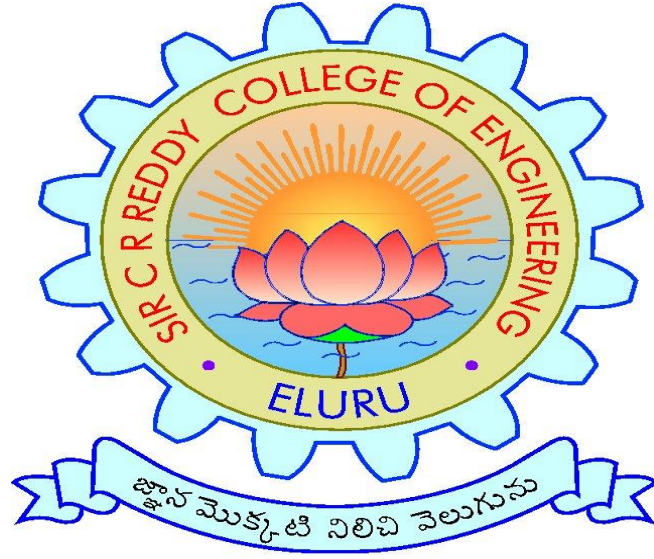


SIR C.R.REDDY COLLEGE OF ENGINEERING, ELURU

DEPARTMENT OF INFORMATION TECHNOLOGY

LESSON PLAN



SUBJECT: CSE 4.1.6 BIGDATA ANALYTICS

CLASS: 4/4 B.Tech., I SEMESTER, A.Y.2019-20

INSTRUCTOR: V.GOPINATH

Sir C R Reddy College of Engineering
DEPT. OF INFORMATION TECHNOLOGY

PROGRAMME: B.Tech
SEMESTER: IV –I Semester
A.YEAR : 2019-20

Course: CSE4.1.6 Bigdata Analytics

Instructor: V.Gopinath

Course Contents

Category of Course	Course Title	Course Code	Credits- 4	Theory Paper
Core Subject	Bigdata analytics	CSE 4.1.6	L-3 T-1	Max.Marks-70 Duration-3hrs.

Course Objectives:

On completing this course student will be able to

1. Understand big data and Apache Hadoop Eco system
2. Understand distributed , parallel, cloud computing and SQL concepts
3. Apply Hadoop concepts
4. Understand concepts of map and reduce and functional programming

Course Outcomes:

1. Gain conceptual understanding of analytics concepts, algorithms and statistical tests
2. Students will be able to look at the core projects used for both batch and real time data processing such as Hadoop
3. Students will be able to look at wider range of problems and data science based solutions

CSE 4.1.6	BIGDATA ANALYTICS	
Instruction: 3 Periods + 1 Tut/week, Univ. Exam: 3 Hours		Credits: 4
Internal: 30 Marks	University Exam: 70 Marks	Total: 100 Marks

