

Project Topic :- Hadoop-Based Business Intelligence Solution

Dataset :- Customer Loan Data

Purpose: The project aims to design and implement a scalable Business Intelligence (BI) solution using the Hadoop ecosystem. The goal is to process large-scale datasets efficiently and transform them into actionable insights to support data-driven decision-making, dashboards, and machine learning (ML) applications.

Insights from the Data: The project will extract insights such as:

1. Loan Distribution Analysis:

Total loan amounts distributed across different regions or customer demographics
Identification of high-performing regions with the most loans disbursed.

2. Customer Segmentation:

Categorization of customers based on loan amounts, repayment status, or credit scores.
Identification of high-risk customers with delayed or defaulted repayments.

3. Loan Repayment Trends:

Average repayment times segmented by loan amount or customer type.
Insights into repayment delays and their frequency over time.

4. Risk Assessment:

Analysis of default rates by loan type, amount, or customer segment.
Predictive modeling to flag customers at risk of default.

5. Revenue Growth Opportunities:

Identification of underserved regions with potential for loan disbursement.
Trends in loan applications over time to forecast demand.

6. Operational Efficiency Metrics:

Monitoring loan processing times to identify bottlenecks.
Optimization opportunities in customer approval pipelines.

9. Demographic-Based Insights:

Trends in loans approved by age, income bracket, or occupation.
Popular loan types among specific customer groups.

Architecture:

- **HDFS:** Distributed storage for raw and processed data.
- **Apache Spark:** In-memory processing for transformations and aggregations.
- **Apache Hive:** Data warehousing and ad-hoc querying.
- **YARN:** Resource management and job scheduling.
- **Hue/Power BI/Tableau:** Visualization tools for insights.