



**HADOOP CLUSTER 1**



**HADOOP CLUSTER 2**

## BIG DATA HADOOP ADMINISTRATOR

*"Cluster Championing: The True Measure of a Hadoop Admin"*

In an article released by IBM, there are 2.5 quintillion bytes of data being generated each day and about 80% of this is considered to be unstructured. With the upswing of Cloud Computing during the last decade paired with the rapid emergence of the "Internet of Things", we are now on the brink of a Data Deluge. Apache Hadoop is a software framework that is used to process and store big unstructured data and this is why it became one of the fastest-growing technologies and has never been so consequential for organizations. Since this technology is considered to be avant-garde, the industry is facing one of the biggest challenges in the current setting, the vast skill gap in the talent pool; the definite lack of professionals equipped with the ideal skillset and competency in managing and facilitating an organization's Hadoop adoption.

This 4-Day Big Data Hadoop Admin training takes your knowledge and involvement in Big Data to the next level through profound lessons on the basics of Hadoop, the 2 core functions along with the various tools coexisting within the Hadoop framework such as Flume, Sqoop, YARN, Hive, Impala and Pig. More importantly, the course is designed to guide the Hadoop Administrator on the industry best practices in the Installation, Configuration, Management, Security, Maintenance, Monitoring and Troubleshooting of Hadoop. The training end goal is to prepare participants for the Cloudera CCA Administrator Certification (CCA131).



NOV 14 - 17, 2017



4 DAYS



MAKATI CITY

### Learning Objectives

At the end of the training, participants are expected to:

- Understand and plan the installation and configuration of Hadoop Clusters, multi-node and single node
- Develop expertise in data handling tools such as YARN, MapReduce and HDFS to name a few
- Ensure that organizational plans of Hadoop Adoption and Maintenance is aligned with the industry best practices.
- Master the Hadoop Troubleshooting and Failure Recovery Techniques
- Gain substantial knowledge in the installation and configuration of other Hadoop Ecosystem components (Pig, Impala, Hive, etc.)

### Benefits

- Select hardware and software environment that would be best fit for Hadoop implementation within your organization
- Optimize your organization's Hadoop cluster through proper configuration, load balancing and tuning
- Monitor Hadoop cluster job performances, connectivity and security, all which would prove useful during capacity planning
- Provide service-level-agreements for multiple users through various tools

### Target Audience

- IT Managers
- Systems Administrators
- IT administrators
- IT Systems Engineer
- Data Engineer
- Cloud Systems Administrator
- Data Analytics Administrator



### Key Clients





## Agenda

### The Case for Apache Hadoop

- Why Hadoop?
- Fundamental Concepts
- Core Hadoop Components

### Hadoop Cluster Installation

- Rationale for a Cluster Management Solution
- Cloudera Manager Features
- Cloudera Manager Installation
- Hadoop (CDH) Installation

### The Hadoop Distributed File System (HDFS)

- HDFS Features
- Writing and Reading Files
- NameNode Memory Considerations
- Overview of HDFS Security
- Web UIs for HDFS
- Using the Hadoop File Shell

### MapReduce and Spark on YARN

- The Role of Computational Frameworks
- YARN: The Cluster Resource Manager
- MapReduce Concepts
- Apache Spark Concepts
- Running Computational Frameworks on YARN
- Exploring YARN Applications Through the Web UIs, and the Shell
- YARN Application Logs

### Hadoop Configuration and Daemon Logs

- Cloudera Manager Constructs for Managing Configurations
- Locating Configurations and Applying Configuration Changes
- Managing Role Instances and Adding Services
- Configuring the HDFS Service
- Configuring Hadoop Daemon Logs
- Configuring the YARN Service

### Getting Data Into HDFS

- Ingesting Data From External Sources With Flume
- Ingesting Data From Relational Databases With Sqoop
- REST Interfaces
- Best Practices for Importing Data

### Planning Your Hadoop Cluster

- General Planning Considerations
- Choosing the Right Hardware
- Virtualization Options
- Network Considerations
- Configuring Nodes

### Installing and Configuring Hive, Impala, and Pig

- Hive
- Impala
- Pig

### Hadoop Clients Including Hue

- What Are Hadoop Clients?
- Installing and Configuring Hadoop Clients
- Installing and Configuring Hue
- Hue Authentication and Authorization

