NSDA Code
2020/ ITES/ NIELIT/03605

QUALIFICATION FILE TEMPLATE

QUALIFICATION FILE –

CONTACT DETAILS OF THE BODY SUBMITTING THE QUALIFICATION FILE

Name and address of submitting body:

NIELIT Chennai Centre,
ISTE Complex, NO. 25, Gandhi Mandapam Road, Chennai – 600025
chennai@nielit.gov.in
Phone: 044-24421445/47

Name and contact details of individual dealing with the submission

Name Shri Sanjeev Kumar Jha
Position in the organisation Joint Director (Technical)
Address if different from above NA
Tel number(s) 044-24421445/47
E-mail address skjha@nielit.gov.in

List of documents submitted in support of the Qualifications File

1. Industry Validation (Annexure I)
2. Detailed Curriculum (Annexure II)

SUMMARY

1.	**Qualification Title:**	PG Diploma in Data Science & Analytics
2.	**Qualification Code**	
3.	**NCO code and occupation**	2512.0300, Programmer Analyst
4.	**Nature and purpose of the qualification:**	**Nature**: ❖ The Post Graduate Diploma in Data Science & Analytics a unique 6-month (840 Hours) program offered by NIELIT Chennai is an excellent blend of knowledge and practice in the field of Data Science and its industrial applications. The program is targeted for creating qualified Data Science professionals. The course progresses through the Operating System, concepts of Data and its storage, programming for data science, Big Data Technology and its implementation. Various advanced tools such as R and Python, along with MySQl, Java Programming and Hadoop Framework are used for achieving the goal of solving critical business and Analytic problems. The course has been designed after proper industry survey and consultation with multiple industry leaders to ensure that participants learn exactly what employers need. ❖ With a strong emphasis on 'learning by doing', our programs are developed with the goal of creating well-rounded, job-ready professionals that can add immediate value to any organization. ❖ On completion of the Course, the Participants will learn the concept of Data Analytics using open source statistical tools like R, Python and some very good visualization tools and techniques. They will be able to implement industry-oriented Data Science Projects. ❖ The PG Diploma Course is targeted for creating qualified Data Science professionals which will help in employment and entrepreneur development. **Purpose:** ❖ To make Data Scientists, Data Engineers and Data Analysts. ❖ This course is meant to sensitize students for computational statistics applications and usage as well as provide hands-on experience for solving real world data science issues.

5.	**Body /bodies which will Awardthe qualification:**	**Examination Cell,** National Institute of Electronics and Information Technology NIELIT Bhawan, Plot no 3, PSP Pocket, Sector-8, Dwarka, New Delhi-110077
6.	**Body which will accredit providers to offer courses leading to the qualification:**	National Institute of Electronics and Information Technology NIELIT Bhawan, Plot no 3, PSP Pocket, Sector-8, Dwarka, New Delhi-110077
7.	**Whether accreditation/affiliation norms are already in place or not, if applicable (if yes, attach a copy)**	N/A
8.	**Occupation(s) to which the qualification gives access:**	Statistical Analysts, Data Scientists, Data Analysts, Big Data Engineer,Hadoop Developer.
9.	**Job description of the occupation**	**Statistical Analysts**: Statistical Analysis **Data Scientists**: Applying data mining techniques, doing statistical analysis, and building high quality prediction systems. **Data Analysts**: Collecting and interpreting data, analysing results, Reporting the results back to the relevant members of the business, identifying patterns and trends in data sets, working alongside teams within the business or the management team to establish business needs, defining new data collection and analysis processes. **Big Data Engineer**: Development of Hadoop Environment, loadingdata from disparate data sets, pre-processing using Hive and Pig,designing solutions independently based on high-level architecture,Manage the technical communication between the survey vendor and internal systems,Maintain the production systems,collaborate with other development and research teams, building a cloud-based platform that allows easy development of new applications. **Hadoop Developer**: Design and develop Hadoop system with strong documentation skills. The job of a Hadoop developer is almost similar to the software developer but in the Big Data domain.
10.	**Licensing Requirements**	N/A [All Open Source Software will be used]

NSQC Approved

11.	**Statutory and Regulatory requirement of the relevant sector (documentary evidence to be provided)**	N/A
12.	**Proposed level of the qualification in the NSQF**	Level-8
13.	**Anticipated volume of training/learning required to complete the qualification**	840 Hours
14.	**Indicative list of training tools required to deliver this qualification**	Hardware: A rack mounted server, Desktop as per batch size, One Projector Software: Linux Operating System, Java,R, Python, Hadoopetc.
15.	**Entry requirements/ Recommendations**	ME/M.Tech/ BE/B.Tech/MCA/M.Sc/DOEACC B Level with Knowledge of Statistics and Computer Programming.
16.	**Progression from the Qualification (Professional and Academic Progression)**	Statistical Analysts, Data Scientists, Data Analysts, Big Data Engineer, Hadoop Engineer, Hadoop Developer, Hadoop Administrator /Business Analyst-Next Job Role **1) <u>Academic</u>** After completion of this course, students can go for higher studiesin specialized courses of NoSQL database like MongoDb Administrator, MongoDB Developer,Hadoop Security etc. Further a candidate can pursue research in Data Science in reputed institutes. **2) <u>Professional</u>** Now-a-days almost every big organizations are using analytics for the growth of their organization. So, there are plenty of opportunities in Data Science field. Many companies are providing a good growth to their employees after a good knowledge of Data Analytics. So, it's a good opportunity for professionals in all domains to upgrade their skills and get a better opportunity in existing company itself. The freshers may join any organization as Data Science professional and gain the experience and upgrade their skills.
17.	**Arrangements for the Recognition of Prior learning (RPL)**	**1.** Presently only candidates who undergo training shall be assessed. **2.** Online examination and certification can be considered in future if the candidate possess necessary knowledge and qualification.
18.	**International Compatibility where Known.**	N/A

19.	**Date of Planned review of the Qualification**	After Every 2 Years

20. **Format Structure of the Qualification**

Mandatory Components

(A)

Title of component and identification code/NOSs/Learning outcomes	**Estimated Size** (Learning hours)	**Level**
Configuring Platform for Data Analytics	120	7
Acquiring Skills on Data Warehousing using MySQL and MongoDB	120	9
Skills on Data Analytics using R & Python	120	9
Acquiring Skills on Fundamentals of Javafor Hadoop	120	7
Implementation of Big Data Technology using Hadoop Eco System	240	9
Mini Project	120	9
Sub Total(A)	**840**	

(B) **Format Structure of the Qualification**

Optional Components

Title of component and identification code/NOSs/Learning outcomes	**Estimated Size** (Learning hours)	**Level**
----------N/A----------	----------	----------
Sub Total(B)	-	

