# BIG DATA ANALYTICS
# WITH
# IBM  CLOUD

**ABSTRACT:**

Big data is a new driver of the world economic and societal changes. The world's data collection is reaching a tipping point for major technological changes that can bring new ways in decision making, managing our health, cities, finance and education. While the data complexities are increasing including data's volume, variety, velocity and veracity, the real impact hinges on our ability to uncover the `value' in the data through Big Data Analytics technologies. Big Data Analytics poses a grand challenge on the design of highly scalable algorithms and systems to integrate the data and uncover large hidden values from datasets that are diverse, complex, and of a massive scale. Potential breakthroughs include new algorithms, methodologies, systems and applications in Big Data Analytics that discover useful and hidden knowledge from the Big Data efficiently and effectively.

**INTRODUCTION:**

Big Data stands for data sets which is usually much larger and complex than the common know data sets which usually handles by RDBMS. Well, know traditional data management applications like RDBMS are not able to manage those data sets. Big Data can be applied to Un-structured, Structured and Semi-structured data sets based on requirements and needs. Whereas Big Data is mostly applied to Un-structured data sets. Many of the well-known simultaneous Computing tools, **Business Analytics Software** require Big Data for handling their large data sets. Nowadays Big Data Analytics has been used in various Sectors like Media, Education, Healthcare, Manufacturing, various Government and non-government sectors and so on.

**PROJECT DESCRIPTION:**

In this project, you will leverage IBM Cloud services and tools to analyze and predict customer churn for a fictional telecom company. Customer churn, or the rate at which customers leave a service, is a critical problem for many businesses, and predictive analytics can help identify at-risk customers early so that proactive measures can be taken to retain them.

**Project Steps:**

1  Data Collection:

Gather a dataset containing customer information, including demographics, usage patterns, and churn status. You can find datasets for customer churn prediction on platforms like Kaggle.

2  Data Preparation:

Clean and preprocess the dataset. Handle missing data, encode categorical variables, and perform feature engineering to create relevant features for analysis.

3  Data Storage:

Utilize IBM Cloud Object Storage to store your prepared dataset securely.

4  Data Exploration:

Use IBM Watson Studio to explore and visualize the data. Generate summary statistics, histograms, and correlation matrices to gain insights into the dataset.

5    Model Development:

Build predictive models using tools available in IBM Watson Studio. You can experiment with various machine learning algorithms such as logistic regression, decision trees, random forests, or XGBoost.

6    Model Evaluation:

Assess the performance of your models using metrics like accuracy, precision, recall, F1- score, and ROC curves. Evaluate the models using cross-validation techniques to ensure robustness.

7    Deployment:

Deploy the best-performing model as a web service using IBM Watson Machine Learning. This will allow you to make real-time predictions.

8    Dashboard and Visualization:

Create a dashboard using IBM Cognos Analytics or other visualization tools available in Watson Studio to monitor customer churn in real-time. Visualize key metrics and insights from your model.

9    Automation and Alerts:

Set up automated alerts and notifications within IBM Cloud to trigger actions when the churn prediction model identifies customers at risk of churning.

10   Documentation and Presentation:

Document your entire project, including data preprocessing steps, model development, deployment procedures, and key findings.

11   Testing and Validation:

Test the deployed solution with new data or simulate scenarios to validate its effectiveness.

12   Scaling and Optimization:

If needed, explore how the solution can be scaled for handling larger datasets or more complex use cases. Optimize for performance and cost-efficiency.

13   Conclusion and Future Work:

Summarize your findings, the impact of your solution, and any recommendations for further improvement or expansion of the project.


**Optional Enhancements:**

➢ Implement automated email notifications to alert customer support teams when a high-value customer is at risk of churning.
➢ Integrate natural language processing (NLP) to analyze customer feedback and reviews for additional insights.
➢ Explore advanced machine learning techniques such as deep learning for improved churn prediction.


# OVERVIEW:

● Credit card companies monitor every purchase their customers make and can identify fraudulent purchases with a high degree of accuracy using rules derived by processing billions of transactions.

● Mobile phone companies analyze subscribers' calling patterns to determine, for example, whether a caller's frequent contacts are on a rival network. If that rival network is offering an attractive promotion that might cause the subscriber to defect, the mobile phone company can proactively offer the subscriber an incentive to remain in her contract.

● For companies such as LinkedIn and Facebook, data itself is their primary product. The valuations of these companies are heavily derived from the data they gather and host, which contains more and more intrinsic value as the data grows.

