05 Local Vs Cluster Mode -KirkYagami

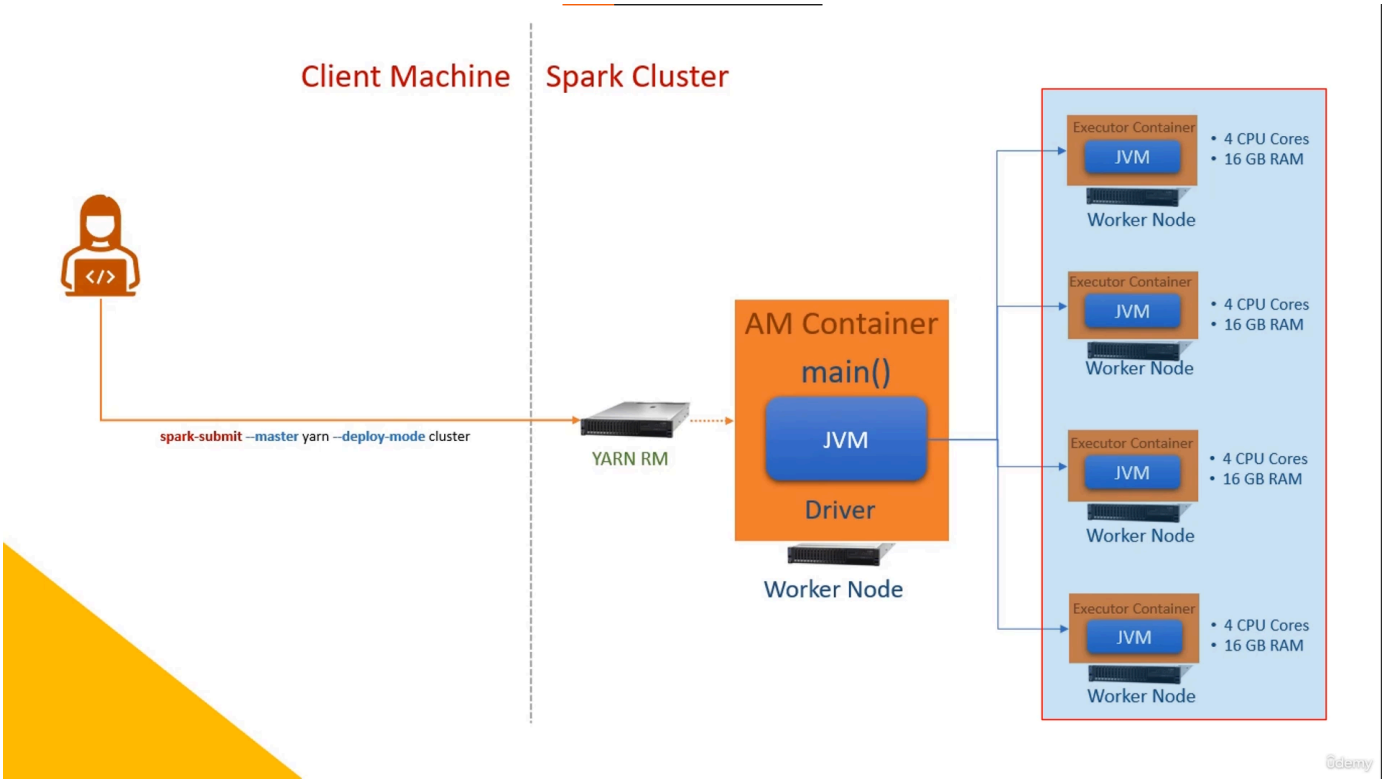
Deploy Modes

Cluster Mode VS Client Mode

1. **Cluster Mode:** Driver program runs **on one of the nodes in the cluster**.

Benefits:

- **Scalability:** Can leverage the processing power of multiple machines in the cluster for large datasets.
- **High Availability:** If a node fails, the driver can be restarted on another node with minimal disruption.
- **Security:** Driver runs within the secure environment of the cluster.
- No network latency
- **When to use:**
  - Production workloads requiring large-scale data processing.
  - Applications demanding high availability and fault tolerance.