Total (A+B) **840**

Please attach any document giving further detail about the structure of the qualification-e.g. a Curriculum or Qualification Pack.
Detail Curriculum attached at **Annexure II.**

SECTION -1

ASSESSMENT

21.	**Body/Bodies which will carry out assessment:**	**Examination Cell,** National Institute of Electronics and Information Technology NIELIT Bhawan, Plot no 3, PSP Pocket, Sector-8, Dwarka, New Delhi-110077
22.	**How will RPL assessment be managed and who will carry it out?**	Presently only candidates who undergo training shall be assessed. In future NIELIT shall consider conducting online examination and certification for those candidates possessing necessary knowledge and qualification.
23.	**Describe the overall assessment strategy and specific arrangements which have been put in place to ensure that assessment is always valid, consistent and fair and show that these are in line with the requirements of NSQF.**	The emphasis is on practical demonstration of skills & knowledge based on the performance criteria. Student is required to pass in all OUTCOMEs individually and marks are allotted. The following assessment methodologies are used. A. Written Examination B. Practical Examination &Assignments C. Project The assessment results are backed by following evidences. 1. The assessor collects a copy of the attendance for the training done under the scheme. The attendance sheets are signed and stamped by the In-charge / Head of the Training Centre. 2. The assessor verifies the authenticity of the candidate by checking the photo ID card issued by the institute as well as any one Photo ID card issued by the Central/Government. The same is mentioned in the attendance sheet. 3. The assessor assigns roll number. 4. The assessor takes photograph of all the students along with the assessor standing in the middle and with the centre name/banner at the back as evidence.

Please attach most relevant and recent documents giving further information about assessment and/or RPL. Give the titles and other relevant details of the document(s) here. Include page references showing where to find the relevant information.

24. ASSESSMENT EVIDENCE

Title of Component:

Outcomes to be assessed/NOSs to be assessed	Assessment Criteria for the outcome	Mode of Assessment		
		Marks	**Mode**	**Module**
Configuring Platform for Data Analytics	Configuring Linux Server for Data Analytics.	10	**Theory**	**Module-1**
	Acquiring skills on shell scripting	20		
	Acquiring skills on Virtualization & Cloud Configuration	20		
Acquiring Skills on Data Warehousing using MySQL and MongoDB.	Acquiring skills on data warehousing using MySQL.	15		
	Acquiring skills on MongoDB Server for NoSQL.	15		
	Acquiring skills on working with Replica Server and Sharding.	20		
	Total Marks	**100**		
Skills on Data Analytics using R & Python	Acquiring skills on Data Manipulation & Distribution Fitting	40	**Theory**	**Module-2**
	Acquiring skills on Data Mining & Bayesian Analysis	40		
	Acquiring Skills on Data Analytics using Python and MongoDB.	20		
	Total Marks	**100**		
Acquiring Skills on Fundamentals of Java for Hadoop	Acquiring basic skills of Java Programming	10		

	Acquiring skills of Inheritance and Exception Handling	10	**Theory**	**Module-3**
	Acquiring skills of Multithreading and File Handling	10		
Implementation of Big Data Technology using Hadoop Eco System	Configuring Hadoop for Big Data Analytics	20		
	Acquiring skills on working with Hadoop Eco System	25		
	Acquiring skills on implementation of Machine Learning with Hadoop Eco System	25		
	Total Marks	**100**		
Configuring Platform for Data Analytics	1. Configuring Linux Server for Data Analytics. 2. Acquiring skills on shell scripting 3. Acquiring skills on Virtualization & Cloud Configuration	20	**Practical**	**Practical-1**
Acquiring Skills on Data Warehousing using MySQL and MongoDB.	1. Acquiring skills on data warehousing using MySQL. 2. Acquiring skills on MongoDB Server for NoSQL. 3. Acquiring skills on working with Replica Server and Sharding.	25		
Skills on Data Analytics using R & Python	1. Acquiring skills on Data Manipulation & Distribution Fitting 2. Acquiring skills on Data Mining & Bayesian Analysis 3. Acquiring Skills on	45		

NSQC Approved

	Data Analytics using Python and MongoDB.				
	Total Marks	**90**			
Acquiring Skills on Fundamentals of Java for Hadoop	1. Acquiring basic skills of Java Programming 2. Acquiring skills of Inheritance and Exception Handling 3. Acquiring skills of Multithreading and File Handling	40	**Practical**	**Practical-2**	
Implementation of Big Data Technology using Hadoop Eco System	1. Configuring Hadoop for Big Data Analytics 2. Acquiring skills on working with Hadoop Eco System 3. Acquiring skills on implementation of Machine Learning with Hadoop Eco System	50			
	Total Marks	**90**			
Internal Assessment		**60**			
Project Presentation		60			
Project		100			
	Total Marks	**100**			
	Grand Total	**700**			
Means of assessment 1	Proctored online assessments (LAN and Web based), carried out using a variety of question formats applicable for the course.				
Pass/Fail					

Following Grading Scheme (on the basis of total marks) will be followed:

Grade	**S**	**A**	**B**	**C**	**D**	**Fail**
Marks Range (in %)	85 to 100	75 to 84	65 to 74	55 to 64	50 to 54	Below 50

SECTION 2

25. EVIDENCE OF LEVEL

Saturday
OPTION A

Title/Name of Qualification/Component: **PG Diploma in Data Science & Analytics**. **Level: 8**			
NSQF Domain	**Outcomes of the Qualification/Component**	**How the Job role relates to the NSQF level descriptor**	**NSQF Level**
Process	**Configuring Platform for Data Analytics**	In configuration of platform for Data Analytics candidate learns theoretical and practical skills to configure Linux operating system for data analytics applications. They will learn the skills of configuring secure shell for secure communication between different nodes. They will also learn the theoretical and practical skills of shell scripting for automating the process in Data Analytics. They learn the theoretical and practical skills for virtualization and cloud techniques to be implemented for Data analytics and Hadoop Eco System in nonroutine context.	**7**
Professional Knowledge		The participants will have in-depth knowledge in a broad context for Linux servers. They will also learn wide ranging theoretical and practical skills in configuration of cloud techniques, information security and virtualization techniques for data analytics platform. After acquiring Professional Knowledge on Configuration and proper working of Linux Operating System for Data Analytics candidate will have wide ranging knowledge (from configuration of server to cloud computing) of facts and principles adopted for configuration of servers for data analytics.	7

