

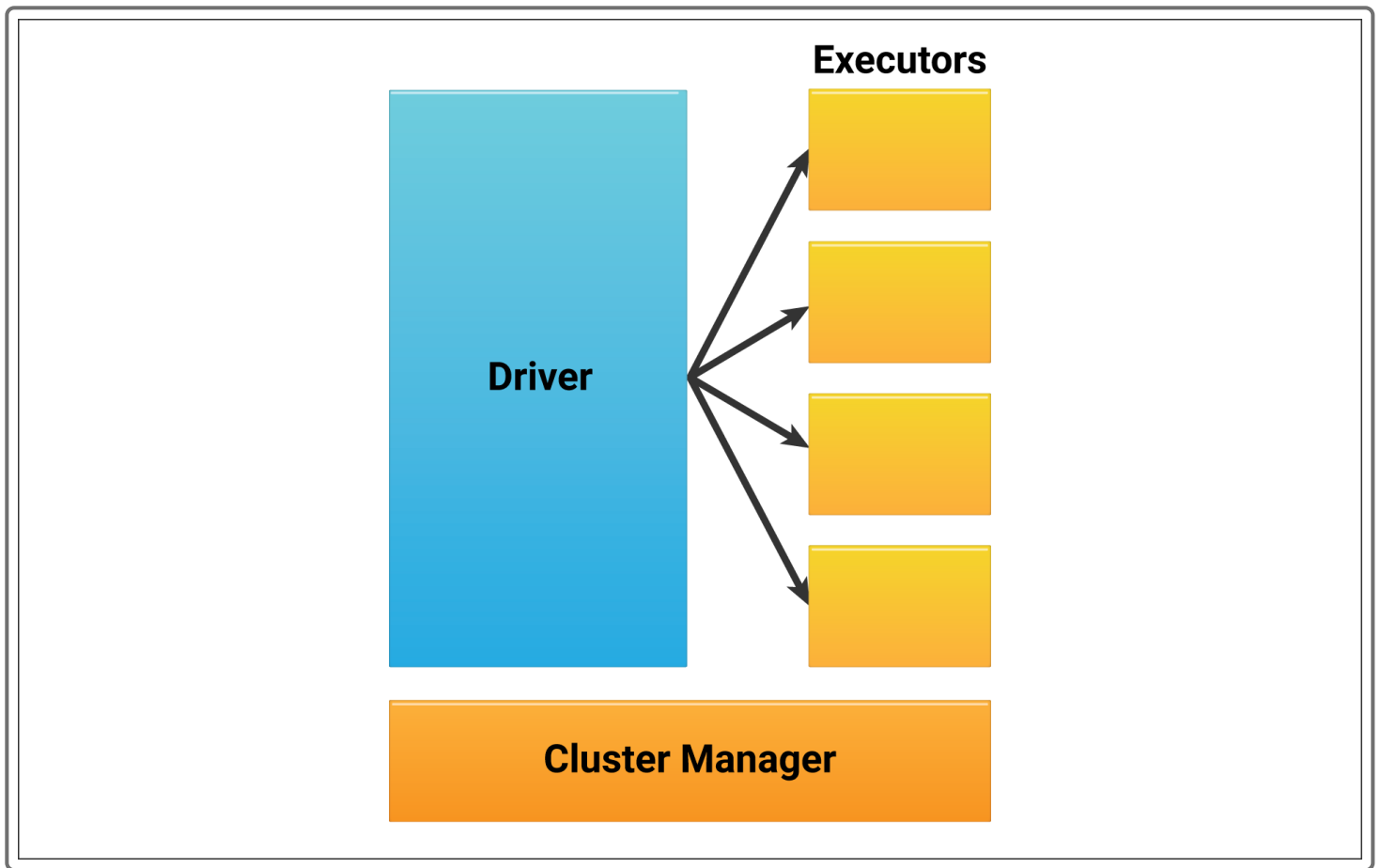
16.3.2

Spark Architecture

Your client is very curious about what goes on behind the scenes. They are familiar with Spark, as it has become so popular in the industry, but they don't understand how it works. In order to be fully prepared for your presentation, the two of you prepare to get into the architecture of how Spark actually works.

The Spark architecture includes the driver, executors, and the cluster manager:

- The **driver** is the heart of the application. It is responsible for maintaining the application information; responding to the code or input; and analyzing, distributing, and scheduling work to the executors.
- The **executors** perform the code assigned by the driver and then report the state of the computation to the driver.
- The **cluster manager** controls the driver and executors and allocates resources to the machines on the Spark applications. The cluster manager is an external service for acquiring resources on the cluster. Spark can either use its own standalone cluster manager that comes standard with Spark or another application (e.g., Apache Mesos, Hadoop YARN).



Think of the driver as a manager who assigns work to employees, who then perform the tasks and report back to the manager whether the task was finished or not. The cluster manager functions as a budget manager who allocates the specific amounts to pay employees for performing tasks—the more they are paid, the more work they do.

© 2020 - 2022 Trilogy Education Services, a 2U, Inc. brand. All Rights Reserved.