### B.TECH. CSE with specialization in Big Data Analytics

<table>
<thead>
<tr>
<th>Departmental Elective-I</th>
<th>Introduction to Big data (BCS 049)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Departmental Elective-II</td>
<td>Cloud computing and virtualization (BCS 059)</td>
</tr>
<tr>
<td>Departmental Elective-III</td>
<td>Analytics and statistical modeling for big data (BCS 069)</td>
</tr>
<tr>
<td>Departmental Elective-IV</td>
<td>Application Development in Cloud (BCS 078)</td>
</tr>
<tr>
<td>Open Elective</td>
<td>Machine Learning (BOE 078)</td>
</tr>
<tr>
<td>Departmental Elective-V</td>
<td>Data Visualization (BCS 088)</td>
</tr>
<tr>
<td>Departmental Elective-VI</td>
<td>Hadoop and MapReduce (BCS 091)</td>
</tr>
</tbody>
</table>
# Department Elective-I

**BCS-049: Introduction to Big data**

<table>
<thead>
<tr>
<th>Unit</th>
<th>Credits</th>
<th>Hours</th>
<th>Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>UNIT I</td>
<td>08</td>
<td>08</td>
<td>Introduction to Big Data Analytics: Big Data overview, State of the practice in analytics role of data scientists, Big Data Analytics in industry verticals.</td>
</tr>
<tr>
<td>UNIT II</td>
<td>08</td>
<td>08</td>
<td>End-to-end Data Analytics Life Cycle: key roles for successful analytic project, main phases of life cycle, developing core deliverables for stakeholders.</td>
</tr>
<tr>
<td>UNIT III</td>
<td>08</td>
<td>08</td>
<td>Basic Analytic Methods: introduction to “R”, analyzing and exploring data with “R”, statistics for model building and evaluation</td>
</tr>
<tr>
<td>UNIT IV</td>
<td>08</td>
<td>08</td>
<td>Advanced Analytics and Statistical Modeling for Big Data: Naïve Bayseian Classifier, K-means Clustering, Association Rules, Decision Trees, Linear and Logistic Regression, Time Series Analysis, Text Analytics;</td>
</tr>
<tr>
<td>UNIT V</td>
<td>08</td>
<td>08</td>
<td>Technology and Tools – MapReduce/Hadoop, In- database Analytics, MADlib and advanced SQL Tools</td>
</tr>
</tbody>
</table>

**References:**

1. Noreen Burlingame ,The little book on Big Data, New Street publisher(eBook)
Department Elective-II

*BCS-059: Cloud Computing and Virtualization*

L T P Credit
3 1 2 4

UNIT-I 08 hours

UNIT-II 08 hours
Cloud Applications: Technologies and the processes required when deploying web services; Deploying a web service from inside and outside a cloud architecture, advantages and disadvantages.

UNIT-III 08 hours
Cloud Services Management: Reliability, availability and security of services deployed from the cloud. Performance and scalability of services, tools and technologies used to manage cloud services deployment; Cloud Economics: Cloud Computing infrastructures available for implementing cloud based services. Economics of choosing a Cloud platform for an organization, based on application requirements, economic Constraints and business needs (e.g Amazon, Microsoft and Google, Salesforce.com, Ubuntu and Redhat)

UNIT-IV 08 hours
Application Development: Service creation environments to develop cloud based applications. Development environments for service development; Amazon, Azure, Google App.

UNIT-V 08 hours
Best Practice Cloud IT Model: Analysis of Case Studies when deciding to adopt cloud computing architecture. How to decide if the cloud is right for your requirements. Cloud based service, applications and development platform deployment so as to improve the total cost of ownership (TCO).

References
Department Elective-III

BCS-069: Analytics and statistical modeling for big data

<table>
<thead>
<tr>
<th>L</th>
<th>T</th>
<th>P</th>
<th>Credit</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

UNIT-I 08 hours

UNIT-II 08 hours

UNIT-III 08 hours
Control charts for attributes: np chart, p chart, c chart and u chart. Basis, construction and interpretation. OC and ARL curve for attribute charts.

UNIT-IV 08 hours
Product Control: Sampling inspection and 100 percent inspection. AQL, LTPD, Producer’s risk and consumer’s risk. Acceptance sampling. Sampling plans-single and double sampling plans by attributes.

UNIT-V 08 hours
Reliability: Reliability concepts. Reliability of components and systems. Life distributions, reliability functions, hazard rate, common life distributions-Exponential, Gamma and Weibull. System reliability, Series, parallel, standby systems, r/n systems. Complex systems

Text Books
Reference Books
Department Elective-IV

**BCS-077: Application Development in Cloud**

**L T P**  
3 1 2  

**Credit**  
4

### UNIT-I  
08 hours

Cloud Based Applications: Introduction, Contrast traditional software development and development for the cloud. Public v private cloud apps. Understanding Cloud ecosystems – what is SaaS/PaaS, popular APIs, mobile;

### UNIT-II  
08 hours


### UNIT-III  
08 hours

Web Development Techniques and Frameworks : Building Ajax controls, introduction to Javascript using JQuery, working with JSON, XML, REST. Application development Frameworks e.g. Ruby on Rails , .Net, Java API's or JSF; Deployment Environments – Platform As A Service (PAAS) ,Amazon, vmForce, Google App Engine, Azure, Heroku, AppForce

### UNIT-IV  
08 hours

Use Case 1: Building an Application using the LAMP stack: Setting up a LAMP development environment. Building a simple Web app demonstrating an understanding of the presentation layer and connectivity with persistence.