Syllabus

1. **Introduction to Big Data:** Big Data-definition, Characteristics of Big Data (Volume, Variety, Velocity, Veracity, Validity), Importance of Big Data , Patterns for Big Data Development, Data in the Warehouse and Data in Hadoop,
2. **Introduction to Hadoop:** Hadoop- definition, Understanding distributed systems and Hadoop, Comparing SQL databases and Hadoop, Understanding MapReduce, Counting words with Hadoop—running your first program, History of Hadoop, Starting Hadoop - The building blocks of Hadoop, NameNode, DataNode, Secondary NameNode, JobTracker and Task Tracker
3. **MapReduce** -A Weather Dataset, Analyzing the Data with Unix Tools, Analyzing the Data with Hadoop, Scaling Out, Hadoop Streaming, Hadoop Pipes, Developing a MapReduce Application - The Configuration API, Configuring the Development Environment, Running Locally on Test Data, Running on a Cluster, Tuning a Job, MapReduce Workflows
4. **HDFS:** Components of Hadoop -Working with files in HDFS, Anatomy of a MapReduce program, Reading and writing the Hadoop Distributed File system -The Design of HDFS, HDFS Concepts, The Command-Line Interface, Hadoop Filesystem, The Java Interface, Data Flow, Parallel Copying with distcp, Hadoop Archives
5. **MapReduce Programming:** Writing basic Map Reduce programs - Getting the patent data set, constructing the basic template of a Map Reduce program, Counting things, Adapting for Hadoop's API changes, Streaming in Hadoop, Improving performance with combiners.
6. **MapReduce Advanced Programming:** Advanced MapReduce - Chaining MapReduce jobs, joining data from different sources, creating a Bloom filter, Passing job-specific parameters to your tasks, probing for task-specific information, Partitioning into multiple output files, Inputting from and outputting to a database, keeping all output in sorted order
7. **Graph Representation in MapReduce:** Modeling data and solving problems with graphs, Shortest Path Algorithm, Friends-of-Friends Algorithm, PageRank Algorithm, Bloom Filter, Parallelized Bloom filter creation in MapReduce, Map-Reduce semi-join with Bloom filters

Textbooks:

1. Dirk deRoos, Chris Eaton, George Lapis, Paul Zikopoulos, Tom Deutsch
“Understanding Big Data Analytics for Enterprise Class Hadoop and Streaming Data”,
1st Edition, TMH,2012.
2. Hadoop: The Definitive Guide by Tom White, 3rd Edition, O’reilly

Reference Books:

1. Hadoop in Action by Chuck Lam, MANNING Publ.
2. Hadoop in Practice by Alex Holmes, MANNING Publishers
3. Mining of massive datasets, Anand Rajaraman, Jeffrey D Ullman, Wiley Publications.

Prerequisite

Java,Data mining,SQL concepts and data structures

Internal Assessment Details:

Internal Test 1& 2: 20Marks

Assignment-1& 2: 10Marks

Total: 30 Marks

ONLINE REFERENCES:

- a. Hadoop:<http://hadoop.apache.org/>
- b. . Hive: <https://cwiki.apache.org/confluence/display/Hive/Home>

Digital Learning Materials:

□ <http://192.168.0.49/videos/videosListing/270#>

SIR C R REDDY COLLEGE OF ENGINEERING :: ELURU
DEPARTMENT OF INFORMATION TECHNOLOGY
COURSE SCHEDULE

The schedule for the whole course/subject is:

Unit No	Description of the Chapter	Description of the Topics	Total no of periods (L+T)
1	Introduction to big data	Big Data-definition, Characteristics of Big Data (Volume, Variety, Velocity, Veracity, Validity), Importance of Big Data , Patterns for Big Data Development, Data in the Warehouse and Data in Hadoop	5+1
2	Introduction to hadoop	Hadoop- definition, Understanding distributed systems and Hadoop, Comparing SQL databases and Hadoop, Understanding MapReduce, Counting words with Hadoop—running your first program, History of Hadoop, Starting Hadoop - The building blocks of Hadoop, NameNode, DataNode, Secondary NameNode, JobTracker and Task Tracker	10+2
3	mapreduce	A Weather Dataset, Analyzing the Data with Unix Tools, Analyzing the Data with Hadoop, Scaling Out, Hadoop Streaming, Hadoop Pipes, Developing a MapReduce Application - The Configuration API, Configuring the Development Environment, Running Locally on Test Data, Running on a Cluster, Tuning a Job, MapReduce Workflows	10+2
4	HDFS	Components of Hadoop -Working with files in HDFS, Anatomy of a MapReduce program, Reading and writing the Hadoop Distributed File system -The Design of HDFS, HDFS Concepts, The Command-Line Interface, Hadoop Filesystem, The Java Interface, Data Flow, Parallel Copying	9+1

