****

Future Skills PRIME

Government Officials Training Course Big Data Analytics

**in collaboration with CDAC, Noida**

**17th July – 21st July 2023**

**Through Offline Mode**

|  |
| --- |
| **Course Schedule** |
| **Day 1: 17th July 2023 (Monday)** |
| **Timings** | **Sessions** |
| **10:00 AM – 10:15 AM** | **Inauguration** |
| 10:15 AM – 1:00 PM | **Pre-requisites: Basics of Java, Linux Faculty: Ms. Chandni**Big Data Concepts, need for analyzing Big Data, Type of data, roles of Big Data Analytics in Business Intelligence and decision making. Introduction to Hadoop framework, History of Hadoop, Hadoop ecosystem |
| **Timings** | **Sessions** |
| 2:00 PM – 5:00 PM | **Pre-requisites: Basics of Java, Linux Faculty: Ms. Chandni**Hadoop Distributed File System (HDFS), Hadoop commands, writing files onto HDFS, reading files from HDFS, Cloudera demonstration |
| **Day 2: 18th July 2023 (Tuesday)** |
| **Timings** | **Sessions** |
| 10:00 AM – 1:00 PM | **Pre-requisites: Basics of Java, Linux Faculty: Ms. Chandni**MapReduce concept, Map Reduce (MR) programming technique, structure of Map-reduce program, Data types, data loading, Map and Reduce Tasks and executing MRExecution of WordCount Program in MapReduce using Cloudera Distribution Hadoop (CDH) |
| **Timings** | **Sessions** |
| 2:00 PM – 5:00 PM | **Pre-requisites: Basics of Java, Linux Faculty: Ms. Chandni**Apache Hive & its History, Architecture, Data Flow in Hive, Hive Data Models, Hive DatatypesHive Practical by creating internal & external tables, insert commands, performing joins, partitioning & bucketing, User Defined Functions in Hive, Calling Python Script |

|  |
| --- |
| **Day 3: 19th July 2023 (Wednesday)** |
| **Timings** | **Sessions** |
| 10:00 AM – 1:00 PM | **Pre-requisites: Basics of Java, Linux Faculty: Ms. Chandni**Apache Spark, Hadoop vs Spark, Spark Features & Components, Spark Architecture, Modes of Operation, RDDs, RDD OperationsSpark installation, creation of RDDs & performing different actions on them, Word Count Program in Scala and Pyspark, PySpark installation in Google Colab |
| **Timings** | **Sessions** |
| 2:00 PM – 5:00 PM | **Pre-requisites: Basics of Java, Linux, Python Faculty: Ms. Chandni**Introduction to Data analytics, Data science and Application, Introduction to Machine Learning. |
| **Day 4: 20th July 2023 (Thursday)** |
| **Timings** | **Sessions** |
| 10:00 AM – 1:00 PM | **Pre-requisites: Basics of Python Faculty: Ms. Chandni**Working with NumPy, Pandas, Matplotlib. |
| **Timings** | **Sessions** |
| 2:00 PM – 5:00 PM | **Pre-requisites: Basics of Python Faculty: Ms. Chandni**Classification Technique with Python ImplementationDecision Tree, KNN -- Data Loading, Preprocessing, Data Splitting, Model Training, Visualization |
| **Day 5: 21st July 2023 (Friday)** |
| **Timings** | **Sessions** |
| 10:00 AM – 1:00 PM | **Pre-requisites: Basics of Python Faculty: Ms. Chandni**Clustering techniques with Python ImplementationK-Means -- Data Loading, Preprocessing, Data Splitting, Model Training, Visualization |
| **Timings** | **Sessions** |
| 2:00 PM – 5:00 PM | **Pre-requisites: Basics of Python Faculty: Ms. Chandni**Ensemble Techniques with Python ImplementationRandom Forest -- Data Loading, Preprocessing, Data Splitting, Model Training, Visualization**GOT Assessment** |
| **Valediction** | **05:00 PM – 05:15 PM** |
| **Tea Break (11:15 AM – 11:45 AM) Lunch (01:00 PM – 02:00 PM) Tea Break (03:15 PM – 03:45 PM)** |