### Advanced Cluster Configuration

- Advanced Configuration Parameters
- Configuring Hadoop Ports
- Configuring HDFS for Rack Awareness
- Configuring HDFS High Availability

### Hadoop Security

- Why Hadoop Security Is Important
- Hadoop's Security System Concepts
- What Kerberos Is and how it Works
- Securing a Hadoop Cluster With Kerberos
- Other Security Concepts

### Managing Resources

- Configuring cgroups with Static Service Pools
- The Fair Scheduler
- Configuring Dynamic Resource Pools
- YARN Memory and CPU Settings
- Impala Query Scheduling

### Cluster Maintenance

- Checking HDFS Status
- Copying Data Between Clusters
- Adding and Removing Cluster Nodes
- Rebalancing the Cluster
- Directory Snapshots
- Cluster Upgrading

### Cluster Monitoring and Troubleshooting

- Cloudera Manager Monitoring Features
- Monitoring Hadoop Clusters
- Troubleshooting Hadoop Clusters
- Common Misconfigurations

### Conclusion

## Pre-Requisites

#### The following are recommended but not mandatory:

1. Basic knowledge of Linux
  2. Knowledge on Java
  3. Basic system administration and troubleshooting skills.
- Participants **MUST** bring their own laptops

## Materials



TRAINING MANUAL



CERTIFICATE



SOFT COPY DOWNLOAD



## Training Highlights



The trainer has almost eighteen years of experience in programs and project management, systems and web development, user supporting and training, IT project management, service delivery and business processes and data processing. He is an effective team leader with record of success managing projects and assuming full responsibility for project management, overseeing IT operations, team building, technical infrastructure planning, and on-time delivery of business solutions. He is highly skilled in systems and network management and information systems planning. He also has related skills in industrial feasibility studies and cross-border business partnership facilitation. He was able to work with international development organizations such as ADB, UNDP, UNCTAD, UNIDO, JICA, EU, DANIDA and DFID.

As the Business Development Manager of i.Can Worldwide Inc., he played the role of developing the strategies, roadmaps and ecosystem partnerships to bring to market i.Can's near and long term emerging payments and mobile commerce vision. He is the innovator in creating mobile device-based solutions that are grounded in customer rooted insight and solve customer and business problems, while delivering growth and transformation to the industry. He became a Project Manager of Expersys ITC and was able to formulate and deploy IT long-term strategic plans in support of business objectives, specifically, in the acquisition of new clients and support of business growth. He was able to establish the Project Management Office (PMO) and the implementation of Project Portfolio Management (PPM) system.

As a trainer, he facilitated trainings on programming, OOAD, DBMS, Big Data foundation and Hadoop, and business intelligence at Asia Pacific College and IBM Learning Services.

## Training Highlights

### EXPERT FACILITATOR

The program is facilitated by one of the leading consultants with years of experience in consulting and training in various industries, giving the participants the advantage to tap on his extensive experience to solve practical problems in their current environment.

### INTERACTIVE LEARNING

From face-to-face interaction to modules and workbooks, the training is an instructor-led course with a combination of presentation modules and interactive sessions with participants, allowing mindshare to proactively address specific needs and issues in their own environment.

### UP-TO-DATE INFORMATION

Provides the latest updates on best practices from industries across the globe with case studies and practical information acquired from the training and resources of our trainer.

### TRAINING PACKAGE

We provide high quality and well-researched training materials to make learning a meaningful experience for the participants. The training aids we provide to our participants are used as a quick reference resource for future use.

## About APEX Global

APEX Global (The Academy for Professional Excellence) is the learning solutions arm of ECCI—the leading process improvement solutions provider in Southeast Asia. Our sole aim is to promote performance excellence among professionals. We help our customers achieve greater success through effective, experiential and results-oriented training delivery. Our professional development solutions cover Trainings and Conferences, Managed Services and Learning on Demand. We organize public trainings and in-house workshops tailored for specific organizations. We help companies manage their non-core yet critical training function through end-to-end training management - starting from planning and needs analysis to program evaluation and records management. The Learning on Demand products we offer provide convenient and innovative ways for learning.

Applying the experience of training over 100,000 professionals in the last decade, a strong pool of expert trainers and facilitators with expertise in a niche array of domains and a strong regional presence, we provide an extensive portfolio of high-quality industry specific and functional programs coupled with high quality training materials to deliver our ultimate "promise"—the R.E.A.L. learning experience.

