

AkiraIT – Hadoop Content

Hadoop

- What is Big Data?
- What is Hadoop?
- Relation between Big Data and Hadoop.
- What is the need of going ahead with Hadoop?
- Scenarios to apt Hadoop Technology in REAL TIME Projects
- Challenges with Big Data
 - » Storage
 - » Processing
- How Hadoop is addressing Big Data Changes
- Comparison with Other Technologies
 - » RDBMS
 - » Data Warehouse
 - » TeraData
- Different Components of Hadoop Echo System
 - » Storage Components
 - » Processing Components

HDFS (Hadoop Distributed File System)

- What is a Cluster Environment?
- Cluster Vs Hadoop Cluster.
- Significance of HDFS in Hadoop
- Features of HDFS
- Storage aspects of HDFS
 - » Block
 - » How to Configure block size
 - » Default Vs configurable block size
 - » Why HDFS block size is so large
 - » Design Principles of block size
- HDFS Architecture 5 demons of Hadoop
 - » NameNode and its functionality
 - » Data node and its functionality
 - » Job Tracker and its functionality
 - » Task Track and its functionality
 - » Secondary Name Node and its functionality
- Replication in Hadoop - Fail Over Mechanism
 - » Data Storage in Data Nodes
 - » Fail Over Mechanism in Hadoop - Replication
 - » Replication Configuration
 - » Custom Replication



AkiraIT, HNO: 3-7, 4th Floor, Syndicate Bank Building, Nizampet, Hyderabad, 500090.

Phone: +91-9553406450, +91-40 40061338.

Email: info@akirait.com

AkiraIT – Hadoop Content

- » Design Constraints with Replication Factor
- Accessing HDFS
 - » CLI (Command Line Interface) and HDFS Commands
 - » Java Based Approach
- Hadoop Archives



AkiraIT, HNO: 3-7, 4th Floor, Syndicate Bank Building, Nizampet, Hyderabad, 500090.

Phone: +91-9553406450, +91-40 40061338.
Email: info@akirait.com

AkiraIT – Hadoop Content

MapReduce

- Why Map Reduce is essential in Hadoop?
- Processing Daemons of Hadoop
 - » Job Tracker
 - > Roles Of Job Tracker
 - > Drawbacks w.r.to Job Tracker failure in Hadoop Cluster
 - > How to configure Job Tracker in Hadoop Cluster
 - » Task Tracker
 - > Roles of Task Tracker
 - > Drawbacks w.r.to Task Tracker Failure in Hadoop Cluster
- Input Split
 - » InputSplit
 - » Need Of Input Split in Map Reduce
 - » InputSplit Size
 - » InputSplit Size Vs Block Size
 - » InputSplit Vs Mappers
- Map Reduce Life Cycle
 - » Communication Mechanism of Job Tracker and Task Tracker
 - » Input Format Class
 - » Record Reader Class
 - » Success Case Scenarios
 - » Failure Case Scenario
 - » Retry Mechanism in Map Reduce
- MapReduce Programming Model
 - » Different phases of Map Reduce Algorithm
 - » Different Data types in Map Reduce
 - > Primitive Data types Vs Map Reduce Data types
 - » How to write basic Map Reduce Program
 - > Driver Code
 - > Mapper Code
 - > Reducer Code
 - » Driver Code
 - > Importance of Driver Code in a Map Reduce program
 - > How to Identify the Driver Code in Map Reduce program
 - > Different sections of Driver code
 - » Mapper Code
 - > Importance of Mapper Phase in Map Reduce
 - > How to Write a Mapper Class?
 - > Methods in Mapper Class
 - » Reducer Code
 - > Importance of Reduce phase in Map Reduce

AkiraIT, HNO: 3-7, 4th Floor, Syndicate Bank Building, Nizampet, Hyderabad, 500090.

Phone: +91-9553406450, +91-40 40061338.
Email: info@akirait.com

AkiraIT – Hadoop Content

- > How to Write Reducer Class?
- > Methods in Reducer Class
- » IDENTITY MAPPER & IDENTITY REDUCER
- » Input Format's in Map Reduce
 - > TextInputFormat
 - > KeyValueTextInputFormat
 - > NLineInputFormat
 - > DBInputFormat
 - > SequenceFileInputFormat
 - > How to use the specific input format in Map Reduce
- » Output Format's in Map Reduce
 - > TextOutputFormat
 - > KeyValueTextOutputFormat
 - > NLineOutputFormat
 - > DBOutputFormat
 - > SequenceFileOutputFormat
 - > How to use the specific Output format in Map reduce
- » Map Reduce API(Application Programming)
 - > New API
 - > Deprecated API
- » Combiner in Map Reduce
 - > Is combiner mandate in Map Reduce
 - > How to use the combiner class in Map Reduce
 - > Performance tradeoffs w.r.to Combiner
- » Partitioner in Map Reduce
 - > Importance of Partitioner class in Map Reduce
 - > How to use the Partitioner class in Map Reduce
 - > hashPartitioner functionality
 - > How to write a custom Partitioner
- » Compression techniques in Map Reduce
 - > Importance of Compression in Map Reduce
 - > What is CODEC
 - > Compression Types
 - > GzipCodec
 - > BzipCodec
 - > LZOCODEC
 - > SnappyCodec
 - > Configurations w.r.to Compression Techniques
 - > How to customize the Compression per one job Vs all the job
- » Joins - in Map Reduce
 - > Map Side Join

