****

**COURSE SYLLABUS**

Data Engineering ( Spring 2020 )

**Instructors**: Kit Menke, Tim Sagona, Will Farrell

**Class Time**: 6:00-9:00 Monday Evening

**Room:** Edward Jones Dome, 3rd Floor CET@CIC

**E-mail**: [wfarrell@1904labs.com](mailto:wfarrell@1904labs.com)

**Course Description:**

The course will teach the fundamentals of Data Engineering, including working with a pub/sub system, building a data pipeline, working with a streaming data framework, and interacting with NoSQL databases.

**Resources:**

All resources are provided by 1904Labs.

Kafka - Tutorial

* <https://www.udemy.com/course/apache-kafka/learn/lecture/11566882#overview>

Spark - Python Tutorial

* <https://www.udemy.com/course/taming-big-data-with-apache-spark-hands-on/> )

Spark - Scala Tutorial

* <https://www.udemy.com/course/apache-spark-with-scala-hands-on-with-big-data/> )

Spark-Streaming Documentation

* <https://spark.apache.org/docs/latest/structured-streaming-programming-guide.html>
* <https://spark.apache.org/docs/latest/structured-streaming-kafka-integration.html>

Hadoop Book

* <https://www.amazon.com/Hadoop-Definitive-Storage-Analysis-Internet-dp-1491901632/dp/1491901632/ref=mt_paperback?_encoding=UTF8&me=&qid=> )

**Schedule**

|  |  |  |  |
| --- | --- | --- | --- |
| **Week** | **Date** | **Topics** | **Readings/Assignments** |
| 1 | Feb. 24 | * Course Overview ( K ) * Intro to Data Engineering |  |
| 2 | Mar. 2 | * Kafka ( W ) |  |
| 3 | Mar. 9 | * Spark Overview ( T ) |  |
| 4 | Mar. 16 | * Spark Streaming ( K ) |  |
| 5 | Mar. 23 | * Interacting with NoSQL ( Hbase ) ( T ) |  |
| 6 | Mar. 30 | * Cold Storage ( Serialization / Hive ) ( Na ) |  |
| 7 | Apr 6 | * Capstone Project Week 1 ( W ) |  |
| 8 | Apr 13 | * Capstone Project Week 2 |  |

**Class Structure:**

Class will meet weekly on Mondays from 6:00 - 9:00. The model for this class is a little flipped. The bulk of classwork and instruction is done outside of class, and then in class lecture and exercises reinforce and fill gaps in the learning. For class time, the first hour will be going over more complex concepts from the assigned reading, asking questions and working through examples to make sure everybody understands. The last two hours of class will be work time, working on specific projects by yourself or often in pairs, with instructors circling and helping work through difficult concepts.

**Working in GCP:**

1904 will run GCP instances to which people can work with. e.g. Kafka servers.

**Communication:**

Slack channel added to 1904Labs.  
Through Email.