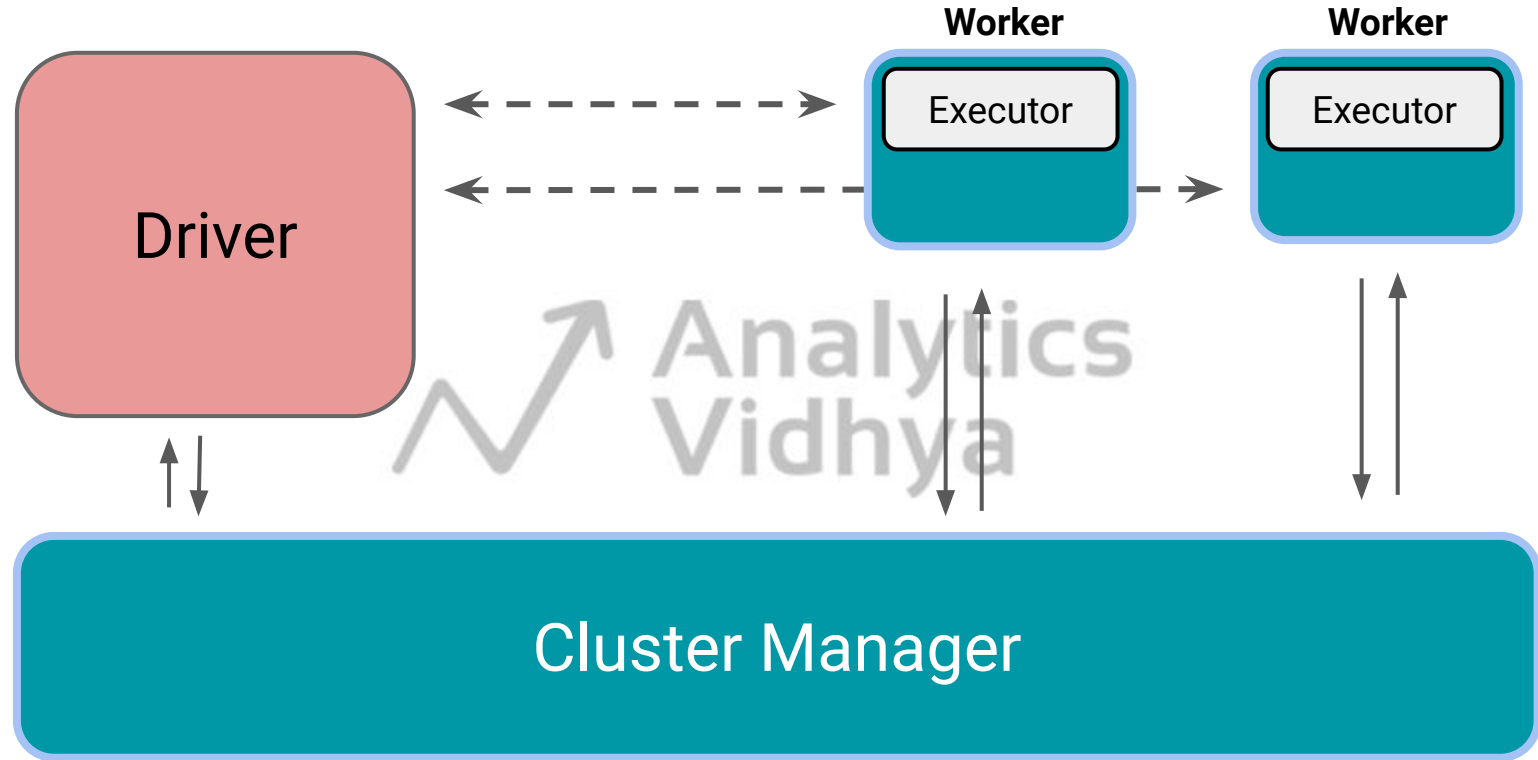
