B9DA102
Data Storage Solutions for Data Analytics

CA1 Assessment Tasks

Part 1 - Business Drivers

Select a subject for analysis using data warehousing techniques from the operational data captured in the Microsoft Northwind database. (The operational ER model is included in this document)

Document your reasons for selecting this subject area, identify key stakeholders, formalise the vision and goals of the data warehouse.

[10 Marks]

Part 2 - Data Modelling

Develop and present a suitable schema for the data warehouse (data mart). Discuss your reasons for the design.

[20 Marks]

Part 3 - Implement Tables and ETL Procedures

Using Microsoft SQL Server, implement your tables and extract, transform, and load data from the operational Northwind database into the data warehouse (datamart). This can be done using any available open source tool such as Talend and/or SQL

[15 Marks]

What would be a motivation for using the Microsoft SQL Server DW solution? Discuss the Microsoft solution and compare with a competitive data warehousing solution from another vendor

[5 Marks]

Part 4 - Reporting and Analysis

Produce reports and analyses in support of the requirements outlined in Part 1. You can use any suitable BI/Visualisation tools including SSDT, Microsoft PowerBI, or Tableau. Explain your selection of tools.

[20 Marks]

Part 5 - Modelling and Graph Data Models

Discuss the limitations of the relational ER and star schema dimensional models. Implement the Northwind ER model as a graph using Neo4j.

[15 Marks]

Part 6 - Graph information retrieval and analysis

Evaluate the use of Neo4j and Cypher Query language for information retrieval of the Northwind data. Discuss the role of graph databases in the Hadoop-Spark Ecosystem.

[15 Marks]

[Total: 100 Marks]