Mi Gao

migao@bu.edu

Term Project Update # 2

MET CS 779

MET CS 779 Term Project Update # 1

Table of Contents

[1. Introduction 2](#_Toc46633337)

[2. Self Questions 2](#_Toc46633338)

[1) What is the proof of concept component in your term project? 2](#_Toc46633339)

[2) What data are you looking to use specifically? 2](#_Toc46633340)

[3. Conclusion 2](#_Toc46633341)

[5. Revision History 4](#_Toc46633342)

# Introduction

After learning from the course which I took from last Saturday about the distributed database, I have learned more knowledge such as the distributed model, implementing fragmentation, replication, and other things; they are fantastic. I am going to answer questions which I did not reply last time.

# Self Questions

## What is the proof of concept component in your term project?

According to M.Tamer Ozsu and Patrick Valduriez stated in Principles of Distributed Database Systems (1999), we define a distributed database as a group of logically interconnected databases distributed on a computer network. A distributed database management system (distributed DBMS) is a software system that supports the management of distributed databases, which makes distribution more important to users. Transparency. Sometimes, Distributed Database System (DDBS) is used to refer to both distributed database and distributed DBMS. In the above expression, a group of distributed on the network and logically interconnected is the essence. A group of logically interconnected databases can be distributed on one or more physical nodes. Of course, it is mainly applied to multiple physical nodes. On the one hand, it is related to the improvement of server cost performance. On the other hand, because the development of the Internet has brought high concurrency and massive data processing needs, the original single physical server node is not enough to meet this demand.

The main points for me I think should be the implementation of data storage no longer limited by the capacity of a single disk, and also the horizontal and vertical fragmentation of the original database. Thus, they should make my database to be improved a lot on the extensibility and response speed at least.

## What data are you looking to use specifically?

Currently, I am still considering about what data I should start with. I have two source can be useful. One of it is my food express system project did in CS 579 (<https://github.com/gaomigithub/cs579_database_management>); I made few trigger functions and views for basic functionality. And the other one is the data of NetFlix which we currently work with; I am keep maintaining its feature through the assignment requests although it is really tough for me so far. I may still need one to support the requirement which listing of at least three potential sources for the term project report.

Moreover, the DBMS I most familiar with is PostgreSQL. However, MongoDB looks is a great choice and I saw a lot of article talked about it; In contrast, the visibility of technical articles related to postgresql is far less than it. Therefore, I am weighing factors such as the learning time cost of applying those two DBMS.

# Conclusion

Distributed or cloud-based database design and implementation proof of concept are really great but heavy target if I really set them to be my term project. I doubt and consider about the completion caused by my limited ability and time after I have actually spent much time on reading more documents and references since I was done the proposal last week. I think I may need some advice or hints from the professor to help me see the possibility of completing this work without the cost of performance. Otherwise, I may change the target on this term project.

# 5. Revision History

A history of things you added and why, not required but nice to have.

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Date** | **Version** | **Description** |
| Mi Gao | 07/26/19 | 1.0 | Initial Document Creation |
|  |  |  |  |