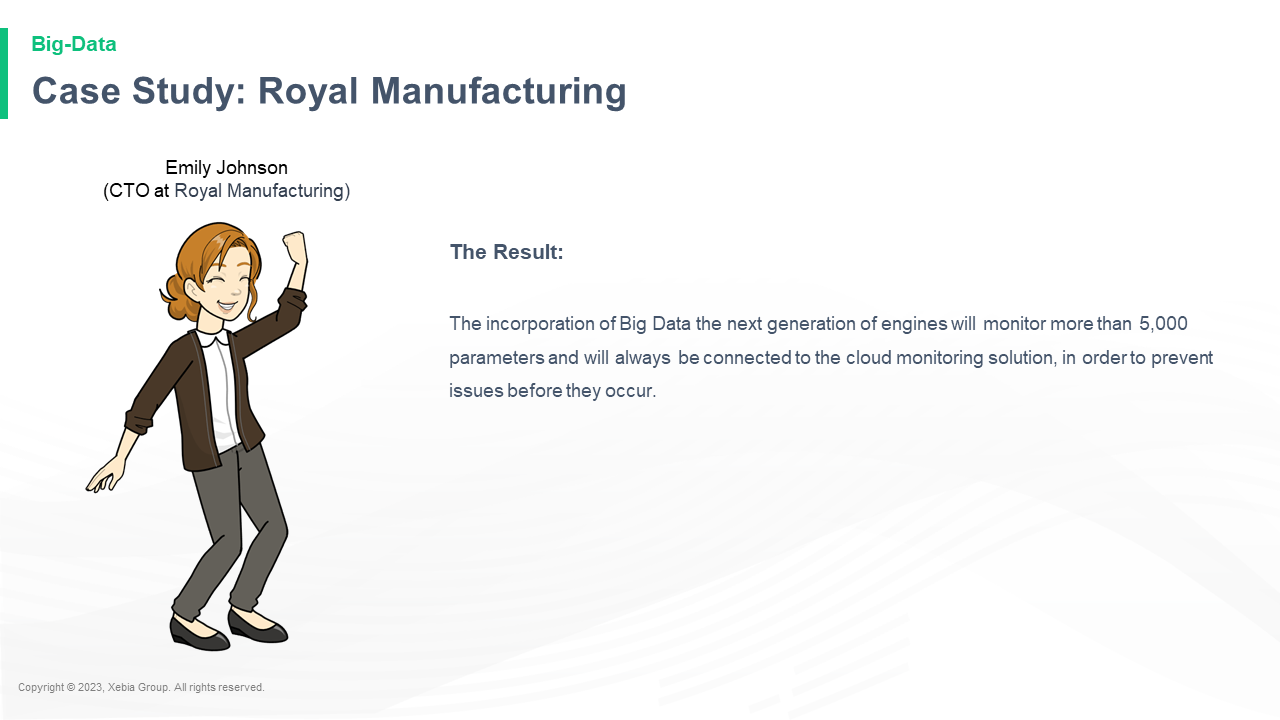
Current Situation Challenges Despite having large amount of data for monitoring Royal manufacturing s analytics teams is only able to analyze the data after the flight With technical advancement in industry there are so much possibilities of mapping large numbers of parameters They are also facing lag delay between data capturing and analysis process Emily Johnson CTO at Royal Manufacturing

**Case Study Royal Manufacturing**



Understanding the pressing need for a solution Emily Johnson CTO Of Royal Manufacturing decided to implement Big data Solutions to address these challenges The Solution Big data Analysis Emily led a series of educational workshops to familiarize analytical teams with the benefits and application of Big Data Analysis The integration of bigdata infrastructure and solution helped them to capture more than 5000 parameters and process it efficiently Their data capturing devices are always connected with cloud monitoring databases Which enables them to capture live data 24 7 Emily Johnson CTO at Royal Manufacturing

**Case Study Royal Manufacturing**



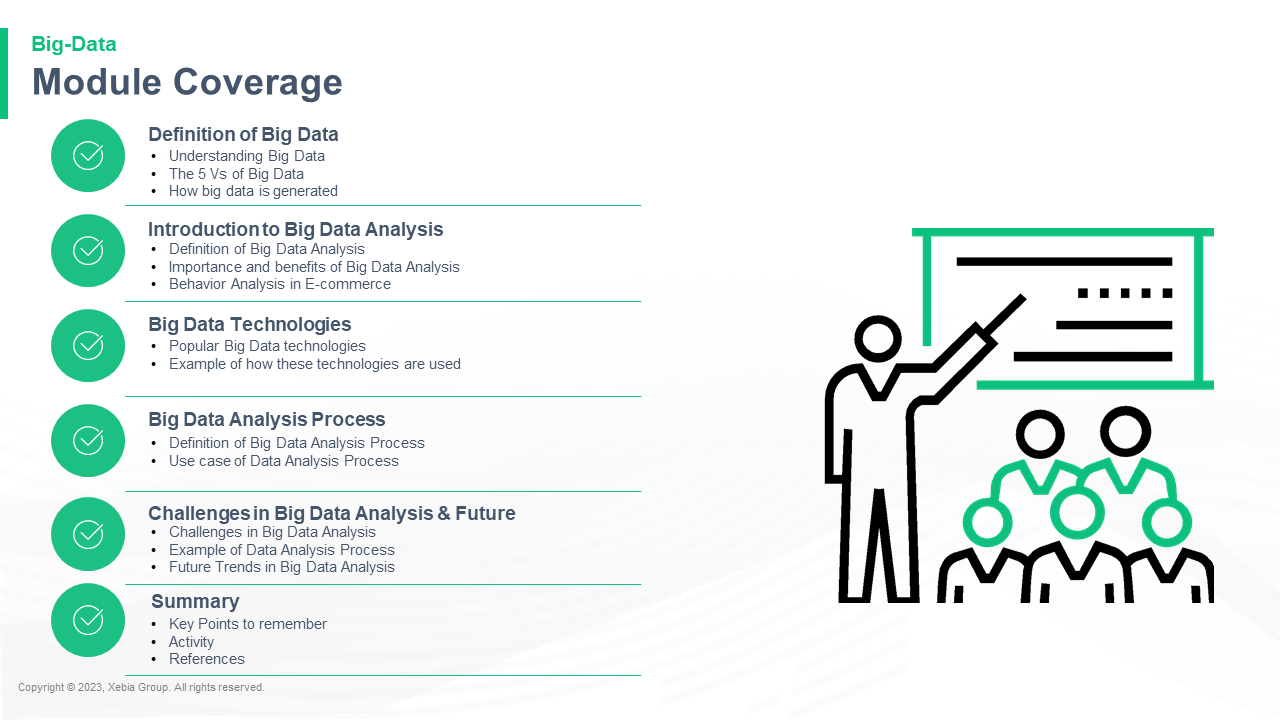
The Result The incorporation of Big Data the next generation of engines will monitor more than 5 000 parameters and will always be connected to the cloud monitoring solution in order to prevent issues before they occur Emily Johnson CTO at Royal Manufacturing

**This emphasizes the importance of Big Data Analytics for organizations across segments**

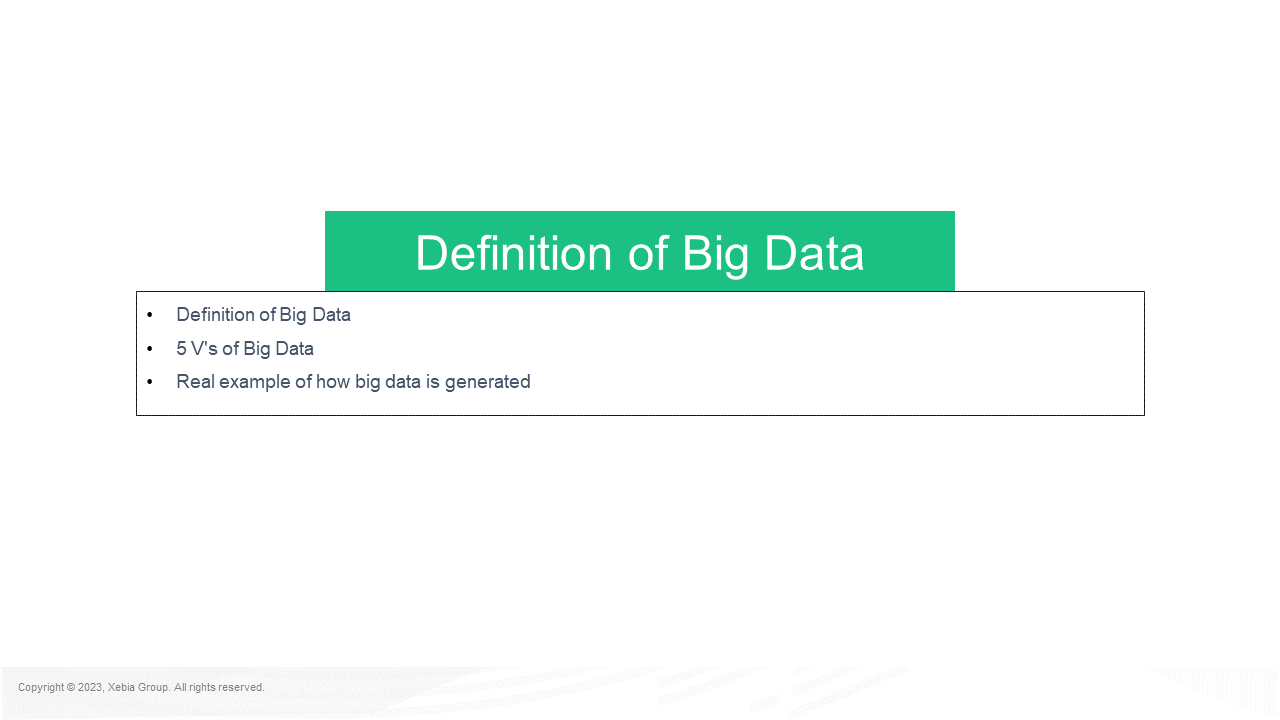


Let s study more about Big Data Analytics in this Module

**Module Coverage**

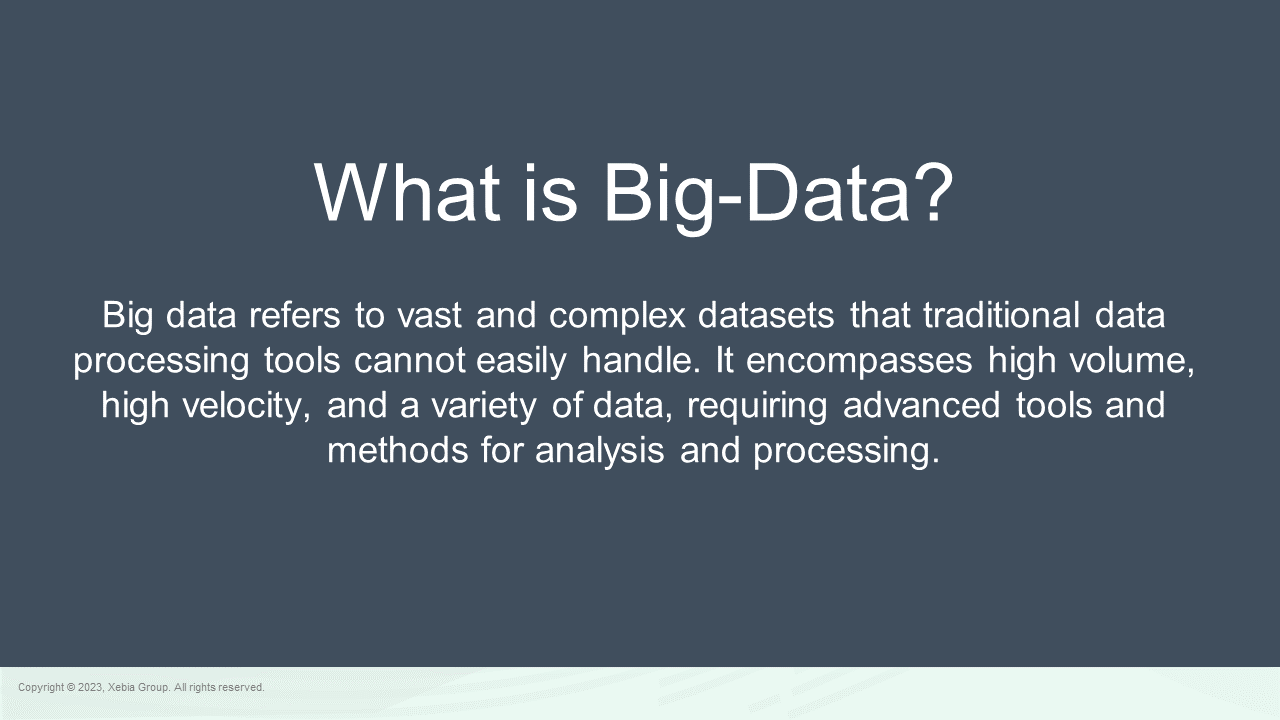


Understanding Big Data The 5 Vs of Big Data How big data is generated Definition of Big Data Definition of Big Data Analysis Importance and benefits of Big Data Analysis Behavior Analysis in E commerce Introduction to Big Data Analysis Popular Big Data technologies Example of how these technologies are used Big Data Technologies Definition of Big Data Analysis Process Use case of Data Analysis Process Big Data Analysis Process Challenges in Big Data Analysis Future Challenges in Big Data Analysis Example of Data Analysis Process Future Trends in Big Data Analysis Key Points to remember Activity References Summary



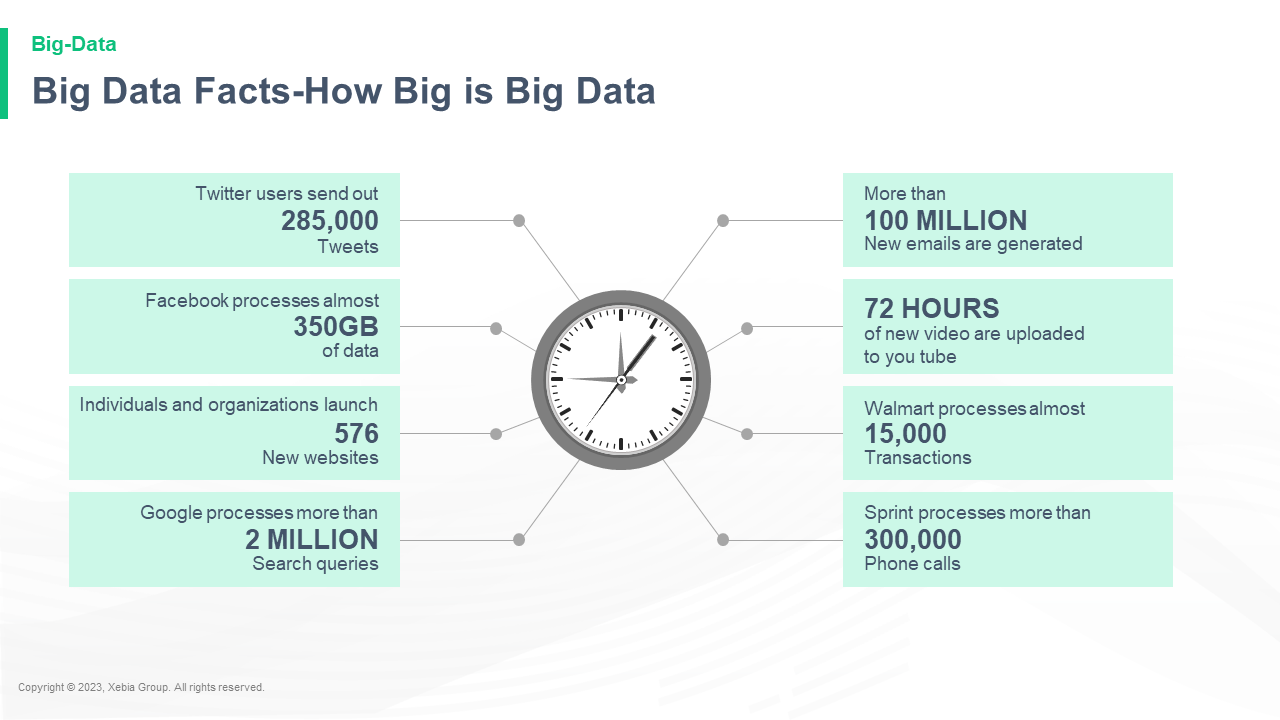
Definition of Big Data 5 V s of Big Data Real example of how big data is generated Definition of Big Data

