**DISCnet Big Data Engineering Course**

**March 2-4, 2020**

**Course Guide**

**Presentations**All the Presentations are available at: https://github.com/julieweeds/BigData/presentations

|  |  |
| --- | --- |
| Presentation | Description |
| 0 | Course Introduction |
| 1 | Big Data Overview and Motivation |
| 2 | Map Reduce |
| 3 | Apache Spark |
| 4 | Intro to SQL |
| 5 | Additional Tools and Spark components |
| 6 | Theory of Big Data Scalability |
| 7 | Data Visualization |
| 8 | Introduction to Machine Learning |
| 9 | NoSQL |
| 10 | Realtime Data Processing |
| 11 | Recap and Conclusions |

**DISCnet Big Data Engineering Course**

**March 2-4, 2020**

**Exercise Guide  
  
All exercises are available at:**

https://github.com/julieweeds/BigData/labs

|  |  |
| --- | --- |
| Exercise | Description |
| 1 | Jupyter, Python and Pandas |
| 2 | Apache Spark and Wordcount |
| 3 | Simple SQL in Spark |
| 4 | Apache Spark on EC2, Flintrock |
| 5 | An undirected Spark problem |
| 6 | Joining datasets and finding correlation |
| 7 | Simple machine learning - clustering with k-means |
| 8 | NoSQL basics - importing data into Cassandra with Spark |