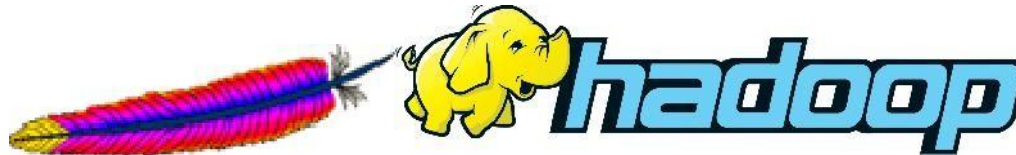




BIG DATA WITH HADOOP ECOSYSTEM



STARTING UP WITH BIG

- Introduction to BIG Data
- Use cases of Big Data
- The Big data core components
- Knowing the requirements, knowledge on Analyst job profile

SETTING UP ENVIRONMENT

- Setting up Linux environment on Commodity Computers
- LINUX commands & various use cases

CASE FOR HADOOP ECOSYSTEM

- ABriefHistoryofHadoop
- CoreHadoopComponents
- What is Hadoop Ecosystems
- Integration Tools
- Analysis Tools
- Data storage & Retrieval Tools
- FundamentalConcepts

HADOOPINSTALLATION

- DeploymentTypes
- InstallingHadoop
- UsingHadoopManagerforEasyInstallation
- BasicConfigurationParameters
- Hands-OnExercise

PLANNING YOURHADOOPCLUSTER

- GeneralplanningConsiderations
- ChoosingtheRightHardware



- NetworkConsiderations
- ConfiguringNodes

TheHadoopDistributedFileSystem (HDFS)

- HDFSFeatures
- HDFSDesignAssumptions
- OverviewofHDFSArchitecture
- WritingandReadingfiles
- DataNode, NameNodeConsideration
- HDFS Federation
- HDFS High-Availability
- The Command-Line Interface
- Basic Filesystem Operations
- Hadoop Filesystems
- Interfaces
- Setting up Various HDFS Commands
- AnOverviewofHDFSSecurity

CLUSTER MAINTENANCE

- Checking HDFS status
- Copying Data Between clusters
- Adding & removing Cluster Notes
- Rebalancing The Cluster
- Name Mode Meta Data Backup

MAPREDUCE

- WhatIsMapReduce?
- FeaturesofMapReduce
- BasicMapReduceConcepts
- ArchitecturalOverview
- MapReduceVersion2
- Hands-OnExercise using Java.

MANAGING& SCHEDULINGJOBS

- Managing & Running Jobs
- The FIFO scheduler



- The Fair Scheduler

MAPREDUCE WITH EXAMPLE

- A Weather Dataset
- Data Format
- Analyzing the Data with Unix Tools
- Analyzing the Data with Hadoop
- Map and Reduce
- Java MapReduce
- Scaling Out
- Data Flow
- Combiner Functions
- Running a Distributed MapReduce Job
- Hadoop Streaming
- Compiling and Running

DEVELOPING A MAPREDUCE APPLICATION

- The Configuration API
- Combining Resources
- Variable Expansion
- Configuring the Development Environment
- Managing Configuration
- GenericOptionsParser, Tool, and ToolRunner
- Writing a Unit Test
- Mapper
- Reducer
- Running Locally on Test Data
- Running a Job in a Local Job Runner
- Testing the Driver
- Running on a Cluster
- Packaging
- Launching a Job
- The MapReduce Web UI
- Retrieving the Results
- Debugging a Job
- Hadoop Logs



- Tuning a Job
- Profiling Tasks
- MapReduce Workflows
- Decomposing a Problem into MapReduce Jobs
- JobControl

HOW MAPREDUCE WORKS

- Anatomy of a MapReduce Job Run
- Classic MapReduce (MapReduce 1)
- Failures
- Failures in Classic MapReduce
- Failures in YARN
- Job Scheduling
- The Capacity Scheduler
- Shuffle and Sort
- The Map Side
- The Reduce Side
- Configuration Tuning
- Task Execution
- The Task Execution Environment
- Speculative Execution
- Output Committers
- Task JVM Reuse
- Skipping Bad Records

MAPREDUCE TYPES & FORMATS

- MapReduce Types
- The Default MapReduce Job
- Input Formats
- Input Splits and Records
- Text Input
- Binary Input
- Multiple Inputs
- Database Input (and Output)
- Output Formats
- Text Output
- Binary Output



- Multiple Outputs
- Lazy Output
- Database Output

MAPREDUCE FEATURES

- Counters
- Built-in Counters
- User-Defined Java Counters
- User-Defined Streaming Counters
- Sorting: Preparation, Partial Sort, Total Sort
- Secondary Sort, Joins

PIG

- Installing and Running Pig
- Execution Types
- Running Pig Programs
- Grunt
- Pig Latin Editors
- An Example

HIVE

- Installing Hive
- The Hive Shell
- An Example
- Running Hive
- Configuring Hive
- Hive Services
- Comparison with Traditional Databases
- Schema on Read Versus Schema on Write
- Updates, Transactions, and Indexes
- HiveQuery Language

Register Yourself at:

www.technospecies.com/training