AkiraIT, HNO: 3-7, 4th Floor, Syndicate Bank Building, Nizampet, Hyderabad, 500090.

Phone: +91-9553406450, +91-40 40061338.

Email: info@akirait.com

AkiraIT – Hadoop Content

- > Reduce Side Join
- > Performance Trade Off
- > Distributed cache
- » How to debug MapReduce Jobs in Local and Pseudo cluster Mode.
- » Introduction to MapReduce Streaming
- » Data localization in Map Reduce



AkiraIT, HNO: 3-7, 4th Floor, Syndicate Bank Building, Nizampet, Hyderabad, 500090.

Phone: +91-9553406450, +91-40 40061338.

Email: info@akirait.com

AkiraIT – Hadoop Content

Apache PIG

- Introduction to Apache Pig
- Map Reduce Vs Apache Pig
- SQL Vs Apache Pig
- Different data types in Pig
- Modes Of Execution in Pig
 - » Local Mode
 - » Map Reduce OR Distributed Mode
- Execution Mechanism
 - » Grunt Shell
 - » Script
- Embedded
- Transformations in Pig
- How to write a simple pig script
- How to develop the Complex Pig Script
- Bags, Tuples and fields in PIG
- UDFs in Pig
 - » Need of using UDFs in PIG
 - » How to use UDFs
 - » REGISTER Key word in PIG
- When to use Map Reduce & Apache PIG in REAL TIME Projects

HIVE

- Hive Introduction
- Need of Apache HIVE in Hadoop
- Hive Architect
 - » Driver
 - » Compiler
 - » Executor(Semantic Analyzer)
- Meta Store in Hive
 - » Importance Of Hive Meta Store
 - » Embedded metastore configuration
 - » External metastore configuration
 - » Communication mechanism with Metastore
- Hive Integration with Hadoop
- Hive Query Language(Hive QL)
- Configuring Hive with MySQL MetaStore
- SQL VS Hive QL
- Data Slicing Mechanisms
 - » Partitions In Hive
 - » Buckets In Hive



AkiraIT, HNO: 3-7, 4th Floor, Syndicate Bank Building, Nizampet, Hyderabad, 500090.

Phone: +91-9553406450, +91-40 40061338.
Email: info@akirait.com

AkiraIT – Hadoop Content

- » Partitioning Vs Bucketing
- » Real Time Use Cases
- Collection Data Types in HIVE
 - » Array
 - » Struct
 - » Map
- User Defined Functions(UDFs) in HIVE
 - » UDFs
 - » UDAFs
 - » UDTFs
 - » Need of UDFs in HIVE
- Hive Serializer/Deserializer SerDe
- HIVE - HBASE Integration

SQOOP

- Introduction to sqoop
- MySQL client and Server Installation
- How to connect to Relational database using Sqoop
- different Sqoop Commands
 - » Different flavours of Imports
 - » Export
 - » Hive-Imports

HBASE

- HBASE Introduction
- HDFS Vs HBase
- Hbase use cases
- Hbase basics
 - » Column families
 - » Scans
- HBase Architecture
- Clients
 - » REST
 - » Thrift
 - » Java Based
 - » Avro
- Map Reduce Integration
- Map Reduce over HBase
- HBase Admin
 - » Schema Definition
 - » Basic CRUD Operations



AkiraIT, HNO: 3-7, 4th Floor, Syndicate Bank Building, Nizampet, Hyderabad, 500090.