**Big data refers to vast and complex datasets that traditional data processing tools cannot easily handle It encompasses high volume high velocity and a variety of data requiring advanced tools and methods for analysis and processing**

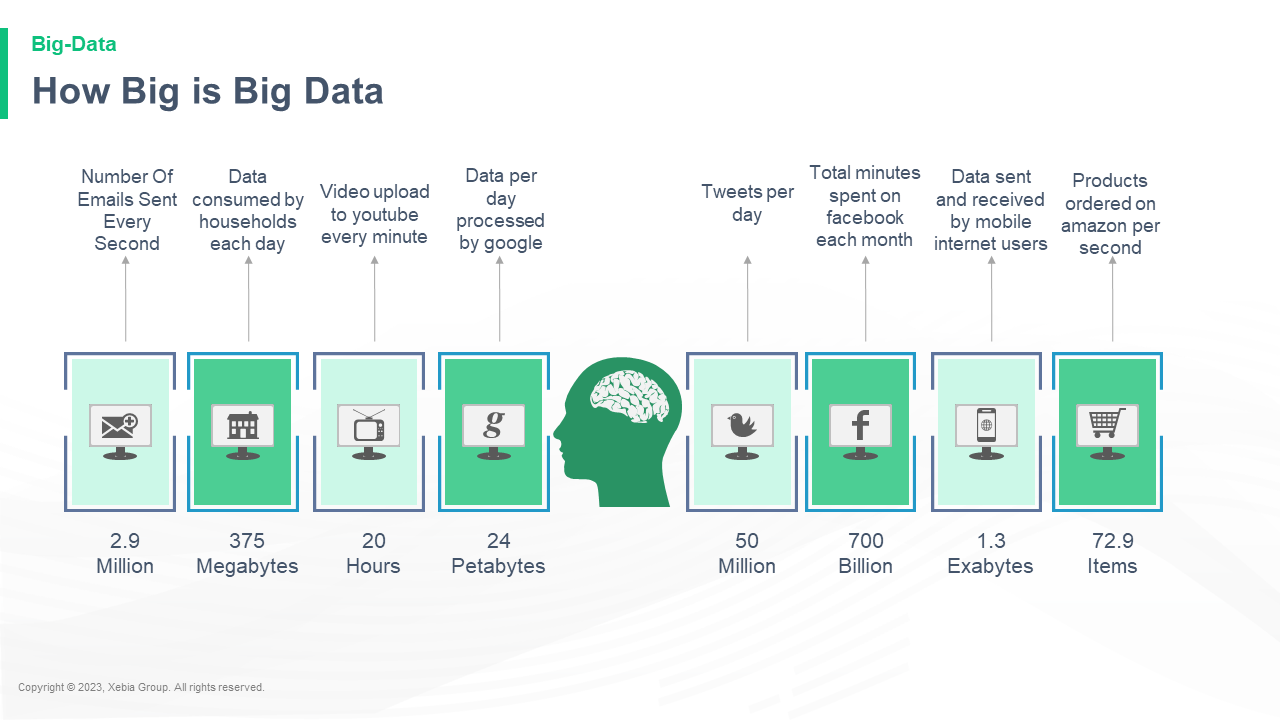


What is Big Data

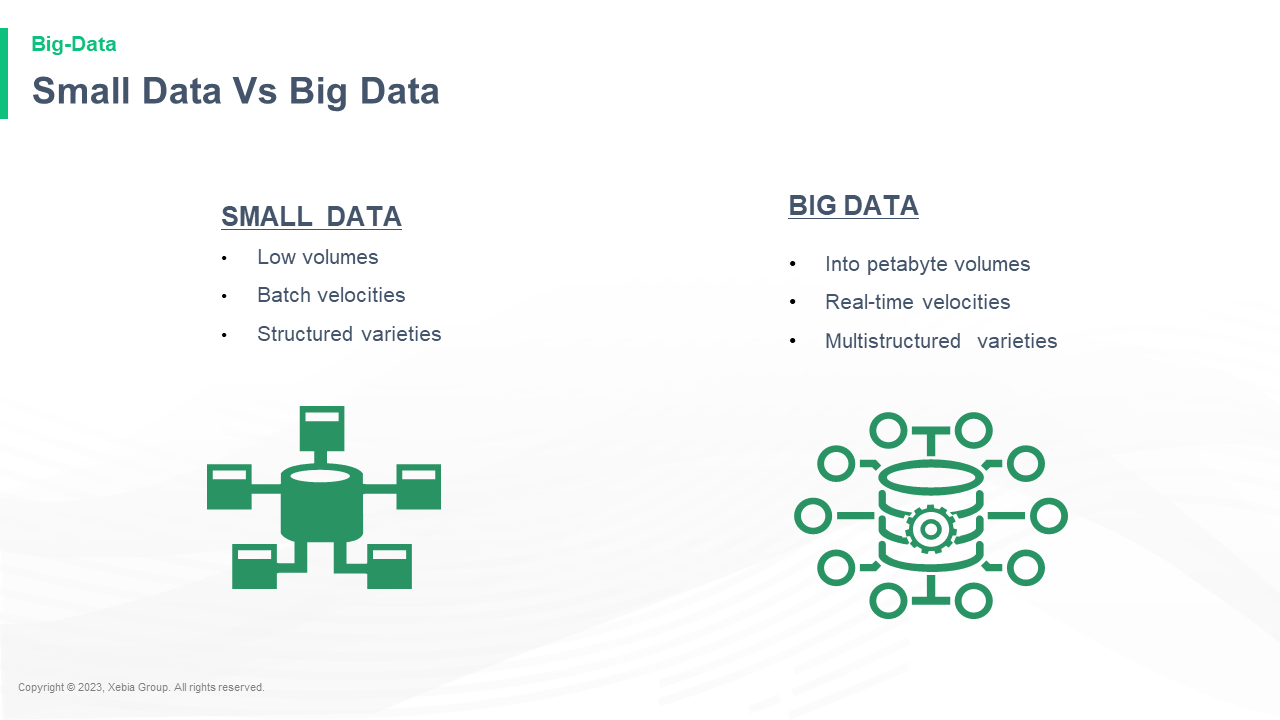
**Big Data Facts How Big is Big Data**



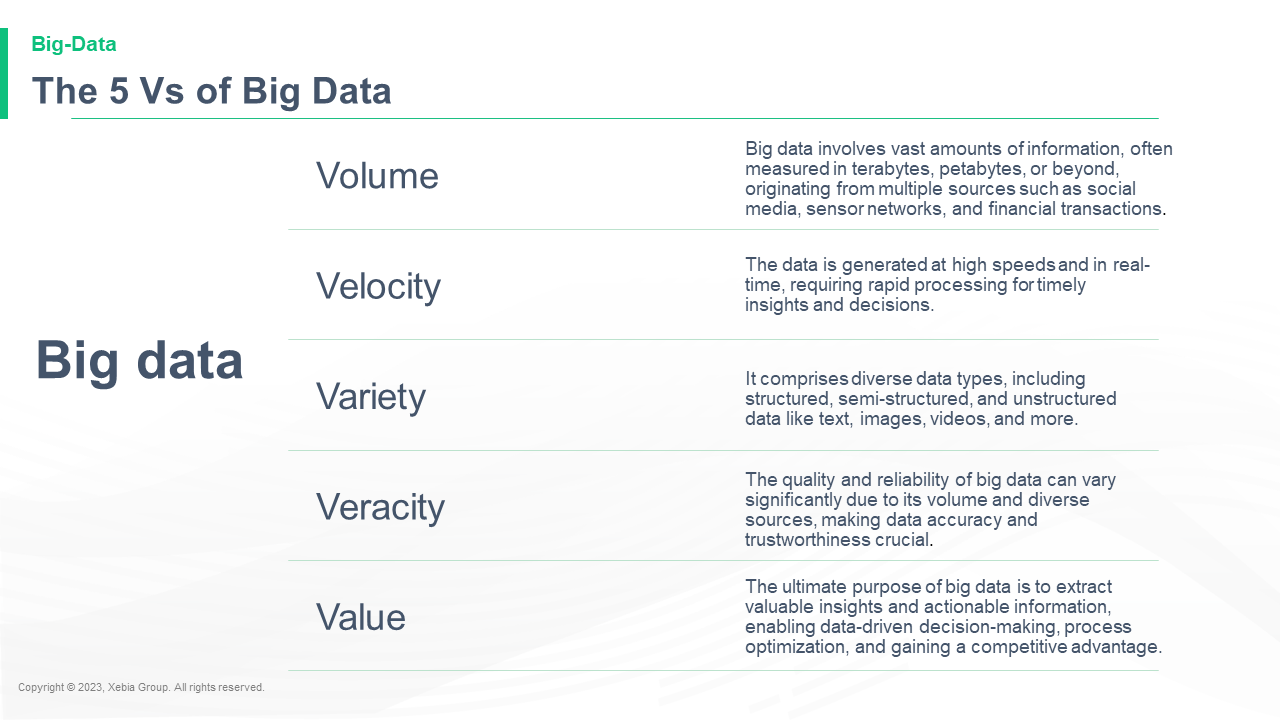
**How Big is Big Data**



**Small Data Vs Big Data**

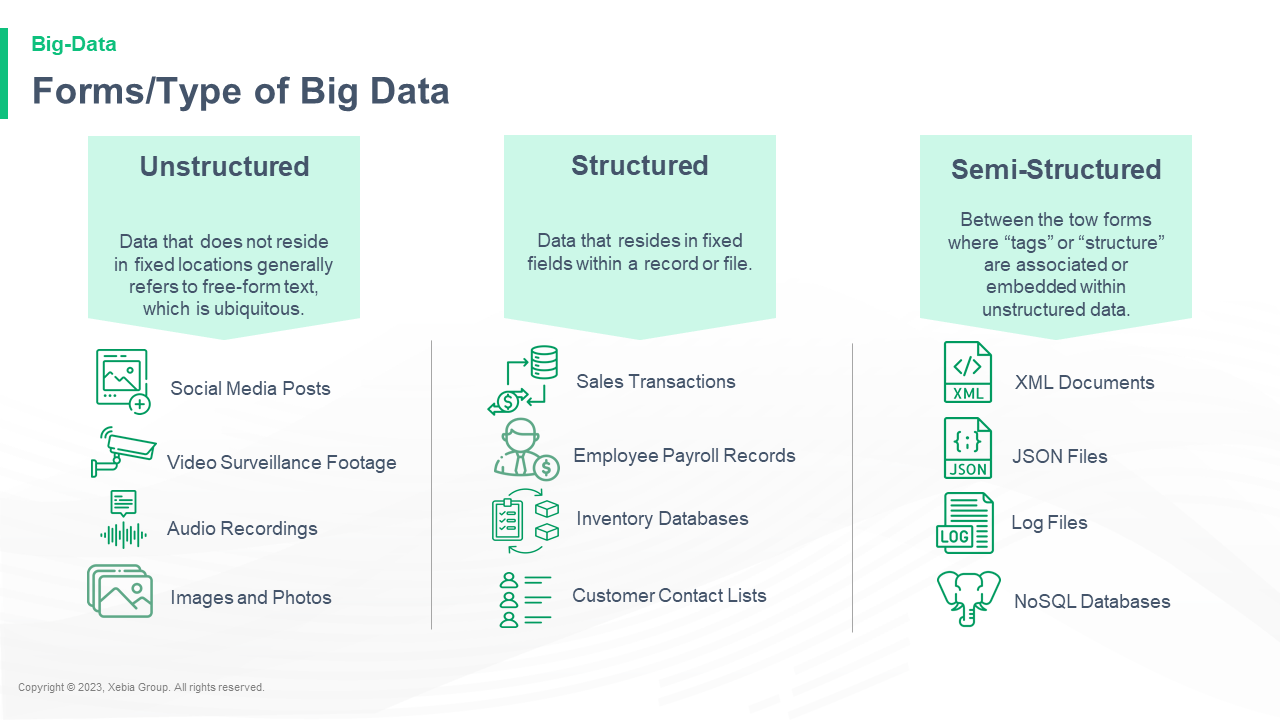


**The 5 Vs of Big Data**



Big data Volume Big data involves vast amounts of information often measured in terabytes petabytes or beyond originating from multiple sources such as social media sensor networks and financial transactions Velocity The data is generated at high speeds and in real time requiring rapid processing for timely insights and decisions Variety It comprises diverse data types including structured semi structured and unstructured data like text images videos and more Veracity The quality and reliability of big data can vary significantly due to its volume and diverse sources making data accuracy and trustworthiness crucial Value The ultimate purpose of big data is to extract valuable insights and actionable information enabling data driven decision making process optimization and gaining a competitive advantage

**Forms Type of Big Data**



**Example of Big Data**



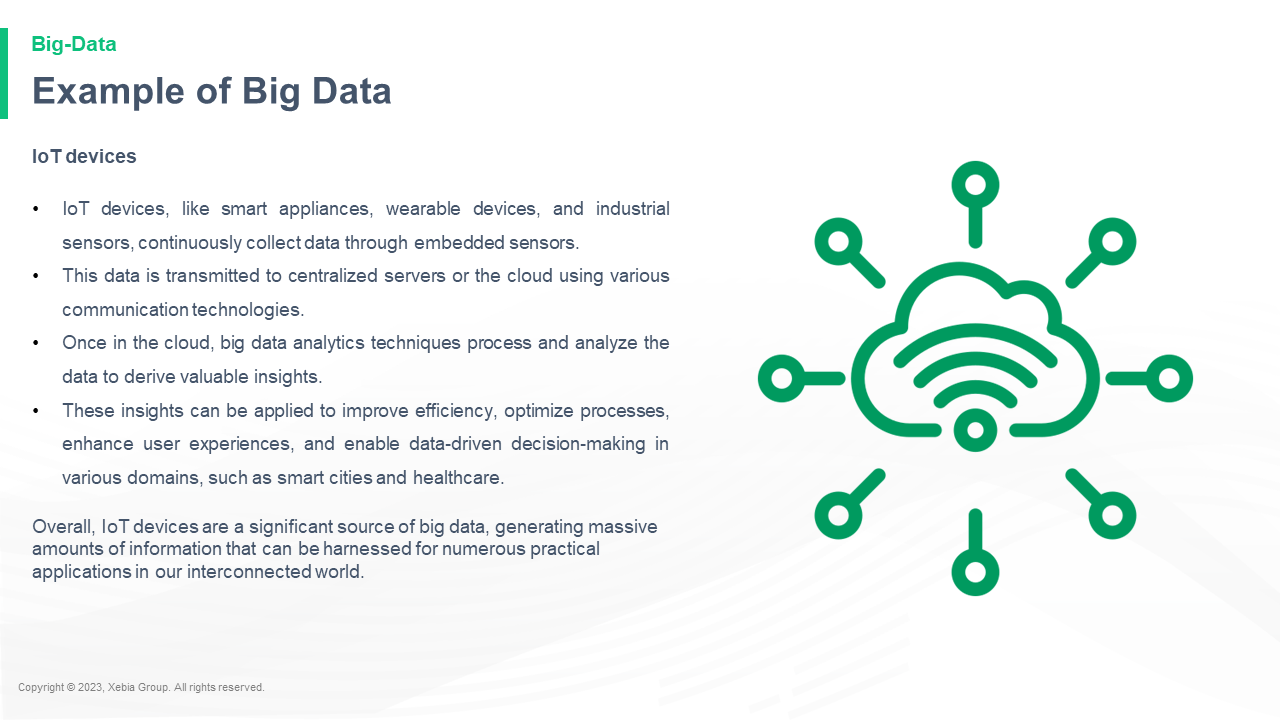
Social Media Interaction Social media platforms like Facebook Twitter Instagram and LinkedIn gather massive amounts of data through user interactions Each time a user posts a status shares a photo writes a comment likes a post follows a profile or sends a message data is generated and stored Content Sharing When users upload photos videos and other media files each piece of content adds to the growing pool of big data The frequency and volume of content sharing contribute significantly to the generation of big data on these platforms On that Ad Interactions Social Graph Sentiment Analysis and Location Data