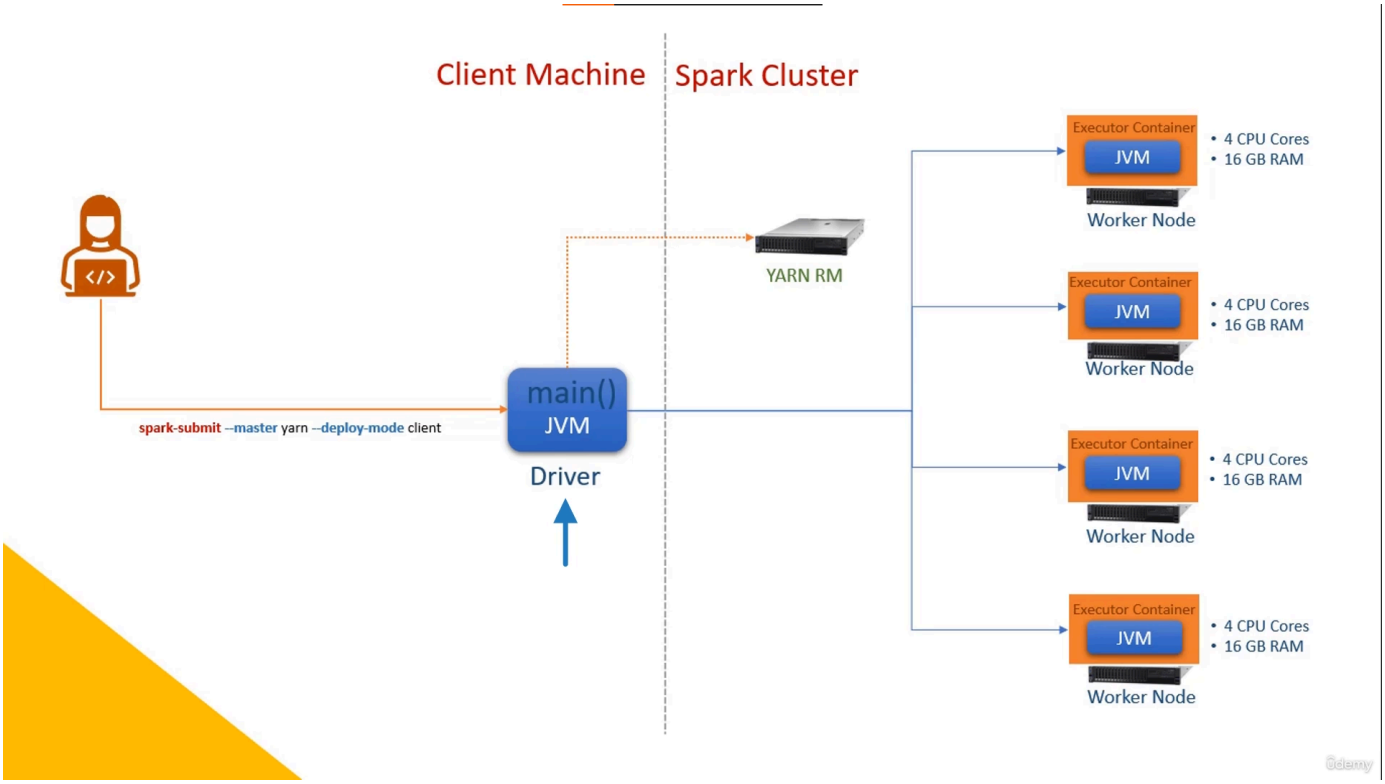


1. **Client Mode:** Driver program runs **in the local machine's JVM (Java Virtual Machine)**.

Benefits:

- **Simplicity:** Easier to set up and use, especially for development and testing.
- **Faster startup:** Driver doesn't need to be deployed on the cluster, reducing initial overhead.
- **Development interaction:** Easier to monitor and interact with the driver program during development.
- **When to use:**
  - Development and testing of Spark applications.
  - Smaller datasets that can be processed efficiently on a single machine.

- Interactive Spark sessions where immediate feedback is necessary.
- PySpark, Spark-Shell, spark-sql



Driver Deploy Modes

There are two deploy modes:

- Client Mode – the application submitter launches the driver process outside the cluster
- Cluster Mode – the framework launches the driver process inside the cluster