Professional Skill		After acquiring knowledge on Configuration of Linux Operating System for Data Analytics candidate will learn wide range of practical skills related to configuration of Linux server for data analytics. They will able to configure a wide range of applications related to data analytics	7
Core Skill		After acquiring Core skill on Configuration of Linux Operating System candidate will learn logical and mathematical skills in (shell scripting and cloud) and learn logical communication and presentation skills.	7
Responsibility		After acquiring knowledge on Configuration of Linux Operating System candidate will learn responsibility for own work and will learn how to help others in group.	5
Process	**Acquiring Skills on Data Warehousing using MySQL and MongoDB.**	After acquiring knowledge of Data warehousing using MySQL candidate will have acomprehensive theoretical knowledge of MySQL and practical skills for writing optimized query using different techniques. Apart from this they will learn theoretical knowledge of configuring MongoDB server and its different components. They will learn practical skills for configuration of standalone MongoDB Server and **server in clustered mode** with replication. They will also learn the advance concept of **Sharding** for distributed processing of big data. The participants will be aware of various techniques used in MongoDB and thus will be able to undertake self-study to configure extra latest functionality included in future versions of MongoDB from time to	8
Professional Knowledge			8
Professional Skill			8

		time. They will be able to demonstrates the entire working of MongoDB environment.	
Core Skill		After acquiring Professional skill on Data Warehousing using MySQL and MongoDBcandidate will have sufficient knowledge such that they can do any unpredictable scaling as well as they will be able to develop a MongoDB cluster server as per requirement of client	8
Responsibility			8
Process	**Skills on Data Analytics using R & Python**	After acquiring knowledge of Data Analytics using R & Python candidate will learn advanced knowledge of statistical analysis. They will be able to understand the various statistical concepts used in Data Analytics. By doing this they will have advanced knowledge and critical understanding of statistical distributions and its fitting in different scenario. They will also learn the implementation techniques for these distributions in R as well as Python. After completion of the course they will able to demonstrate the hypothesis to be used for analysis. They can develop their own statistical model for a specific analysis using existing statistical distributions.	**9**
Professional Knowledge			9
Professional Skill			9
Core Skill		After acquiring the skills on data analytics using R & Python, candidates will enhance their skills related to distribution fitting. They will be able create a data frame from existing data sources. They will also be able to identify critical region and acceptance region for a particular analysis. They will learn various packages in R and Python for making decisions more efficient	9
Responsibility			9

		After acquiring Professional skill on Data Analytics using R & Pythoncandidate can apply their skills to identify a proper test method depending upon the current situations. They will get confidence of handling complex situations related to technical activities in Data Analytics	
Process	**Acquiring Skills on Fundamentals of Java for Hadoop**	In Acquiring skills on Fundamentals of Java for Hadoopthe participant will learn fundamental concepts of Java. They will enhance their skills such that they can identify the suitable techniques of Java for smooth working of Hadoop Eco System.	5
Professional Knowledge		The participants will learn basic concepts of java useful for configuration of Hadoop Eco System. They will learn concepts of Java and their implementation in Hadoop Eco System.	5
Professional Skill		In this module participants will learn the cognitive and practical skills of object-oriented concepts and its implementation in JAVA. It will help in accomplishment of tasks related to setup of Hadoop Eco System.	5
Core Skill		In this module participants will learn inheritance, exception handling, serialization and multithreading which will help them in implementation of their mathematical skills further.	5
Responsibility		After acquiring knowledge on Basics of Java candidate will enough confident to complete own work and they will help others to understand the concepts.	5
Process	**Implementation of Big Data Technology using Hadoop Eco System**	After acquiring knowledge of Big Data Technology using Hadoop Eco System candidate will learn the detail of Hadoop Eco System. They will be able to understand the various components of Hadoop Eco System. By acquiring the detailed	9
Professional Knowledge			9
Professional Skill			9

		knowledge of Hadoop Eco System, they will get advanced knowledge and critical understanding of Hadoop architecture and usage of various components in different scenario. They will also learn the implementation techniques for these components in Big Data. After completion of the course they will able to demonstrate the working of Hadoop Eco System. They can develop their own Big Data Hadoop Cluster for a specific task.	
Core Skill		After acquiring the skills on Big Data Technology using Hadoop Eco System, candidates will enhance their skills related to analytics for big data. They will be able to create a Hadoop eco system with all related tools for big data analytics. They will be able to create a data base server for unstructured data and will be skilled with analysis of unstructured data. They will also be able to identify the suitability of various Hadoop eco system tools. After acquiring the professional skill on Hadoop eco system, they will be able to work on Pig and Hive in Combined form. After acquiring Professional skill on Hadoop Eco System, candidate can apply the techniques of Deep Learning Algorithms to optimize Data Analytics. They will get confidence of handling complex situations related to technical activities in Hadoop Environment.	9
Responsibility			9
Process	**Mini Project**	After acquiring skills on Data Analytics candidate will learn specialised techniques for design of a Hadoop cluster. They will be able to understand criticality of analytics problem. They will be skilled with Advanced Knowledge of Data Analytics. During development of project on Data Analytics a candidate will be able to implement their innovative	9
Professional Knowledge			9
Professional Skill			9

		ideas and new analytical techniques for optimization of the analytic model. They will have mastery on Hadoop Eco System and analytics. They will be able to apply new research-oriented techniques in Big Data Analytics.	
Core Skill		After acquiring skills of developing a live project in Big Data domain, candidate will enhancetheir skills of decision making in various complex scenario. They will be able to take an appropriate decision in any unpredictable situations. Their decision making and analytical capabilities will be enhanced manifold.	8
Responsibility			8

SECTION 3

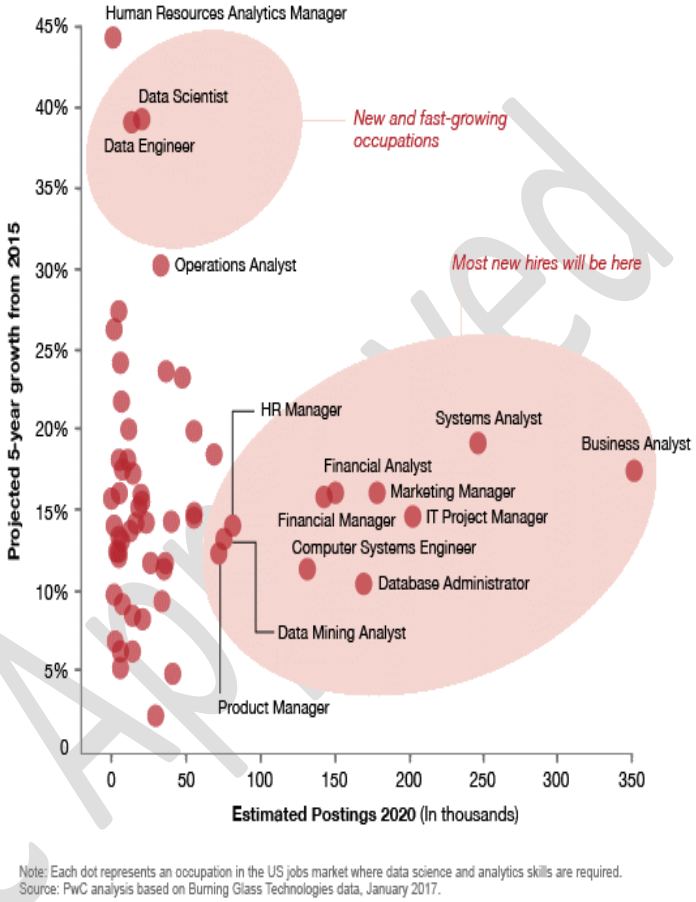
EVIDENCE OF NEED

26. What evidence is there that the qualification is needed? What is the estimated uptake of this qualification and what is the basis of this estimate?