		with distcp, Hadoop Archives	
5	Mapreduce programming	Writing basic Map Reduce programs - Getting the patent data set, constructing the basic template of a Map Reduce program, Counting things, Adapting for Hadoop's API changes, Streaming in Hadoop, Improving performance with combiners.	7+1
6	Mapreduce advanced programming	Advanced MapReduce - Chaining MapReduce jobs, joining data from different sources, creating a Bloom filter, Passing job-specific parameters to your tasks, probing for task-specific information, Partitioning into multiple output files, Inputting from and outputting to a database, keeping all output in sorted order	8+1
7	Graph representation in map reduce	Modeling data and solving problems with graphs, Shortest Path Algorithm, Friends-of-Friends Algorithm, PageRank Algorithm, Bloom Filter, Parallelized Bloom filter creation in MapReduce, Map-Reduce semi-join with Bloom filters	7+1

Total no of instructional periods available for the course : 75 periods

Total no of estimated periods : 65 periods

Signature of the H.O.D

Signature of the Faculty

Date:

<u>LECTURE PLAN</u>	
DEPARTMENT	INFORMATION TECHNOLOGY
NAME OF LECTURER	V.Gopinath

Sl.No	Topics to be covered	No. of Lecture hours	Teaching method	Text book	CO Mapping
1	Introduction to Big Data:	1	BB	TB1	
2	Big Data-definition, Characteristics of Big Data	1	BB	TB1	
3	Importance of Big Data ,	1	PPT with LCD	TB1	
4	Patterns for Big Data Development	1	BB	TB1	
5	, Data in the Warehouse and Data in Hadoop	1	PPT with LCD	TB1	
6	Introduction to Hadoop: Hadoop- definition,	1	BB	TB3	
7	Understanding distributed systems and Hadoop, Comparing SQL databases and Hadoop,	2	PPT with LCD	TB3	
8	Understanding MapReduce,	1	PPT with LCD	TB3	
9	Counting words with Hadoop— running your first program	1	BB	TB3	
10	, History of Hadoop, Starting Hadoop - The building blocks of Hadoop,	1	BB	TB3	
11	NameNode, DataNode, Secondary NameNode,	1	BB	TB3	

12	JobTracker and Task Tracker	1	BB	TB3	
13	MapReduce -	1	BB	TB2	
14	A Weather Dataset, Analyzing the Data with Unix Tools, Analyzing the Data with Hadoop,	2	PPT with LCD	TB2	
15	Scaling Out, Hadoop Streaming, Hadoop Pipes,	1	BB	TB2	
16	Developing a MapReduce Application - The Configuration API,	1	BB	TB2	
17	Configuring the Development Environment, Running Locally on Test Data, Running on a Cluster	2	PPT with LCD	TB2	
18	, Tuning a Job, MapReduce Workflows	1	BB	TB2	
19	HDFS: Components of Hadoop -Working with files in HDFS,	2	BB	TB3	
20	Anatomy of a MapReduce program,	1	BB	TB3	
21	Reading and writing the Hadoop Distributed File system -	1	BB	TB3	
22	The Design of HDFS, HDFS Concepts,	2		TB2	
23	The Command-Line Interface, Hadoop Filesystem,	1	BB	TB2	
24	The Java Interface, Data Flow,	1	BB	TB2	
25	Parallel Copying with distcp, Hadoop Archives	2	BB	TB2	

26	MapReduce Programming:.	1	BB	TB3	
27	Writing basic Map Reduce programs - Getting the patent data set,	1	BB	TB3	
28	constructing the basic template of a Map Reduce program,	2	PPT with LCD	TB3	
29	Counting things,	1	BB	TB3	
30	Adapting for Hadoop's API changes,	1	BB	TB3	
31	Streaming in Hadoop,	1	BB	TB3	
32	Improving performance with combiners	1	BB	TB3	
33	MapReduce Advanced Programming: Advanced MapReduce -	1	BB	TB3	
34	Chaining MapReducejobs,	2	BB	TB3	
35	joining data from different sources,	1	BB	TB3	
36	creating a Bloom filter	1	PPT with LCD	TB3	
37	, Passing job-specific parameters to your tasks,	1	PPT with LCD	TB3	
38	probing for task-specific information,	1	BB	TB3	
39	Partitioning into multiple output files,	1	BB	TB3	
40	Inputting from and outputting to a database,	1	BB	TB3	
41	keeping all output in sorted order	1	BB	TB3	

