Project Design Phase-I
Solution Architecture

Date	1 November 2023
Team ID	Team-591161
Project Name	Data-driven insights on Olympic sports for performance and participation
Maximum Marks	4 Marks

Designing a solution architecture for a project focused on data-driven insights on Olympic sports participation and performance involves several key components. Here is a high-level overview of the solution architecture for this project

Solution Architecture Diagram:

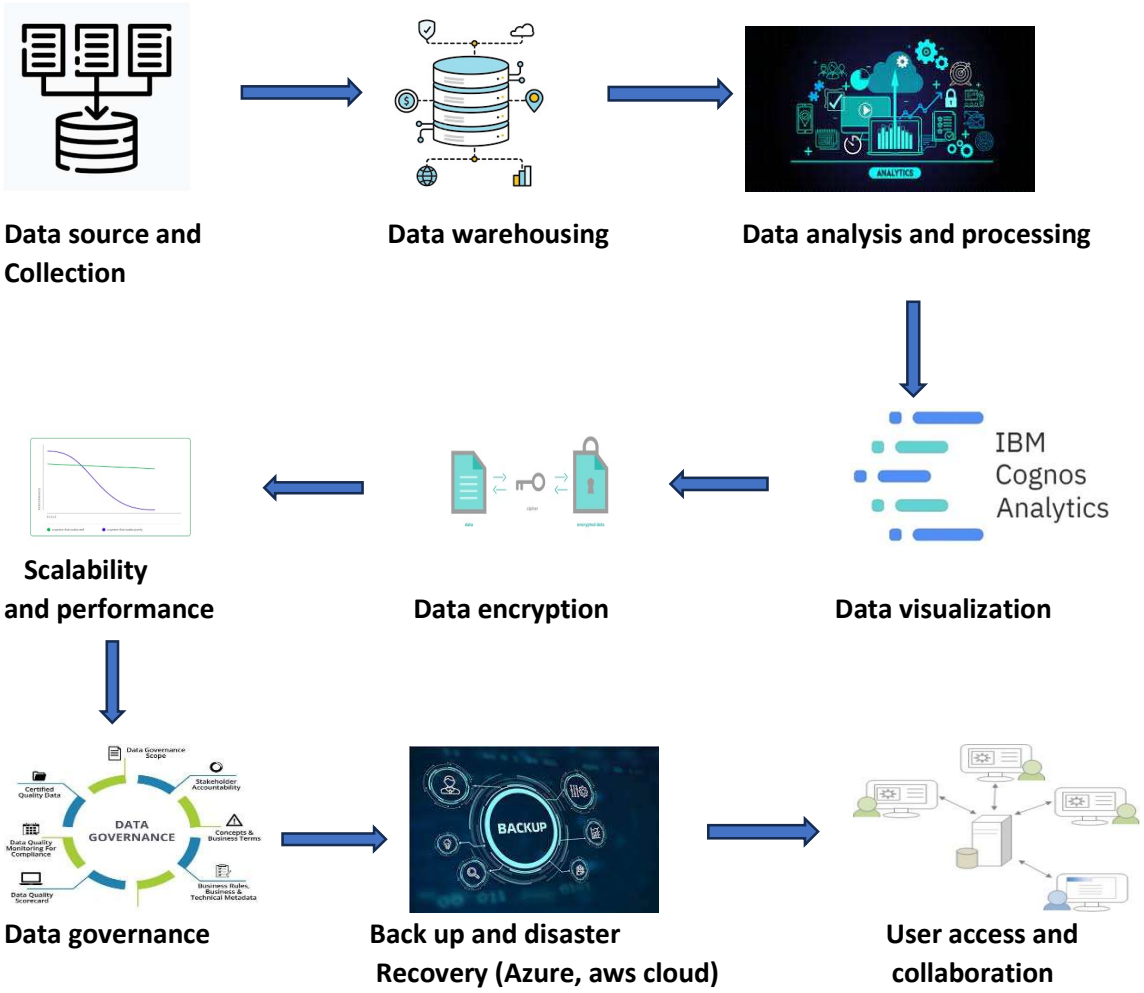

Figure 1: Architecture and data flow for the Data-driven insights on Olympic sports for performance and participation

Data Source and Collection:

* Olympic Organizing Committees: Historical data from previous Olympic Games, including participant information, results, and event details.
* National Olympic Committees: Data on athletes and their backgrounds, participation, and performance.
* ETL (Extract, Transform, Load) processes to collect and consolidate data from various sources.

Data Warehousing:

* Data Warehouses: Centralized storage for structured data, including athlete profiles, medal counts, and event histories.
* Data Lakes: Storage for unstructured and semi-structured data, such as media coverage, social media sentiment, and historical documents.

Data Analysis and Processing:

* Data Analysis Tools: Implement tools and frameworks like Python, R, and Jupyter notebooks for exploratory data analysis.
* Big Data Technologies: Use platforms like Hadoop and Spark for processing large volumes of data.
* Machine Learning and AI: Implement algorithms for predictive analytics, sentiment analysis, and performance modelling.

Data Visualization:

* Dashboard Creation: Use visualization tools like Tableau, Power BI, or custom web-based dashboards to present insights in a user-friendly manner.
* Interactive Charts and Maps: Display trends, patterns, and geographical insights for a comprehensive view of the data.

Data Encryption:

* Implement data encryption, access control, and data anonymization techniques to protect sensitive information.
* Ensure compliance with data privacy regulations, especially when dealing with athlete profiles and personal data.

Scalability and Performance:

* Choose a scalable infrastructure that can handle data growth over time and during peak usage (e.g., during Olympic events).
* Implement load balancing and data replication for performance optimization.

Data Governance:

* Define data ownership and stewardship roles to maintain data quality and integrity.
* Implement version control and data lineage tracking for data governance.

Backup and Disaster recovery (Azure, AWS Cloud):

* Implement regular backups and disaster recovery plans to ensure data availability and business continuity.
* Choose between cloud-based solutions (e.g., AWS, Azure, GCP) for scalability and ease of maintenance or on-premises infrastructure for data control and compliance.

User access and Collaboration:

* Provide role-based access to the data and insights for different stakeholders, including researchers, analysts, and IOC officials.
* Enable collaboration and knowledge sharing among users.