**Data examples from social media**



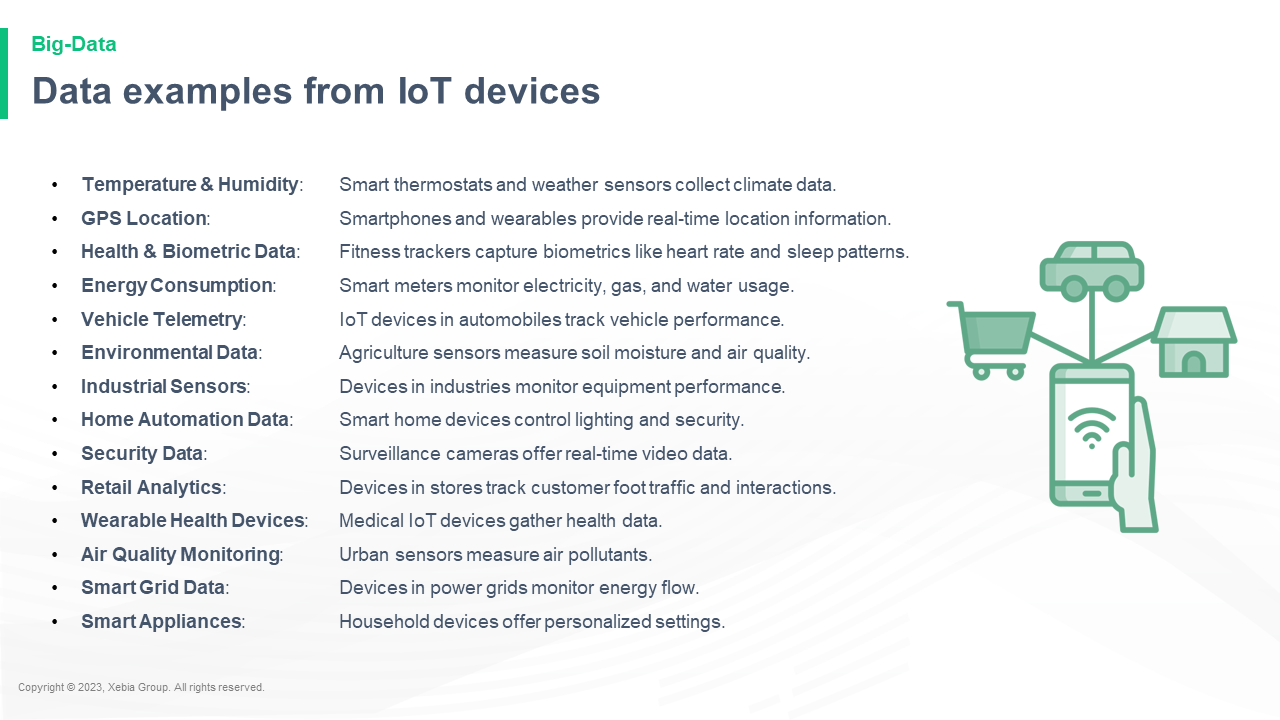
Text Data Posts comments tweets and messages including status updates reviews Images Photos and visual content uploaded by users Videos Video content shared on social media platforms User Profiles Demographic data interests and preferences collected from user profiles Location Data Geotagged information revealing user locations Engagement Metrics Likes shares comments retweets and reactions used to measure content performance Social Network Data Information on user connections and interactions forming social network graphs Hashtags and Keywords Tracking popular trends and brand mentions Sentiment Data Analyzing user sentiment towards topics brands or events Time Series Data Timestamped data for trend analysis and event impact assessment

**Example of Big Data**



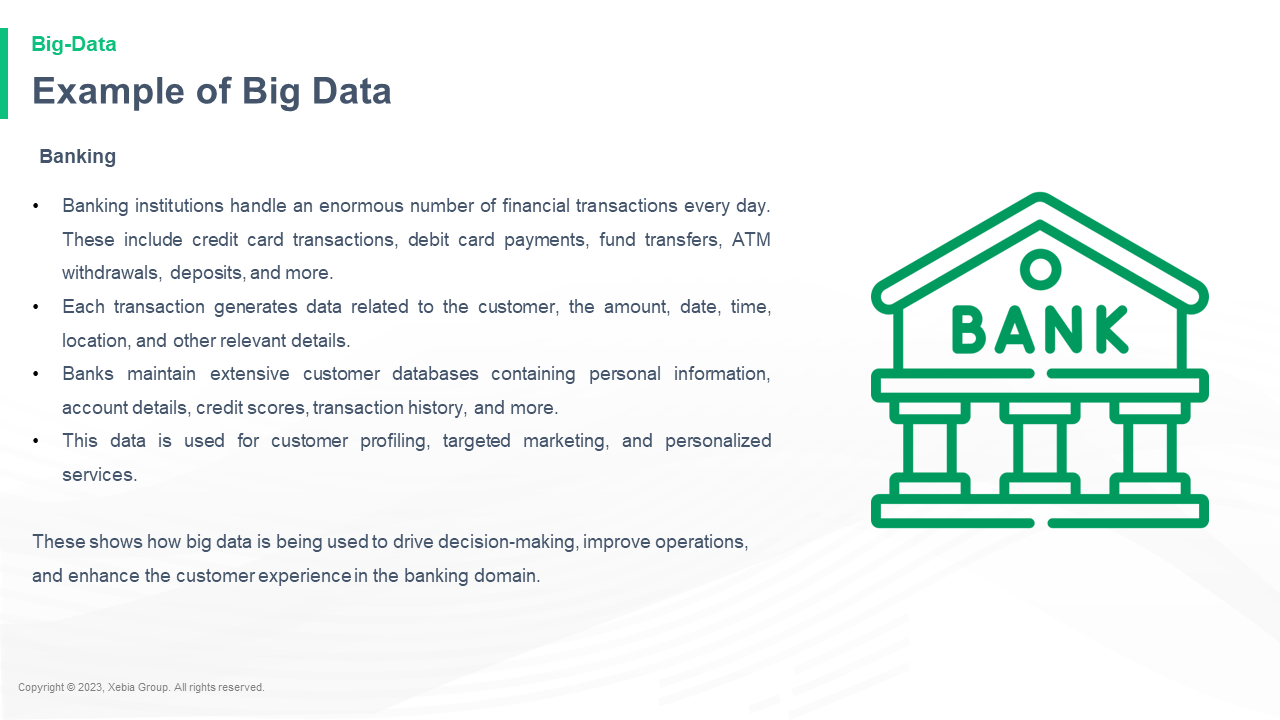
IoT devices IoT devices like smart appliances wearable devices and industrial sensors continuously collect data through embedded sensors This data is transmitted to centralized servers or the cloud using various communication technologies Once in the cloud big data analytics techniques process and analyze the data to derive valuable insights These insights can be applied to improve efficiency optimize processes enhance user experiences and enable data driven decision making in various domains such as smart cities and healthcare Overall IoT devices are a significant source of big data generating massive amounts of information that can be harnessed for numerous practical applications in our interconnected world

**Data examples from IoT devices**



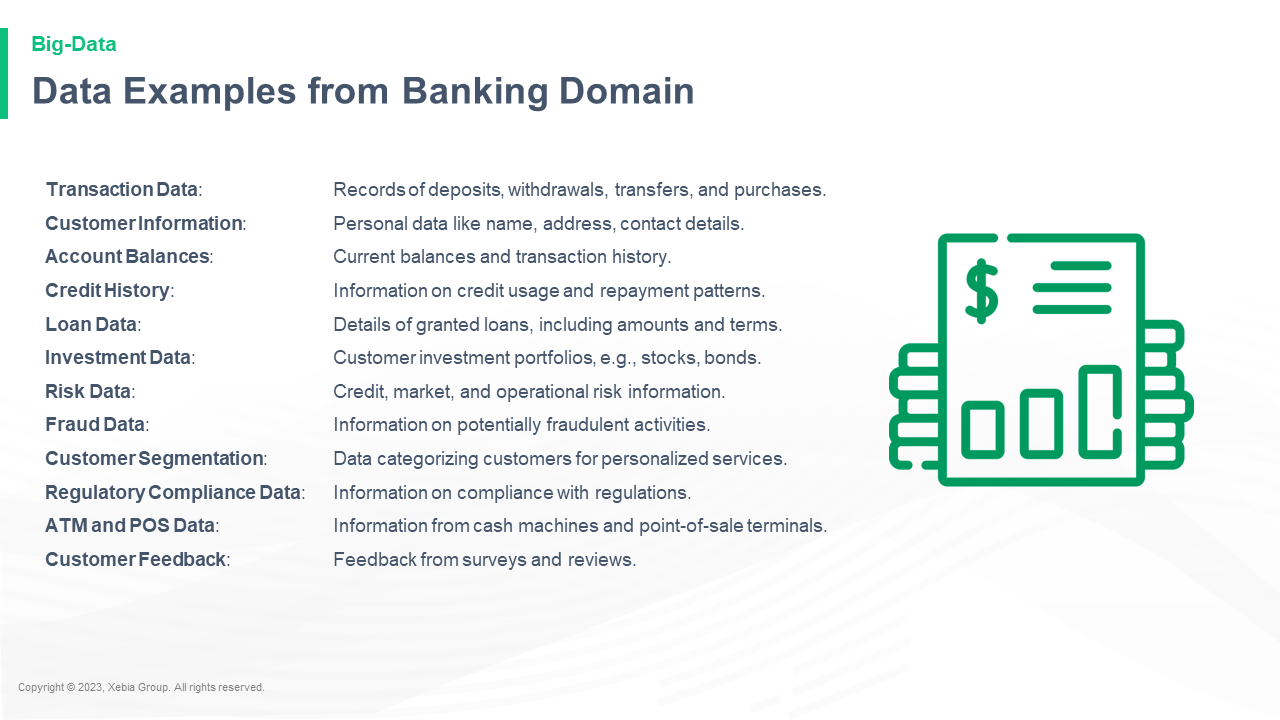
Temperature Humidity Smart thermostats and weather sensors collect climate data GPS Location Smartphones and wearables provide real time location information Health Biometric Data Fitness trackers capture biometrics like heart rate and sleep patterns Energy Consumption Smart meters monitor electricity gas and water usage Vehicle Telemetry IoT devices in automobiles track vehicle performance Environmental Data Agriculture sensors measure soil moisture and air quality Industrial Sensors Devices in industries monitor equipment performance Home Automation Data Smart home devices control lighting and security Security Data Surveillance cameras offer real time video data Retail Analytics Devices in stores track customer foot traffic and interactions Wearable Health Devices Medical IoT devices gather health data Air Quality Monitoring Urban sensors measure air pollutants Smart Grid Data Devices in power grids monitor energy flow Smart Appliances Household devices offer personalized settings

**Example of Big Data**



Banking institutions handle an enormous number of financial transactions every day These include credit card transactions debit card payments fund transfers ATM withdrawals deposits and more Each transaction generates data related to the customer the amount date time location and other relevant details Banks maintain extensive customer databases containing personal information account details credit scores transaction history and more This data is used for customer profiling targeted marketing and personalized services These shows how big data is being used to drive decision making improve operations and enhance the customer experience in the banking domain Banking

**Data Examples from Banking Domain**



Transaction Data Records of deposits withdrawals transfers and purchases Customer Information Personal data like name address contact details Account Balances Current balances and transaction history Credit History Information on credit usage and repayment patterns Loan Data Details of granted loans including amounts and terms Investment Data Customer investment portfolios e g stocks bonds Risk Data Credit market and operational risk information Fraud Data Information on potentially fraudulent activities Customer Segmentation Data categorizing customers for personalized services Regulatory Compliance Data Information on compliance with regulations ATM and POS Data Information from cash machines and point of sale terminals Customer Feedback Feedback from surveys and reviews

**Quiz**



What is the primary characteristic of Big Data Small Volume Structured Format Low Velocity Massive quantity Answer D Explanation The primary characteristic of Big Data is its massive volume Big Data refers to datasets that are too large and complex for traditional data processing applications to handle efficiently It involves vast amounts of data generated from various sources such as social media sensors and transaction records The sheer volume of data necessitates specialized tools and techniques for storage processing and analysis

**Quiz**

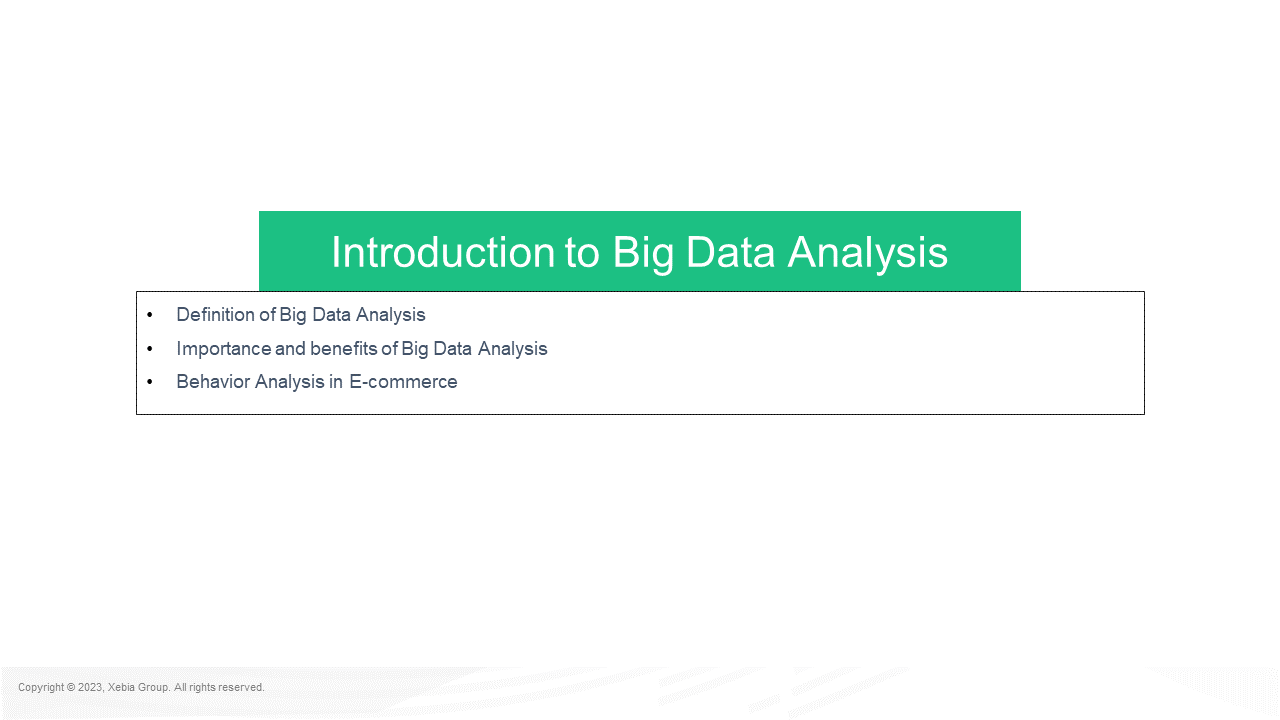


What is the goal of Big Data analytics To store and manage large volumes of data efficiently To analyze data to gain valuable insights and make informed decisions To develop complex algorithms for data processing To create visually appealing dashboards for data representation Answer B Explanation the primary goal of Big Data analytics is to analyze large volumes of data to extract valuable insights and patterns enabling informed decisions and driving business improvements It involves using advanced tools to process and interpret data efficiently helping organizations identify opportunities detect risks enhance customer experiences optimize processes and improve overall performance

