

**Nirma University**  
**Institute of Technology**  
**Computer Engineering Department**

**Practical List**

**2CS702**

**Big Data Analytics**

**[2-0-2-3]**

**Laboratory details: (List of Experiments, Schedule, assessment policy)**

<b>Sr. NO</b>	<b>Practical Title</b>	<b>Hours</b>	<b>CLO</b>
1.	Study and explore various applications of big data in different domains. Choose one of it and study in detail, Also write down the report on different types of digital data generated in selected application. For eg: <ul style="list-style-type: none"><li>• Big Data in Retail</li><li>• Big Data in Healthcare</li><li>• Big Data in Education</li><li>• Big Data in E-commerce</li><li>• Big Data in Media and Entertainment</li><li>• Big Data in Finance</li><li>• Big Data in Travel Industry</li><li>• Big Data in Telecom</li><li>• Big Data in Automobile</li></ul>	02 Hours	3
2	Learning limitation of data analytics by applying Machine Learning Techniques on large amount of data. Write a program to read data set from any online website, excel file and CSV file and to perform <ul style="list-style-type: none"><li>a) Linear regression and logistic regression on iris dataset.</li><li>b) K-means clustering.</li><li>• Students will learn the limitation of platform and algorithm.</li></ul>	02 Hours	3
3.	Setup single node Hadoop cluster and apply HDFS commands on single node Hadoop Cluster.	04 hours	3
4.	Design MapReduce algorithms to take a very large file of integers and produce as output: <ul style="list-style-type: none"><li>a) The largest integer</li><li>b) The average of all the integers.</li><li>c) The same set of integers, but with each integer appearing only once. *</li><li>d) The count of the number of distinct integers in the input.*</li></ul>	04 hours	3
5	Apply MapReduce algorithms to find phrase frequency from given dataset. <ul style="list-style-type: none"><li>• Prepare a report to guide design of mapper and reducer.</li></ul>	02 Hours	3

6	Analyse impact of different number of mapper and reducer on same definition as practical 4. <ul style="list-style-type: none"> <li>• Prepare a conclusive report on analysis.</li> </ul>	02 Hours	3
7	Implement any one of the analytic algorithm using mapreduce by handling larger datasets in main memory. <ul style="list-style-type: none"> <li>• PCY/Multi-Hash/SON algorithm</li> <li>• Regression</li> <li>• K-means Clustering</li> </ul>	04 Hours	3
8	Setup MongoDB environment in your system. Import Restaurant Dataset and perform CRUD operation.	02 Hours	3
9	Setup Cassandra environment in your system and apply Create, Update, Read and Delete operations.	04 Hours	3
10	Case study: Use following platforms for solving any big data analytic problem of your choice. (1) Amazon web services,(2) Microsoft Azure, (3)Google App engine	02 Hours	3
11*	Extend MongoDB functionality for MapReduce on document collection	02 Hours	3
12*	Extend Cassandra functionality for Map Reduce on restaurant dataset.	02 Hours	3

**Nirma University**  
**Institute of Technology**

**Computer Engineering Department**

**Lesson Planning**

**2CS702**

**Big Data Analytics**

**[2-0-2-3]**

<b>Lecture No.</b>	<b>Topic</b>	<b>Mapped CLO</b>
<b>1</b>	<b>Introduction to Data Analytics, teaching scheme , Evaluation methodology and overall instructions</b>	-
<b>2</b>	Nature of Data, Types of Digital Data, Classification of Digital Data, Structured Data, Semi-Structured Data , Unstructured Data, Characteristics of Data, need for data analytics	<b>CLO1</b>
<b>3</b>	Introduction to Big Data, Evolution of Big Data , Definition of Big Data Challenges of Conventional Systems	CLO2
<b>4</b>	Intelligent Data Analysis, Challenges of Big Data Analytic Processes and Tools, Analysis vs Reporting, ,	<b>CLO3</b>
<b>5</b>	Statistical Concepts	<b>CLO4</b>
<b>6</b>	Sampling Distributions, Re-Sampling, Statistical Inference - Prediction Error	<b>CLO1</b>
<b>7</b>	importance of Big data analytics, Sudden Hype Around Big Data Analytics, ,	CLO2
<b>8</b>	Classification of Analytics	<b>CLO2</b>
<b>9</b>	Top Challenges Facing Big Data, Kind of Technologies to meet the Challenges Posed by Big Data	<b>CLO2</b>
<b>10</b>	Data Science, Role of data scientist, Terminologies Used in Big Data Environment	<b>CLO1</b>
<b>11</b>	Hadoop: Introducing Hadoop, comparisons of RDBMS and Hadoop, ,	CLO3
<b>12</b>	Distributed Computing Challenges, A Brief History of Hadoop, Hadoop Overview	<b>CLO2</b>
<b>13</b>	Business Value of Hadoop, Hadoop Distributors, Map Reduce	<b>CLO3</b>
<b>14</b>	Hadoop Distributed File System	<b>CLO3</b>
<b>15</b>	Processing Data with Hadoop ,	<b>CLO4</b>
<b>16</b>	Applications in Hadoop	<b>CLO4</b>
<b>17</b>	Introduction to Map reduce, working of Map reduce, , ,	<b>CLO3</b>
<b>18</b>	Hadoop YARN , Hadoop Ecosystem	<b>CLO3</b>
<b>19</b>	HDFS, Hadoop in the Cloud	<b>CLO2</b>
<b>20</b>	The Big data technology landscape, NoSQL Vs SQL	<b>CLO4</b>
<b>21</b>	NoSQL, Types of No SQL databases	<b>CLO4</b>

<b>22</b>	why No SQL for big data Analytics	<b>CLO1</b>
<b>23</b>	<b>NoSQL databases</b>	<b>CLO3</b>
24	Introduction to MongoDB, Introduction to MongoDB	CLO4
25	Apache Cassandra, Features of Cassandra,	CLO4
26	Applications on Big Data Using Pig and Hive,	CLO4
27	Data Processing Operators in Pig, Pig Scripting language	<b>CLO4</b>
28	Hive Services, HiveQL , Querying Data in Hive,	<b>CLO4</b>
29	Fundamentals of HBase	<b>CLO4</b>
30	Fundamentals of Zookeeper	<b>CLO4</b>
	<b>Total Hours:</b>	<b>30 hrs</b>