

BIG DATA TECHNOLOGIES

Presented by: Taibat Nofiu



OVERVIEW

01

INTRODUCTION

02

STORAGE TECH

03

ANALYSIS & QUERY

04

ML & AI FRAMEWORK

05

VISUALIZATION TOOLS

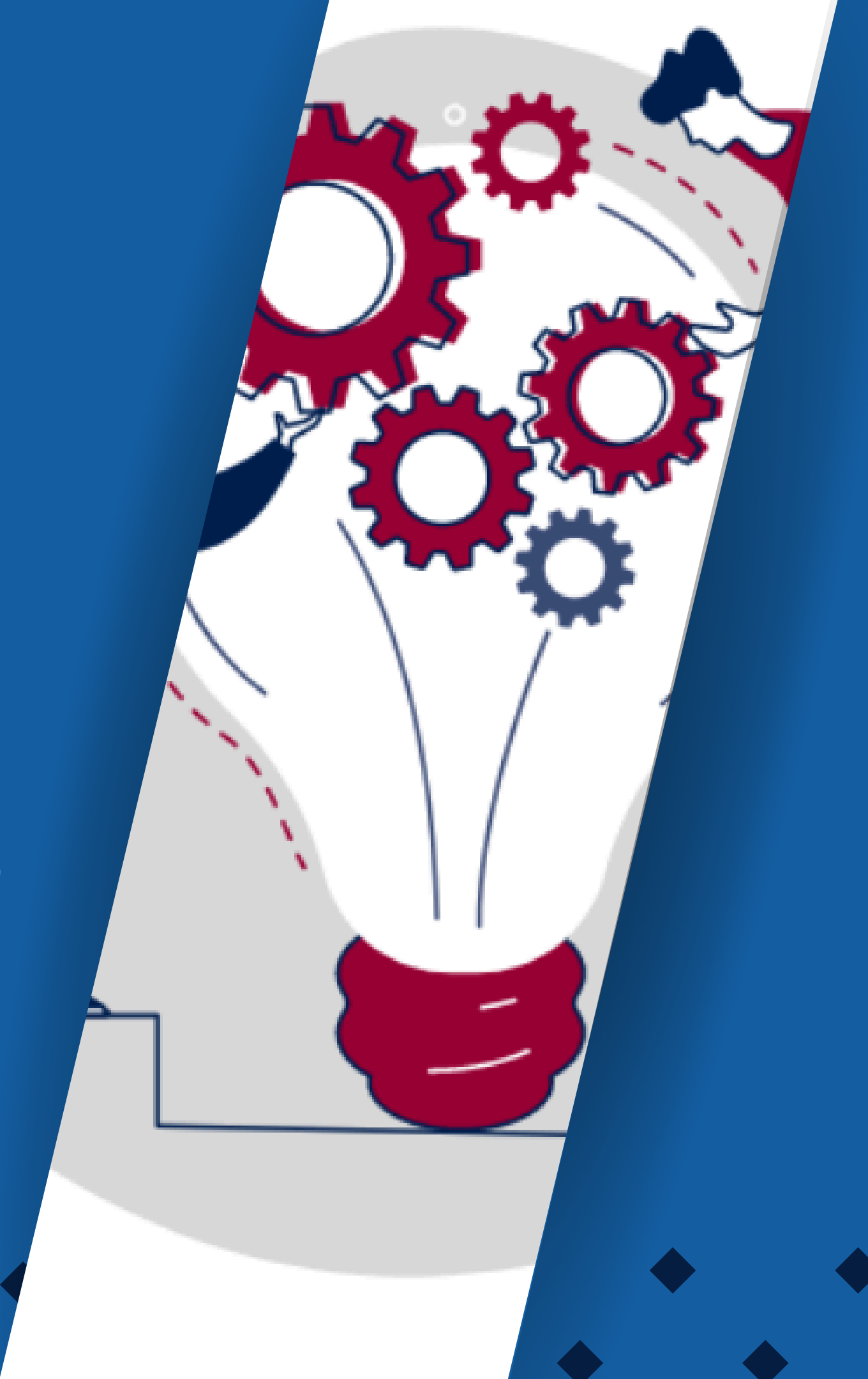
06

THANK YOU

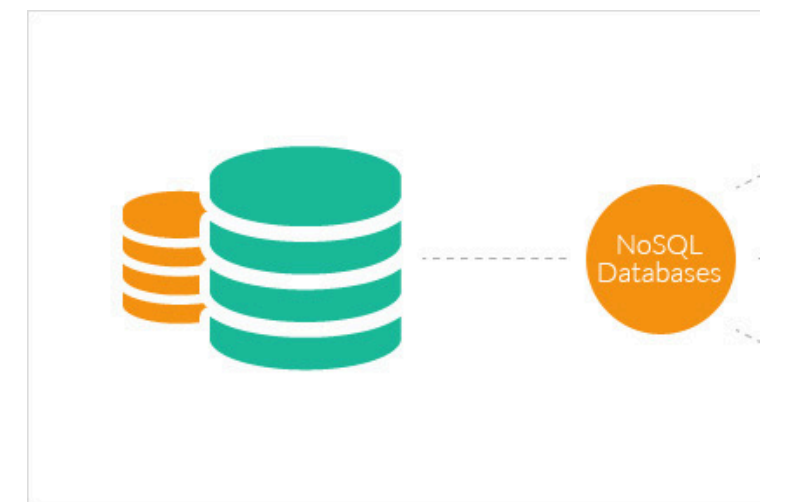


INTRODUCTION

Big Data technologies refer to tools, frameworks, and platforms designed to process, analyze, and manage large and complex datasets.



STORAGE TECHNOLOGIES



DATA ANALYSIS & QUERY ENGINES

Apache Hive

an open-source data warehouse software that enables querying and managing large datasets stored in distributed storage like Hadoop Distributed File System (HDFS).

Presto

Presto is an open-source, distributed SQL query engine designed for running interactive analytics on large datasets.

Elasticsearch

