Emilija Dimikj

March 27, 2020

BIDD 210 A

Final Project

Lessons Learned from taking the course “Relational and Non-Relational Databases”

In this course we covered fundamentals of normalized and non-normalized databases, learned about data warehousing, relational data modeling (star schema, and snowflake schema), scratched the surface of data extraction and transformation and covered OLAP multidimensional and tabular models and document database modeling.

It was interesting to learn the differences between relational OLTP databases and OLAP data warehouses, design and implement star schema data warehouse.

Key Differences Between OLTP and OLAP

|  |  |
| --- | --- |
| OLTP | OLAP |
| Online transaction system and manages database modification | Online data retrieval and analysis system |
| Insert, Update, Delete information from the database | Extract data for analyzing that helps in decision making |
| OLTP and its transactions are the original source of data | Different OLTPs database becomes the source of data for OLAP |
| Short but frequent transactions | Long and less frequent transactions |
|  |  |
| Normalized database | May not be normalized (star, snowflake schema) |
| Maintains data integrity constraint | Data integrity is not affected |

As far as choosing between Multidimensional and Tabular solution, it’s better to conduct a detailed study based on project requirements.

|  |
| --- |
| Why tabular solution can be chosen:   * easier for understanding and creating the model * works quicker than multidimensional cubes for queries based on columns * hardware, such as disks, is not important. However, tabular is a memory dependent solution and more memory will ensure better performance * more efficient data compression about one-tenth of the size, whereas the compression of multidimensional is a third of the size of the original database |
| Why multidimensional solution can be chosen:   * works better with a large amount of data (terabytes), it’s better to go with the multidimensional database. If your database requires more than 5 terabytes, multidimensional is the only option * performs better in terms of scalability * some features, such as aggregations or actions, are supported in the multidimensional model only |

Grasping the concepts of document-oriented database was challenging because required shift of my thinking approach from relational logic to concept with no internal structure and relationships. A key difference between the document-oriented and relational models is that the data formats are not predefined in the document case. In most cases, any sort of document can be stored in any database, and those documents can change in type and form at any time.

Key Differences between SQL and NoSQL

|  |  |
| --- | --- |
| SQL | No-SQL |
| Vertically Scallable | Horizontally Scalable |
| Fixed or predefined schema | Dynamic schema |
| Not suitable for hierarchical data storage | Best suitable for hierarchical data storage |
| Can be used for complex queries | Not good for complex queries |

There might be different preferences, distinct requirements to prefer one over some other database. Figuring out the requirements and taking into consideration all strengths and weaknesses of these database types is a key for project success.

## What went well

The class was taught extremely well and provided grateful information. The material was extremely interesting. I had the required basic skills needed to understand the material and didn’t found the material difficult. I’ve learned a lot for a short time and gained hands-on experience while working on the assignments.

## What didn’t go well

It would’ve been nice to get different kind of assignment requirements, like description of the assignment and expected result without step-by-step instructions. At times, it was like class and learning material repetition.

Final project opened a question about the best way of implementing ETL process. I am looking forward to the next semester and learning all about ETL Processing.