

Branch Name:	Bachelor of Science (Information Technology)
Program Code:	ET110
Semester/Year:	Semester - VI/ Third Year
Subject Title:	Big Data on Cloud
Subject Code:	1ET1100602
Pre-requisite:	SQL, Data Warehousing/Data Processing, and Database Knowledge

Course Objective:

- To study the basic technologies that forms the foundations of Big Data.
- To study the security and management aspects of cloud computing with a view to rapid prototyping of complex applications.
- To understand the specialized aspects of Data Extraction, Transformation, and loading.
- To study different types of algorithms and data processing techniques.

Teaching and Examination Scheme:

Teaching Scheme (Hours per week)				Evaluation Scheme (Marks)				
Lecture (L)	Tutorial (T)	Practical (P)	Credit	Theory (Marks)		Practical (Marks)		Total (Marks)
				University Assessment	Continuous Assessment	University Assessment	Continuous Assessment	
03	01	02	05	70	30	30	20	150

Subject Contents			
Sr. No	Topic	Total Hours	Weight (%)
1	Introduction to Cloud Computing & Big Data • Cloud characteristics • Cloud delivery models • Cloud deployment models • Cloud infrastructure and management • Cloud security • Prominent cloud vendors and their services • Big Data characteristics and benefits • Data storage technologies and storing data on Amazon S3	3	10
2	Data Extraction, Data Processing and Knowledge Extraction • Data extraction process and tools • Change Data Capture (CDC) • Distributed data processing • Different data processing engines • ETL using Amazon Glue • Data processing using Amazon	10	20

	• Data querying using Amazon Athena • Redshift data warehouse • Association rule mining • Apriori algorithm • Noise and outlier detection • Data pre-processing techniques • Sampling techniques • Similarity/dissimilarity measures		
3	Classification and Clustering of Big Data • Supervised and unsupervised learning • Collaborative filtering • Big Data analytics • Streaming data clustering • Feature extraction • Subspace clustering	6	15
4	In-Memory Computation • In-memory computation using Spark • Spark context • RDD and Data Frames • Transformations and actions • Data lineage	6	15
5	Real-time Data Processing • Real-time vs batch processing • Advantages and tools • Use cases • Using Amazon Kinesis and Spark streaming	5	10
6	Data Governance • Purpose of data governance • Metadata management • Data quality • Privacy and security • Data orchestration using AWS Data Pipeline	4	10
7	Data Architectures and Modelling • Characteristics of modern data architecture • Enterprise data architecture • Data modelling • Conceptual, logical and physical data modelling	3	10

8	Business Intelligence (BI) and Data Visualization Tools • Benefits of BI • Introduction to Tableau • Data sources • Connecting to Redshift • Worksheets, operators, functions, sorting and filtering • Tableau charts	4	10
---	---	---	----

List of Reference Books:

1. Cloud Computing: Principles and Paradigms by Rajkumar Buyya, James Broberg, Andrzej Goscinski. Released March 2011. Publisher(s): Wiley. ISBN: 9780470887998
2. Cloud Computing: A Practical Approach. Front Cover · Toby Velte, Anthony Velte, Robert C. Elsenpeter. McGraw Hill Professional, Oct 22, 2009
3. Cloud Application Architectures. by George Reese. Released April 2009. Publisher(s): O'Reilly Media, Inc. ISBN: 9780596156367
4. John Rhoton, "Cloud Computing Explained: Implementation Handbook for Enterprises," Recursive Press, 2009, ISBN:0956355609
5. Cloud Computing(1st Edition) Fundamentals, Industry Approach and Trends by Rishabh Sharma Paperback, Published 2015 by Wiley India
6. Kailash Jayaswal, Jagannath Kallakurchi, Donald J Houde, Dr. Deven Shah Cloud Computing: Black Book Dreamtech Publications 2012 (ISBN 978-93-5119418-7)
7. Programming Google App Engine with Java. by Dan Sanderson. Released June 2015. Publisher(s): O'Reilly Media, Inc. ISBN: 9781491903452
8. Cloud Computing For Dummies. Front Cover. Judith S. Hurwitz, Robin Bloor, Marcia Kaufman, Fern Halper. John Wiley & Sons, Nov 16, 2009
9. Cloud Computing: SaaS, PaaS, IaaS, Virtualization, Business Models, Mobile, Security and More,. Kris Jamsa, Jones & Bartlett Publishers, Paper back edition,2013.
10. Learning Tableau 10 - Second Edition: Business Intelligence and data visualization that brings your business into focus 2nd Edition, Kindle Edition by Joshua N. Milligan (Author)

List of Suggested titles of Experiments:

1. Develop a hello world program web application and deploy it on the Google app engine
2. Develop a web application which displays the current date and time in a formatted way.
3. Develop a web application which is customized version of the practical no 2, which relates the clock with Google accounts. Each user will get different view based on the preferences and the user's time zone.
4. Develop a web application for task management. Create a form to enter summary, URL and description of task and a button to insert a task in the task list. Use JPA (Java Persistence API) for the persistence and Sends the selected todo to yourself via email as a reminder
5. Develop a simple Java web application for Google App Engine--a guestbook that lets users post

messages to a public message board.

6. Create a book class as entity and make the object persistence using JPA. Demonstrate the saving, fetching and deleting objects operations.
7. Create an employee class as entity and makes the objects persistence using JPA. Demonstrate transaction processing in the data store using JPA.
8. JPA includes a SQL-like query language called JPQL, perform the practical no 6 with JPQL also.
 1. Perform the practical no 7 with JPQL[list the employees having salary between a particular range]

Course Learning Outcomes (CLO): On completion of this course, the students will be able to:

CLO	Description	Bloom's Taxonomy Level
CLO1	Understand concepts of Cloud Computing	1 Remembering 2. Understanding
CLO2	Understand virtualization techniques and Service Level Agreement	2 Understanding 3 Applying
CLO3	Estimate cost for Cloud Computing implementation	2 Understanding 3 Applying 4 Analyze
CLO4	Understand and analyze architecture of cloud computing	2 Understanding 3 Applying 4 Analyze
CLO5	Deploying application in cloud environment	2. Understanding 3 Applying 4 Analyze 5 Evaluate

Mapping of CLOs with POs & PSOs

Course	Program Outcomes (POs)												Program Specific Outcomes(PSOs)			
Learning Outcomes	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO 1	PSO 2	PSO 3	PSO 4
CLO1	M			M	M							M		M	L	
CLO2	M	H	L	M	H							L	M	M		
CLO3		L	L			M			M	H	H					L
CLO4	H	H	M	M	M							M	M	L	M	
CLO5	M	M	H								H		M	M	L	H

H: High, M: Medium, L: Low