Phone: +91-9553406450, +91-40 40061338.
Email: info@akirait.com

AkiraIT – Hadoop Content

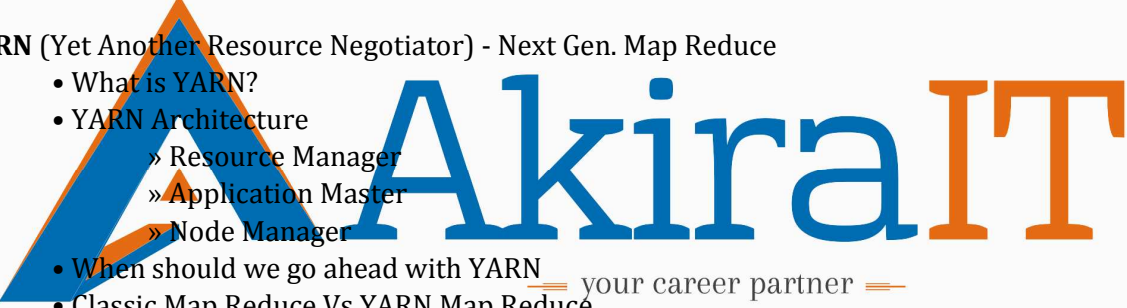
Flume

- Flume Introduction
- Flume Architecture
- Flume Master, Flume Collector and Flume Agent
- Flume Configurations
- Real Time Use Case using Apache Flume

Oozie

- Oozie Introduction
- Oozie Architecture
- Oozie Configuration Files
- Oozie Job Submission
 - » Workflow.xml
 - » Coordinator.xml
 - » job.coordinationproperties

YARN (Yet Another Resource Negotiator) - Next Gen. Map Reduce

- What is YARN?
 - YARN Architecture
 - » Resource Manager
 - » Application Master
 - » Node Manager
 - When should we go ahead with YARN
 - Classic Map Reduce Vs YARN Map Reduce
 - Different Configuration Files for YARN
- 
- your career partner

Impala

- What is Impala?
- How can we use Impala for Query Processing
- When should we go ahead with Impala
- HIVE Vs Impala
- REAL TIME Use Case with Impala

MongoDB (NoSQL Database)

- Need of NoSQL Databases
- Relational Vs Non-Relational Databases
- Introduction to MongoDB
- Features of MongoDB
- Installation of MongoDB
- Mongo DB Basic operations

AkiraIT, HNO: 3-7, 4th Floor, Syndicate Bank Building, Nizampet, Hyderabad, 500090.

Phone: +91-9553406450, +91-40 40061338.

Email: info@akirait.com

AkiraIT – Hadoop Content

Mahout (As a part of BIGDATA ANALYTICS)

- Introduction to Machine Learning (ML) Languages
- Types of Machine Learning
- Introduction to Apache MAHOUT
- Categories of Mahout Algorithms
- Real Time Use case using Classifier Algorithm of Mahout - Naives Bayes

Introduction to Hadoop 'R'

Introduction to Scala



AkiraIT, HNO: 3-7, 4th Floor, Syndicate Bank Building, Nizampet, Hyderabad, 500090.

Phone: +91-9553406450, +91-40 40061338.

Email: info@akirait.com

AkiraIT – Hadoop Content

Hadoop Administration

- Hadoop Single Node Cluster Set Up (Hands on Installation on Laptops)
 - » Operating System Installation
 - » JDK Installation
 - » SSH Configuration
 - » Dedicated Group & User Creation
 - » Hadoop Installation
 - » Different Configuration Files Setting
 - » Name node format
 - » Starting the Hadoop Daemons
- Multi Node Hadoop Cluster Set Up (Hands on Installation on Laptops)
 - » Network related settings
 - » Hosts Configuration
 - » Password less SSH Communication
 - » Hadoop Installation
 - » Configuration Files Setting
 - » Name Node Format
 - » Starting the Hadoop Daemons
- PIG Installation (Hands on Installation on Laptops)
 - » Local Mode » Clustered Mode » Bashrc file configuration
- SGOOP Installation (Hands on Installation on Laptops)
 - » Sqoop installation with MySQL Client
- HIVE Installation (Hands on Installation on Laptops)
 - » Local Mode » Clustered Mode
- Hbase Installation (Hands on Installation on Laptops)
 - » Local Mode » Clustered Mode
- OOZIE Installation (Hands on Installation on Laptops)
- Mongo DB Installation (Hands on Installation on Laptops)
- Commissioning of Nodes In Hadoop Cluster
- Decommissioning of Nodes from Hadoop Cluster

Offerings:

- Proof Of Concepts (POCs) - En End Execution and demonstration by POC Groups.
- Hand Written Hard Copy & Soft Copy Materials for all the Components.
- Detailed Assistance in RESUME Preparation with Real Time Projects based on your technical back ground.
- Guidance in Resume preparation.
- All the Real Time interview Questions will be provided.
- Linux and SQL concepts will be covered as part of the course.
- Discussing the Interview Questions on a daily basis

AkiraIT, HNO: 3-7, 4th Floor, Syndicate Bank Building, Nizampet, Hyderabad, 500090.

Phone: +91-9553406450, +91-40 40061338.
Email: info@akirait.com