MCDA 5540 Hadoop & Spark Workshop Fall 2018

Instructor: William Wells(wellsww@hotmail.com)

Agenda

9:00am – 10:00am	Intro to Hadoop – Hortonworks HDP
10:00am – 10:10am	Break
10:10am – 11:00am	Apache Spark
11:00am – 11:10am	Break
11:10am – 11:30am	Q&A
11:30am – 12:30pm	Lunch
12:30am – 1:30pm	Spark Exercise
1:30am – 2:00pm	Assignment

Free Training Resources:

https://www.businessnewsdaily.com/10774-free-hadoop-training-online.html4

Free Hadoop Sandboxes:

https://hortonworks.com/downloads/#sandbox

https://www.cloudera.com/downloads.html

https://www.cloudera.com/downloads/quickstart_vms/5-13.html

Hortonworks Tutorials:

https://hortonworks.com/tutorials/

https://hortonworks.com/tutorial/learning-the-ropes-of-the-hortonworks-sandbox/

Hive Tutorial:

https://hortonworks.com/tutorial/how-to-process-data-with-apache-hive/

Cloudera Hadoop Tutorials:

https://ondemand.cloudera.com/

Databricks Tutorials:

https://docs.databricks.com/spark/latest/training/index.html

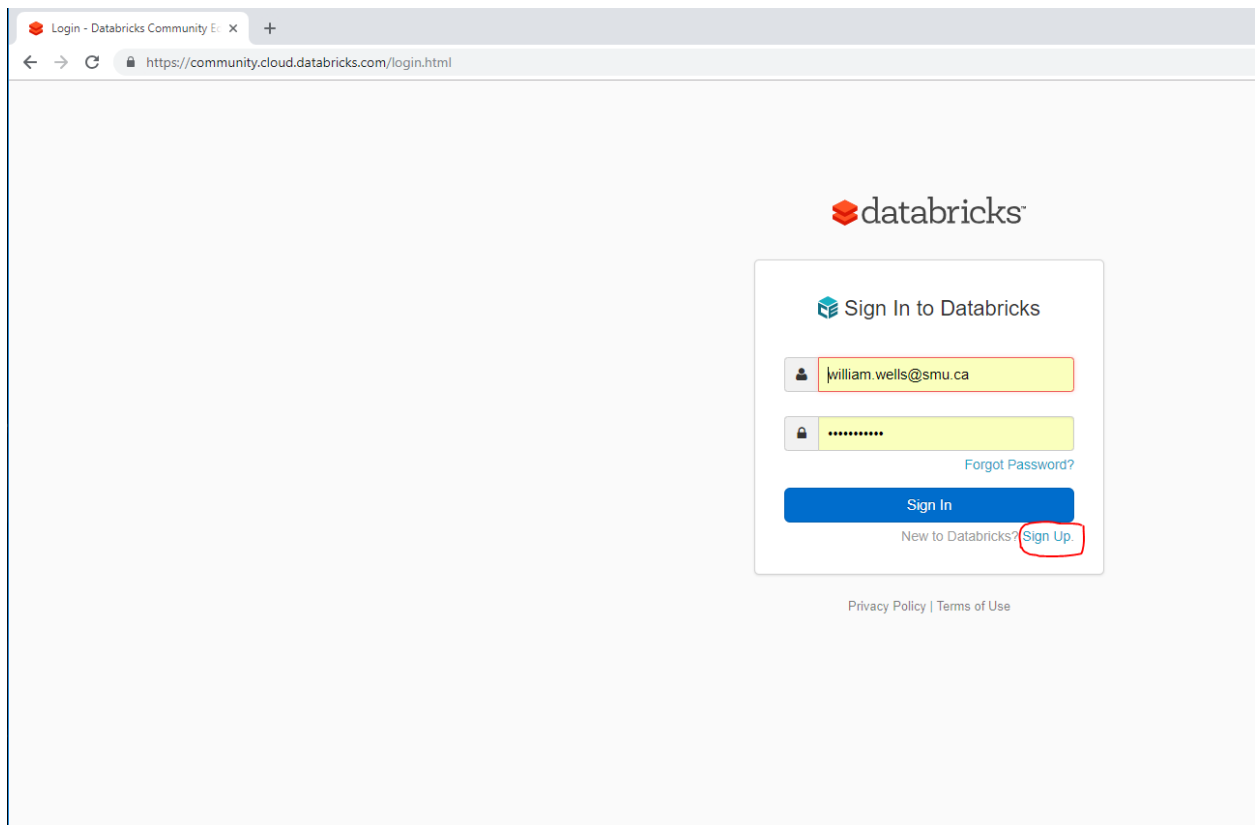# Practical session using Spark

Practical session:

Tasks:

- **1) Querying and analyzing open data source with Apache Spark.**

    - **Steps:**
        - 1) Extract the Data.zip file
        - 2) Setup an account in Data bricks: https://databricks.com/try-databricks.

    Spark exercise steps with the Databricks Community Environment:

- 3) Create a cluster in Databricks.
- 4) Import data files into workspace.
- 5) Step through notebook