Course Details

Day 1

* **What is Big Data and why Hadoop**
  - Big Data characteristics
  - Challenges with traditional system
  - RDBMS/SQL vs. Hadoop
  - *** Hadoop Overview and Ecosystem**
  - Architecture of Hadoop cluster
  - Installing and configuring Hadoop
  - Hands-on Exercise

* **Hadoop Distributed file System - HDFS**
  - Name Node and Data Node
  - Virtual Machine Setup
  - Hands-on Exercise

* **Map Reduce Architecture**
  - How does it work?
  - The Mapper and Reducer
  - Input & Output Formats, Data Type

Day-3

* **Sqoop**

Importing and exporting data from RDBMS

Using Pig & Hbase basics

- Pig philosophy and architecture
- Pig Latin and the Grunt shell
- Loading Data
- Data types and schemas
- Intro to UDF and Scripts
- Pig Latin Details: structure, functions, expressions, relational operators
- HBase vs. RDBMS
- HBase Master and Region Servers
- Intro to ZooKeeper
- Data Modeling
- Column Families and Regions
- Write pipeline / Read pipeline
- Catalog Tables
- Hands-on Exercise

Day- 2

* **Developing Map Reduce Programs**
  - Setting up environment
  - Creating Map Reduce programs
  - Hands-on Exercise

* **Advanced Map Reduce concepts**
  - Combiner, Partitioner, Compression
  - Hands-on Exercise

***Monitoring and Management**
  - Managing HDFS with tools like FSCK and DFSADMIN
  - Using HDFS & Job Tracker Web UI
  - Commissioning and decommissioning of nodes
  - Hands-on Exercise

TRAINER'S PROFILE

Biswajyoti Kar
(Big Data, C, CPP, Data Structures, Unix expert)

Career Summary

Biswajyoti Kar is a Senior Architect with over 19 years of rich experience with proven record in architecture, designing and implementing systems software. He has experience of BIG Data Analytics, UNIX/Linux kernel mode development, Data structures and algorithm development in C.
His current area of working is building solutions around Big Data and Analytics. He also is working towards IP creation in Big Data space.

Academic Background

- A.M.I.E from Institution of Engineers(India), Gokhale Road Calcutta.
- B.Sc in Physics from University Of Calcutta
- Big Data, Hadoop Distributed file systems in Dell.
- Algorithm and Data Structures in C/C++, UNIX/Linux advanced programming, shell scripting in Dell
- Algorithm and Data Structures in C Proton solutions

Selection of Project Experiences

1. BIG Data Work
   - ✓ Leading a project that involved setting up of Hadoop distributed file system (HDFS) on Linux box to test the elasticity part of cloud computing.
   - ✓ Bench-marking the hadoop system for crunching terabytes of data using macro-programming model called PIG Latin.
   - ✓ Statistical analysis was done using R language.
2. Parallel network file system
   - ✓ Leading a project that involved setting up a pNFS client-server file and block layout configuration.
   - ✓ Figuring out pros and cons of each configuration in HPC and NAS environments.
3. Big Data Analytics
   - ✓ Providing consulting in the area of Big Data Analysis to credit rating agency.