Basis	In case of other Awarding Bodies (Institutes under Central Ministries and states departments)
Need of the qualification	Employers in New York, San Francisco, Washington D.C., Chicago,Los Angeles and India have the highest demand for candidates with data skills. Businesses in Dallas, Atlanta and Philadelphia are actively hiring for this skillset as well. And data scientists are consistently in demand in Seattle, San Jose and Boston. Companies with good streams of interesting problems and continual flow of data will have the best chance to land qualified candidates. But when the best companies search within a small community of candidates, the market gets distorted and demand pushes salaries up. We expect more employers to look beyond the hottest markets as part of their overall talent strategy. A plus: large, diverse metro areas are the likely places for performance-based workforce and education initiatives, meaning they will design their programs based on employer needs and measure success on the number of candidates they match to jobs. By 2020 there is a requirement of more than 2.7 million job postings for data science and analytics roles only in US. **Source**: *https://www.pwc.com/us/en/library/data-science-and-analytics.html* It is predicted that it will increase year by year. Now a day in industry many courses are available in Big Data Analytics. None of these courses includes the concepts of statistical analysis in their course plan. Thus, to fulfil the industry demand and gap there is a need of a course which can fulfil the industry demand. Keeping in view the above facts after an iterative procedure of syllabus modifications a course with the name "**Post Graduate**

Diploma in Data Science and Analytics" is designed by NIELIT Chennai.

The course covers detailed study in statistical analysis, storage of data on MongoDB(unstructured) & MySQL (structured) as well as machine learning and big data analytics.



Note: Each dot represents an occupation in the US jobs market where data science and analytics skills are required.
Source: PwC analysis based on Burning Glass Technologies data, January 2017.

NIELIT Chennai has started this course to meet the ever-increasing skilledmanpower requirements of the Information Communication Technologies (ICT) industry in the field of Data Analytics as well as supplement its intellectual resource base for cutting-edge research and development. Over the years NIELIT Chennai has designed and delivered various postgraduate and undergraduate diploma programmes. Inaddition, NIELITChennai imparts ICT training to state and national governments and agencies, strategic sectors,corporate and industriesbased on specific requirements.

The Education and Training activities of NIELIT Chennai are governed and steered by Academic Committee (AC). As per the Academic Committee

		direction, a syllabus is formed by data science group.After minutes of the meeting with draft syllabuscontents were circulated across all the members for any suggestion and comments. If anysuggestions come through discussion of all concerned members, we incorporate the same and circulate againfor finalization.
	Industry Relevance	Annexure I
	Estimated uptake	As it is the new course, estimated uptake of this qualification will vary as per demand in Industry. Perhaps, estimated uptake on the basis of current scenario is 150 to 200 per year. Basis of this estimate is 30% of current average count of enrolments in Php Course in all NIELIT Centres. The Data Science industry is one of the fastest growing sectors of the Indian economy and the need for trained manpower is growing. As per the reports data analytics sector in India is expected to witness eight-fold growth to reach $16 billion by 2025 from the current level of $2 billion, the National Association of Software and Services Companies (Nasscom). It strengthens the uptake for the course.
27	Recommendation from the concerned Line Ministry of the Government/Regulatory Body. To be supported by documentary evidences. **NA**	
28	What steps were taken to ensure that the qualification(s) does (do) not duplicate already existing or planned qualifications in the NSQF? Give justification for presenting a duplicate qualification The Qualification does not exist as per information available in public domain.	
29	**What arrangements are in place to monitor and review the qualification(s)? What data will be used and at what point will the qualification(s) be revised or update?** The Qualification is to be monitored and reviewed every two years. The following data will be used ❖ Results of assessments ❖ Employer feedback regarding student skill after conducting a placement drive ❖ Employer feedback will be sought post-placement ❖ Student feedbacks ❖ Workshops and seminar for reviewing the qualifications ❖ Consultation/ Tie-up with Industries or Expert for review of the Curriculum so as to meet the changing pace of technology and Industry requirements. **Please attach any documents giving further information about any of the topics**	

	above.: NIL

SECTION 4

EVIDENCE OF PROGRESSION

30	**What steps have been taken in the design of this or other qualifications to ensure that there is a clear path to other qualifications in the sector?** Please attach any documents giving further information about any of the topics above: **NA** This qualification has been designed in consultation with industry and domain expert keeping in mind today's need. Evaluation criteria have been added to ensure progression to related path ways identified as per career path.

Please attach most relevant and recent documents giving further information about any of the topics above. Give the titles and other relevant details of the document(s) here. Include page references showing where to find the relevant information.

Annexure I

No. of Students Registered: Nil (New Course)

Placement Status: Nil (New Course).

Industry Validations (Detail enclosed)

Sr. No	Name of the Organization
1	SUREKHA IT SERVICES HYFDERABAD
2	WEBTOALL PASUMALAI, MADURAI

Annexure -II

Detailed Curriculum

Name of Unit of Qualification: Fundamental of Linux & Server Configuration for Data Science and Analytics

Topics to be covered	Theory	Practical	Total Duration (In Hours)
Installation and Initialization	**01**	**03**	**04**
Basic Linux Commands	**01**	**03**	**04**
Package Management and process Monitoring	**04**	**04**	**08**
Important Files, Directories and Utilities	**01**	**03**	**04**
Shell Programming	**22**	**26**	**48**
System Services	**02**	**02**	**04**
User Administration	**03**	**05**	**08**
Virtualization	**08**	**08**	**16**
Basics of Information Security & Cloud	**06**	**18**	**24**
Total Duration	**48**	**72**	**120**

1. Installation and Initialization

Learning Objectives - In this section participants will understand Linux architecture in general, booting process, how to install Ubuntu OS, system initialization and some basic commands to get started.

