

University of Gujrat
Faculty of CS & IT

Title	High Performance Computational Infrastructures								
Code									
Credit Hours	3 Theory/week: Weight3 Cr. Hrs. Contact Hours3 Hrs. Theory								
Prerequisite	None								
Follow Up	Introduction to Computing, Database Systems								
Category	Elective								
Aims and Objectives	The aim of the course is to develop knowledge and skills necessary for working effectively with the large-scale data storage and processing infrastructures that underpin data science. Again, you will develop both practical skills and an ability to reflect critically on concepts, theory and appropriate use of infrastructure. Content here covers, highly-scalable data-storage paradigms (e.g. NoSQL data stores) alongside cloud computing tools (e.g. Amazon EC2) and in-memory approaches.								
Learning Outcomes	LO1: Demonstrate the appropriate practical skills/abilities required to implement solutions using modern large-scale data storage and processing infrastructures. LO2: Reflect critically on the concepts, theory and appropriate use of large-scale data storage and processing infrastructures (commonly used in modern organisational environments).								
Syllabus	<ul style="list-style-type: none">Relational database systems and SQL.Highly-scalable data-storage paradigms, such as NoSQL data stores (MongoDB, Cassandra, HBase etc) and distributed hash tables.Cloud Computing tools for large-scale data storage and processing (e.g., Amazon S3, EC2 and Elastic MapReduce).In-memory storage and processing solutions (e.g., SAP HANA, Spark).								
Text Book	Hadoop: the definitive guide by Tom White, Third Edition, O'Reilly, 2012								
Reference Material	1. Hadoop in action by Chuck Lam, Manning Publications. 2011 2. Business database systems by Thomas Connolly, Carolyn E. Begg, Richard Holowczak 2008								
Assessment Criteria		Sessional	25%	Mid	25%	Final	50%	Total	100%
		Quizzes and Tests							
		Assignments		Paper	25	Paper	50		
		Project and Presentation							
Resource Person	Dr. Nauman Riaz Chaudhry								

Framework				
Week	Lecture	Topic	Source (Book-Chapter No. Section No.)	Recommendations for Learning Activities (Mention Assignments, Test, Quizzes, Practical, Case Study, Projects, Lab Work or Reading Assignments)
1	1	Introduction	Ch. 1	
	2	Revisit Relational Database Systems	Ch. 1	
2	3	Normalization and Structure Query Language (SQL)	Ch. 1	
	4	Cloud Computing Tools	Ch. 1	
3	5	Hadoop Fundamentals 1	Ch. 1	Quiz/Test 1
	6	MapReduce 1 Hello World	Ch. 2	Assignment 1
4	7	Hadoop Fundamentals 2	Ch. 2	
	8	MapReduce 2	Ch. 2	
5	9	Hadoop Distributed File System (HDFS)	Ch. 3	
	10	Hadoop Distributed File System (HDFS) Practical Work	Ch. 3	
6	11	Yet Another Resource Negotiator (YARN)	Ch. 6	

	12	Yet Another Resource Negotiator (YARN) Practical Work	Ch. 6	Assignment 2
7	13	MapReduce Program Design 1	Ch. 5	Quiz/Test 2
	14	MapReduce Program Design 1 Practical Work	Ch. 5	
8	15	MapReduce Program Design 2	Ch. 5	
	16	MapReduce Program Design 2 Practical Work	Ch. 5	
9	17	Mid-Term Exam		
	18	Revisit Hadoop and MapReduce		Term Projects Topic Assigned
10	19	Pig 1	Ch. 11	
	20	Pig 2	Ch. 11	
11	21	Hive 1	Ch. 12	Assignment 3
	22	Hive 2	Ch. 12	Quiz/Test 3
12	23	NoSQL 1	Ch. 13	
	24	NoSQL 2	Ch. 13	
13	25	HBase 1	Ch. 13	
	26	HBase 2	Ch. 13	

14	27	Introduction to Big Data		Assignment 4
	28	Storing Big Data		
15	29	Processing Big Data		
	30	Tools and Techniques to Analyze Big Data		
16	31	Developing a Big Data Strategy		Quiz/Test 4
	32	Implementing a Big Data Solution		
Final-Term Exam				