### UNIT-V  
08 hours

Use Case 2: Developing and Deploying an Application in the Cloud: Building on the experience of the first project students will study the design, development, testing and deployment of an application in the cloud using a development framework and deployment platform.

### References:

Open Elective

Machine Learning

L T P Credit
3 1 2 4

UNIT-I 08 hours
Basics: Introduction to machine learning - different forms of learning; Basics of probability theory, linear algebra and optimization.

UNIT-II 08 hours
Regression Analysis: Linear regression, ridge regression, Lasso, Bayesian regression, regression with basic functions.

UNIT-III 08 hours
Classification Methods: Linear Discriminant Analysis, Logistic regression, Perceptrons, Large margin classification, Kernel methods, Support Vector Machines. Classification and Regression Trees, Multi-layer Perceptrons and Back propagation
Graphical Models: Bayesian Belief Networks, Markov Random Fields, Exact inference methods, approximate inference methods.

UNIT-IV 08 hours
Ensemble Methods: Boosting - Adaboost, Gradient Boosting; Bagging - Simple methods, Random Forest.
Computational Learning Theory: PAC Learning, VC Dimension, Bias/Variance Tradeoff.

UNIT-V 08 hours
Clustering: Partitional Clustering - k-means, k-medoids; Hierarchical Clustering - Agglomerative, Divisive, Distance measures; Density based clustering - DBScan; Spectral clustering.
Frequent Pattern Mining: Apriori Algorithm; FP-Growth

References:

1. Elements of Statistical Learning. Hastie, Tibshirani, and Friedman. Springer
3. Data Mining: Tools and Techniques, 3rd Edition. Jiawei Han and Michelline Kamber.
Department Elective-V

BCS-098: Data Visualization

<table>
<thead>
<tr>
<th>L</th>
<th>T</th>
<th>P</th>
<th>Credit</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

UNIT-I


UNIT-II


UNIT-III


UNIT-IV


UNIT-V

**Text Visualization**, Information Visualization for Search Interfaces, Information Visualization for Text Analysis, Interpretation and Trust: Designing Model-Driven Visualizations for Text Analysis.

**References**

2. Interactive Data Visualization for the Web by Scott Murray, O'Reilly Media, 2012.
Department Elective-VI

**BCS-092: Hadoop and MapReduce**

<table>
<thead>
<tr>
<th>L</th>
<th>T</th>
<th>P</th>
<th>Credit</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

**UNIT-I**
08 hours

Introduction to big data and Hadoop, Hadoop Architecture, Installing Ubuntu with Java 1.8 on VM Workstation 11, Hadoop Versioning and Configuration Single Node Hadoop, installation on Ubuntu, Multi Node Hadoop, Linux commands and Hadoop commands, Cluster architecture and block placement, Modes in Hadoop, Local Mode, Pseudo Distributed Mode, Fully Distributed Mode, Hadoop Daemon, Master Daemons (Name Node, Secondary Name Node, Job Tracker), Slave Daemons (Job tracker, Task tracker) Task Instance, Hadoop HDFS Commands, Accessing HDFS, CLI Approach, Java Approach.

**UNIT-II**
08 hours

Map-Reduce, Understanding Map Reduce Framework, Inspiration to Word-Count Example, Developing Map-Reduce Program using Eclipse Luna, HDFS Read-Write Process, Map-Reduce Life Cycle Method, Serialization (Java), Datatypes, Comparator and Comparable (Java), Custom Output File, Analysing Temperature dataset using Map-Reduce, Custom Partitioner & Combiner, Running Map-Reduce in Local and Pseudo Distributed Mode.

**UNIT-III**
08 hours

Advanced Map-Reduce, Enum (Java), Custom and Dynamic Counters, Running Map-Reduce in Multi-node Hadoop Cluster, Custom Writable, Site Data Distribution, Using Configuration, Using DistributedCache, Using stringifie, Input Formatters, NLine Input Formatter, XML Input Formatter,

**UNIT-IV**
08 hours

Sorting, Primary Reverse Sorting, Secondary Sorting, Compression Technique, Working with Sequence File Format, Working with AVRO File Format, Testing MapReduce with MR Unit, Working with NYSE DataSets, Working with Million Song DataSets, Running Map-Reduce in Cloudera Box

**UNIT-V**
08 hours

Hive Introduction & Installation, Data Types in Hive, Commands in Hive, Exploring Internal and External Table, Partitions, Complex data types, UDF in Hive, Built-in UDF, Custom UDF, Thrift Server, Java to Hive Connection, Joins in Hive, Working with HWI, Bucket Map-side Join, More commands, View, SortBy, Distribute By, Lateral View, Running Hive in Cloudera.

**References**