## CHARACTERISTICS:

● Huge volume of data: Rather than thousands or millions of rows, Big Data can be billions of rows and millions of columns.

● Complexity of data types and structures: Big Data reflects the variety of new data sources, formats, and structures, including digital traces being left on the web and other digital repositories for subsequent analysis.

● Speed of new data creation and growth: Big Data can describe high velocity data, with rapid data ingestion and near real time analysis.

## BIG DATA ANALYTICS TOOLS:

- Hadoop - helps in storing and analyzing data
- MongoDB - used on datasets that change frequently
- Talend - used for data integration and management
- Cassandra - a distributed database used to handle chunks of data
- Spark - used for real-time processing and analyzing large amounts of data
- STORM - an open-source real-time computational system
- Kafka - a distributed streaming platform that is used for fault-tolerant storage
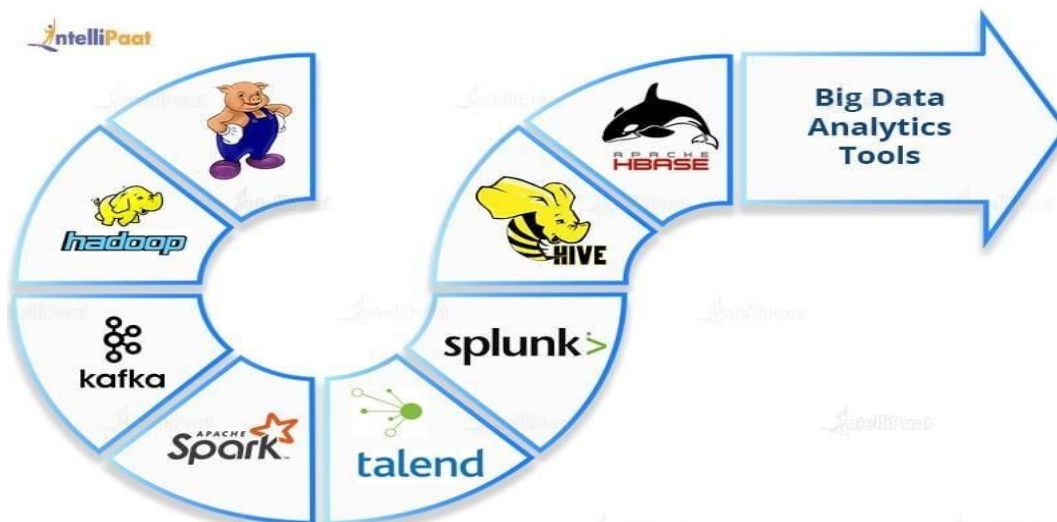


## USES AND EXAMPLES OF BIG DATA ANALYTICS:

There are many different ways that Big Data analytics can be used in order to improve businesses and organizations. Here are some examples:

- Using analytics to understand customer behavior in order to optimize the customer experience
- Predicting future trends in order to make [better business decisions](#)
- Improving marketing campaigns by understanding what works and what doesn't
- Increasing operational efficiency by understanding where bottlenecks are and how to fix them
- Detecting fraud and other forms of misuse sooner

## HISTORY OF BIG DATA ANALYTICS:

The history of Big Data analytics can be traced back to the early days of computing, when organizations first began using computers to store and analyze large amounts of data. However, it was not until the late 1990s and early 2000s that Big Data analytics really began to take off, as organizations increasingly turned to computers to help them make sense of the rapidly growing volumes of data being generated by their businesses.

## BENEFITS AND ADVANTAGES OF BIG DATA ANALYTICS

➢ Risk Management

➢ Product Development and Innovations

➢ Quicker and Better Decision Making Within Organizations

➢ Improve Customer Experience

## DISADVANTAGE OF BIG DATA ANALYTICS:

➢ A talent gap

➢ Security hazard

➢ Adherence

➢ High Cost

➢ Data quality

## CONCLUSION:

Finally, we can say using Big Data Analytics Examples we can add a big value to various sectors and business, where we can easily find out the result of any complex query simply from a massive data set, also can predict the future analysis which will help to take more accurate business decisions.

By completing this project, you'll not only gain valuable experience in big data analytics and machine learning but also create a practical solution that can benefit businesses by reducing customer churn and increasing customer retention.

## REFERENCE:

1."Learning Spark: Lightning-Fast Big Data Analysis" by Holden Karau

2."Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data" by EMC Education Services

3."Big Data: Does Size Matter?" by Timandra Harkness