Project Design Phase-II
Technology Stack (Architecture & Stack)

Date	22 October 2023
Team ID	NM2023TMID07292
Project Name	Project – The future of work :Data analytics of glass door jobs
Maximum Marks	4 Marks

Technical Architecture:

The Deliverable shall include the architectural diagram as below and the information as per the table1 & table 2

Example: Order processing during pandemics for offline mode

Reference: https://developer.ibm.com/patterns/ai-powered-backend-system-for-order-processing-during-pandemics/

Guidelines:

1. Include all the processes (As an application logic / Technology Block)
2. Provide infrastructural demarcation (Local / Cloud)
3. Indicate external interfaces (third party API's etc.)
4. Indicate Data Storage components / services
5. Indicate interface to machine learning models (if applicable)

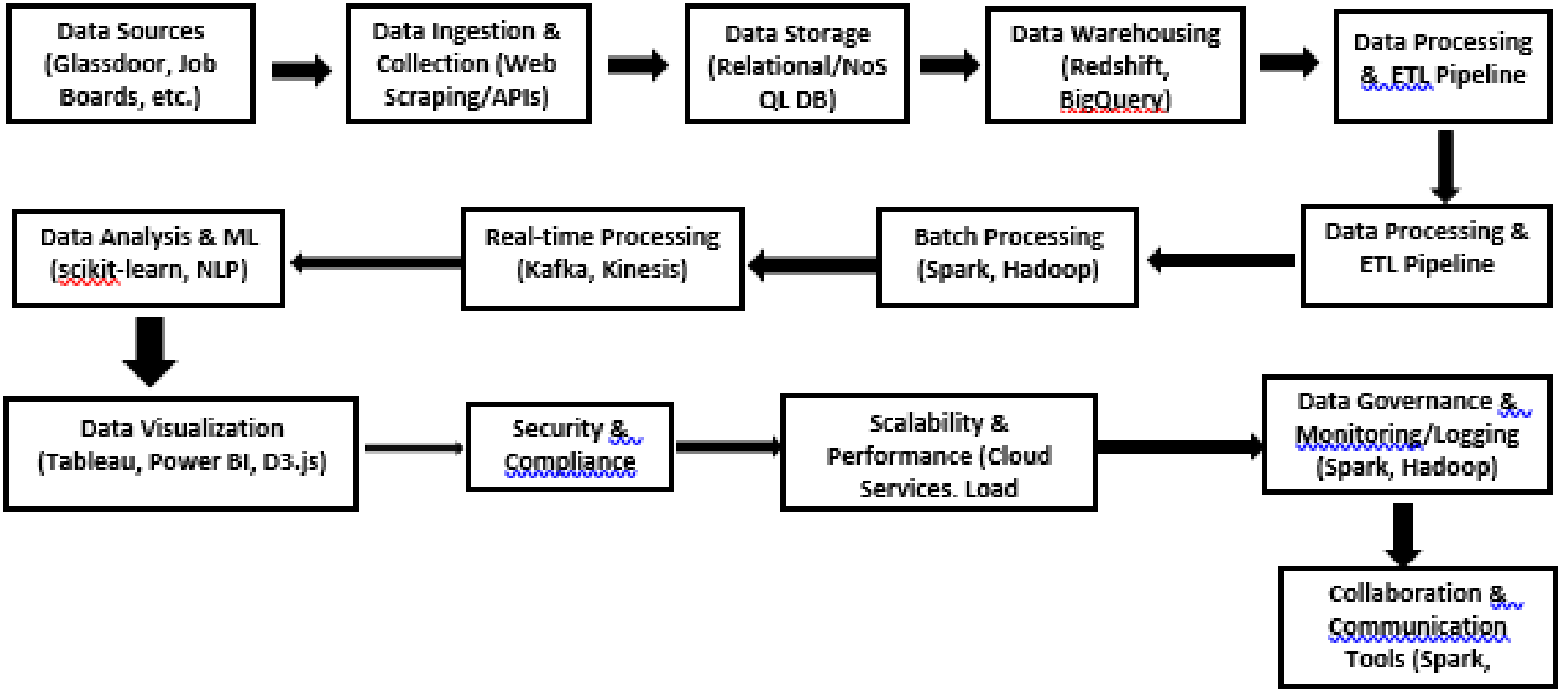
Data Sources (Glassdoor, Job Boards, etc.)
Data Ingestion & Collection (Web Scraping/APIs)
Data Storage (Relational/NoS QL DB)
Data Warehousing (Redshift, BigQuery)
Data Processing & ETL Pipeline
Data Processing & ETL Pipeline
Batch Processing (Spark, Hadoop)
Real-time Processing (Kafka, Kinesis)
Data Analysis & ML (scikit-learn, NLP)
Data Visualization (Tableau, Power BI, D3.js)
Security & Compliance
Scalability & Performance (Cloud Services. Load
Data Governance & Monitoring/Logging (Spark, Hadoop)
Collaboration & Communication Tools (Spark,

Table-1 : Components & Technologies:

S.No	Component	Description	Technology
1.	User Interface	users can interact with the application, view analyses	React Js ,D3.js
2.	Application Logic-1	Handles user requests, processes data, and returns results to the UI.	Python(flask/Django)
3.	Application Logic-2	manage more specialized processes like data transformations, advanced analytics	Java (using Spring Boot) or Node.js
4.	Application Logic-3	integration and orchestration of machine learning models.	Python (especially with libraries such as TensorFlow or Scikit-learn).
5.	Database	Stores structured data, e.g., job listings, company profiles, user data.	PostgreSQL
6.	Cloud Database	A scalable database solution in the cloud, ensuring accessibility and scalability.	Amazon RDS
7.	File Storage	Storage of larger files, e.g., resumes, company logos, or larger datasets.	Amazon S3
8.	External API-1	Integration with Glassdoor for job data extraction.	RESTful API calls using Python's **requests** library or JavaScript's **axios**.
9.	External API-2	integration with a service providing supplementary data	RESTful API.
10.	Machine Learning Model	Analyzes and predicts job trends based on historical data.	Python with TensorFlow or Scikit-learn
11.	Infrastructure (Server / Cloud)	application is hosted, ensuring uptime, scalability, and performance	Amazon Web Services (AWS),Microsoft Azure.

Table-2: Application Characteristics:

S.No	Characteristics	Description	Technology
1.	Open-Source Frameworks	List the open-source frameworks used	Python pandas numpy
2.	Security Implementations	List all the security / access controls implemented,use of firewalls etc.	SSL/TLS
3.	Scalable Architecture	Justify the scalability of architecture (3 – tier, Micro-services)	Amazon Redshift, Google BigQuery, or Snowflake

S.No	Characteristics	Description	Technology
4.	Availability	Justify the availability of application (e.g. use of load balancers, distributed servers etc.)	Python's BeautifulSoup or Scrapy
5.	Performance	Design consideration for the performance of the application (number of requests per sec, use of Cache, use of CDN's) etc.	Tableau,powerbi

References:

https://c4model.com/

https://developer.ibm.com/patterns/online-order-processing-system-during-pandemic/

https://www.ibm.com/cloud/architecture

https://aws.amazon.com/architecture

https://medium.com/the-internal-startup/how-to-draw-useful-technical-architecture-diagrams-2d20c9fda90d