RIT جامعة روتشستر الأمريكية للتكنولوجيا في دبي
A Global American University in Dubai

PROF741- Enterprise Infrastructure for Data Analytics

Weeks	Topics	Assigned Readings	Assessments
W1	Course Introduction • Enterprise Infrastructure • Enterprise Infrastructure Components Big Data • Introduction • Architecture & Patterns Data Integration • ETL vs ELTSYLLA Cloud Computing Infrastructure • Cloud Deployment Models (Public/Private/Hybrid) • Cloud Service Models (Iaas, Paas, SaaS) • Amazon AWS • Microsoft Azure • Google Engine	• Lecture notes • BC1- ch1 • BC1-ch4	• Homework Assignment 1
W2	Data Access Connectors • Raw Data exploration • Data Access connectors (Kafka/Flume/Sqoop) Storage • HDFS concepts • HBase • Intro Hive/Pig Data Engineering Storage Abstractions Data Lake Design • Kappa and Lambda Data Fabric Architecture • Data Fabric Layers • Data Fabric Functional Architecture Data Mesh Layers Data Product Architecture	• Lecture notes • BC6-ch1&2 • BC1-ch7 • BC8- ch6	• Homework Assignment2 • Project Assignment

W3	Hadoop ● Introduction Hadoop ● History of Hadoop ● Apache Hadoop ecosystem Map Reduce ● Map Reduce concepts ● Map Reduce Data flow ● Map Reduce Process	● Lecture notes	● Homework Assignment 3
W4	Map Reduce ● Map Reduce Execution Jobs ● Map Reduce Features ● Counter ● Sorting ● Joins SPARK ● Introduction to Spark ● RDDs ● Spark Shell ● Spark on Yarn	● Lecture notes ● BC2-ch3 ● BC1-ch7	
W5	Hive ● Hive Client ● Comparison with traditional DBs ● Partitions and Buckets Zookeeper ● The Zookeeper services ● Zookeeper implementation Big Data Infrastructure (Topologies) ● Stack topologies ● Cloudera CDH ● Hortonworks ● EMR Amazon Analytics ● Azure Analytics Data Architecture ● Data Warehouse ● Data Lake ● Data Fabric ● Data Mes	● Lecture notes ● BC3-ch2 ● BC4-ch1 ● BC5-ch2	● Homework Assignment 3
W6	● Spark MLib ● Serverless computing and architecture ● SQL/noSQL Databases ● Final Exam	● Lecture notes ● BC2-ch3-5 ● BC1-ch9	● Project Report Submission ● Project Presentation