highly versatile, designed for full-text search, structured search, and data analytics across a wide variety of use cases.

ClickHouse

ClickHouse is an open-source, columnar database management system designed for high-performance analytical queries.

ML & AI FRAMEWORKS

This tools apply machine learning algorithms to big data.



Tensorflow & Pytorch

Both frameworks support distributed training across multiple GPUs or nodes, enabling them to handle Big Data workloads effectively.



Apache Mahout

It integrates seamlessly with Apache Hadoop and Apache Spark, making it highly suitable for distributed and parallel processing.



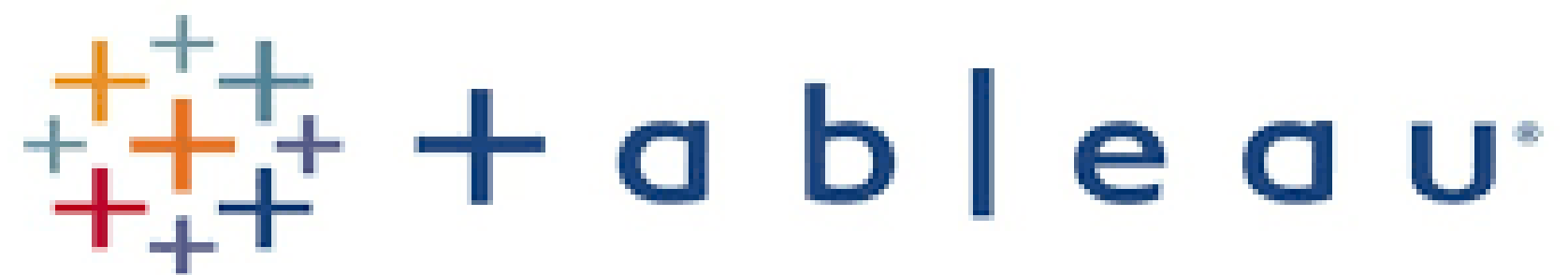
MLlib

It provides a wide range of algorithms and tools for implementing machine learning workflows on Big Data.

DATA VISUALIZATION TOOLS



Power BI



THANK
YOU

