**Course Code: 18CS0538** **R18**

**SIDDHARTH INSTITUTE OF ENGINEERING & TECHNOLOGY:: PUTTUR**
**(AUTONOMOUS)**
Siddharth Nagar, Narayanavanam Road – 517583
**QUESTION BANK (DESCRIPTIVE)**

**Subject with Code:** BIG DATA ANALYTICS(18CS0538) **Course & Branch**: B.Tech - CSE

**Regulation:** R18 **Year & Sem:** IV-B.Tech & I-Sem

## UNIT –I
### Introduction To Big Data And Hadoop

| | | | | |
|---|---|---|---|---|
| **1** | **a** | Define big data. | [L1][CO1] | **[2M]** |
| | **b** | What are the four V's of big data. | [L1][CO1] | **[2M]** |
| | **c** | What is the significance of Apache Hadoop | [L1][CO1] | **[2M]** |
| | **d** | Define Hadoop streaming | [L1][CO1] | **[2M]** |
| | **e** | What is Hadoop ecosystem | [L1][CO1] | **[2M]** |
| **2** | | Examine the different types of digital data with examples? | [L4][CO1] | **[10M]** |
| **3** | | Discuss Big Data in terms of three dimensions, volume, variety and velocity. | [L2][CO1] | **[10M]** |
| **4** | | Explain the Evolution of Hadoop ecosystem with neat diagram. | [L2][CO1] | **[10M]** |
| **5** | | a) List the Top challenges facing big data. | [L1][CO1] | **[5M]** |
| | | b) What is the Significance of big data analytics | [L1][CO1] | **[5M]** |
| **6** | | Distinguish between Analysis of data through Unix tools and Hadoop Ecosystem | [L2][CO1] | **[10M]** |
| **7** | | What is big data analytics? Explain Classification of Analytics | [L1][CO1] | **[10M]** |
| **8** | | Illustrate in detail about Hadoop streaming | [L2][CO1] | **[10M]** |
| **9** | | What is Big Sheets? What can be done with big sheets? | [L1][CO1] | **[10M]** |
| **10** | | Explain in detail about Infosphere Big Insights ? | [L2][CO1] | **[10M]** |

**Course Code: 18CS0538** **R18**

# UNIT –II
## HDFS(Hadoop Distributed File System)

| | | | | |
|---|---|---|---|---|
| **1** | **a** | What is the Hadoop file system | [L1][CO2] | **[2M]** |
| | **b** | What is name node and data node | [L1][CO2] | **[2M]** |
| | **c** | Define data injection in sqoop | [L1][CO2] | **[2M]** |
| | **d** | Distinguish between horizontal file format and vertical File format | [L4][CO2] | **[2M]** |
| | **e** | Define compression operation in HDFS. | [L1][CO2] | **[2M]** |
| **2** | Explain block, name node and data node in Hadoop file system | | [L2][CO2] | **[10M]** |
| **3** | What are the basic commands in Hadoop command line interface. | | [L1][CO2] | **[10M]** |
| **4** | What is an interface? Explain Hadoop system interfaces | | [L1][CO2] | **[10M]** |
| **5** | Discuss about the Hadoop Archives and its Limitations | | [L2][CO2] | **[10M]** |
| **6** | Describe the File read and File write operations in HDFS | | [L1][CO2] | **[10M]** |
| **7** | Discuss about the data injest operation using sqoop and flume | | [L2][CO2] | **[10M]** |
| **8** | Explain compression and serialization operation in Hadoop I/O. | | [L1][CO2] | **[10M]** |
| **9** | Elaborate the AVRO file format with a diagram | | [L6][CO2] | **[10M]** |
| **10** | Explain in detail about File Based Data structures. | | [L2][CO2] | **[10M]** |