Topics – Introduction, Installation, Linux Architecture, Boot Process, Kernel, System Initialization, GRUB(Modify the system boot loader), GUI, CLI(Access a shell prompt and issue commands with correct syntax, Use input-output redirection (>, >>, |, 2>, etc.), create and edit text files, delete, copy, and move files and directories, Introduction to Bash Shell, Basic Commands, Editors, Man Pages.

2. Basic Linux Commands

Learning Objectives - In this section participants will understand the commands used in Linux operating System.

Topics – vi editor, file handling commands, sort, tr, cut, find, grep, egrep, using filters, cat, mkdir, who and other basic commands.

3. Package management and Process Monitoring

Learning Objectives - In this section, participants will learn about a) Package management which includes installing, updating and removing software. b) Topics on basic networking, tools needed etc. c) CMD's to monitor system processes and resources.

Topics - Securing single-user mode (su login), Shutting down and rebooting the system, init process, /etc folder and its importance, Using apt, RPM Package Manager, Installing and Removing Software, rpm Queries, rpm Verification, About yum, Using yum, Searching packages/files, Configuring local Repositories, Process, ftp, sftp, wget. Parent processes and child processes, killing processes and sending signals to a process (kill, killall, xkill), background processes, how to start processes, monitor them, Identify CPU/memory intensive processes, adjust process priority.

4. Important Files Directories and utilities

Learning Objectives - In this section, participants will learn about how to manage services, important directories, files and configurations. Also included are topics on different data processing utilities.

Topics - Control services and Daemons, Start and stop services and configure services to start automatically at boot, /dev & /proc directories, system documentation under /usr/share/doc, use grep and regular expressions to analyze text, Archive, compress, unpack, and uncompress files using tar, star, gzip, and bzip2, dump, restore, Locate and interpret log files.

5. Shell Programming

Learning Objectives-In this section, participants will learn skills to increase their performance using shell scripts. They will learn how to create, debug, and read Bash shell scripts to customize their Linux Systems. This section will help build the requisite expertise through shell scripting to manage, operate and maintain the Linux Server.

Topics: Introduction to Shell & Script Basics: Variable basics, Predefined variables , System variables , Special shell variable, Customizing shell prompt, Variable attributes , Arrays variable , local variable , Exporting variables, Use of quotations ,Escape characters, Advance Bash Shell Scripting ,The eval command, Reading keyboard input , Basic redirection and pipe, File descriptors , Expressions, Operators, Parameter substitution and expansion, Brackets and Extended brackets test construct, Curly brackets construct , Parentheses and double parentheses, Regular expression , Extended regular expression , POSIX character classes , Globbing options , Here Documents , Logical expressions , Compound Commands , Command Status Codes , if Command , case Command , while Loop , until Loop , for Loops , Embedded let , Grouping Commands , Debugging and Version Control , Shell Debugging Features , Shell Debugging Options , Version Control (CVS) , Watching Running Scripts , Timing Execution with Time , Creating Man Pages , Shell Archives , tee Command Switches , Linux Time Command Switches , Bash and Linux Time Command Format Codes , Parameters and getopts , Positional Parameters , The getopts Command , getopts internal variables , The getopt Command , Job Control and Signals , Job Control , Signals , The suspend Command , Traps , Exit Handlers , The killall Command , Being Nice , Process

Status , Text File Basics , Working with Pathnames , File Truncation , Identifying Files , Creating and Deleting Files , Moving and Copying Files , More Information About Files , Verifying Files , Splitting Large Files , Tabs and Spaces , Temporary Files , Lock Files , Named Pipes , Process Substitution , Opening Files , Using head and tail , File Statistics , Cutting , Pasting , Columns , Folding , Joining , Merging , Text File Processing , Finding Lines , Locating Files , Finding Files , Sorting , Character Editing (tr) , Compressing Files , Console Scripting , The Linux Console , The Console Keyboard , The Console Display , select Menus , Custom Menus , Functions and Script Execution , Shell Functions , Local Variables , Recursion and Nested Functions , Function Attributes , Running Scripts , The Linux Execution Environment , The Source Command (.) , Switching Scripts with exec , Shell Security , The Basic Linux Security Model , Knowing Who You Are (id) , Transferring Ownership (chown/chgrp) , Changing Access Rights (chmod) , Default Access Rights (umask) , setuid/setgid and Scripts , The chroot Command , Resource Limits (ulimit) , Restricted Shells , Secure File Deletion

6. System Services

Learning Objectives - In this Section, participants will learn how to configure Network, Secure Communication, Schedule jobs, Secure communication and copying of files, etc.

Topics – Configure LAN on CentOS, SSH: Secure Shell(Access remote systems using ssh, SCP, Passwordless SSH, Configure key-based authentication for SSH), Cron, Controlling Access to cron, System crontab Files, Daily Cron Jobs, , crontab, Anacron Scheduler, Schedule tasks using at, batch etc.

7. User Administration

Learning Objectives - In this section, participants will learn the concepts of user administration in Linux, how to add new user account, modify and delete existing user accounts etc. We will also discuss network users and file permission techniques.

Topics - Adding a New User Account, User Private Groups, Modifying / Deleting User Accounts, Group Administration, Password Aging Policies, Switching Accounts, sudo access, Network Users, Authentication Configuration, SUID and SGID Executable, SGID Directories, The Sticky Bit, Default File Permissions, , Changing file ownership (chown), Changing file group ownership (chgrp), Permissions on files, Permissions on directories, How permissions are applied, Changing permissions (chmod).

8. Virtualisation

Learning Objectives - In this module, you will learn about Linux Virtualization.

Topics: Introduction to KVM Virtualization, Virtual Machine installation, Configuring Virtual Machines, Install Ubuntu/Centos systems as virtual guests, configure systems to launch virtual machines at boot. , Creating Clone of a Virtual Machine and its restoration, IP addressing, virtual LAN, Memory addressing, Paging, Memory mapping, virtual memory,

complexities and solutions of memory virtualization, VM configurations, VM migrations, Migration types and process.

9. Basics of Information Security & Cloud

Learning Objectives - In this section, participant will learn about basics of Information Security and Cloud

Topics: Basic of Information Security, Cloud Architecture, Cloud Storage Infrastructure, Managing Virtual Environments on Cloud, Introduction to AWS.

