National Institute of Electronics & Information Technology Near IIT Patna, Amhara, Bihta, Patna(Bihar) -801106

Details of the Course

Name of The Course: Industrial Training & Internship in Big Data Analysis.
Duration: 4 Weeks
Fee (in Rs.): Rs 1983/-
Eligibility: Diploma/B.Sc./B.Tech In Electronics, Electrical, Instrumentation Engineering, Computer Science, IT or its equivalent. (Completed or Pursuing).

Course Coordinator: Anurag Kumar
Contact No.: 8095695750

Email ID: anurag.etl.mca@gmail.com
Apply Online:
http://nielitpatnaonline.in/onlinecourse/Certificate_Course.php?fbclid=IwAR3a0xB-VpOGOnUhGwfQSdPuDrQvsIRr56stjObTfKq8YN3PUJGEEC1qqek

Course Content:

UNIT I: Introduction To Big Data and Hadoop Frame Work (1st week)			
Day	**Topic**	**Sub Topic**	**Duration (in Min.)**
1	Big Data Introduction	Data Storage and Analysis, Characteristics of Big Data	120
		Big Data Analytics, Typical Analytical Architecture, Requirement for new analytical architecture	
		Challenges in Big Data Analytics	
		Need of big data frameworks	
2	Hadoop Framework	Hadoop, Requirement of Hadoop Framework, Design principle of Hadoop, Comparison with other system	120
		Hadoop Components, Hadoop 1 vs Hadoop 2, Hadoop Daemon's, HDFS Commands	
3	Map Reduce Programming	I/O formats, Map side join, Reduce Side Join	120
		Secondary sorting, Pipelining MapReduce jobs	
4 & 5	Hadoop Ecosystem	Introduction to Hadoop ecosystem technologies, Serialization: AVRO	240
		Co-ordination: Zookeeper, Databases: HBase, Hive	
		Scripting language: Pig, Streaming: Flink, Storm	

UNIT II : Spark Framework (2nd week)			
Day	**Topic**	**Sub Topic**	**Duration (in Min.)**
1	Spark Framework	Introduction to GPU Computing	120
		CUDA Programming Model	
		CUDA API	
2	CUDA in Detail	Simple Matrix, Multiplication in CUDA,	120
		CUDA Memory Model, Shared Memory Matrix Multiplication,	
		Additional CUDA API Features.	
3	Data Analysis with Spark Shell	Writing Spark Application	120
		Spark Programming in Python	
4	Spark SQL	SQL Context – Importing and Saving data	120
		Data frames – using SQL	
5	GraphX	GraphX overview	120
		Creating Graph	
		Graph Algorithms	

UNIT III Spark Framework continued (3rd Week)			
Day	**Topic**	**Sub Topic**	**Duration (in Min.)**
1	Spark Streaming	Overview, Errors and Recovery	120
		Streaming Source	
		Streaming live data with spark	
2 & 3	Mining Social-Network Graphs	Clustering of Social-Network Graphs	240
		Direct Discovery of Communities	
		Partitioning of Graphs Finding Overlapping Communities	
		Counting Triangles using MapReduce	
		Neighbourhood Properties of Graphs	
4 & 5	Web scraping	Web scraping using python, Combining and merging	240

UNIT IV List of Challenging Experiments (4th week)		
Day	**Topic**	**Duration (in Min.)**
1	HDFS Commends Map Reduce Program to show the need of Combiner	120
2	Map Reduce I/O Formats-Text, key-value Map Reduce I/O Formats – Nline, Multiline	120
3	Sequence file Input/output Formats Secondary sorting	120
4 & 5	Distributed Cache & Map Side Join, Reduce side Join Building and Running a Spark Application Word count in Hadoop and Spark Manipulating RDD	240