

## Solution Architecture

Date	30 oct 2023
Team id	NM2023TMID05062
Project name	Data driven insights on Olympic sports participation and performance
Maximum marks	4 marks

### Solution architecture :



#### 1. Data Sources:

- Collect the data from the kaggle website.

#### 2. Data Ingestion:

- Use data ingestion tools to gather and integrate data from diverse sources into a centralized data repository. Tools in IBM cognoss custom scripts can be used for this purpose.

#### 3. Data Storage:

- Store the integrated data in IBM cloud storage

#### 4. Data Processing:

- Perform data cleansing, transformation, and enrichment to ensure data quality and consistency. Utilize ETL (Extract, Transform, Load) processes in IBM Cognos analytics.

#### 5. Data Modeling:

- Create a data model that defines how data entities are related and organized. Use a schema or a NoSQL database like MongoDB, depending on the data structure.

#### 6. Analytics Engine:

- Set up an analytics engine for running complex queries, statistical analyses, and machine learning models. Tools like Python with libraries such as Pandas, Scikit-Learn, and TensorFlow can be used for this purpose.

#### 7. Visualization:

- Develop a user-friendly interface for data visualization and exploration. Tools like Tableau, Power BI, or custom web applications using D3.js and React can help create interactive dashboards.

#### 8. Machine Learning and Predictive Modeling:

- Implement machine learning algorithms to derive insights and predictions. Techniques like regression, classification, clustering, and time series analysis can be applied.

#### 9. Security and Compliance:

- Ensure data security and compliance with relevant regulations, such as GDPR or HIPAA. Implement authentication and authorization mechanisms to protect sensitive data.

#### 10. Scalability:

- Design the architecture to scale as data volumes grow. Utilize cloud-based infrastructure for elasticity and scalability as needed.

#### 11. Monitoring and Logging:

- Implement monitoring and logging systems to track system performance, detect issues, and troubleshoot.

#### 12. Reporting and Insights:

- Generate automated reports and alerts to keep stakeholders informed about the latest insights.

Use scheduled tasks or notification services.

#### 13. Documentation:

- Maintain comprehensive documentation of the architecture, data sources, processes, and analytics models for future reference and on boarding of new team members.

#### 14. Continuous Improvement:

- Regularly update the data, analysis models, and visualizations to reflect changes over time and improve the accuracy of insights.

#### 15. Disaster Recovery and Backup:

- Implement a disaster recovery plan to ensure data integrity and availability in case of unexpected events or data loss.