Coursework Overview: Assignment 1

2019

Music Database Website

Task: *Investigate whether a NoSQL data store is a suitable replacement for your company's relational database. Trials on small subset of the database – the methods used will need to be scalable to work with millions / billions of records.*

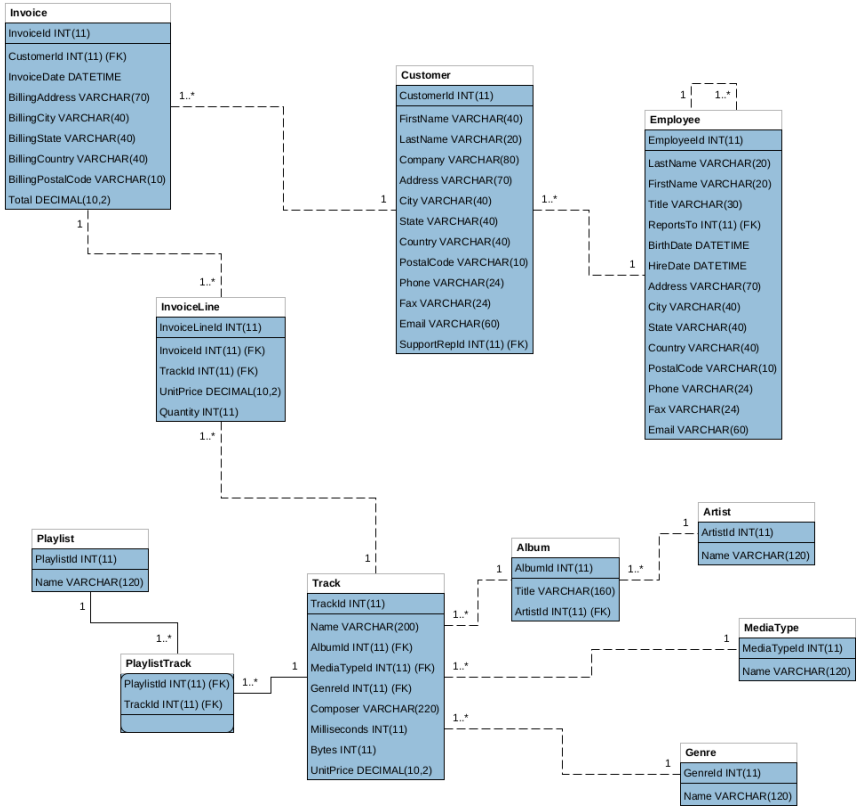
- Tasks
  - Choose 2 NoSQL systems (from the list of MongoDB, Neo4J, Cassandra)
  - Design data models
  - Extract-Transform-Load data into the NoSQL stores
  - Analysis queries over the NoSQL stores
  - Compare the NoSQL solutions and suggest if one would be good to use
- Work in Groups of 3
- Submission through Vision
  - 1 report submission per group as a **PDF**
  - **Deadline: Thu 14th Feb 2019 - 11am (GMT) (Week 6)**
  - Ensure you are in the correct Vision group Edinburgh/Dubai / EdinburghUG/DubaiUG

Dataset: Chinook Music Download Database

- Data supplied as SQL statements for mySQL
- Load the data on to mySQL server (MACS or VM)

SAMPLE PROVIDED

Album:	347
Artist:	275
Customer:	59
Employee:	8
Genre:	25
Invoice:	412
InvoiceLine:	2240
MediaType:	5
Playlist:	18
PlaylistTrack:	8715
Track:	3503



Report

All	**Data Modelling** • Description of the 2 NoSQL systems chosen. Details of the data storage paradigm and consistency/replication model for each of the systems (4) • Presentation of the data model in JSON schema (or alternative) together with a discussion of the resulting data model including details of data types and relationships, and suitability for the music website task(6)	(10)
F21BD	**Comparison of NoSQL Systems** Provide a comparison of the 4 different types of NoSQL system and their suitability for the given task of a music download website. Compare the performance/limitations of 2 systems you implemented for this task.	(10)
All	**Data Extract-Transform-Load** Presentation and explanation of the ETL pipeline developed for the coursework together with evidence that data has been loaded into each of the NoSQL systems.	(10)
F21BD	**Scalability of approach** Discuss the limitations of your approach with respect to its scalability for each of the systems implemented.	(5)
All	**5 Analysis Queries (implemented for each NoSQL system - 1 mark per query)** Describe the analysis queries that you have run over the dataset. Provide an English statement of the intent of the query, the query for the system that you have used, and the results of running the query. Queries must show a variety of language features. Implement the same queries on each NoSQL solution (i.e. 5 queries but 2 implementations).	(10)

Please Do Not:

Collude with other groups:
Examples:

- Showing your work to a student from another group
- Giving a copy of your work to another group
- Work together with another group

- Discussing ideas and then implementing them yourself is fine

Plagiarise: taking the ideas, writings or inventions of another person and using these as if they were your own, whether intentionally or not

- Cite your sources
- Provide a reference list of cited articles