BIG DATA ANALYTICS

IV Year – I Semester

Lecture	: 3	Internal Marks	: 30
Credits	: 3	External Marks	: 70

Course Objectives

- To optimize business decisions and create competitive advantage with Big Data analytics.
- To introduce the architectural concepts of Hadoop, HDFS and Map Reduce paradigm.
- To introduce programming tools Pig , Hive in Hadoop echo system.

Course Outcomes

Upon successful completion of the course, the students will be able to

- illustrate the importance of big data and challenges of conventional systems.
- outline the building blocks of hadoop and basic file system operations.
- analyze data with hadoop Map Reduce framework.
- process the data in hadoop environment using Pig and Hive to solve real world and industrial problems.
- enumerate the open source frameworks used to efficiently store and process large data sets.

Course Content

UNIT – I: Introduction to Big Data

What is big data, Meet Hadoop – Data, Characteristics of Big Data , Data Storage and Analysis, Comparison with other systems: Relational Database Management Systems, Grid computing and Volunteer Computing.

UNIT – II: Hadoop and HDFS

Introduction to Hadoop, Brief history of Hadoop, Apache Hadoop eco system. The design of Hadoop Distributed File System (HDFS), Architecture, Building blocks of Hadoop: Namenode, Datanode, Secondary Namenode, JobTracker, TaskTracker, Basic File System Operations.

UNIT – III: Map Reduce

Java Map Reduce, Introduction to Weather Dataset, Analyzing weather data with UNIX tools, Analyzing weather data with Map and Reduce, Word Count Program using Map Reduce, Combiner Functions, Running a Distributed Map Reduce Job, Anatomy of a Map Reduce Job Run, Shuffle and Sort.

UNIT – IV: Pig - Hadoop Programming Made Easier

Admiring the Pig Architecture, Going with the Pig Latin Application Flow, Working through the ABCs of Pig Latin, Uncovering Pig Latin structures, Looking at Pig

data types and syntax, Evaluating Local and Distributed Modes of Running Pig scripts, Checking Out the Pig Script Interfaces, Scripting with Pig Latin.

UNIT – V: Hive –A Data Warehouse in Hadoop

What is Hive? The Hive shell, Hive Services, The Metastore, Comparison with traditional Databases, HiveQL, Data types, Operators and Functions, Tables, Managed tables and External Tables, Partitions and Buckets, Importing data, Altering Tables, Dropping Tables,. Querying Data, Sorting and Aggregating, Joins, Subqueries, Views, What is UDF ? Types of Hive UDFs.

Text Books

1. Tom White, "Hadoop: The Definitive Guide", O'reilly Media, Fourth Edition, 2015.
2. Dirk deRoos, Paul C. Zikopoulos, "Hadoop for Dummies" John Wiley & Sons, Inc., 2014.

Reference Books

1. Paul Zikopoulos, Chris Eaton, "Understanding Big Data Analytics for Enterprise Class Hadoop and Streaming Data", 1st edition, TMH.
2. Chuck Lam, "Hadoop in Action", 1st edition, Manning Publications.

Web Links

1. Hadoop:http://hadoop.apache.org/
2. Hive:https://cwiki.apache.org/confluence/display/Hive/Home
3. Piglatin: http://pig.apache.org/docs/r0.7.0/tutorial.html

* * *

DEEP LEARNING

IV Year – I Semester

Lecture	: 3	Internal Marks	: 30
Credits	: 3	External Marks	: 70

Course Objectives

- To introduce the basic concepts of deep learning.
- To explore various deep learning algorithms.

Course Outcomes

Upon successful completion of the course, the students will be able to

- outline a feed forward neural network to solve classification problems.
- train the neural network using backpropagation algorithm.
- use deep neural networks to solve real life problems.
- solve classification and pattern problems using Probabilistic Neural Networks.
- apply neural networks in Object recognition, sparse coding, computer vision, and natural language processing.

Course Content

UNIT – I: Feed Forward Neural Network

Introduction: Various paradigms of learning problems, Perspectives and Issues in deep learning framework, review of fundamental learning techniques. Feed forward neural network: Artificial Neural Network, activation function, multi-layer neural network.

UNIT – II: Training Neural Network

Training Neural Network: Risk minimization, loss function, back propagation, regularization, model selection, and optimization. Deep Neural Networks: Difficulty of training deep neural networks, Greedy layer wise training.

UNIT – III: Deep Learning

Deep Feed Forward network, regularizations, training deep models, dropouts, Convolution Neural Network, Recurrent Neural Network, and Deep Belief Network.

UNIT – IV: Probabilistic Neural Networks

Hopfield Net, Boltzmann machine, RBMs, Sigmoid net, Auto encoders.

UNIT – V: Applications

Applications: Object recognition, sparse coding, computer vision, natural language processing. Introduction to Deep Learning Tools: Tensor Flow, Caffe, Theano, Torch.

Text Books

1. Goodfellow, Bengio Y., and CourvilleA., Deep Learning, MIT Press, 2016.
2. Michael Neilson, Neural network and deep learning , e-book, D Press.