**In a nutshell we learnt**

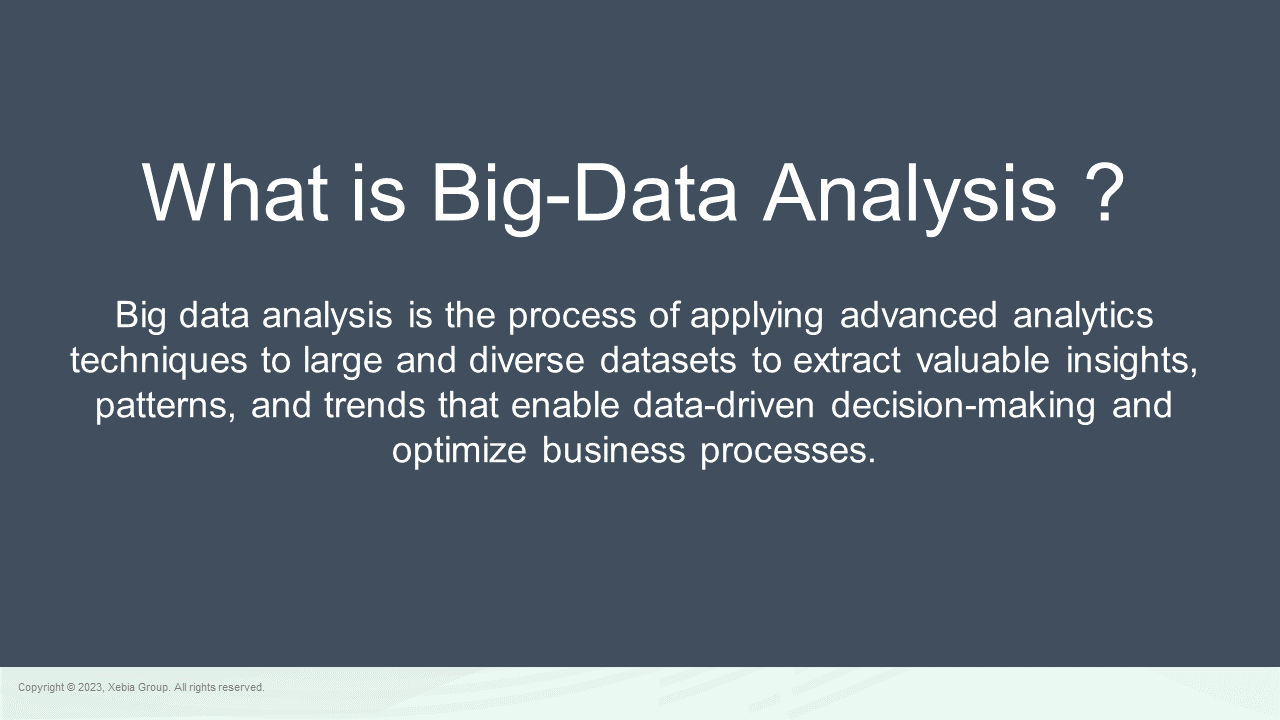


Big data refers to extremely large and complex data sets that surpass the capabilities of traditional data processing methods There are 5 Vs of Big Data Volume Velocity Variety Veracity and Value Big data is generated across multiple domains banking IoT devices e commerce and social media Often transactional data and user interaction data are huge in volume and are termed as big data



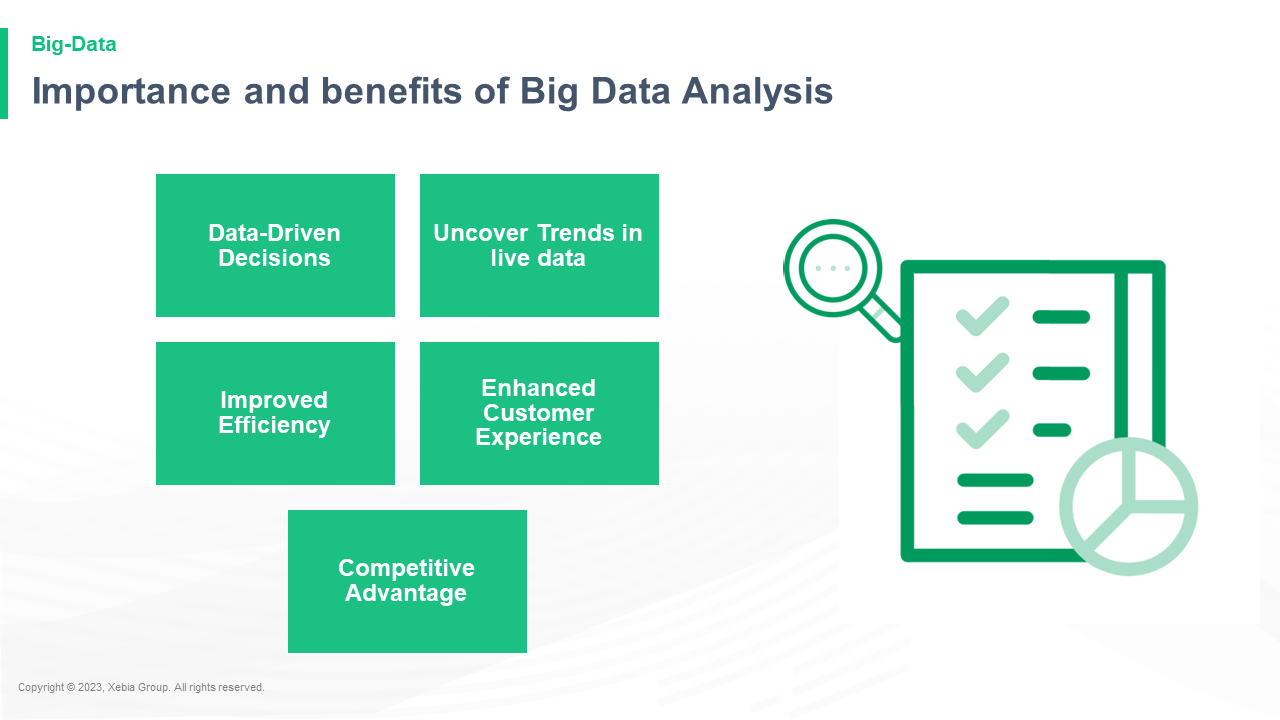
Definition of Big Data Analysis Importance and benefits of Big Data Analysis Behavior Analysis in E commerce Introduction to Big Data Analysis

**Big data analysis is the process of applying advanced analytics techniques to large and diverse datasets to extract valuable insights patterns and trends that enable data driven decision making and optimize business processes**

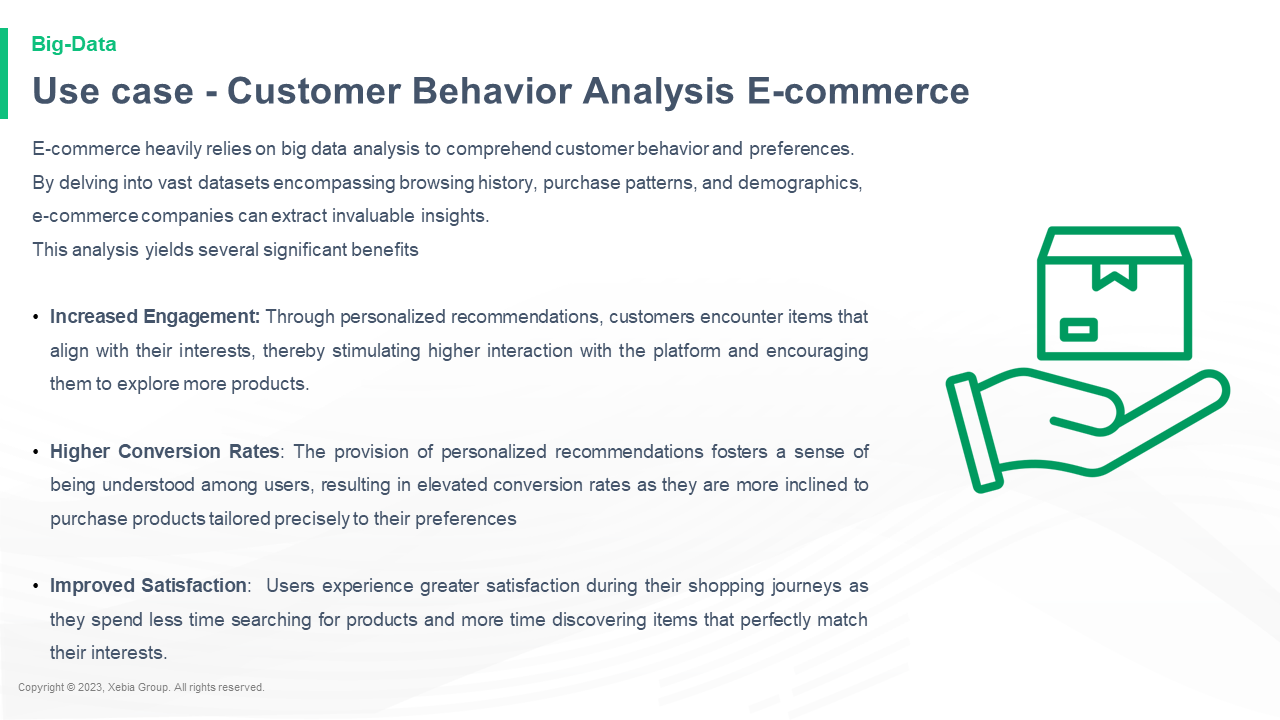


What is Big Data Analysis

**Importance and benefits of Big Data Analysis**



**Use case Customer Behavior Analysis E commerce**



E commerce heavily relies on big data analysis to comprehend customer behavior and preferences By delving into vast datasets encompassing browsing history purchase patterns and demographics e commerce companies can extract invaluable insights This analysis yields several significant benefits Increased Engagement Through personalized recommendations customers encounter items that align with their interests thereby stimulating higher interaction with the platform and encouraging them to explore more products Higher Conversion Rates The provision of personalized recommendations fosters a sense of being understood among users resulting in elevated conversion rates as they are more inclined to purchase products tailored precisely to their preferences Improved Satisfaction Users experience greater satisfaction during their shopping journeys as they spend less time searching for products and more time discovering items that perfectly match their interests

**Use case Supply Chain Optimization**



Current Scenario Efficient supply chain management can significantly impact success The company faced difficulties with managing inventory recurrent stockouts suboptimal transportation routes and operational inefficiencies that adversely affected their profit margin and customer satisfaction Solution The company embraced Big Data Analysis to evaluate various data including sales statistics stock levels logistics information and external variables like weather and demand variations This data centric approach enabled proficient inventory management reduced stockouts through demand forecasting optimized transport routes and enhanced operational efficiency Impact The supply chain s operational efficiency saw a substantial improvement leading to cost reduction and waste minimization This optimization also enhanced the overall customer experience by ensuring products were readily available according to customer needs

**Quiz**



How does the Big Data Analysis is utilized in understanding customer behavior and preferences By enabling real time analysis of customer interactions and feedback By offering personalized product recommendations based on historical data Identifying patterns and trends in customer behavior from large datasets Automating customer service interactions using AI algorithms All Of the Above Answer E Explanation Big Data Analysis is crucial for understanding customer behavior and preferences It enables real time analysis of customer interactions and feedback offers personalized product recommendations based on historical data identifies patterns and trends from large datasets and automates customer service interactions using AI algorithms This valuable insight helps improve products enhance experiences and drive business growth

**Quiz**

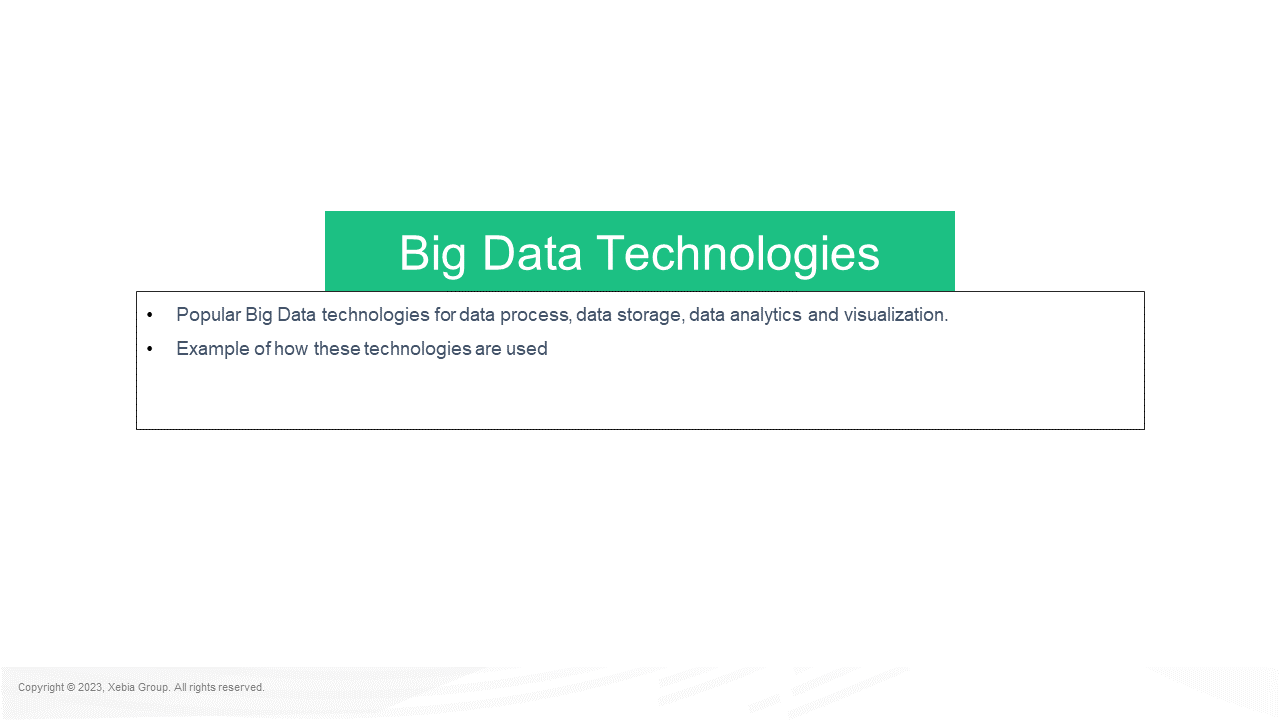


What are the importance and benefits of Big Data Analysis Improved decision making enhanced customer experiences and reduced operational costs Slower data processing increased data complexity and higher infrastructure costs Limited scalability decreased data insights and reduced business agility No significant impact on businesses outdated technology and increased data security risks Answer A Explanation Big Data Analysis refers to the process of analyzing and extracting valuable insights from large and complex datasets The importance and benefits of Big Data Analysis are immense

**In a nutshell we learnt**

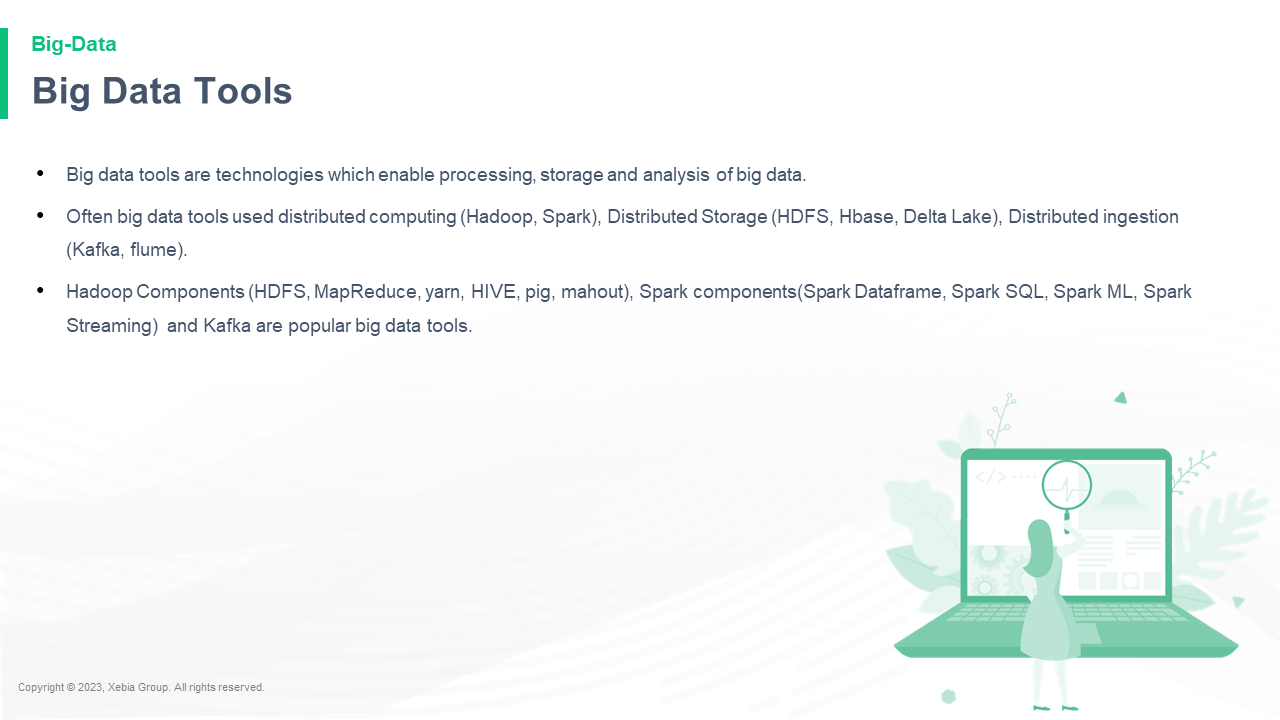


Big Data Analysis involves examining large datasets to extract valuable insights for informed decision making Its importance lies in gaining deeper insights into customer behavior market trends and operational efficiency enabling data driven decision making personalized services and improved business strategies In e commerce behavior analysis analyzes customer interactions to understand consumer behavior optimize user experiences and enhance targeted marketing efforts



