

Data Engineering

Question 1:

Use case 1:

Student academic performance analysis

Analyze a student's performance based on their previous academic performance, highest education, gender and studied credits. The goal of this use case is to analyze if these features have an impact on a student's academic performance.

Use case 2:

Student region detection analysis

Analyze if a student's region has an impact on several features like: student assessment performance, the courses they are following (is a region known for/provides better lectures for certain courses), highest education among other features.

Question 2:

This data architecture is applicable to both use cases mentioned above.

Data Sources and Data Ingestion

- Student information datasets: contains demographics, enrollment details and academic records.
- Extracurricular activities: storing information about a student's participation in extracurricular activities
- Assessment and exam evaluations: stores scores and performance data for each assessment and exam.

ETL Design Pattern

- Gather data from various sources, transform the data and ensure that it fits the data model, then proceed to load the data

Operational Database

- Centralized repository for storing data containing student information

- Make use of database management to create structured database storage environment
- Analysis database and operational database are not tied. The analysis database is brought up to date around every month.

Data Security and Privacy

- Implement encryption on data to ensure that the data is safely stored and kept.
- Ensure that no sensitive data is being accessed by unqualified personnel.
- Make sure that privacy regulations are being followed and respected.

Scalability and Performance

- Implement scalable data processing frameworks to handle large loads of data and datasets
- Ensure the use of distributed data storage to enhance data availability, scalability and fault tolerance