

## **Azure HDInsight – 5 days**

### **Objectives**

- Participants will be able to organize HDInsight Clusters and load the data into HDInsight
- They can Troubleshoot HDInsight and as well Analyze Data with Hive, Spark SQL, and Pheonix
- Participants will be able to create the Big Data Real-Time Processing Solutions by using Apache Storm and define Stream Analytics.

### **Prerequisite**

- Participants must have strong grasp over Relational databases
- Its good to have basic knowledge of the Microsoft Windows Operating System and its main functionalities.
- Experience in Programming using R and knowledge of common R packages.
- Understanding of common statistical techniques and knowledge of the best practices used in Data Analysis is needed

### **Administer and Provision HDInsight Clusters**

- Deploy HDInsight clusters
- Deploy and secure multi-user HDInsight clusters
- Ingest data for batch and interactive processing
- Configure HDInsight clusters
- Manage and debug HDInsight jobs

### **Implement Big Data Batch Processing Solutions**

- Implement batch solutions with Hive and Apache Pig
- Design batch ETL solutions: big data with Spark
- Operationalize Hadoop and Spark

### **Implement Big Data Interactive Processing Solutions**

- Implement interactive queries used for big data with Spark SQL
- Perform exploratory data analysis by using Spark SQL
- Implement interactive queries used for big data with Interactive Hive
- Perform exploratory data analysis by using Hive
- Perform interactive processing by making use of Apache Phoenix on HBase

### **Implement Big Data Real-Time Processing Solutions**

- Create Spark streaming applications using DStream API
- Create Spark structured streaming applications
- Develop the big data real-time processing solutions with Apache Storm
- Build solutions that use Kafka
- Build solutions that use HBase