Name of Unit of Qualification: Data Warehousing using MySQL and MongoDB

Outline

Sl. No	Topics to be covered	Theory	Practical	Total Duration (In Hours)
1	Database Design using MySQL	**03**	**05**	**08**
2	Relational Model and SQL	**03**	**05**	**08**
3	Database design using the relational model	**03**	**05**	**08**
4	Transaction Processing and Concurrency Control in MySQL	**06**	**10**	**16**
5	Introduction to NoSQL and MongoDB	**02**	**02**	**04**
6	Creating, Updating, and Deleting Documents.	**08**	**12**	**20**
7	Query	**04**	**08**	**12**
8	Index, Special Index and Collection Types	**04**	**04**	**08**
9	Aggregation	**02**	**02**	**04**
10	Replication	**04**	**08**	**12**
11	Connecting to a Replica Set from Your Application	**02**	**02**	**04**
12	Sharding	**02**	**02**	**04**
13	Backups	**02**	**02**	**04**
14	Deploying MongoDB	**04**	**04**	**08**
Total Duration		**49**	**71**	**120**

1. Database Design using MySQL

Learning Objective: In this Section Candidate will learn concept of Database and basic Data Types used in MySQL. After completion of the section participants will be equipped with skill of Normalization and basic concepts in MySQL.

Topics: Concept of RDBMS, Normalization, Storage Engine, Structure of MySQL, Installation of MySQL, Creating Database, Data Types, Table etc.

2. Relation Model and SQL

Learning Objective: In this Section Candidate will be skilled with Relational Model, Operators keys etc. in MySQL. After completion of the section participants will be equipped with the skills of relational queries in MySQL.

Topics: Relation Model, Enum, MySQL Query, MySQL Expressions,Creating and Using a Database, Creating and Selecting a Database, Literal Values,Select, Operators,group by were,order by , Primary Key, etc.

3. Database Design using the Relational Model

Learning Objective: In this Section Candidate will be established relation between multiple tables and will be skilled with foreign key implementation.

Topics: Making Relation between tables, Foreign Key, Cascading,join etc.

4. Transaction Processing and Concurrency Control

Learning Objective: In this Section candidate will be skilled with exporting and importing data, Store Procedure, Cursor etc. They will be able to write small application dependent store procedures.

Topics: Working of Storage Engine, Export and Import of External Data, Implementation of Concurrency Control in RDBMS through different techniques, Interacting with different tables, Store Procedure, Cursors etc., Backup and Recovery.

5. Introduction to NoSQL and MongoDB

Learning Objective: In this Section candidate will be skilled with introduction of NoSQL and use of MongoDB in NoSQL. They will be able to configure Mongo Server and equipped with basic elements of MongoDB.

Topics:

Introduction: Introduction to NoSQL and MongoDB, Installation of MongoDB and GUI of MongoDB

Basic Data Types: Documents, Collections, Dynamic Schemas, Mongo Shell, Mongo Server and Client, Data Types, Embedded Documents, Creating Configuration file for Mongo.

6. Creating, Updating, and Deleting Documents

Learning Objective: In this Section candidate will be skilled for adding new documents to a collection, removing documents from a collection, updating existing documents.

Topics: Inserting and Saving Documents, Batch Insert, Insert Validation, Removing Documents, Updating Documents, Document Replacement, Using Modifiers, Upserts, Updating Multiple Documents, Returning Updated Documents.

7. Query

Learning Objective: In this Section candidate will be skilled with writing Query for MongoDB. After completion of the section participants can perform queries on the database using the find or findOne. They will be skilled for set inclusion, inequalities, cursors etc.

Topics: Introduction to find, Query Criteria, Query Conditionals, Conditional Semantics, Type-Specific Queries, Regular Expressions, Querying Arrays, Querying on Embedded Documents, Cursors, Limits, Skips, Advanced Query Options,Getting Consistent Results Immortal Cursors.

8. Index, Special Index and Collection Types

Learning Objective: In this Section candidate will be skilled with indexing and its usage. After completion of the section participants will be able to use special indexes and various capped Collections and GridFs for storing images and other large unstructured data.

Topics: Introduction to Indexing, Introduction to Compound Indexes, Using Compound Indexes, Indexing Objects and Arrays , Index Cardinality, Using explain() and hint(), The Query Optimizer, Index Administration, Changing Indexes, Capped Collections, Geospatial Indexing Storing Files with GridFS , Getting Started with GridFS: mongofiles, Working with GridFS from the MongoDB Drivers

9. Aggregation

Learning Objective: In this Section candidate will be skilled with the aggregation framework and MapReduce techniques in MongoDB

Topics: The Aggregation Framework, Pipeline Operations, $match, $project, $group, $unwind, $sort, $limit, $skip, Using Pipelines, MongoDB and MapReduce, Aggregation Commands.

10. Replication

Learning Objective: In this Section candidate will be skilled with the configuration of replica server on MongoDB. After completion of the section participants will learn replica management on MongoDB.

Topics: Introduction to Replication, Configuring a Replica Set, Networking, Elections, Member Configuration Options, Creating Election Arbiters, Priority, Heartbeats.

11. Connecting to a Replica Set from Your Application

Learning Objective: In this Section candidate will be skilled with working of connections, failovers and routing in replica server. After completion of the section participants will be equipped with the skills of replica customization.

Topics: Client-to-Replica-Set, Custom Replication Guarantees, Guaranteeing One Server per Data Centre, Sending Reads to Secondaries, Consistency Considerations, Load Considerations and reasons to read from Secondaries 206

12. Sharding

Learning Objective: In this Section candidate will be skilled with introduction and importance of sharding in MongoDB. After completion of the section participants will be able to configure sharding on MongoDB.

Topics: Introduction to Sharding, Config Servers, The mongos Processes,Adding a Shard from a Replica Set, Shard Keys,Hashed Shard Keys for GridFS , Shard Key.

13. Backups

Learning Objective: In this Section candidate will be skilled with the concept of backup for MongoDB. After completion of the section participants will be skilled with different backup techniques on MongoDB.

Topics: Backing Up a Server, Filesystem Snapshot, Copying Data Files,Using mongodump, Backing Up a Replica, Backing Up a Sharded Cluster, Backing Up and Restoring an Entire Cluster, Backing Up and Restoring a Single Shard, Creating Incremental Backups with mongooplog.

14. Deploying MongoDB

Learning Objective: In this Section candidate will be skilled with the methods for deploying MongoDB. After completion of the section participants will able to select suitable hardware to and its set it up using virtualized environments with proper Network setup.

Topics: Designing the System, Choosing a Storage Medium, Swap Space, Filesystem, Virtualization, Handling Network Disk IO Issues, Configuring proper Network.