Reference Books

1. Yegnanarayana, B., Artificial Neural Networks PHI Learning Pvt. Ltd, 2009.
2. Bishop, C. ,M., Pattern Recognition and Machine Learning, Springer, 2006.
3. Satish Kumar, Neural Networks: A Classroom Approach, Tata McGraw-Hill Education, 2004.
4. Raúl Rojas ,Neural Networks: A Systematic Introduction, 1996.

* * *

Professional Elective - III

CLOUD COMPUTING

IV Year – I Semester

Lecture	: 3	Internal Marks	: 30
Credits	: 3	External Marks	: 70

Course Objectives

- To provide the architectural concepts of Cloud computing.
- To familiarize with cloud service models and cloud based applications.

Course Outcomes

Upon successful completion of the course, the students will be able to

- describe the stages in historical evolution of cloud computing.
- use suitable cloud services to define cloud for the enterprise.
- demonstrate hardware level and OS level virtualization to implement virtual machines.
- design machine images, web applications and databases for virtual machines.
- apply data, network and host security for the cloud.

Course Content

UNIT – I: Cloud Computing

Introduction, cloud computing: What it is and what it isn't, from collaboration to the cloud : A short history of cloud computing, the network is the computer: How cloud computing works, understanding cloud architecture , storage, services; The pros and cons of cloud computing. Who benefits from cloud computing? who shouldn't be using cloud computing.

UNIT – II: Defining Clouds for the Enterprise

Storage-as-a-Service, Database as-a-Service, Information-as-a-Service, Process as-a-Service, Application-as-a-Service, Platform-as-a-Service, Security-as-a service, Infrastructure-as-a-Service.

UNIT – III: Virtual Machines and Virtualization

Implementation levels of virtualization: levels of virtualization implementation, VMM design requirements and providers, virtualization support at the OS level, virtualization structures/tools and mechanisms: Hypervisor and Xen architecture, binary transition with full virtualization, para- virtualization with compiler support.

UNIT – IV: Hardware Virtualization and Ready for the Cloud

Hardware Virtualization: Virtualization of CPU, memory and I/O devices: Hardware support for virtualization, CPU virtualization, memory virtualization, I/O virtualization.

Ready for the cloud: Web application design, machine image design, privacy design, database management: clustering or replication? primary key management, database backups.

UNIT – V: Security

Data Security: data control, encrypt everything, regulatory and standards compliance; Network Security: firewall rules, network intrusion detection; Host Security: system hardening, antivirus protection, host intrusion detection, data segmentation, credential management, Compromise response.

Text Books

1. Kai Hwang, Jack Dongarra and Geoffrey C.Fox, "Distributed and Cloud Computing: From Parallel Processing to the Internet of Things", 1st edition, Morgan Kaufman Publications.
2. George Reese, "Cloud Application Architectures: Building Applications and Infrastructure in the Cloud", 1st edition, O'Reilly.

Reference Books

1. Michael Miller, "Cloud Computing- Web Based Applications That Change the Way You Work and Collaborate Online", 1st edition, Que publications.
2. David S. Linthicum, "Cloud Computing and SOA Convergence in Your Enterprise: A Step-by-Step Guide" Addison Wesley.

* * *

Professional Elective - IV

REINFORCEMENT LEARNING

IV Year – I Semester

Lecture	: 3	Internal Marks	: 30
Credits	: 3	External Marks	: 70

Course Objectives

- To provide the fundamentals of Reinforcement learning.

Course Outcomes

Upon successful completion of the course, the students will be able to

- enumerate the elements of Reinforcement Learning.
- solve the n-armed Bandit problem.
- compare different Finite Markov Decision Process.
- discuss about Monte Carlo Methods in solving real world problems.
- list the Applications and Case Studies of Reinforcement Learning.

Course Content

UNIT – I: The Reinforcement Learning Problem

Reinforcement Learning, Examples, Elements of Reinforcement Learning, Limitations and Scope, An Extended Example: Tic-Tac-Toe, Summary, History of Reinforcement Learning.

UNIT – II: Multi-arm Bandits

An n-Armed Bandit Problem, Action-Value Methods, Incremental Implementation, Tracking a Nonstationary Problem, Optimistic Initial Values, Upper-Confidence-Bound Action Selection, Gradient Bandits, Associative Search (Contextual Bandits).

UNIT – III: Finite Markov Decision Processes

The Agent–Environment Interface, Goals and Rewards, Returns, Unified Notation for Episodic and Continuing Tasks, The Markov Property, Markov Decision Processes,

Value Functions, Optimal Value Functions, Optimality and Approximation.

UNIT – IV: Monte Carlo Methods

Monte Carlo Prediction, Monte Carlo Estimation of Action Values, Monte Carlo Control, Monte Carlo Control without Exploring Starts, Off-policy Prediction via Importance Sampling, Incremental Implementation, Off-Policy Monte Carlo Control, Importance Sampling on Truncated Returns.

UNIT – V: Applications and Case Studies

TD-Gammon, Samuel's Checkers Player, TheAcrobot, Elevator Dispatching, Dynamic Channel Allocation, Job-Shop Scheduling.

