

---

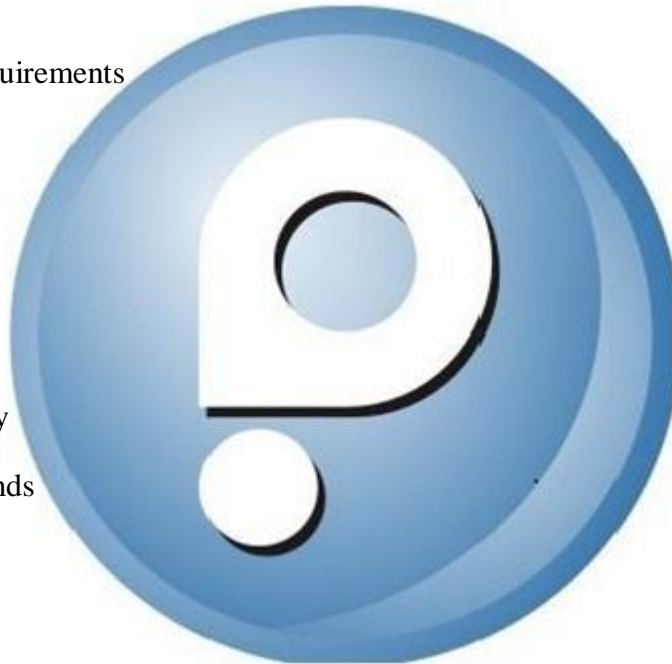
## HADOOP COURSE CONTENT

### **INTRODUCTION**

- Big Data
- Hadoop in Big data roles
- Use Cases of Hadoop
- Hadoop ecosystem
- 3Vs
- Hadoop Requirements

### **HDFS**

- Web HDFS
- Data Flow
- Design
- Data Integrity
- CLI Commands
- Architecture
- API Java
- Archives Data Flow
- Compression



### **MAPREDUCE**

- Theory
- Streaming API
- Writable

- Data Flow
- Input Format
- Programming
- Output format

### **MAPREDUCE ARCHITECTURE**

- Hadoop 1.0 Map reduce Architecture
- Hadoop 2.0 Map reduce Architecture

### **MAPREDUCE PROGRAMMING BASICS**

- Programming Concepts of Map Reduce
- Map Reduce Program steps

### **MAPREDUCE ADVANCED PROGRAMMING**

- Counters
- Distributed Cache
- Sorting
- Debugging
- Side Data Distribution
- Custom Input Format
- Joins
- Tool Runner
- Performance Fine tuning

### **ADMINISTRATION**

- Considerations of Hardware

- Balancers
- Name Node Recovery
- Schedulers

### **H Base**

- CAP Theorem
- Mapreduce Integration
- Architecture
- Role of Zookeeper
- Java Based APIs
- Configuration
- Performance Tuning

### **HIVE**

- Architecture of hive
- Hive Server
- Tables
- Hive-Hbase Integration
- Partitioning
- Bucketing
- Hive Web Interface

