**Specialized Training Program in Big Data Analytics**

|  |  |  |
| --- | --- | --- |
| A. | **Name of the Institute** | Centre for Development of Advanced Computing, Mohali |
| B. | **Name/title of the Course** | **Specialized Training Program in Big Data Analytics** |
| C. | **Proposed Dates and Duration of the Course in weeks/ months** | Duration: **Eight weeks**  From: **Dec 16, 2019 to Feb 7, 2020** |
| D. | **Eligibility Criteria for Participants** | |
| 1. Educational Qualification | Technical Graduate (Computer Science/ Electronics/ Telecommunications/ or equivalent) with working knowledge of computers. |
| 1. Work Experience | As per MEA guidelines |
| 1. Age Limit | As per MEA guidelines |
| 1. Target group (Level of participants and target ministry/department etc. may be identified) | Working Professional with knowledge of computers and mobile devices with some programming language exposure. |
| E. | Aims & Objectives of the Course | To develop in-depth knowledge and understanding of the big data domain. Provides an introduction to one of the most common frameworks, Hadoop, that has made big data analysis easier and more accessible. |
| F. | Details / Content of the Course ***(please attach detailed Course Profile****)* | As per sheet attached |
| G. | Mode of Evaluation of Performance of the ITEC Participant | Theory, viva voce & Practical |

**Specialized Training Program in Big Data Analytics**

**Course Content:**

1. **Introduction to Big Data**

* Big Data: Why & Where?
* Big Data Ecosystem, Industry outlook
* Big Data Applications
* Types of big data
* Why analyze Big Data?
* The need to analyze new more complex data sources
* Industry use cases  - Popular big data analytic applications
* Data Warehousing and BI Versus Big Data
* Overview of high value Big Data Use cases - Examples

1. **Big Data Components**

* Characteristics of Big Data
* Four V’s in Big Data
* Challenges when Managing & Analyzing Big Data
* Key Components in Big Data Analytics Environment
* Text Analytics & Streams
* Cloud & Big Data

1. **Introduction to Hadoop**

* What is Hadoop
* Big Data Hadoop Stack
* Open Source Software related to Hadoop
* Hadoop Hardware and Software Requirements

1. **Hadoop Architecture**

* Main Hadoop Components
* Overview of Hadoop based Applications and Services
* Hadoop Architecture & Cluster
* Resource Management: YARN
* Resource Management: YARN Architecture
* Resource Management: Working with YARN
* Step by step Hadoop Installation

1. **Hadoop Applications & HDFS**

* HDFS Architecture & Configuration
* HDFS Clusters – Name node, Data node & Client
* HDFS Design & role in Hadoop
* Metadata, Web-based administration
* HDFS: Performance & Tuning
* HDFS Access, Commands, APIs and Applications

1. **Hadoop MapReduce Framework**

* Describe the MapReduce philosophy
* MapReduce Architecture & Framework
* Processing & Generating large Data sets
* Data Flow in MapReduce
* MapReduce Life cycle

1. **Hadoop Components**

* Significance of Pig & Hive in Hadoop environment
* Analyzing data with Pig – Pig architecture, program structure & executionprocess, Joins & filtering using Pig, Group & co-group, Schema merging &redefining functions.
* Use of Flume &Sqoop to move data into Hadoop
* What is Apache Flume and its Architecture
* Flume Sources, Sinks, Channels and its Configuration
* What is HBase and its Architecture , Data storage in HBase, HBase vs RDBMS, Working with HBase
* Use of Oozie to schedule

1. **Introduction to Spark**

* What is Spark & what is its purpose?
* Components of Spark Unified Stack
* Resilient Distributed Dataset (RDD)
* Downloading & installing Spark standalone

1. **Resilient Distributed Dataset &DataFrames**

* Create parallelized collection & external datasets
* Resilient Distributed Dataset (RDD) operations
* Utilize shared variables and key-value pairs

1. **Spark application programming**

* Purpose & usage of the Spark Context
* Describe & run some spark examples
* Pass functions to spark
* Create & run a Spark standalone application
* Submit application to the cluster

1. **Spark Configuration, monitoring and tuning**

* Components of the Spark Cluster
* Configure Spark to modify the Spark properties, environment variables, or logging properties
* Monitor Spark, metrics and external instrumentation