DEPARTMENT CIS

AUTHOR(S) Peter Chapin

COURSE NUMBER CIS-4620 (temporary)

COURSE TITLE Big Data Processing

SHORT TITLE Big Data Processing

COURSE LEVEL 4000

DATE CREATED 9/27/2017

CHECKED/CHANGED 9/27/2017

PREREQUISITES CIS-3030, CIS-2230, CIS-3010 recommended

COREQUISITES Click here to enter corequisites (ex. **UWB 9998**, **UWB 9999**); if NA, leave blank

RESTRICTIONS Click here to enter any restrictions on enrollment; if NA, leave blank

SPECIAL FEES No

CREDITS 3

HOURS 3 hours of lecture

SEMESTER Fall?

COURSE DESCRIPTION This course describes techniques for processing very large data sets that are typically stored across multiple machines in a cluster. It is primarly a programming course, although some topics in cluster administration and configuration are also discussed. Technologies covered include, as a sample, Hadoop (MapReduce), Apache Spark, Apache Kafka, and other specialized technologies as time allows (e.g. Pig). Fluency with Java is required; experience with Scala is helpful but not essential.

SUGGESTED TEXTS There is no text.

OPTIONAL TEXTS Click here to enter optional texts; if NA, leave blank

COURSE OUTCOMES The student should be able to: 1) Write programs using MapReduce on Hadoop to compute simple query results on a clustered data set, 2) Use Scala to write programs using Apache Spark to compute simple query results on a clustered data set, 3) Understand the basics of distributed file systems, 4) Understand the basic configuration of Hadoop in both pseudo-distributed mode and fully distributed mode, 5) Write programs using Apache Kafka to process a large streamed data set.

COURSE CONTENT 1) Hadoop Basics & MapReduce (9 hours); 2) Introduction to Scala (4 hours); 3) Apache Spark (8 hours); 4) Cluster Configuration (6 hours); 5) Apache Kafka (9 hours); 6) Other Technologies (6 hours)

LAB OUTCOMES Click here to enter lab outcomes; if NA, leave blank

LAB CONTENT Click here to enter lab content; if NA, leave blank

LECTURE CAPACITY Click here to enter capacity (standard VTC lecture cap is 32)

LAB CAPACITY Click here to enter capacity (standard VTC lab cap is 16)

GRADED OR P/NP Graded

EVALUATION Homework, Projects, Exams

DELIVERY METHOD HYB

ROOM REQUIREMENTS Click here to enter room requirements (ex. **Telepresence**)

AUTHOR’S NOTES Click here to enter optional author’s notes; if NA, leave blank