Popular Big Data technologies for data process data storage data analytics and visualization Example of how these technologies are used Big Data Technologies

**Big Data Tools**



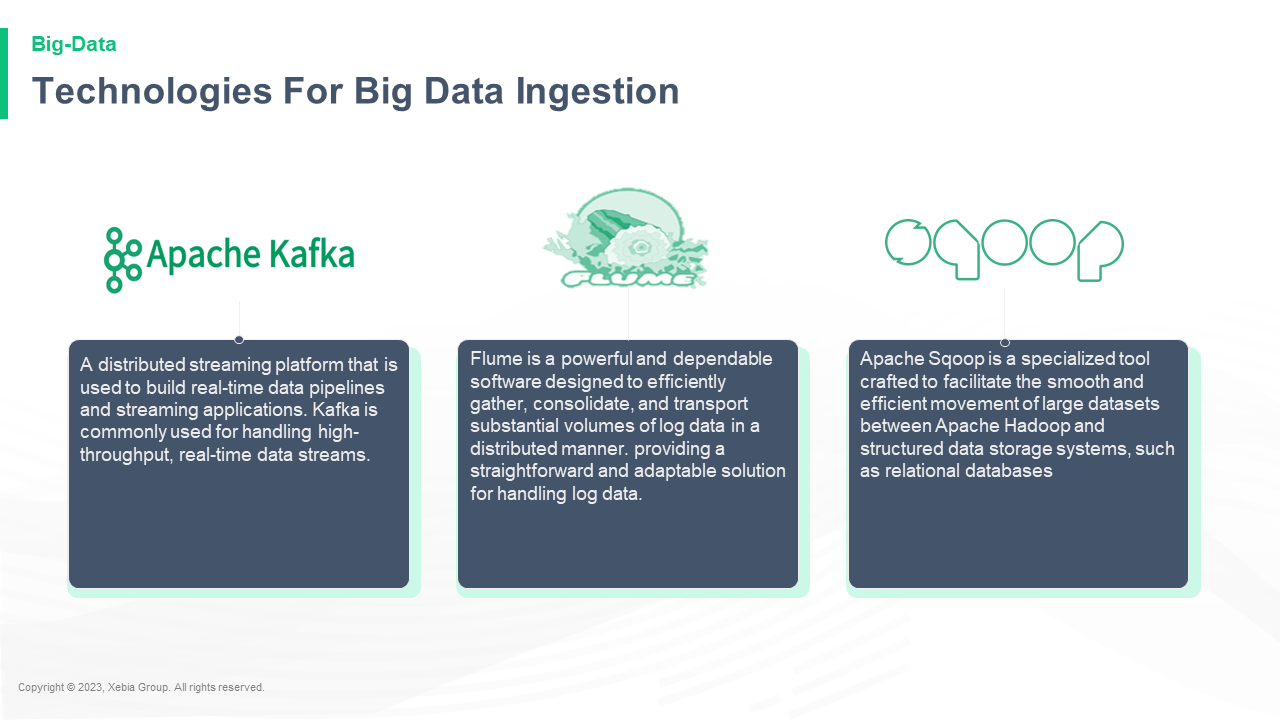
Big data tools are technologies which enable processing storage and analysis of big data Often big data tools used distributed computing Hadoop Spark Distributed Storage HDFS Hbase Delta Lake Distributed ingestion Kafka flume Hadoop Components HDFS MapReduce yarn HIVE pig mahout Spark components Spark Dataframe Spark SQL Spark ML Spark Streaming and Kafka are popular big data tools

**Big Data Tools**

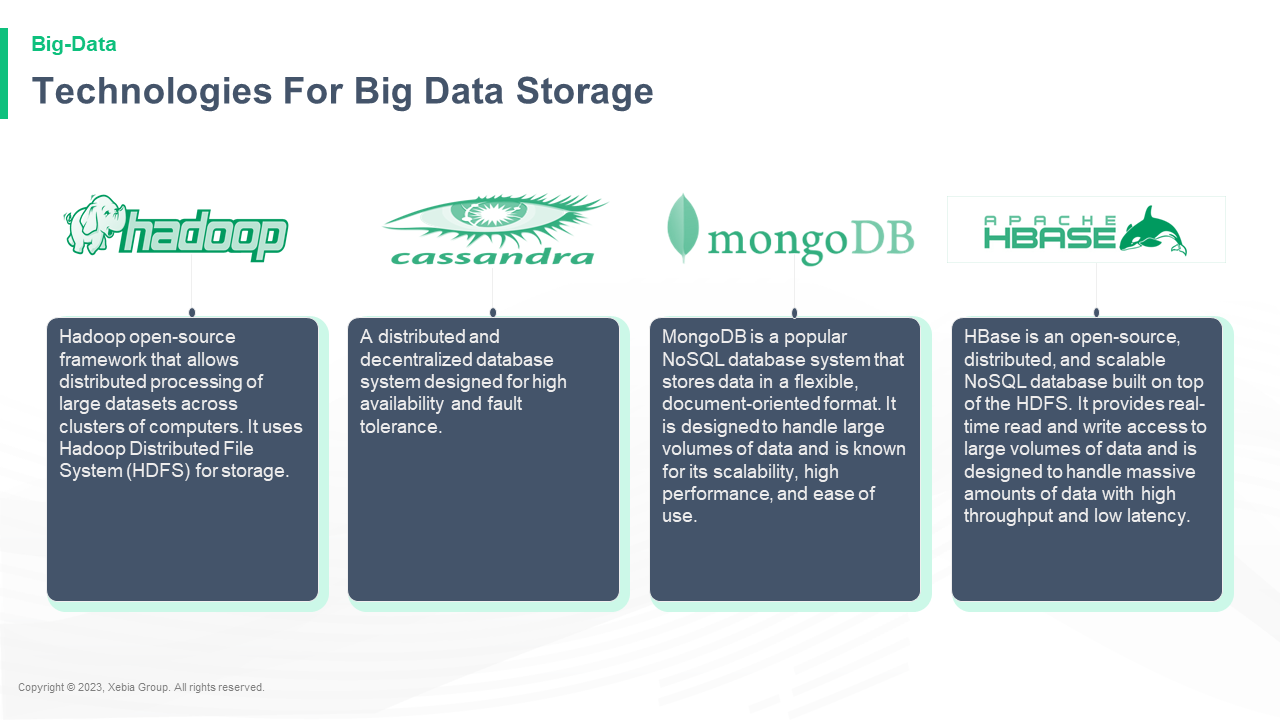


Big Data Ingestion Big Data Storage Big Data Processing Big Data Analytics Big Data Visualization Apache Kafka Apache Flume Sqoop HDFS Cassandra MongoDB Hbase Apache Spark MapReduce Apache Flink Apache Beam Apache Hive Apache Pig Tableau Power BI Grafana

**Technologies For Big Data Ingestion**



**Technologies For Big Data Storage**



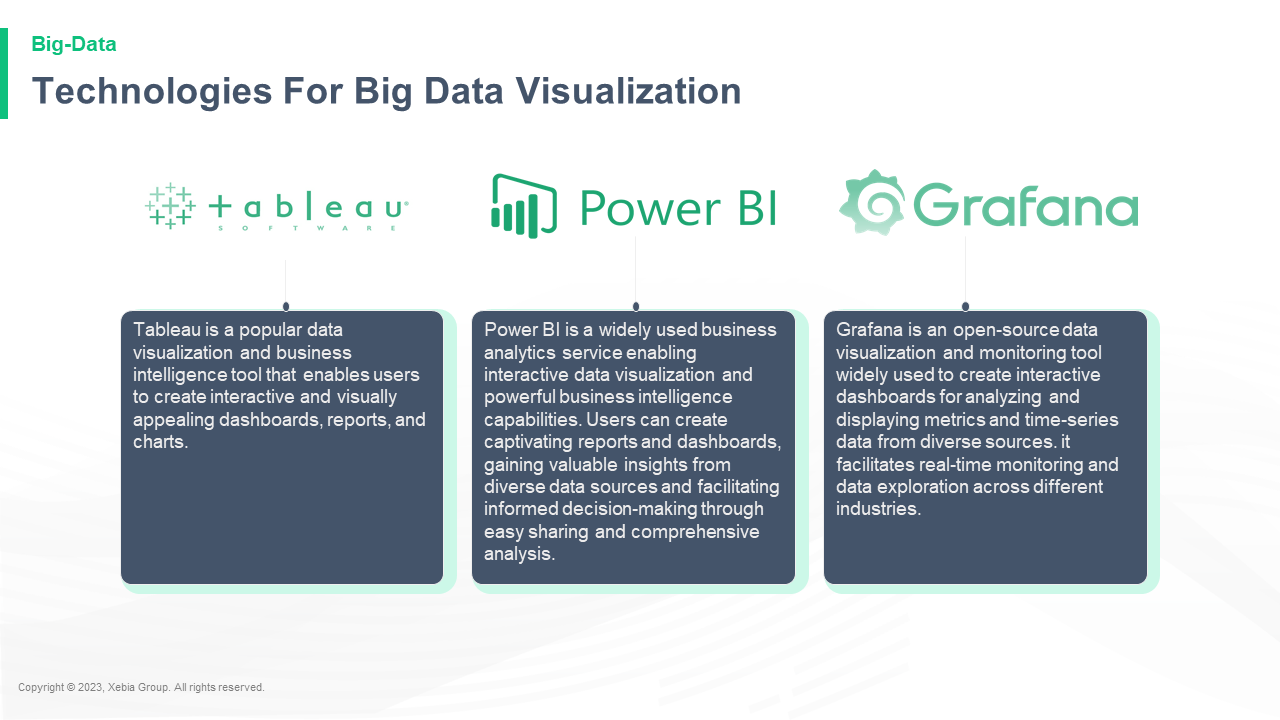
**Technologies For Big Data Processing**



**Technologies For Big Data Analytics**



**Technologies For Big Data Visualization**



**How Yahoo utilise Hadoop**



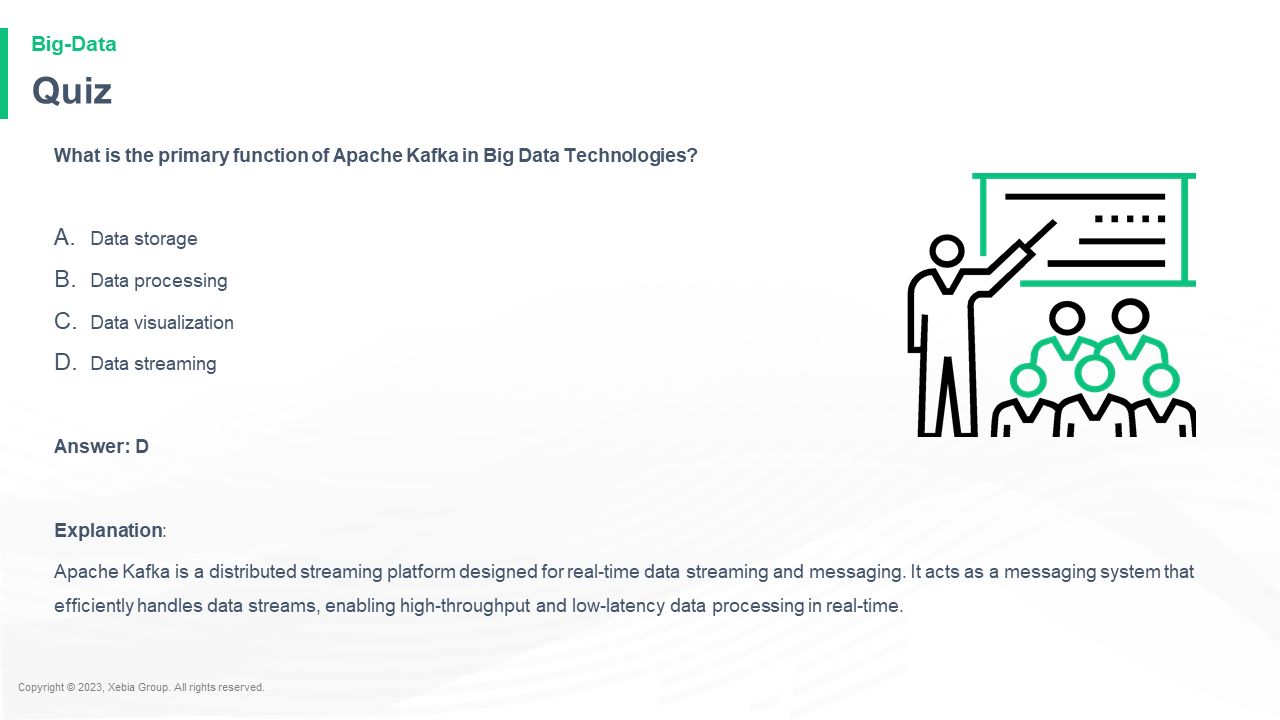
40 Yahoo a global technology company manages extensive user data from services such as search email news and advertising To handle this massive data efficiently Here s how Yahoo utilized Hadoop Data Storage Yahoo used Hadoop s Hadoop Distributed File System HDFS to store petabytes of data across a cluster of commodity servers HDFS allows for data replication ensuring data redundancy and fault tolerance Data Processing Yahoo employed Hadoop s MapReduce programming model to process and analyze the stored data MapReduce enables parallel processing of data across multiple nodes in the Hadoop cluster making it possible to crunch vast datasets quickly

**Yahoo utilise Hadoop**



Data Analysis Hadoop s ability to handle unstructured data allowed Yahoo to perform in depth analysis of user behavior preferences and interactions with their services This data analysis provided valuable insights for improving their products personalizing user experiences and optimizing advertising strategies Scalability As Yahoo s data continued to grow Hadoop s scalability allowed them to add more nodes to the cluster seamlessly accommodating the increasing data volume without disruptions

**Quiz**



What is the primary function of Apache Kafka in Big Data Technologies Data storage Data processing Data visualization Data streaming Answer D Explanation Apache Kafka is a distributed streaming platform designed for real time data streaming and messaging It acts as a messaging system that efficiently handles data streams enabling high throughput and low latency data processing in real time

