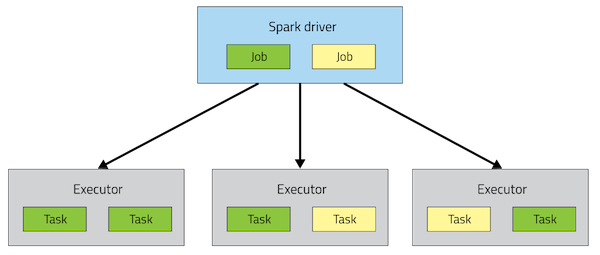
**Learning Objective 1**

To be able to describe Apache Spark with Apache Yarn as a solution in a given context.

* The Yarn cluster manager has two components.
* A central master service the YARN ResourceManager.
* A slave service running on every node the YARN NodeManager.
* YARN is the only cluster manager for Spark that supports security.
* When running Spark on YARN, each Spark executor runs as a YARN container.
* Spark supports two modes for running on YARN, “yarn-cluster” mode and “yarn-client” mode
* “yarn-cluster” mode is more like the traditional hadoop job
* In yarn-client mode you can use using Spark interactively with the Spark Shell.

A bit deeper

How spark runs



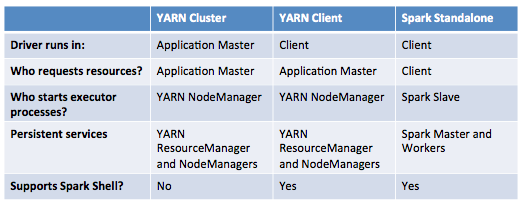
issues

**yarn-cluster mode,** the driver runs in the Application Master. This means that the same process is responsible for both driving the application and requesting resources from YARN, and this process runs inside a YARN container. The client that starts the app doesn’t need to stick around for its entire lifetime.

The yarn-cluster mode, however, is not well suited to using Spark interactively. Spark applications that require user input, like spark-shell, need the Spark driver to run inside the client process that initiates the Spark application.

**yarn-client mode**, the Application Master is merely present to request executor containers from YARN. The client communicates with those containers to schedule work after they start

**Summary**



**Spark Context**

