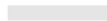# WEEKLY PROGRESS REPORT

Directed Research Projects

Zohaib Ahmad & Waleed Irfan

Student ID : 170902110 & 170406830

ahma2110@mylaurier.ca & irfa6830@mylaurier.ca

Immediate Supervisor : Dr. Safaa Bedawi
Assistant : Professor Sukhjit Singh Sehra

WEEKLY REPORT #2

DIRECTED PROJECTS

Student Names : Zohaib Ahmad & Waleed Irfan
Department: Computer Science (Honors)
Date: January 22,2021
Immediate Supervisor: Dr. Safaa Bedawi
Assistant : Professor Sukhjit Singh Sehra

TASK	STATUS	DURATION
Working on integrating hive ontop of Hadoop	Finishing integrating Hive	1 week

WORK DONE

Second and third meeting have been done for updating supervisors on progress.

1)Installed Hive
2)Got jps command working

Student ID : 170902110 & 170406830

Hive

- Apache hive is a data warehouse software project built on top of apache hadoop for providing data query and analysis.

- It translates input programs written in HiveQL to one or more Java MapReduce, Tez or spark jobs.

- It then proceeds to organize data into tables for Hadoop (HDFS) and runs on a cluster to produce an answer.

- This week both students have gotten jps command working and Hive successfully installed.

- Immediate goals going into next week are to get a well formatted database for queries to execute optimally.

- As this was our first time installing Hive we had looked up a guide online to make sure everything works well with Hadoop.

References
Big Data Analytics with R and Hadoop, By : Prajapati Vignesh eText ISBN : 9781782163299,1782163298

Guide for Installing Apache Hive on Machine
https://towardsdatascience.com/installing-apache-hive-3-1-2-on-windows-10-70669ce79c79

Guide for Loading CSV into Hive
https://www.youtube.com/watch?v=4bdKxq7G5s8