**Quiz**

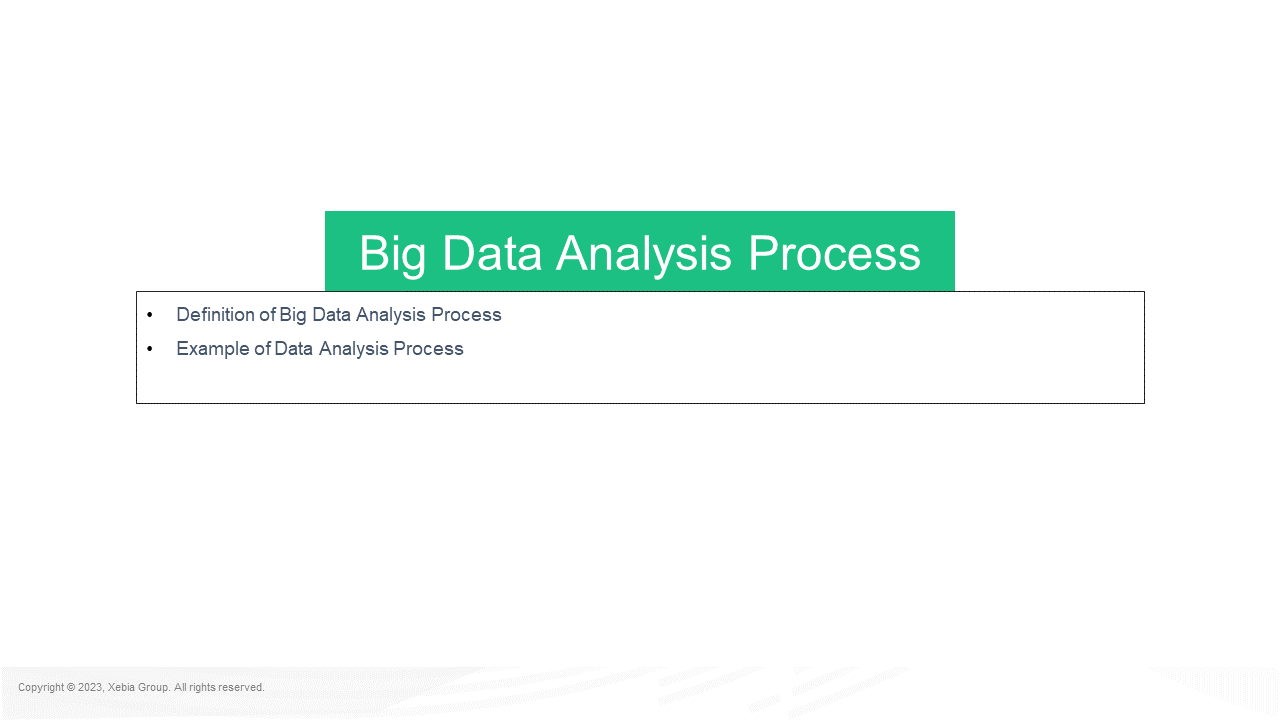


Which technology is commonly used for distributed storage and processing of large datasets in Big Data applications SQL Server Hadoop Microsoft Excel Oracle Answer B Explanation Hadoop is a widely used technology in the Big Data domain for distributed storage and processing of large datasets It is an open source framework that allows the processing of massive amounts of data across clusters of commodity hardware

**In a nutshell we learnt**

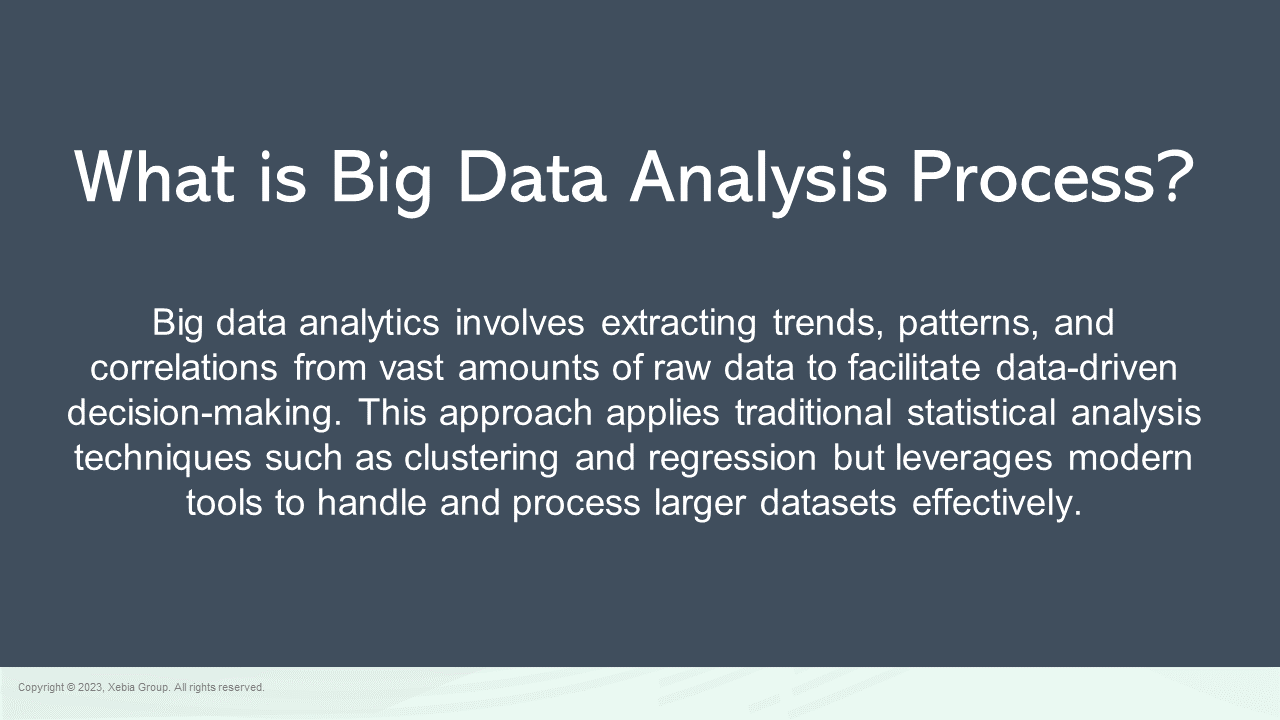


well known high level big data technologies such as Hadoop an open source framework that allows distributed processing of large datasets across clusters of computers Kafka A distributed streaming platform that is used to real time data streams The practical application of these technologies in real world scenarios showcasing their efficiency in handling massive datasets



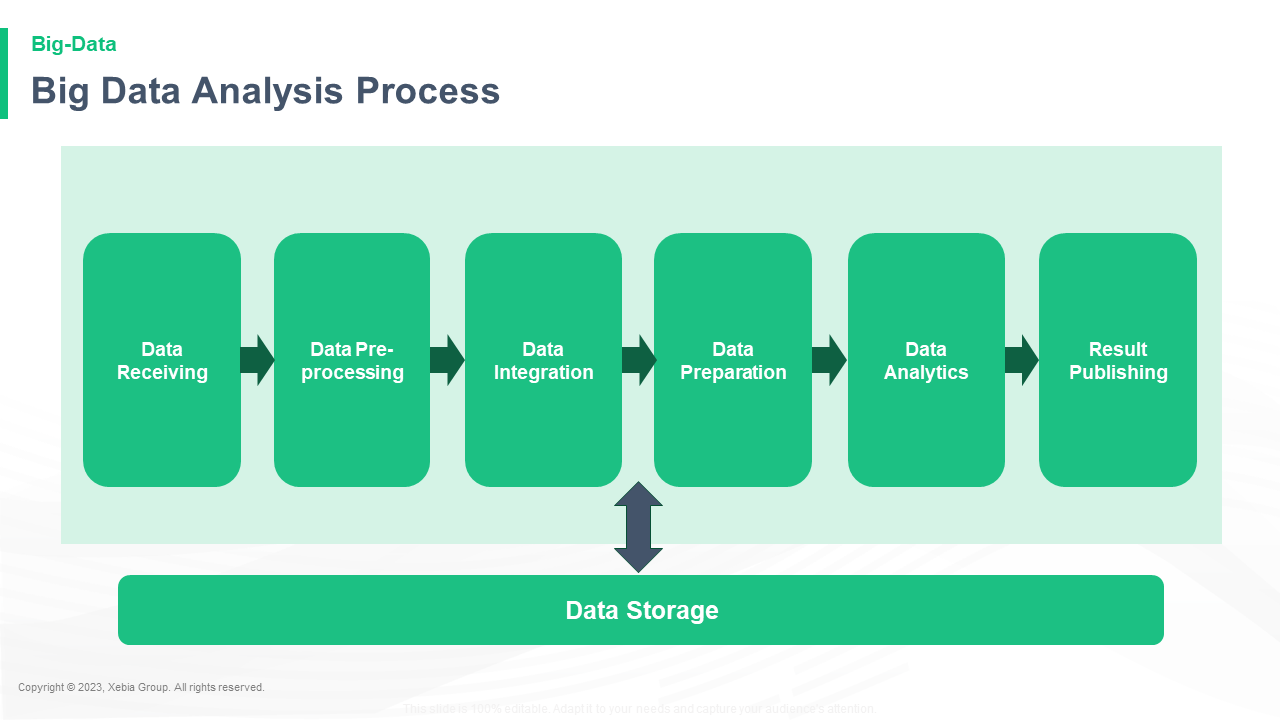
Definition of Big Data Analysis Process Example of Data Analysis Process Big Data Analysis Process

**Big data analytics involves extracting trends patterns and correlations from vast amounts of raw data to facilitate data driven decision making This approach applies traditional statistical analysis techniques such as clustering and regression but leverages modern tools to handle and process larger datasets effectively**



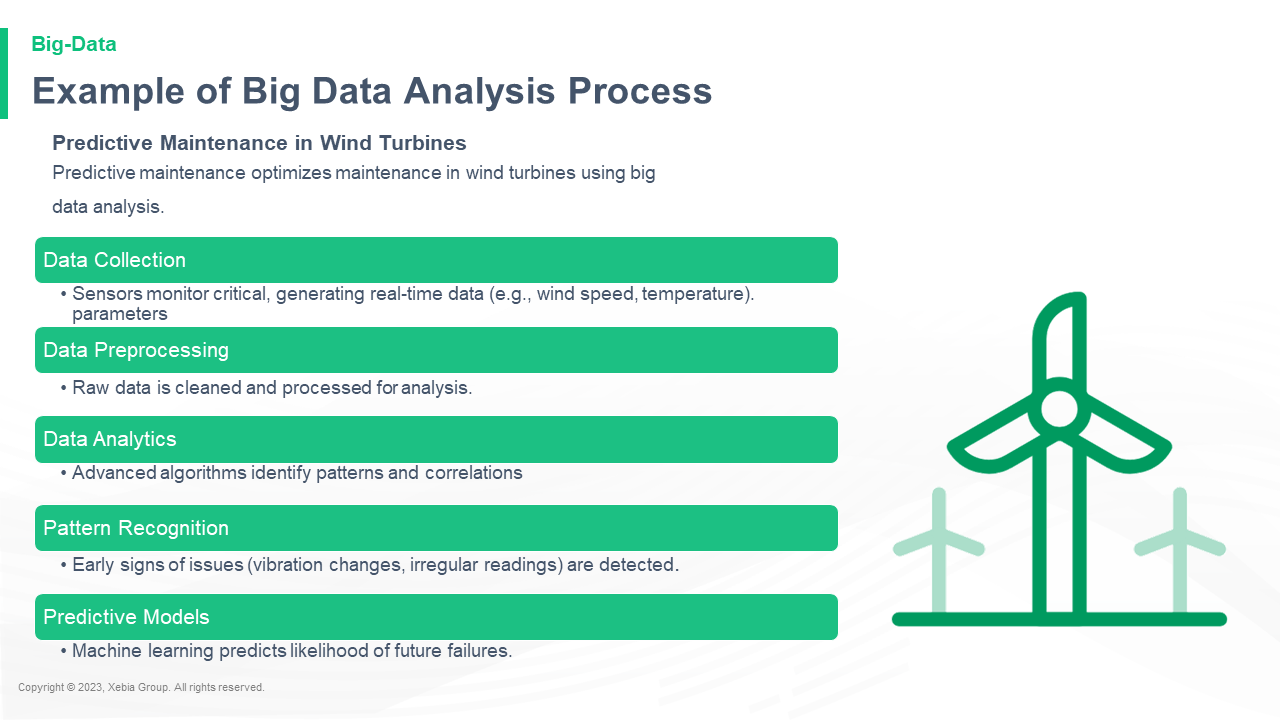
What is Big Data Analysis Process

**Big Data Analysis Process**



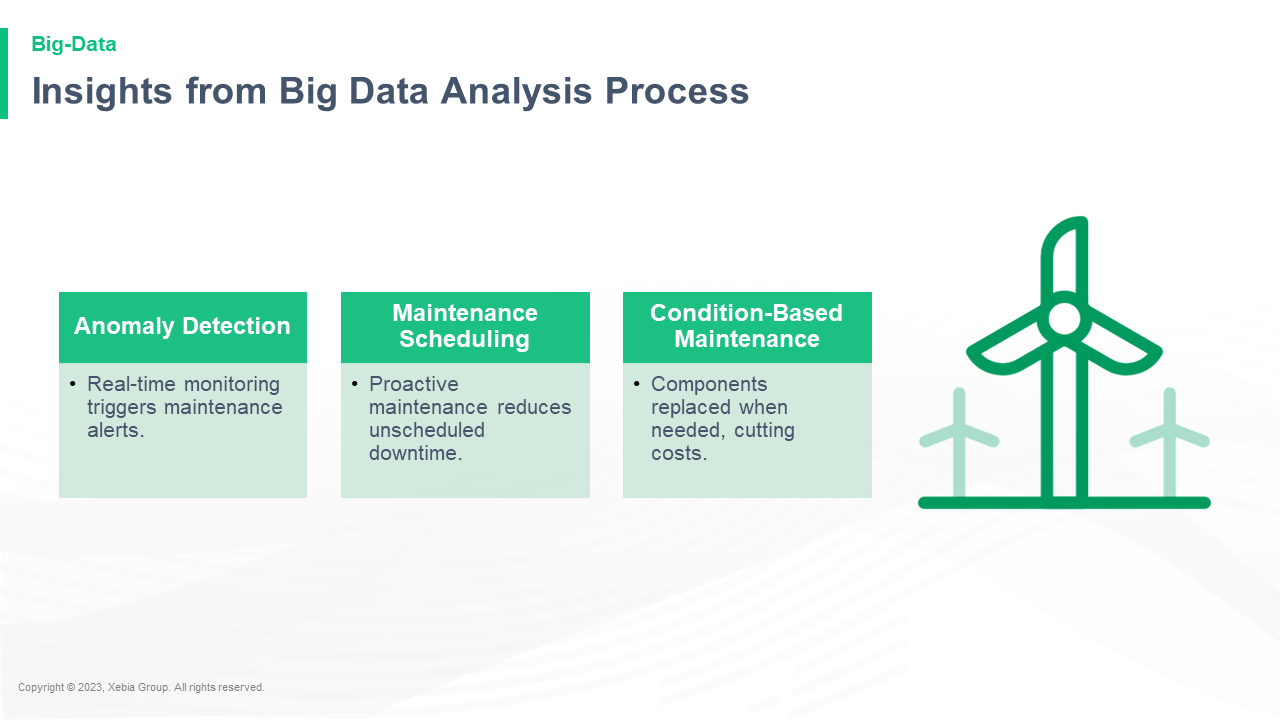
This slide is 100 editable Adapt it to your needs and capture your audience s attention Data Storage Data Receiving Data Pre processing Data Integration Data Preparation Data Analytics Result Publishing

