APACHE OOZIE

Ms. Kavita D. Mahajan
Department of Computer Application, MAEER's MIT Arts, Commerce and Science College, Alandi (D.), Pune.

Abstract

This paper explains Apache Oozie, which is the tool in which all sort of programs can be pipelined in a desired order to work in Hadoop's distributed environment. Oozie also provides a mechanism to run the job at a given schedule. This Paper explains the scheduler system to run and manage Hadoop jobs called Apache Oozie. It is tightly integrated with Hadoop stack supporting various Hadoop jobs like Hive, Pig, Sqoop, as well as system specific jobs like Java and Shell. With this paper you will have enough understanding on architecture of Oozie.This paper explores the Applications of Apache Oozie like workflow, coordinatorand bundle.

Keywords: Oozie, Sqoop, Oozie workflow, Oozie bundle, Oozie coordinator, Hadoop, Cron job.

What is Apache Oozie?

Apache Oozie is a scheduler system to run and **manage Hadoop jobs** in a distributed environment. It allows to combine multiple complex jobs to be run in a sequential order to achieve a bigger task. Within a sequence of task, two or more jobs can also be programmed to run parallel to each other.

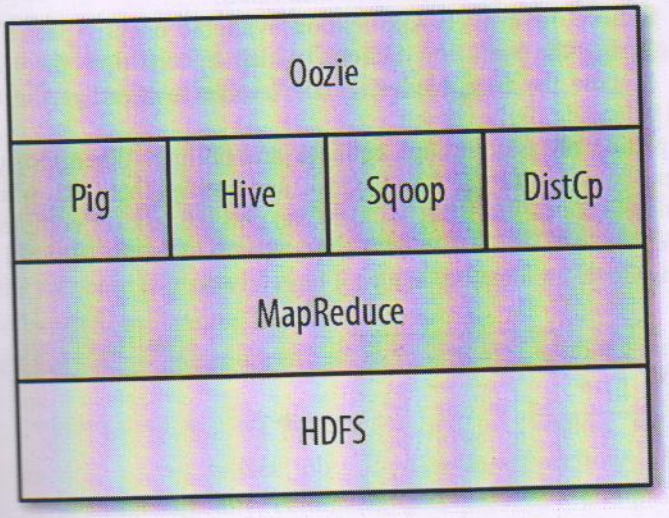

Figure1:Oozie in the Hadoop ecosystem

What exactly is Oozie?

Oozie is an orchestration system for Hadoop jobs. Oozie is designed to run multistage Hadoop jobs as a single job: an Oozie job. Oozie jobs can be configured to run on demand or periodically.

Oozie jobs running on demand are called *workflow jobs*.