



Name:	Prerna Sunil Jadhav
Sap Id:	60004220127
Class:	T. Y. B. Tech (Computer Engineering)
Course:	Big Data Infrastructure Laboratory
Course Code:	DJ19CEEL6011
Experiment No.:	01

**AIM:** Case study on any organization using big data technologies.

### WHAT IS BIG DATA & WHY IS IT IMPORTANT?

Big data analytics refers to the methods, tools, and applications used to collect, process, and derive insights from varied, high-volume, high-velocity data sets.

This ability to derive insights to inform better decision making is why big data is important. It's how a retailer might hone their targeted ad campaigns, or how a wholesaler might resolve bottlenecks in the supply chain. Big data analytics enables a more holistic, data-driven approach to decision-making, in turn promoting growth, efficiency, and innovation.

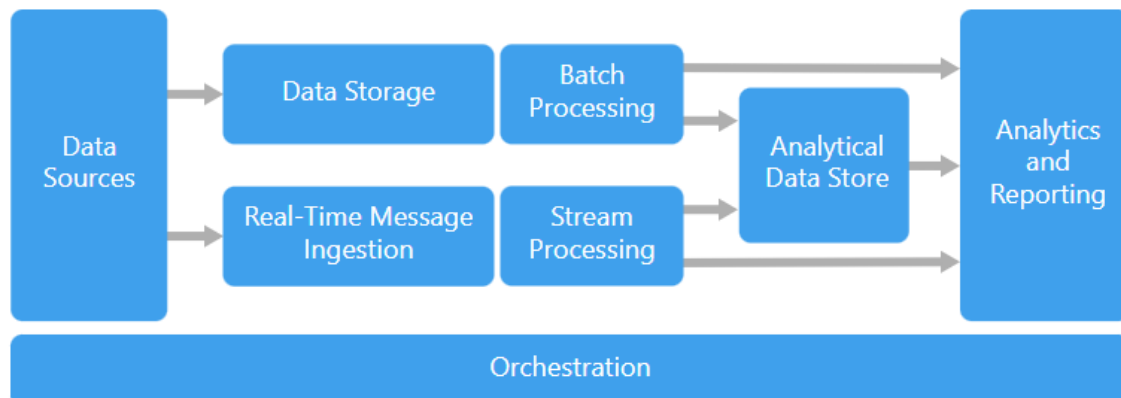
## MICROSOFT

### Introduction:

Microsoft makes extensive use of big data to improve its products and services. A well-known case study is Microsoft Azure, a cloud computing platform that uses various information technologies for data processing and analysis.

### Big data applications:

- Microsoft Azure leverages many aspects of big data, including Apache Hadoop for distribution and processing, Apache Spark for big data processing, and Apache Spark for big data.
- Azure Data Lake Storage for scalable and secure data lakes.
- Azure also offers services like Azure Databricks and Azure HDInsight to simplify big data.
- A big data architecture is designed to handle the ingestion, processing, and analysis of data that is too large or complex for traditional database systems.
- Components of a big data architecture:



- Microsoft leverages a variety of big data across its platforms and services, demonstrating how to process big data and providing insights from it.



Some of the important information technologies used by Microsoft are:



1. Azure Data Lake Storage (ADLS): ADLS is designed to store big data in a distributed and secure manner. It provides a perfect solution for storing big data by supporting structured and unstructured data.

2. Azure HDInsight: This is a cloud-based big data analytics service that makes it easy to process big data using popular open source services such as Apache Hadoop, Spark, Hive, and more. HDInsight allows this process to be implemented on demand, providing a scalable and cost-effective solution.



3. Apache Hadoop: Microsoft integrates open-source Hadoop into its big data to store and process big data. Hadoop's Distributed File System (HDFS) & MapReduce programming model allow Microsoft to store & process data efficiently across clusters of computers.

4. Apache Spark: Apache Spark is a fast and versatile computing solution for big data processing. Microsoft is integrating Spark into real-time data analytics, machine learning, and image processing services to provide faster and more flexible data processing.



5. Azure Databricks: This is an analytics platform based on Apache Spark that is available as part of the Azure cloud service. Databricks simplifies the Spark cluster creation process and provides a unified platform for data scientists and architects to work together on big data analytics projects.

6. Azure Synapse Analytics (formerly SQL Data Warehouse): A cloud-based business data warehouse service that can process large amounts of data and provide close analysis. It integrates with various big data technologies to enable data storage and analysis of different data.



7. Azure Streaming Analytics: Microsoft leverages Azure Streaming Analytics for real-time data processing. The service enables analysis of data streams from multiple sources, providing rapid insights and actionable information.

Together, these technologies enable Microsoft to deliver a powerful ecosystem for big data that supports the storage, processing and analysis of diverse and large data sets. Listed as scalable and efficient.



✚ Advantages:

1. Scalability: Large data sets allow Microsoft to scale its processes based on data needs.
2. Analytics: Advanced analytics capabilities help Microsoft derive insights from large amounts of data to help make informed decisions.
3. Real-time processing: Microsoft can help respond to user needs faster by processing and analysing data in real-time through technologies such as Apache Spark.

✚ Disadvantages:

1. Complexity: Using and managing large data sets can be complex and require special skills and resources.
2. Security issues: Processing large amounts of data can cause security issues and leaks can have serious consequences.

Conclusion:

In conclusion, Microsoft's use of big data exemplified in the Azure ecosystem demonstrates the company's commitment to using data to create updated content. Although this technology has the potential to provide powerful and robust analysis, issues such as complexity and security need to be carefully managed to achieve good results. Overall, Microsoft's big data strategy has helped bolster its leadership in technology.