**Example of Big Data Analysis Process**

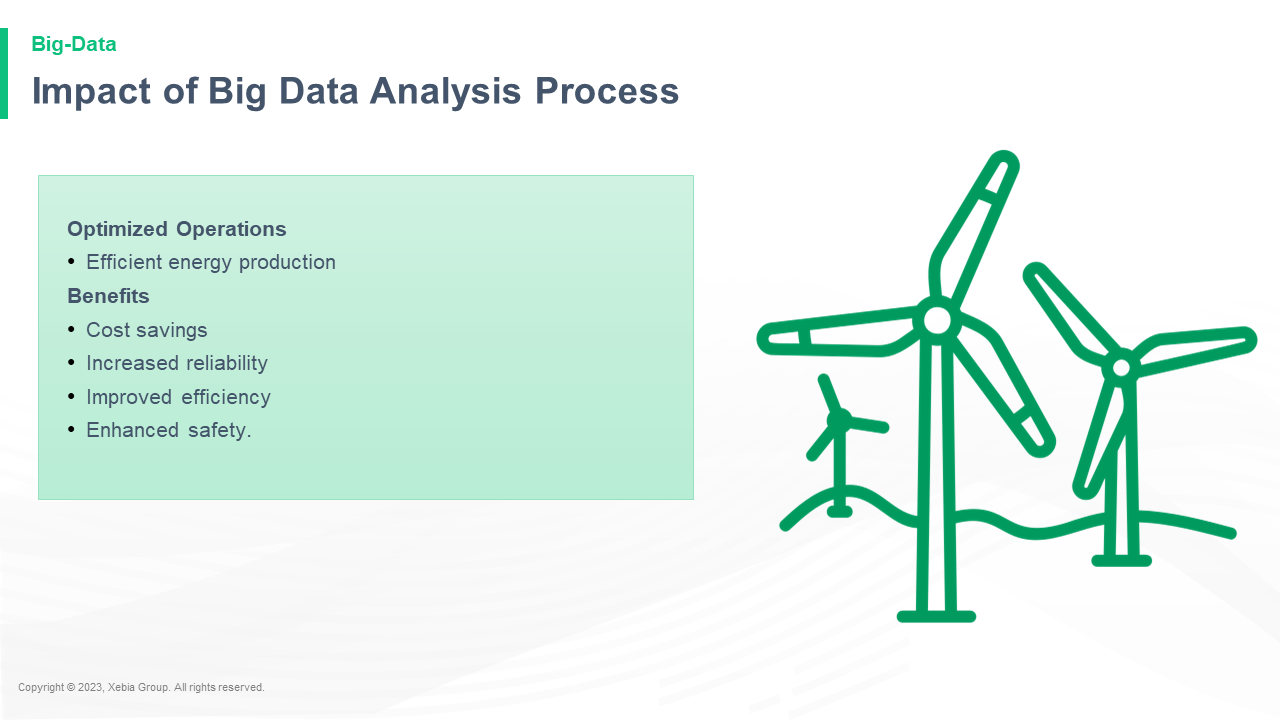


Predictive Maintenance in Wind Turbines Predictive maintenance optimizes maintenance in wind turbines using big data analysis

**Insights from Big Data Analysis Process**

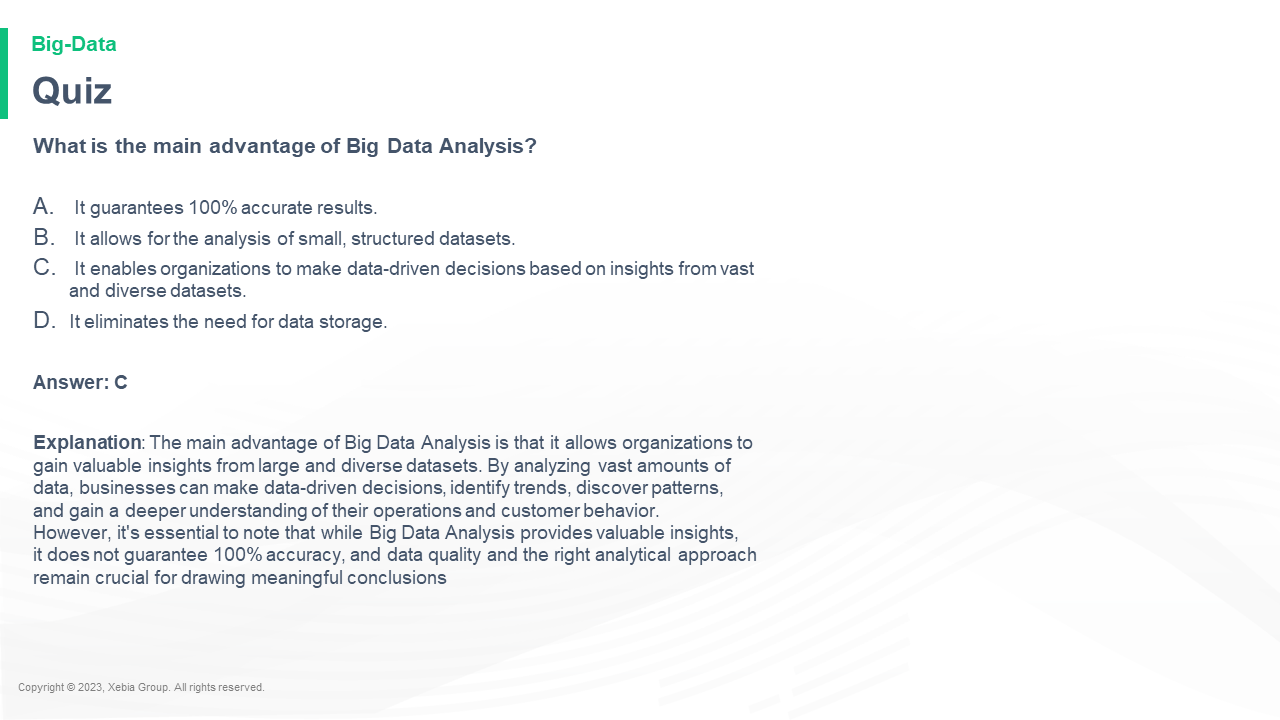


**Impact of Big Data Analysis Process**



Optimized Operations Efficient energy production Benefits Cost savings Increased reliability Improved efficiency Enhanced safety

**Quiz**



What is the main advantage of Big Data Analysis It guarantees 100 accurate results It allows for the analysis of small structured datasets It enables organizations to make data driven decisions based on insights from vast and diverse datasets It eliminates the need for data storage Answer C Explanation The main advantage of Big Data Analysis is that it allows organizations to gain valuable insights from large and diverse datasets By analyzing vast amounts of data businesses can make data driven decisions identify trends discover patterns and gain a deeper understanding of their operations and customer behavior However it s essential to note that while Big Data Analysis provides valuable insights it does not guarantee 100 accuracy and data quality and the right analytical approach remain crucial for drawing meaningful conclusions

**Quiz**

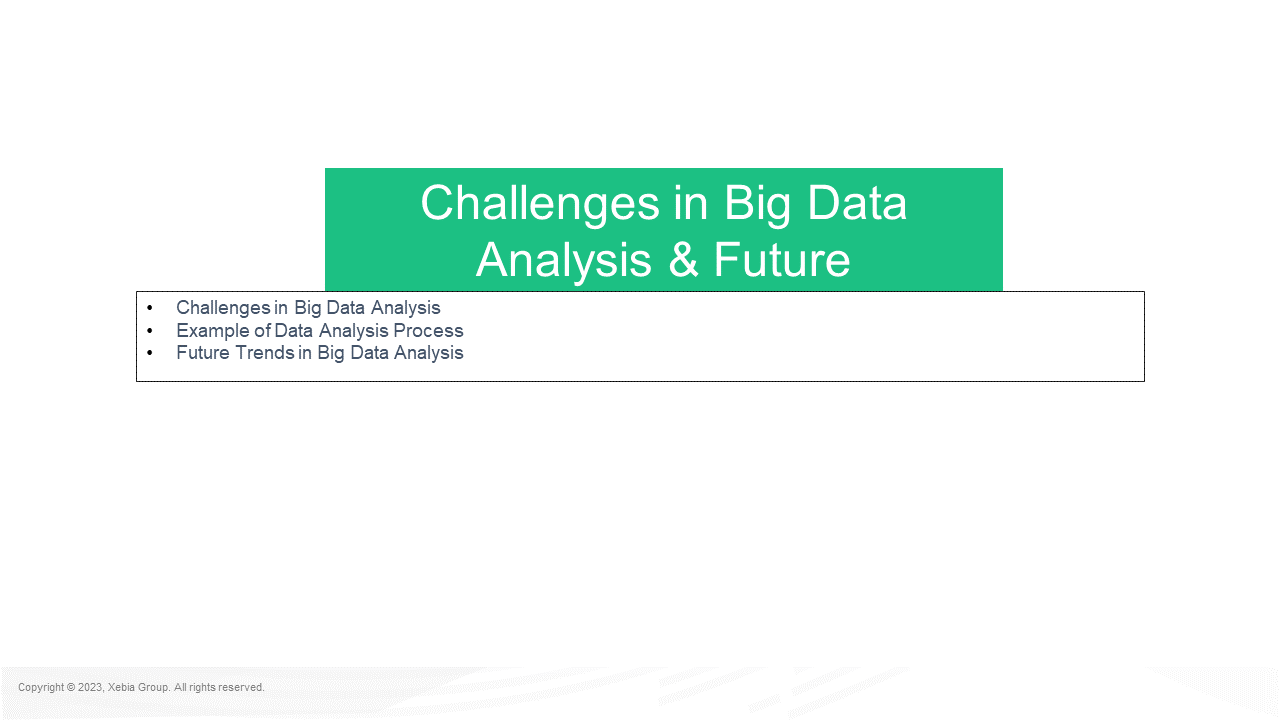


What is the correct sequence of steps in the Big Data Analysis process Data Collection Data Visualization Data Cleaning Data Modeling Insights Extraction Data Cleaning Data Visualization Data Collection Insights Extraction Data Modeling Data Collection Data Cleaning Data Modeling Data Visualization Insights Extraction Data Modeling Data Cleaning Data Visualization Data Collection Insights Extraction Answer C Explanation Data Collection Data Cleaning Data Modeling Data Visualization Insights Extraction represents the correct sequence of steps in the Big Data Analysis process ensuring a systematic approach to derive valuable insights from vast and complex datasets

**In nutshell we learnt**

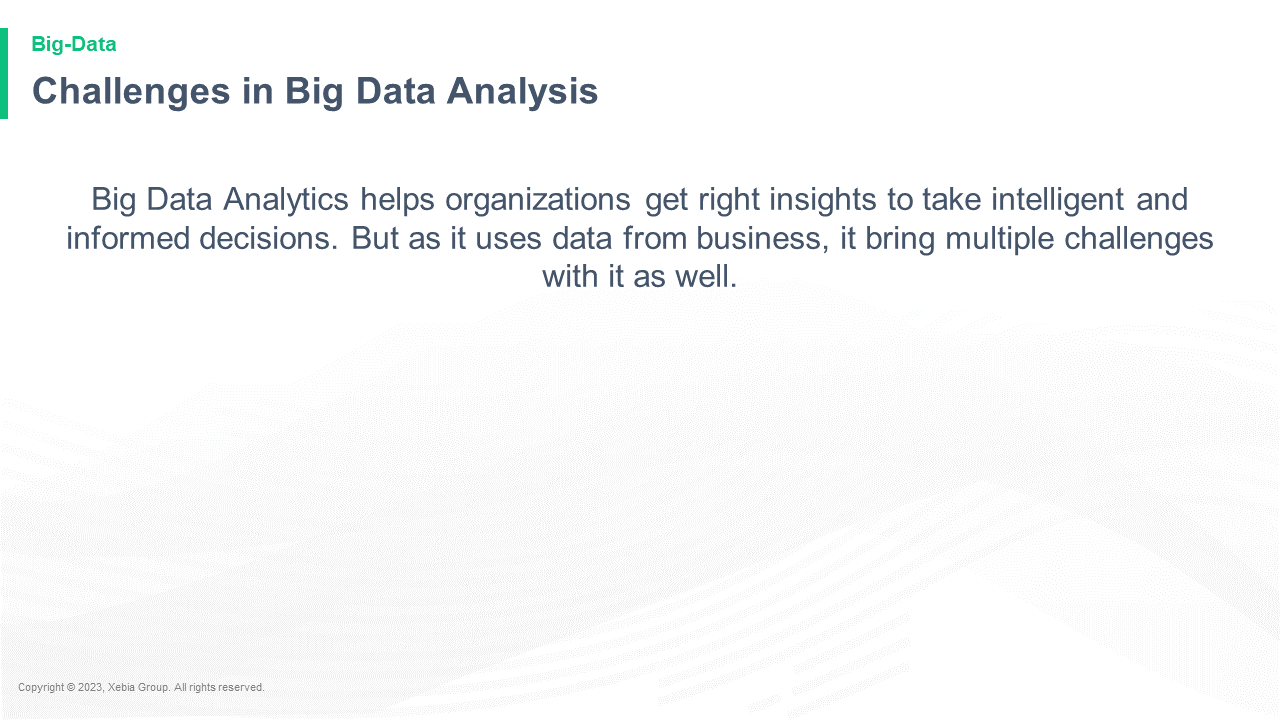


Definition of big data analysis process and how it s working Example of data analysis process in a wind turbines the data analysis process involves collecting data information and performing analysis optimization to improve operational efficiency



Challenges in Big Data Analysis Example of Data Analysis Process Future Trends in Big Data Analysis Challenges in Big Data Analysis Future

**Challenges in Big Data Analysis**



Big Data Analytics helps organizations get right insights to take intelligent and informed decisions But as it uses data from business it bring multiple challenges with it as well

**Challenges in Big Data Analysis**



56 Data Privacy Security and Compliance Data Quality Variety and Integration Infrastructure and Scalability Real Time Analysis and Data Processing Speed Ethics and Bias in Data

**Challenges and Solutions Twitter**



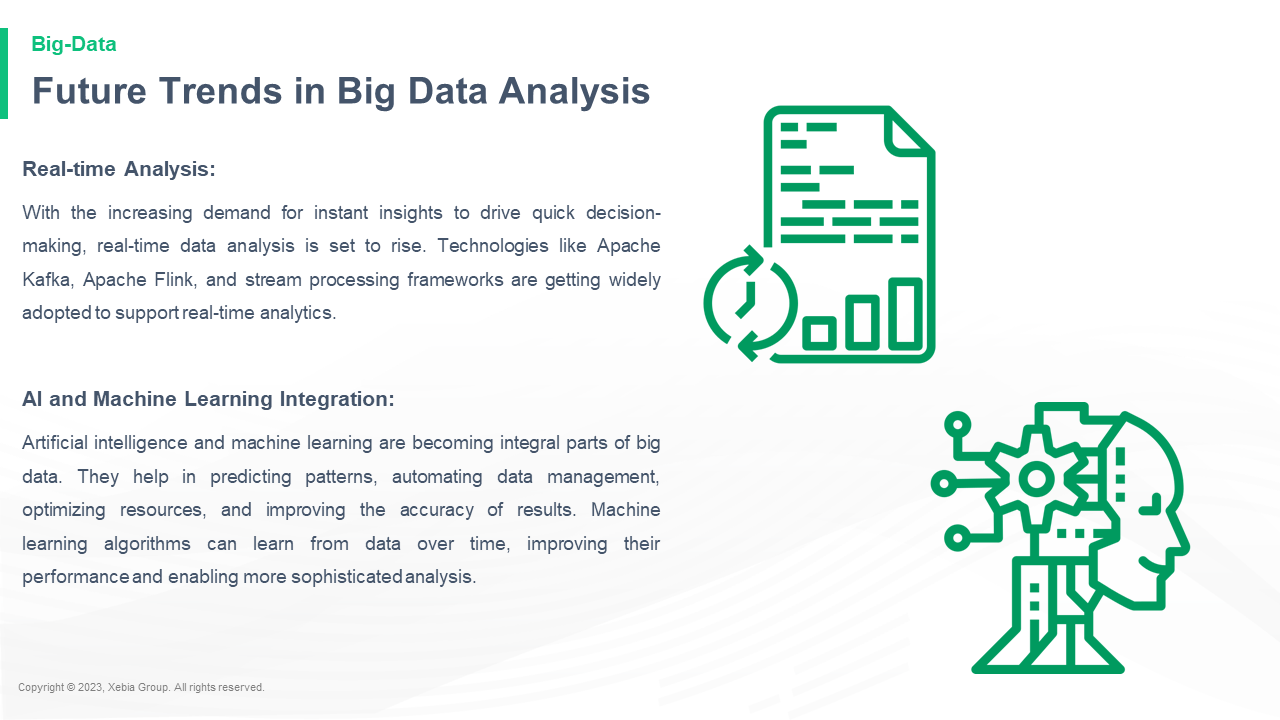
Challenge Twitter faced the challenge of analyzing and processing an enormous stream of real time data generated by millions of users Traditional relational databases and single server solutions were insufficient to handle the scale and complexity of the data in a timely manner The company needed a system that could process analyze and extract insights from a continuous stream of tweets which could potentially reach thousands of tweets per second Solution Twitter overcame this challenge by adopting Apache Storm a distributed real time computation system Apache Storm allows the processing of streaming data in a fault tolerant and scalable manner Instead of relying on a single server Apache Storm distributes the computational load across a cluster of machines enabling horizontal scalability and high throughput

