CMPU4003 Advanced Databases

Continuous Assessment

Part I (Deadline Thursday 9th November 2023 @ 14.00)

Setup

- A description of the initial dataset you need to use is available in a companion document named CMPU4003 CA Data Description.pdf.

- Download the script MusicComp.SQL from Brightspace.

- Run this against in your MariaDB instance. It will create and populate a new database called MusicCompDB.

Tasks

- MariaDB:

  - Design a star dimensional model to facilitate answering a number of questions related to viewers' activity. The questions are available in the companion document CMPU4003 CA Part I Questions.pdf.

  - Create a database based on this.

  - Build an ETL pipeline to populate this new database with data from the original database undertaking any data transformations needed (you can use SQL or Python to achieve this). You must implement staging as part of this pipeline.

  - Write the queries needed to answer the questions using this new database.

  - Identify and implement appropriate indexes to improve query performance.

  - Write the SQL needed to generate evidence about the performance of your queries pre and post the index creation.

- CouchDB:

  - Design a document model version of your dimensional model in CouchDB.

  - Build a pipeline to populate this database with data from your MariaDB database. This does not require staging.

  - Implement a simple query in CouchDB.

What you need to submit
(before the lab class on Thursday 9th November @ 14.00):

- SQL to create a database in MariaDB based on the dimensional model you derived.

- Code for a pipeline to populate this database from the original data (this can be SQL or Python).

- SQL to implement the queries to answer the questions posed.

- SQL to implement any indexes introduced to improve query performance.

- SQL needed to demonstrate the performance pre and post implementation of any indexes.

- Code needed to create and populate an equivalent CouchDB database (this can be Python or a shell script).

- Code needed to implement a simple query against the equivalent of the fact table in this CouchDB.

What you need to demonstrate:
(in the lab class on Thursday 9th November @ 14.00):

- That you have created and populated a database based on a dimensional model in MariaDB.

- That you can execute at least one of the required queries against that database.

- That you have implemented at least one index and that it is being used.

- That you have created and populated a database in CouchDB which implements the dimensional model.

- That you can execute at least one query against that CouchDB database.

Marking Scheme

This part of the assessment will be marked out of 100 but weighted to 40% of the module marks.

Aspect	Marks From Assessment of Submission		Marks From Demo
Creation of MariaDB database	5		
Population of MariaDB database	15		5
Implementation of Queries in MariaDB	15		3
Implementation of Indexes (including generation of evidence of impact)	15		2
Creation of CouchDB database	10		
Population of CouchDB database	10		3
Implementation of a simple query in CouchDB	10		2
Appropriate levels of comments explaining and justifying choices	5		
Total Marks (100)	**85**		**15**