Nina Nguyen

November 5, 2019

CPSC 5890

**Progress Report**

**Title:** Scalable Architecture for Massively Parallel Data Processing

My initial plan is still on track, I found a lot of articles relevant to my topic. In my paper, I want to start off with an introduction on Big Data and some current architecture. Then base on the articles, mention some of the problems that people are facing with managing big data. Finally, propose some solutions for those problems.

I have identified some interesting articles (below), and will need to connect them all. I am still in the process of finding 2 more articles and well as reading and taking notes on the papers. My timeline might be pushed back a few days due to work load from other classes but I should still be able to complete the project and the presentation on time.

**Research Papers:**

[**Challenges for MapReduce in Big Data**](https://ieeexplore-ieee-org.proxy.seattleu.edu/document/6903263)

* This will give a good overview of Big Data as well as MapReduce

[**Big Data: Framework and Issues**](https://ieeexplore-ieee-org.proxy.seattleu.edu/document/7519647)

[**BigData Oriented Open Scalable Relational Data Model**](https://ieeexplore-ieee-org.proxy.seattleu.edu/document/6906808)

* This will give some overview about traditional databases and compare that with Open Scalable Relational Data Model. It has the characteristics to handle big data such as volume, variety and velocity.

[**A Scalable and Robust Framework for Data Stream Ingestion**](https://ieeexplore-ieee-org.proxy.seattleu.edu/document/8622360)

* With increased in data volumes, this will be a good paper to understand how to make data stream ingestion more scalable and robust. It will go hand in hand with Big Data Processing.

[**Near Real-Time Big-Data Processing for Data Driven Applications**](https://ieeexplore-ieee-org.proxy.seattleu.edu/document/8316297)

* This paper talks about Auto-scaling and adjustment platform for cloud-based systems to address problems with data integration and design of data driven application.

[**Oracle parallel RDBMS on massively parallel systems**](https://ieeexplore-ieee-org.proxy.seattleu.edu/document/253071)

* Scalable relational database superservers

[**Big Data Scalability, Methods and its Implications: A Survey of Current Practice**](https://dl-acm-org.proxy.seattleu.edu/citation.cfm?id=2743121)