Text Books

1. Richard S. Sutton and Andrew G. Barto, “Reinforcement Learning-An Introduction”, 2nd Edition, The MIT Press, 2018
2. Marco Wiering, Martijn van Otterlo Reinforcement Learning: State-of-the-Art (Adaptation, Learning, and Optimization (12)) 2012th Edition.

Reference Books

1. Vincent François-Lavet , Peter Henderson , Riashat Islam, An Introduction to Deep Reinforcement Learning (Foundations and Trends(r) in Machine Learning) , 2019.

* * *

Professional Elective - V

DATA VISUALIZATION TECHNIQUES

IV Year – I Semester

Lecture	: 3	Internal Marks	: 30
Credits	: 3	External Marks	: 70

Course Objectives

- Familiarize students with the basic and advanced techniques of information visualization and scientific visualization.
- To learn key techniques of the visualization process
- A detailed view of visual perception, the visualized data and the actual visualization, interaction techniques.

Course Outcomes

Upon successful completion of the course, the students will be able to

- describe the key techniques and theory behind data visualization
- classify various visualization systems and describe the methods used to spatial data to graphical depictions.
- use effectively the various visualization structures like tables, trees ,graphs/ networks, Text and Document.
- identify a wide variety of available visualization systems and its key features.
- analyze and Distinguish between visualization techniques for Line, Point and Region.

Course Content

UNIT – I: Introduction

Introduction: What is Visualization, Relationship between Visualization and Other Fields, The Visualization Process, The Scatterplot.

Data Foundations: Types of Data, Structure within and between Records, Data Pre-processing.

UNIT – II: Visualization Techniques for Spatial Data

Visualization Foundations: Semilogy of Graphical Symbols, The Eight Visual Variables, Taxonomies.

Visualization Techniques for Spatial Data: One-Dimensional Data, Two-Dimensional Data, Three-Dimensional Data, Dynamic Data.

UNIT – III: Visualization Techniques for Trees, Graphs , Networks, Text and Document

Visualization Techniques for Trees, Graphs andNetworks : Displaying Hierarchical Structures, Displaying Arbitrary Graphs / Networks.

Text and Document Visualization: Levels of Text Representations, Single Document Visualizations, Document Collection Visualizations.

UNIT – IV: Interaction Concepts and Techniques

Interaction Concepts: Interaction Operators, Interaction Operands and Spaces.

Interaction Techniques: Screen Space, Object Space, Data Space.

UNIT – V: Visualization Techniques for Time-Oriented Data and Multivariate Data

Visualization Techniques for Time-Oriented Data: Characterizing Time-Oriented Data, Visualizing Time-Oriented Data.

Visualization Techniques for Multivariate Data: Point-Based Techniques, Line-Based Techniques, Region-Based Techniques.

Text Books

1. Ward, Grinstein Keim, Interactive Data Visualization: Foundations, Techniques, and Applications. Natick: A K Peters, Ltd.
2. E.Tufte, The Visual Display of Quantitative Information, Graphics Press.

Reference Books

1. Chun-houh Chen Wolfgang Hardle Antony, Handbook of Data Visualization, Springer.

* * *

BIG DATA ANALYTICS LAB

IV Year – I Semester

Practical	: 2	Internal Marks	: 15
Credits	: 1	External Marks	: 35

Course Objectives

- To familiarize the basic concepts of Hadoop and its eco system.
- To develop programs using Map Reduce, PIG and HIVE.

Course Outcomes

Upon successful completion of the course, the students will be able to

- apply suitable LINUX commands to work in Hadoop environment.\
- use HDFS file structure and Map Reduce framework to solve complex problems.
- analyze data using Pig and Hive.

List of Experiments

1. Practice on basic Linux commands.
2. Implement the following file management tasks in Hadoop:
   a. Adding files and Directories
   b. Retrieving files
   c. Deleting files
   d. Copying files from local filesystem to HDFS and vice versa.
   e. Moving files
3. Write driver code, mapper code, reducer code to count number of words in a given file. (Hint: WordCount Map- Reduce Program)
4. Write a MapReduce program that mines weather data. Weather sensors collecting data every hour at many locations across the globe gather a large volume of log data, which is a good candidate for analysis with MapReduce, since it is semi structured and record-oriented.
5. Implement Matrix Multiplication with Hadoop Map Reduce
6. Install Pig and write Pig latin scripts to Load , Store and Filter data
7. Write Pig Latin scripts to perform data processing operations
   a. Grouping and joining data
   b. Sorting data
   c. Combining and Splitting data
8. Implement user defined functions in PIG

9. Install Hiveand use Hive to create databases and tables
    a. Create and drop databases
    b. Create, alter, and drop tables
    c. Insert, Update and delete records
10. Perform data processing operations using Hive
    a. Sort and Aggregation of data
    b. Joins
11. Perform data processing operations using Hive
    a. Views
    b. Indexes
12. Implement user defined functions in Hive

Text Book

1. Tom White , “Hadoop: The Definitive Guide “, 3rd edition, O’Reilly. Chuck Lam, “Hadoop in Action”, Manning Publications.

Reference Book

1. Alex Holmes, “Hadoop in Practice “, Manning Publications.

* * *