**Course Code: 18CS0538** **R18**

# UNIT –III
## Map Reduce

| | | | | |
|---|---|---|---|---|
| **1** | **a** | What is Shuffling in MapReduce?. | [L1][CO3] | **[2M]** |
| | **b** | Define MapReduce. | [L1][CO3] | **[2M]** |
| | **c** | What is and Sorting in MapReduce? | [L1][CO3] | **[2M]** |
| | **d** | What are the parameters of mappers and reducers? | [L1][CO3] | **[2M]** |
| | **e** | What is the role of combiner and partitioner in map reduce application? | [L1][CO3] | **[2M]** |
| **2** | Explain the Classic MapReduce Job Run with a neat diagram. | | [L2][CO3] | **[10M]** |
| **3** | Describe the Significance of YARN over Classic MapReduce Job Run. | | [L1][CO3] | **[10M]** |
| **4** | What are the different types of failures in<br>a) Classic MapReduce<br>b) YARN | | [L1][CO3] | **[10M]** |
| **5** | Examine the different types of Job Scheduling process in Map Reduce. | | [L3][CO3] | **[10M]** |
| **6** | Describe the Shuffle and Sort operations in Map side and Reduce side | | [L1][CO3] | **[10M]** |
| **7** | a) What are the Properties in Task Execution Environment.<br>b) Discuss about Speculative Execution and its Properties | | [L1][CO3]<br>[L2][CO3] | **[5M]**<br>**[5M]** |
| **8** | Discuss the different types of input formats in MapReduce. | | [L2][CO3] | **[10M]** |
| **9** | Explain the different types of output formats in MapReduce. | | [L2][CO3] | **[10M]** |
| **10** | Contrast the below features in MapReduce.<br>a) Counters b) Sorting c) Joins | | [L4][CO3] | **[10M]** |

**Course Code: 18CS0538** **R18**

# UNIT –IV
## Hadoop Eco System-Pig

| | | | | |
|---|---|---|---|---|
| **1** | **a** | Define Pig Latin. | [L1][CO4] | **[2M]** |
| | **b** | Illustrate and Give two examples of user defined functions. | [L2][CO4] | **[2M]** |
| | **c** | What is Grunt? | [L1][CO4] | **[2M]** |
| | **d** | Compare any two execution modes of pig? | [L2][CO4] | **[2M]** |
| | **e** | What are pig Latin relational operators | [L1][CO4] | **[2M]** |
| **2** | | What is PIG? How to Install and execute PIG on Hadoop Cluster | [L1][CO4] | **[10M]** |
| **3** | | Compare the PIG with Databases with an Example | [L2][CO4] | **[10M]** |
| **4** | | Explain in detail about how to Run PIG Programs. | [L2][CO4] | **[10M]** |
| **5** | | Illustrate User Define Functions in Pig Latin. | [L2][CO4] | **[10M]** |
| **6** | | Explain about Arithmetic Operators in Pig Latin . | [L2][CO4] | **[10M]** |
| **7** | | Describe about Relational Operators in Pig Latin . | [L1][CO4] | **[10M]** |
| **8** | | Explain Briefly about Schemas and Functions in Pig Latin | [L2][CO4] | **[10M]** |
| **9** | | Explain about the data types in Pig Latin. | [L2][CO4] | **[10M]** |
| **10** | | Explain Briefly about Structures, Statements in Pig Latin | [L1][CO4] | **[10M]** |

**Course Code: 18CS0538** **R18**

# UNIT –V
## Hive, Hbase, Big SQL

| | | | | |
|---|---|---|---|---|
| **1** | **a** | list out 5 hive shell commands. | [L1][CO5] | **[2M]** |
| | **b** | What is Hbase. | [L1][CO5] | **[2M]** |
| | **c** | Define Metadata. | [L1][CO5] | **[2M]** |
| | **d** | What is Big SQL. | [L1][CO5] | **[2M]** |
| | **e** | What are the advantages of Hive query language | [L1][CO5] | **[2M]** |
| **2** | Explain about Hive shell command line interface. | | [L2][CO5] | **[10M]** |
| **3** | Describe about Hive architecture and its components | | [L1][CO5] | **[10M]** |
| **4** | Examine the various services offered by Hive. | | [L3][CO5] | **[10M]** |
| **5** | What are the advantages of Hive over traditional databases? | | [L1][CO5] | **[10M]** |
| **6** | Explain about Hive query language? What is Metastore in Hive? | | [L2][CO5] | **[10M]** |
| **7** | Describe about the user defined functions in Hive. | | [L1][CO5] | **[10M]** |
| **8** | Explain with a neat diagram the architecture of Hbase. | | [L2][CO5] | **[10M]** |
| **9** | Discuss the advantage of Hbase over RDBMS. | | [L2][CO5] | **[10M]** |
| **10** | Explain about IBM Big SQL? | | [L2][CO5] | **[10M]** |

**Prepared by**:
N.Siva, S.Shreevignesh, G. Ravi Kumar