Advanced Diploma in Big Data Analytics
(Certified Big Data Analyst)

Preamble:

In today's world there is data available in abundance from variety of sources like web server logs, social media, and large databases and from diverse domains like Ecommerce, Medical, Scientific etc. There are be a lot of useful and meaningful information lies in this voluminous data but retrieving the useful information is itself a challenge. Various government bodies, financial institutions, MNC, Corporate houses, consulting firms etc needs this extracted information for various decision making and improving the quality of the services.

Objective:

Big data analytics is the process of examining the voluminous data to uncover hidden patterns, unknown correlations and other useful information that can be used to make better decisions Business people, Doctors, Scientists can use this to improve their services. The main challenge to the analysis of big data comes because of the 4 V's– volume, velocity, variety and veracity. For effective analytics, we need to deal with high volume of data of different variety which is being generated in high velocity. The data what is available from such sources is highly unstructured which calls for analytics on the same.

The objective is to make the participants capable of identifying and applying appropriate techniques and tools to solve problems in managing huge quantity of data.

Expected Job Roles:

- Data Scientist
- Data Analyst
- Data Administrator
- Data Engineer Duration

Duration:

480 Hours - (Theory: 190 hrs + Practical: 200 hrs+ Project: 90hrs)

Course Outline:

Sl. No	Module Title	Duration (Hours)		
		Theory	Lab	Total
1	Linux tools and scripting	15	15	30
2	Java programming	15	15	30
3	Python programming	60	70	130
4	Hadoop	70	70	140
5	Hadoop Sub projects	25	35	60
6	Project work	5	85	90

	Total Duration	190	290	480
	Total Credits	13	10	23

Prerequisites:

Having good computer programming knowledge

Eligibility:

a) BE/BTech/BSc (IT/Computer Science/Electronics), BCA, 3 year Diploma (IT/Computer Science/Electronics), Degree holders with PGDCA, DOEACC A, B level Or equivalent of any of these.

b) Candidates who have appeared in the qualifying examination and awaiting results.

Detailed Syllabus and Learning Outcome:

Sno	Module Title	Topics	Duration (Hours)		Learning Outcome
			Theory	Lab	
1	Linux tools and scripting	1.0 Linux shell and kernel 1.1 Basic commands 1.2 filters 1.3 grep,sed,awk 1.4 Shell scripting	15	15	The candidates will be able to do basic administration of Linux machines
2	Java programming	2.0 Basic syntax and environment 2.1 Classes & Objects 2.2 Datatypes,operators,arrays,strings 2.3 Inheritance,Overriding 2.4 Polymorphism,Abstraction 2.5 Packages 2.6 Collection	15	15	This module prepares the participants to develop java applications
3	Python programming	3.0 Python environment 3.1 Control and data structures 3.2 OOP,exceptions,modules 3.3 re,gui,dbaccess,xml 3.4 numpy,matplotlib 3.5 pandas 3.6 scipy,sklearn	60	70	This is one among the three core modules of this course and it makes the participants to develop applications using python in almost all areas like GUI, Database, Data analysis etc.
4	Hadoop	5.0 Bigdata ,HadoopConcepts 5.1 Configuring Hadoop	70	70	These modules steel the students to develop

					and run applications in Hadoop ecosystem. They will be able to administer Hadoop environment also.
		5.2Hdfs 5.3Mapreduce,YARN			
5	Hadoop Sub projects	6.0Sqoop,flume 6.1Oozie,Hbase 6.2Hive,Pig 6.3Hadoopwith python and R 6.4SPARK	25	35	
6	Project work	The participants will be doing an industry relevant project	5	85	
	Total Hours = 480		190	290	

Examination & Certification:

NIELIT's NSQF Examination pattern will be followed for Examination & Certification.

Sl No	Examination Pattern	Modules Covered	Duration in Minutes	Maximum Marks
1	Theory Paper – 1	1,2,3	90	100
2	Theory Paper – 2	4,5	90	100
3	Practical -1	1,2,3,4,5	180	90
4	Internal Assessment	1,2,3,4,5	-	50
5	Project/Presentation /Assignment	1,2,3,4,5	-	60
6	Major Project/Dissertation	6	-	100
	Total			**500**

Note:
1. Pass percentage would be 50% marks in each component, with aggregate pass percentage of 50% and above.
2. Grading will be as under:

Grade	S	A	B	C	D
Marks Range (in %)	>=85%	>=75%- <85%	>=65%- <75%	>=55%- <65%	>=50%- <55%

3. Theory examination would be conducted online and the paper comprise of MCQ and each question will carry 1 marks.

4. Practical examination/Internal Assessment/ Project/Presentation/Assignment would be evaluated internally.
5. Major Project/Dissertation would be evaluated preferably by External / Subject Expert including NIELIT Officials.
6. Candidate may apply for re-examination within the validity of registration.
7. The examinations would be conducted in English Language only.

Recommended hardware/software tools:

1. High end Servers and client machines
2. Linux based Software infrastructure with Hadoop and subprojects, Python with all the required modules (numpy,pandas,matplotlib,sklearn etc.)

Faculty & Support / Lab Instructor:

1. B.Tech (CS/IT)/MCA/M.Sc. Computer Science/NIELIT B Level with experience and knowledge in Hadoop, Hadoop subprojects, R, and Python or PG/Adv Diploma certification in BigData from NIELIT
2. One Support / Lab Instructor with at least Graduateion/Diploma in computer science with knowledge in Hadoop, Hadoop subprojects, R, and Python or PG/Adv Diploma certification in BigData from NIELIT

References:

1. Learning Python By Mark Lutz, David Ascher
2. Hadoop The Definitive Guide By TOM WHITE
3. R for Everyone By Jared P. Lander
4. http://hadoop.apache.org/
5. https://spark.apache.org/
6. https://docs.python.org/
7. https://docs.scipy.org/doc/
8. https://pandas.pydata.org/pandas-docs/stable/
9. https://scikit-learn.org/stable/documentation.html

Course Name	Advanced Diploma in Bigdata Analytics (Certified Bigdata Analytics)	Vertical	Advanced Diploma in Bigdata Analytics
Course Code		Rev No	R4
Prepared By	PrasoonKumar KG	Aligned NSQF Level	7
NIELIT Centre	Calicut	Last Revised on	03.06.2019