**Challenges and Solutions Twitter**



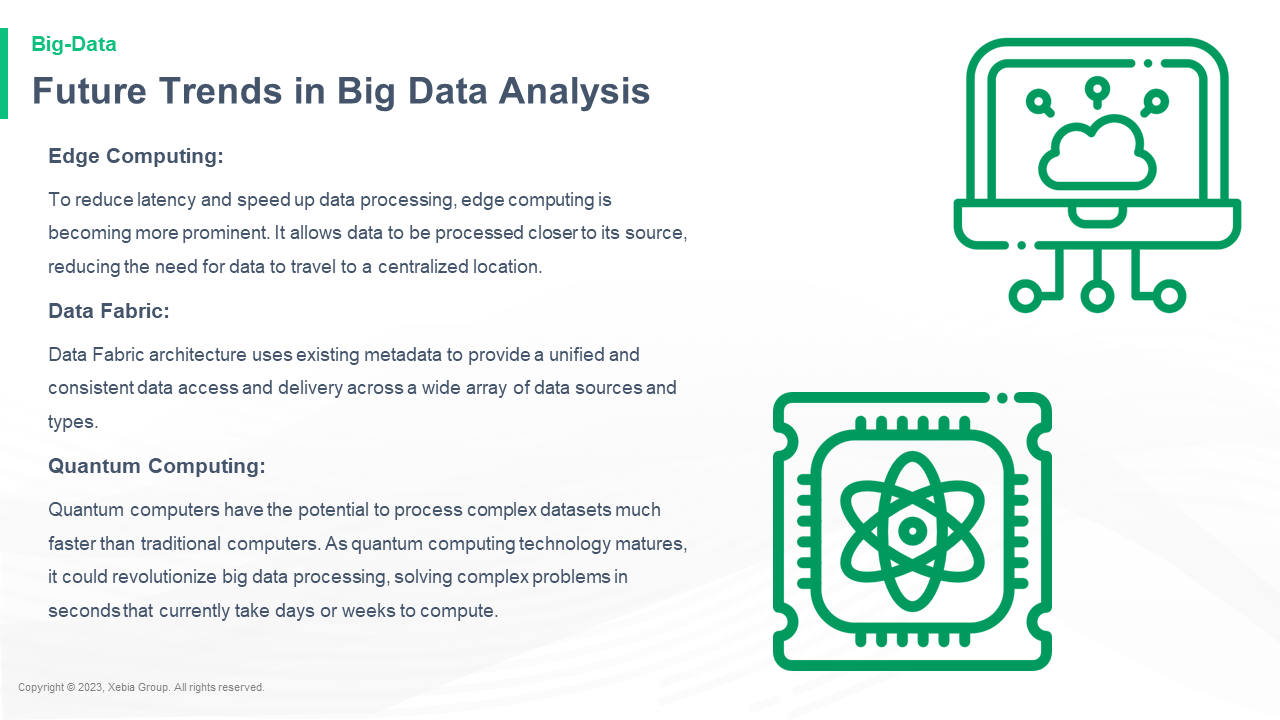
Here s how Apache Storm helped Twitter with real time analytics Distributed Architecture Enables parallel processing across multiple worker nodes optimizing resource utilization and speeding up data analysis Fault Tolerance Built in mechanism ensures processing continues on available nodes if a cluster node fails maintaining system reliability for massive data streams Scalability Easily scales the cluster by adding more worker nodes to handle increasing data volumes without compromising performance Real time Insights Low latency processing allows Twitter to derive real time insights from tweet streams enabling trend monitoring and quick response to critical events Extensibility Supports multiple programming languages giving Twitter developers flexibility in writing data processing logic

**Future Trends in Big Data Analysis**



Real time Analysis With the increasing demand for instant insights to drive quick decision making real time data analysis is set to rise Technologies like Apache Kafka Apache Flink and stream processing frameworks are getting widely adopted to support real time analytics AI and Machine Learning Integration Artificial intelligence and machine learning are becoming integral parts of big data They help in predicting patterns automating data management optimizing resources and improving the accuracy of results Machine learning algorithms can learn from data over time improving their performance and enabling more sophisticated analysis

**Future Trends in Big Data Analysis**



Edge Computing To reduce latency and speed up data processing edge computing is becoming more prominent It allows data to be processed closer to its source reducing the need for data to travel to a centralized location Data Fabric Data Fabric architecture uses existing metadata to provide a unified and consistent data access and delivery across a wide array of data sources and types Quantum Computing Quantum computers have the potential to process complex datasets much faster than traditional computers As quantum computing technology matures it could revolutionize big data processing solving complex problems in seconds that currently take days or weeks to compute

**Quiz**



Which technology is expected to revolutionize data storage in Big Data Analysis Cloud Computing Traditional databases Magnetic tapes Optical disks Answer A Explanation Cloud computing is expected to revolutionize data storage in Big Data Analysis Cloud platforms offer scalable storage solutions allowing organizations to efficiently handle large datasets without the need for significant physical infrastructure investments Cloud based storage also ensures data accessibility and flexibility making it an ideal choice for the future of Big Data Analysis

**Quiz**

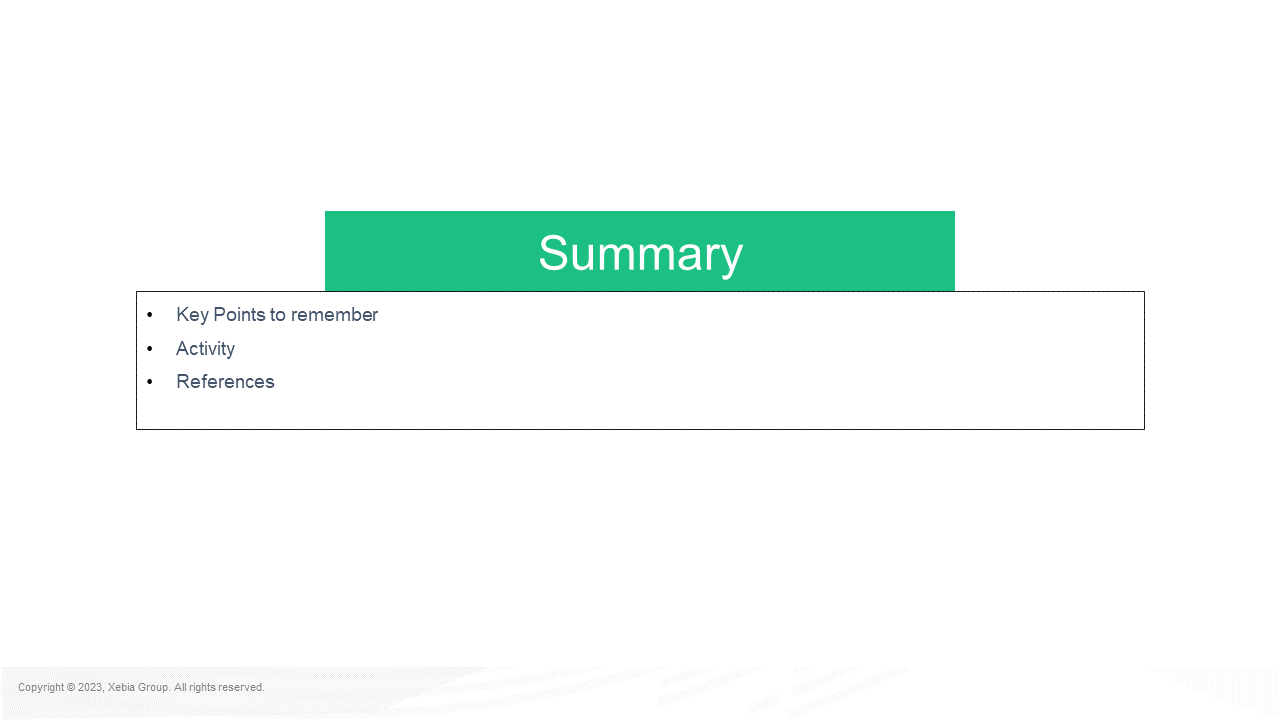


Which of the following is a future trend in Big Data Analysis Increased reliance on traditional relational databases Decreased adoption of cloud computing for data storage Growing emphasis on real time data processing Reduced importance of data security and privacy Answer C Explanation Real time data processing is a future trend in Big Data Analysis As the volume and velocity of data continue to grow the need for real time or near real time analysis becomes crucial Organizations increasingly require the ability to analyze and act upon data as it is generated to make informed decisions promptly Real time data processing enables businesses to respond quickly to changing conditions identify opportunities and mitigate risks in real time

**In a nutshell we learnt**



Challenges in Big Data Analysis Dealing with data volume velocity variety and veracity to extract meaningful insights efficiently Example of Data Analysis Process Cleaning transforming and visualizing large datasets to uncover patterns and trends for business decision making Future Trends in Big Data Analysis Advancements in AI and machine learning integration of edge computing and increased emphasis on data privacy and security



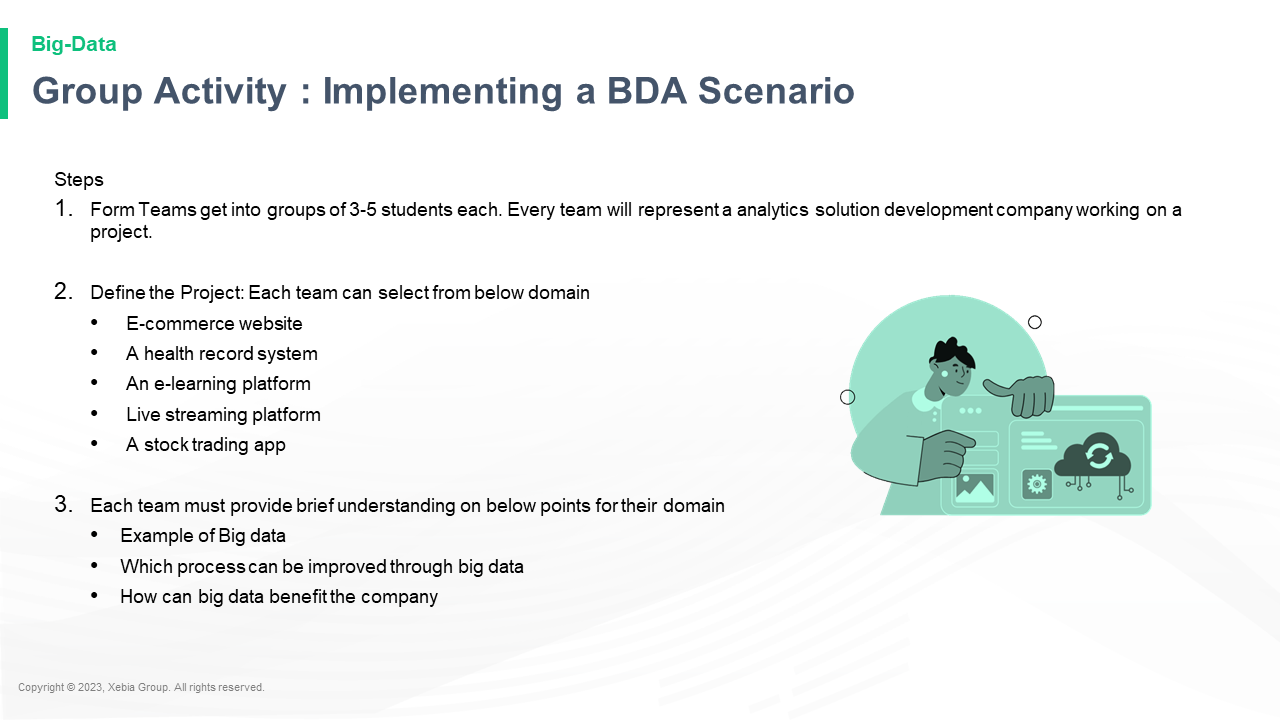
Key Points to remember Activity References Summary

**Key Points to remember**



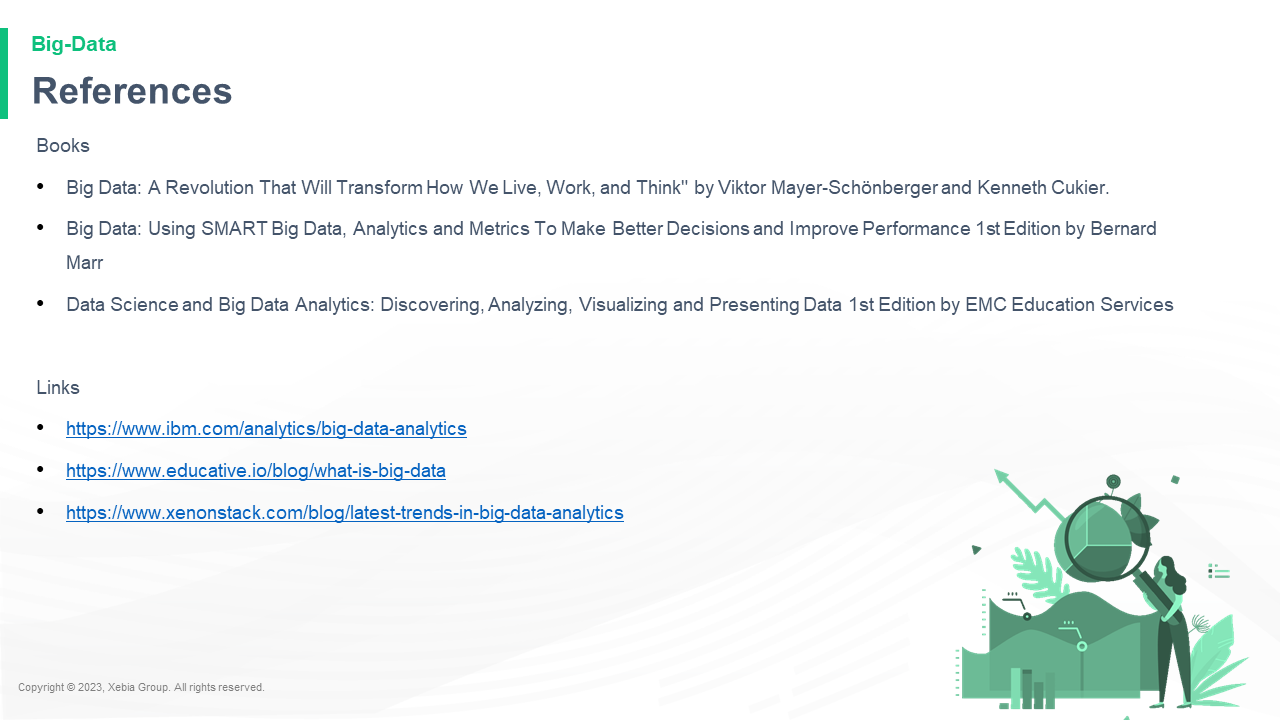
Big Data We defined Big Data as data sets that are characterized by the 5 Vs Volume Velocity Variety Veracity and Value These massive and diverse datasets pose unique challenges and opportunities for organizations Importance of Big Data Analysis We discussed the importance and benefits of Big Data Analysis enabling organizations to gain valuable insights make data driven decisions and stay ahead of the competition Big Data Technologies An overview of popular Big Data technologies like Hadoop which empower organizations to process and analyze vast datasets efficiently Challenges in Big Data Analysis We acknowledged the common challenges faced in Big Data Analysis such as data privacy quality issues and data integration hurdles

**Group Activity Implementing a BDA Scenario**



Steps Form Teams get into groups of 3 5 students each Every team will represent a analytics solution development company working on a project Define the Project Each team can select from below domain E commerce website A health record system An e learning platform Live streaming platform A stock trading app Each team must provide brief understanding on below points for their domain Example of Big data Which process can be improved through big data How can big data benefit the company

**References**



Books Big Data A Revolution That Will Transform How We Live Work and Think by Viktor Mayer Schönberger and Kenneth Cukier Big Data Using SMART Big Data Analytics and Metrics To Make Better Decisions and Improve Performance 1st Edition by Bernard Marr Data Science and Big Data Analytics Discovering Analyzing Visualizing and Presenting Data 1st Edition by EMC Education Services Links https www ibm com analytics big data analytics https www educative io blog what is big data https www xenonstack com blog latest trends in big data analytics

**Thank You**

