INTRODUCTION TO BIG DATA ANALYSIS(IEE3593-01)

CREDIT	3	**INSTRUCTOR**	Jongwook Woo
OFFICE		**OFFICE HOURS**	
TIME	Mon-Fri 13:30-16:00[Period2]	**CLASSROOM LOCATION**	TBA
E-MAIL	jwoo5@exchange.calstatela.edu		

* Please leave the fields blank which haven't been decided yet.

[COURSE INFORMATION]

COURSE DESCRIPTION & GOALS	• Understand the genesis of Big Data Systems • Understand practical knowledge of Big Data Analysis using Hive, Pig, Sqoop • Provide the student with a detailed understanding of effective behavioral and technical techniques in Cloud Computing on Big Data • Demonstrate knowledge of Big Data in industry and its Architecture • Learn data analysis, modeling and visualization in Big Data systems
PREREQUISITE	a. Mastery over Microsoft Windows and its File Management (Windows Explorer) facilities b. Basic knowledge of any programming language (SQL, Python, Java) c. Basic knowledge of BI tools such as Excel, Tableau, Power BI, Google Spread Sheet
COURSE REQUIREMENTS	Students are expected to attend every class session. Since Cloud computing and Big Data concepts are presented during class time, class attendance is essential for successful completion of assignments and tests. As a large part of the course involves work on cloud computing, it is essential that you utilize the time in class for discussion and exercises on the computer. If attendance is not possible for one of the class meetings, please contact the instructor beforehand. Students are expected to use the equipment of the computer labs at Yonsei University if you do not have a personal computer nor internet.
GRADING POLICY	· Class Activities (Pop quizzes, Attendance, Participation in Class): 10% · Labs: 20% · 1 - 2 Homework: 15% · Midterm Exam 20% · Final Team Project Presentation: 35% · Total 100%

TEXTS & REFERENCES	1. Instructional materials (Lecture and Lab) from the instructor 2. Hadoop: The Definitive Guide by Tom White 3. https://hadoop.apache.org/ 4. https://www.cloudera.com/tutorials.html
INSTRUCTOR'S PROFILE	Dr Jongwook Woo received his Ph.D from USC at USA. He is a Professor at CIS Department of California State University Los Angeles, and has served as a Teradata Academic Ambassador, a president at KSEA-SC, and a Council Member of IBM Spark Technology Center. He has consulted companies in Hollywood: CitySearch, ARM, E!, Warner Bros, SBC Interactive. He published more than 70 papers and his research interests include Big Data AI Analysis and Prediction. He awards Teradata TUN faculty Scholarship and received grants for distributed deep learning and Big Data from Amazon, IBM, Oracle, MicroSoft and partnered with Intel, Databricks, Cloudera, Hortonworks, SAS, QlikView, Tableau. He is a founder of Hemosoo Inc and The Big Link. He run BigDAI center (http://www.calstatela.edu/centers/hipic).

[WEEKLY SCHEDULE]

* Your detailed explanation would be very helpful for prospective students to get a pre-approval for credit-transfer from their home university in advance.

WEEK	DAILY TOPIC & CONTENTS	COURSE MATERIAL & ASSIGNMENTS	REFERENCE
1	Course Overview		
	Lecture 1 An Introduction to Big Data and Cloud Computing Systems	Reading Instructor's material about the systems of Big Data and Cloud Computing	
2	Lecture 2 Big Data system development a. Introduction to Hadoop b. Motivation for Hadoop	Reading Instructor's material about Hadoop	
	Lab 1: set up cloud computing accounts such as Oracle Big Data Compute Edition		
3	Lecture 3 Basic Concepts: HDFS, MapReduce, Hive	Reading Instructor's material about HDFS, MR, Hive	
	Lab 2: HDFS and Hive in Oracle Big Data		
4	Lecture 4 Initiating Phase: Basic Concepts: MR cont'd; Cluster; Ecosystems, Hive	Reading Instructor's material about MR, Cluster, Ecosystems, Hive	
	Lab 3: Hive Web Log Analysis in Oracle Big Data	Team Build for term project: email the preferred team members	
5	Lecture 5 Hive Data Processing (Join, Union)	Reading Instructor's material about Join in Hive	
	Lab 4: IoT Sensor Data Analysis using Hive in Oracle Big Data	Team built and choose topics for the team project	
6	Lecture 6 Text Analysis in Hive	Reading Instructor's material about Hive Text Analysis	
	Lab 5: Twitter Data Text Analysis using Hive in Oracle Big Data		
7	Midterm Exam	Lectures and Labs in Week 1 through Week 6	

WEEK	DAILY TOPIC & CONTENTS	COURSE MATERIAL & ASSIGNMENTS	REFERENCE
8	Lecture 7 Sqoop and Hive	Reading Instructor's material about Sqoop and Hive	
	Lab 6: Movie Data Analysis using Sqoop and MySQL DB		
9	Lecture 8 NGram, Text Processing Functions in Hive	Reading Instructor's material about NGram and Functions of Hive	
	Lab 7: NGram Sentiment Text analysis of Twitter social media data		
10	Lecture 9 Pig Fundamentals	Reading Instructor's material about Pig and PigLatin	
	Lab 8: Pig Fundamentals in Oracle Big Data		
11	Lecture 10 Pig Data Filtering and Process	Reading Instructor's material about Data Filtering and Process in Pig	
	Lab 9: Pig Data Process in Oracle Big Data		
12	Lecture 11 Pig and Hive using HCatalog	Reading Instructor's material about Pig and Hive using HCatalog	
	Lab 10: HCatalog with IoT data of TruckEvent in Oracle Big Data		
13	Lecture 12 Pig UDF and Streaming Data	Reading Instructor's material about UDF and Streaming Data	
	Lab 11: Pig ETL Processing and visualization using Tableau in Oracle Big Data		
14	Lecture 13 Big Data Trend with Data Science	Reading Instructor's material about Big Data Analysis and Big Data Science	
	Lab 12: Hands-on Exercise with Spark QL		
15	Term Project Presentation	Present Group Term Project: Topics of Big Data Analysis and	

WEEK	DAILY TOPIC & CONTENTS	COURSE MATERIAL & ASSIGNMENTS	REFERENCE
		Visualization using Cloud Computing	