



Session 5: MapReduce

Assignment 2

Session 5: MapReduce

Assignment

Table of Contents

1. Introduction	3
2. Objective	3
3. Prerequisites	3
4. Associated Data Files	3
5. Problem Statement	3
6. Approximate Time to Complete Task	3

Big Data and Hadoop Development

1. Introduction

In this assignment you need explain about Hadoop demons.

2. Objective

This assignment will help you to understand all the Hadoop demons and their importance.

3. Prerequisites

Acadgild's VM, or linux operating system with Hadoop installed in it

4. Associated Data Files

No data files

5. Problem Statement

1. Explain the importance of below 4 demons in job execution with minimum of 5 points

- Name node :---

1. The namenode manages the filesystem namespace.

2. It maintains the filesystem tree and the metadata for all the files and directories in the tree.

3. This information is stored persistently on the local disk in the form of two files: the namespace image and the edit log.

4. The namenode also knows the datanodes on which all the blocks for a given file are located, [however, it does not store block locations persistently, since this information is reconstructed from datanodes when the system starts.]

5. Without the namenode, the filesystem cannot be used.

- Data node :---

1. Datanodes are the workhorses of the filesystem.

2. They store and retrieve blocks when they are told to (by clients or the namenode), and they report back to the namenode periodically with lists of blocks that they are storing.

- Resource Manager :---

1. resource manager, which coordinates the allocation of compute resources on the cluster.

2. The application master and the MapReduce tasks run in containers that are scheduled by the resource manager.

3. An application master sends periodic heartbeats to the resource manager, and in the event of application master failure, the resource manager will detect the failure and start a new instance of the master running in a new container

- Node manager :-----

1. node managers, which launch and monitor the compute containers on machines in the cluster.

2. The application master and the MapReduce tasks run in containers that are managed by the node managers.

3. Once a task has been assigned a container by the resource manager's scheduler, the application master starts the container by contacting the node manager.

Note: Forward the solutions for review