42	Graph Representation in MapReduce:	1	BB	TB4	
43	Modeling data and solving problems with graphs,	1	PPT with LCD	TB4	
44	Shortest Path Algorithm,	1	BB	TB4	
45	Friends-of-Friends Algorithm,	1	BB	TB4	
46	PageRank Algorithm,	1	PPT with LCD	TB4	
47	Bloom Filter, Parallelized Bloom filter creation in MapReduce,	1	BB	TB4	
48	Map-Reduce semi-join with Bloom filters	1	BB	TB4	
	Total classes	65			

BIGDATA ANALYTICS

Unit Wise Question Bank

UNIT-1

SHORT ANSWER QUESTIONS

1. Define big data
2. Write characteristics of bigdata

LONG ANSWER QUESTIONS

1. Explain the importance of bigdata
2. Describe patterns for bigdata development
3. compare the behavior of data in warehouse and data in hadoop

UNIT-2

SHORT ANSWER QUESTIONS

1. Define hadoop
2. what is distributed system
3. what is mapreduce
4. what is name node, data node
5. what is job tracker and task tracker

LONG ANSWER QUESTIONS

1. compare sql databases with hadoop
2. write a counting words program in hadoop
3. explain the building blocks of hadoop
4. illustrate job tracker and task tracker with a program
5. identify name node, data node and secondary name node in a hadoop program.
6. Explain Hadoop components
7. Explain Hadoop architecture

UNIT-3

SHORT ANSWER QUESTIONS

1. What is scaling out
2. what is streaming in hadoop
3. what is role of pipes in hadoop
4. write mapreduce workflows.

LONG ANSWER QUESTIONS

1. Analyze the weather dataset with unix tools and hadoop
2. How do you develop a mapreduce application with configuration api and development environment.

3.How do you run a mapreduce application locally on test data and running on a cluster.

UNIT-4

SHORT ANSWER QUESTIONS

- 1.what are the components of hadoop
- 2.what are the working files in HDFS.
- 3.what is the hadoop file system
- 4.write hadoop archives.
- 5.what is the role command line interface in HDFS

LONG ANSWER QUESTIONS

- 1.How do you read and write a file in HDFS
- 2.Explain the design of HDFS
- 3.How do you perform parallel copying with distcp.

UNIT-5

SHORT ANSWER QUESTIONS

- 1.How do you change hadoop APIs.
- 2.what is streaming in hadoop
- 3.what is the role of combiners.
- 4.how do you count the things.

LONG ANSWER QUESTIONS

- 1.How do you get the patent data set
- 2.Construct a template of map reduce program
- 3.how do you improve performance with combiners

UNIT-6

SHORT ANSWER QUESTIONS

1.What is the purpose of bloom filter.

LONG ANSWER QUESTIONS

- 1.Explain the chaining mapreduce jobs.
- 2.How do you join data from different sources
- 3.How do you pass job specific parameters to your tasks.
- 4.How do you probe task specific information.
- 5.How do you partition the output files and sort them in order.
- 6.How do you give and take input and output from databases..

UNIT-7

SHORT ANSWER QUESTIONS

- 1.what is graph
- 2.What is importance of graphs in BDA
- 3.differ bloom filter and parallised bloom filter.

LONG ANSWER QUESTIONS

- 1.Implement shortest path algorithm
- 2.Implement of FOF
- 3.Implement page rank algorithm
- 4.how do you perform semi join with bloom filters.