Name of Unit of the Qualification: Data Analytics using R & Python

Outline

Sl.No	Topics to be covered	Theory	Practical	Total Duration (In Hours)
R				
1	Basic Concept of Data Analytics &Data Manipulation in R	**06**	**10**	**16**
2	Statistical Distribution using R	**08**	**08**	**16**
3	Testing of Hypothesis and Goodness of Fit Test using R	**04**	**04**	**08**
4	Data Mining using R	**16**	**20**	**36**
5	Bayesian Analysis in R	**02**	**02**	**04**
Total		**36**	**44**	**80**
Python				
6	Python Basics	02	02	04
7	OOPs concept &Exception Handling in Python	02	04	06
8	Data Analysis in Python	04	04	08
9	Inferential Statistics in Python	04	04	08
10	Data Visualisation using Python	02	04	06
11	MongoDB - Python Interaction	02	02	04
12	Time Series Analysis using Python	02	02	04
		18	**22**	**40**

1. Basic Concept of Data Analytics & Data Manipulation in R

Learning Objective: In this section candidate will learn the concept of Data Analysis and R. After completion of the section participants will be skilled with data manipulation techniques of R.

*Topics:*Data Analysis, process of data analysis,Basic Features of R,Installation R Studio and method of accessing through URL, Basic Data Sets: Vector, Matrices, List, Array, Factors, Data Frames, Data Types, Operators, Basic Constructs, R Functions, String Handling, R Packages, Data Reshaping, Data Pipelines, Data Manipulation

2 **Statistical Distribution using R**

Topics: Random Variable, Distribution Functions (Discrete and Continuous), Measure of Central Tendency, Deviations, Central Limit Theorem, Proportions, Covariance, Correlation, Estimation, Interval Estimation

3**Testing of Hypothesis and Goodness of Fit Test using R**

*Topics:*Null and Alternative Hypothesis, Critical and Acceptance Region, P Value, Type I and Type II error, Testing Mean, Chi Squared Test of Independence, Goodness of Fit Tests.

4 **Data Mining using R**

Topics:

4.1 *Introduction:* Basic Concepts of Data Mining with R.

4.2 *Data Import and Export*: Save and Load R Data, import from and Export to .csv Files, Import/Export via ODBC: Read from Databases, Output to and Input from EXCEL Files

4.3 *Data Exploration:*Explore Individual Variables, Explore Multiple Variables, More Explorations, Save Charts into Files

4.4 *Regression Analysis in R*: Linear Regression, Generalized Linear Regression,Logistic Regression, Multiple regression, Poisson Regression,

4.5 *Clustering:* The k-Means Clustering, The k-Medoids Clustering, Hierarchical Clustering, Density-based Clustering

4.6*Outlier Detection:* Univariate Outlier Detection, Outlier Detection by Clustering, Outlier Detection from Time Series

4.7 *Time Series Analysis and Mining:*Time Series Data in R, Time Series Decomposition, Time Series Forecasting, Time Series Clustering, Dynamic Time Warping, Synthetic Control Chart Time Series Data, Hierarchical Clustering,Time Series Classification.

4.8 *Association Rules:* Basics of Association Rules, The Titanic Dataset, Association Rule Mining, Removing Redundancy, Interpreting Rules, Visualizing Association Rules

4.9 *Text Mining:* Retrieving Text from Twitter, Transforming Text, Stemming Words, Building a Term-Document Matrix, Frequent Terms and Associations, Word Cloud, Clustering Words, Clustering Tweets with the k-means Algorithm, Clustering Tweets with the k-medoids Algorithm

4.10 *Social Network Analysis:*Network of Terms, Network of Tweets, Two-Mode Network

4.11 *Decision Tree:* Tree, Decision Tree, Regression and Classification Decision Tree, Gini Index, Splits, Entropy, Reduction in Variance, CART, Implementation of Decision Tree

4.11 *Case Study I:* Analysis and Forecasting of House Price Indices: Importing HPI Data, Exploration of HPI Data, Trend and Seasonal Components of HPI, HPI Forecasting, The Estimated Price of a Property

4.12 *Case Study II*: Customer Response Prediction and Profit Optimization: Introduction, The Data of KDD Cup 1998, Data Exploration, Training Decision Trees, Model Evaluation, Selecting the Best Tree, Scoring

5 **Bayesian Analysis in R**

Topics: Conditional Probability, Bayesian Prediction

6**Python Basics**

Topics: Features of Python, Configuration of Python on Ubuntu, setting up path, Working with Python, Basic Syntax, Variable and Data Types, Operators, Collections, Conditional Statement, Loops, Functions, Python Lambda, File Handling in Python

7**OOPs concept & Exception Handling in Python**

Topics:

*7.1 OOPs:*Concept of class, object and instances, Constructor, class attributes and destructors, Real time use of class in live projects, Inheritance, overlapping and overloading operators, Adding and retrieving dynamic attributes of classes, Programming using Oops support

*7.2 Exception Handling:*Avoiding code break using exception handling, Safe guarding file operation using exception handling, Handling and helping developer with error code, Programming using Exception handling.

8 **Data Analysis in Python**

*Topics:*Working with Pandas data structures: Series and Data Frames, accessing your data: indexing, slicing, fancy indexing, Boolean indexing, Data wrangling, including dealing with dates and times and missing datas, Adding, dropping, selecting, creating, and combining rows and columns, Pandas:XLS, Pandas: JSON, Missing Value, Data Aggregation, group by, Data Frame Operations and Joins,

DS 9**Inferential Statistics in Python**

*Topics:*SciPy package, Measure of Central tendency, Discrete Distributions, Z Score, Level of Significance, P Value, Testing of Hypothesis, Goodness of fit,Split-apply-combine with Data Frames, Data summarization and aggregation methods, Pandas group by method, Reshaping and transforming your data, Correlation,Outlier Identification, Regression analysis, Distribution Fitting, Testing of Hypothesis, Simple and rolling statistics

10 **Data Visualization using Python**

*Topics:*Understanding on Data Visualization, Graphical functions present ,Pie Chart, Histogram, Box Plot etc.

11**MongoDB - Python Interaction**

*Topics:*Insert, Read, Update, delete in MongoDB using PyMongo,Python Data Operations

Python Data cleansing, Python Processing CSV Data, Python Processing JSON Data, Python Processing XLS Data, Python Relational databases, Python NoSQL Databases, Python Processing Unstructured Data

12 **Time Series Analysis using Python**

Topics: Time Series Introduction, Indexing, Fitting a Time Series Model, Forecasting

Name of Unit of the Qualification: Object Oriented Programming: Java

Sl.No	Topics to be covered	Theory	Practical	Total Duration (In Hours)
1	Basic Java	04	04	08
2	Arrays, Objects and Classes	06	10	16
3	Control Flow Statements	04	04	08
4	Inheritance and Interfaces	08	08	16
5	Exception Handling& Serialization	08	08	16
6	Multithreading in Java	12	12	24
7	Collections	08	08	16
8	Reading and Writing files	08	08	16
Total Duration		**58**	**62**	**120**

1. Basic Java

Topics: Concept of OOPS, Characteristics of OOPS, Lexical Tokens, Identifiers, Keywords, Literals, Comments, Primitive Datatypes, Operators Assignments

2. Arrays, Objects and Classes

Topics: Class Fundamentals , Object & Object reference , Object Life time &Garbage Collection, Creating and Operating Objects , Constructor & initialization code block, Access Control, Modifiers, methods Nested ,Inner Class &Anonymous Classes , Abstract Class & Interfaces Defining Methods, Argument Passing Mechanism , Method Overloading, Recursion, Dealing with Static Members, Finalize() Method, Native Method, Use of Modifiers with Classes & Methods, Generic Class Types, Defining an Array, Initializing & Accessing Array, Multi –Dimensional , Array, Operation on String, Mutable & Immutable String.

3. Control Flow Statements

Topics: Conditional Statements, Loops: While, for, switch case etc.

4. Inheritance and Interfaces

Topics: Concept of Parent and Child Class in Java, Inheritance, types of Inheritance, Use of extends and Implements, Concept of Interfaces etc.

5. Exception Handling and Serialization

Topics: Exception Handling in Java – Overview, Java Exception Handling Keywords, Java Exception Hierarchy, Exception Handling in Java – Useful Methods, Automatic Resource Management and Catch block improvements, Exception Handling in Java – Creating Custom Exception Classes, Serialization and De-Serialization.

6. Multithreading in Java

Topics: Introduction, Thread Life Cycle, Thread Priority, Thread Methods, Thread Class, creating a thread, Joining a thread, Synchronization, Interthread Communication

7. Collections

Topics: Collection Framework in Java, Core Collection in Java, Core Collection framework, Types of Collection, Hierarchy of Collection Framework,Commonly used methods of Collection interface, Iterator Interface, Methods of Iterator interface.

8. Reading and Writing files

Topics: The Classes for Input and Output, The Standard, Streams, Working with File Object, File I/O Basics, Reading and Writing to Files, Buffer and Buffer Management, Read/Write Operations with File Channel, Serializing Objects.

Name of Unit of the Qualification: Big Data Technology using Hadoop and Spark

Sl. No.	Topics to be covered	Theory	Practical	Total Duration (In Hours)
1	Introduction to Big Data and Hadoop Eco System	04	-	04
2	Configuring Hadoop	08	08	16
3	HDFS Architecture	04	08	12
4	Hadoop MapReduce	08	16	24
5	Working with Sqoop	08	16	24
6	Working with Pig and HIVE	08	16	24
7	Configuring HBase	16	24	40
8	Implementation of Machine Learning for Big Data using Python	16	24	40
9	Deep Learning Applications	16	24	40
10	Apache Spark	06	10	16
Total		**94**	**146**	**240**

1.Introduction to Big Data and Hadoop Eco System

Introduction to Big Data

Topics: Big Data Introduction,Attributes of Big Data, Types of data, other technologies vs Big Data,Compare Hadoop vstraditional systems, Limitations and Solutions of existing Data.

Hadoop: Eco System

Topics: Introduction to Hadoop Ecosystem,Hadoop Cluster, Pseudo Distributed mode, Type of clusters, Hadoop Ecosystem, Introduction of Pig, Hive, Oozie, HBase, Flume, SQOOP etc.

2. Configuring Hadoop

*Topics:*InstallingJava, InstallingHadoop, Standalone mode, Pseudo Distributed Mode, Fully Distributed, Monitoring the Cluster Health, Starting and Stopping the Nodes

3. HDFS Architecture

Topics: Distributing Processing System, Core Components of Hadoop, HDFS Architecture, HDFS Design, HDFS role in Hadoop,Features of HDFS,Daemons of Hadoop and its functionality - Name node, Data node, Secondary Name Node**,** Job Tracker, Task Tracker,Anatomy of File Write, Anatomy of File Read, Network Topology, Heartbeat Signal, How to Store the Data into HDFS, How to Read the Data from HDFS, CLI commands (Hadoop FS shell)

4. Hadoop MapReduce

Topics: HDFS Java API, Overview of MapReduce Framework, MapReduce Architecture, learn about Job tracker and Task tracker, use cases of MapReduce, Anatomy of MapReduce Program, Basic MapReduce API Concepts, Writing MapReduce Driver, Mappers, andReducers in Java, Unit TestingMapReduce Programs etc, MapReduce Programming using Python.

5. Working with Sqoop

Topics: Overview of Sqoop, Sqoop Installation and configuration, importing data from RDBMS to HDFS, Conditional based import, Sqoop – to hive import, Sqoop – Export to RDBMS, Sqoop – Meta store, Sqoop – Jobs configure and execution.

6. Working with Pig and HIVE

Pig:

Topics: Installation, Architecture, Datatypes (scalar, complex), Running Pig (interactive, Batch), Pig Operators – Local, Store, Dump,Distinct, Filter, For Each, generate, Limit, Union,join, order by, Describe, Group by, Avg Default UDFs available (Built in function) REG EX, EXPLAIN, Parallel processing, Custom UDF

Hive:

Topics: Installation, Hive Services, Architecture, Comparing Hive to traditional Databases, Relational Data Analysis – (data types (primitive,complex)databasestables,create,alter, delete Hive Schema & Data storage Loading data into Hive views Storing query results (store), Text processing - Built in functions, string functions, regular expressions, Managed vs External Tables.

7. HBase

*Topics:*HBase Introduction, HBase architecture, HBase vs RDBMS (fixed Vs flexible schema), Master and Region servers, HBase commands

8 . Implementation of Machine Learning for Big Data using Python

Topics:

Introduction of Machine Learning: Basic Conceptsof Machine Learning,End-to-end Process of Investigating Datathrough a Machine Learning Len**s,**Evolution and Trends,Application ofMachine Learning, Best Practices of Machine Learning

Machine Learning Algorithms: Classification,Regression, Collaborative Filtering, Clustering Principal Component Analysis

Neural Networks: Understanding Neural Networks, The Biological Inspiration,Perceptron Learning & Binary Classification, Backpropagation Learning, Learning Feature Vectors for Words, Object Recognition

9. Deep Learning Applications

Topics: KERAS for Classification and Regression in Typical Data Science Problems, setting up KERAS, Different Layers in KERAS, Creating a Neural NetworkTraining Models and Monitoring, Artificial Neural Networks, using ANN on KERA., Introducing Tensorflow, Neural Networks using Tensorflow, Debugging and Monitoring, Convolutional Neural Networks , Unsupervised Learning, using CNN on TensorFlow

10. Apache Spark

Topics: Introduction of Apache Spark, Features, Data Source, Data Sets, Data Frames,Spark Configuration and Cluster Modes,Spark Transformers, Spark ML Lib,Spark ML Pipeline,SparkAlgorithms for Machine Learning, Integration with Hadoop etc