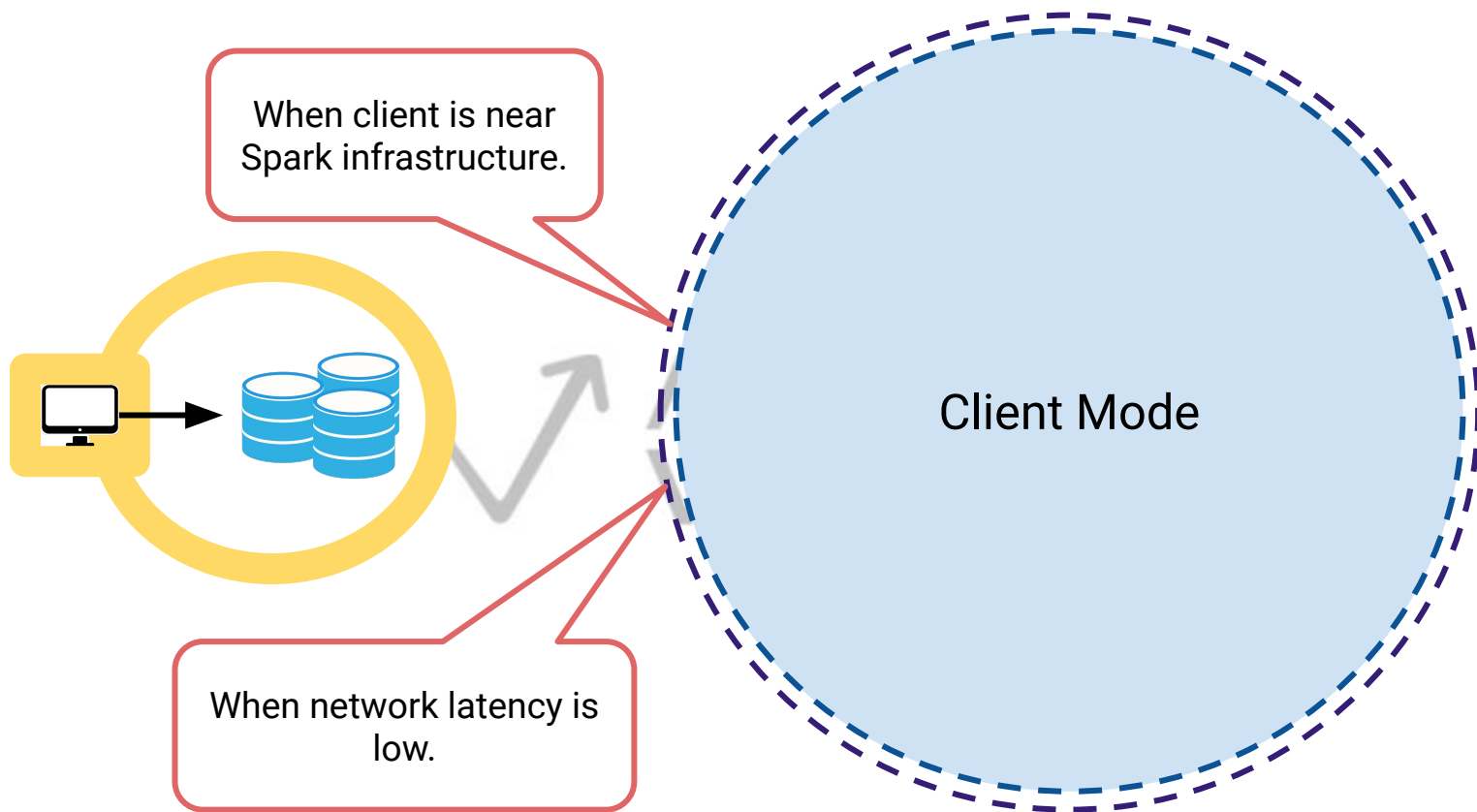


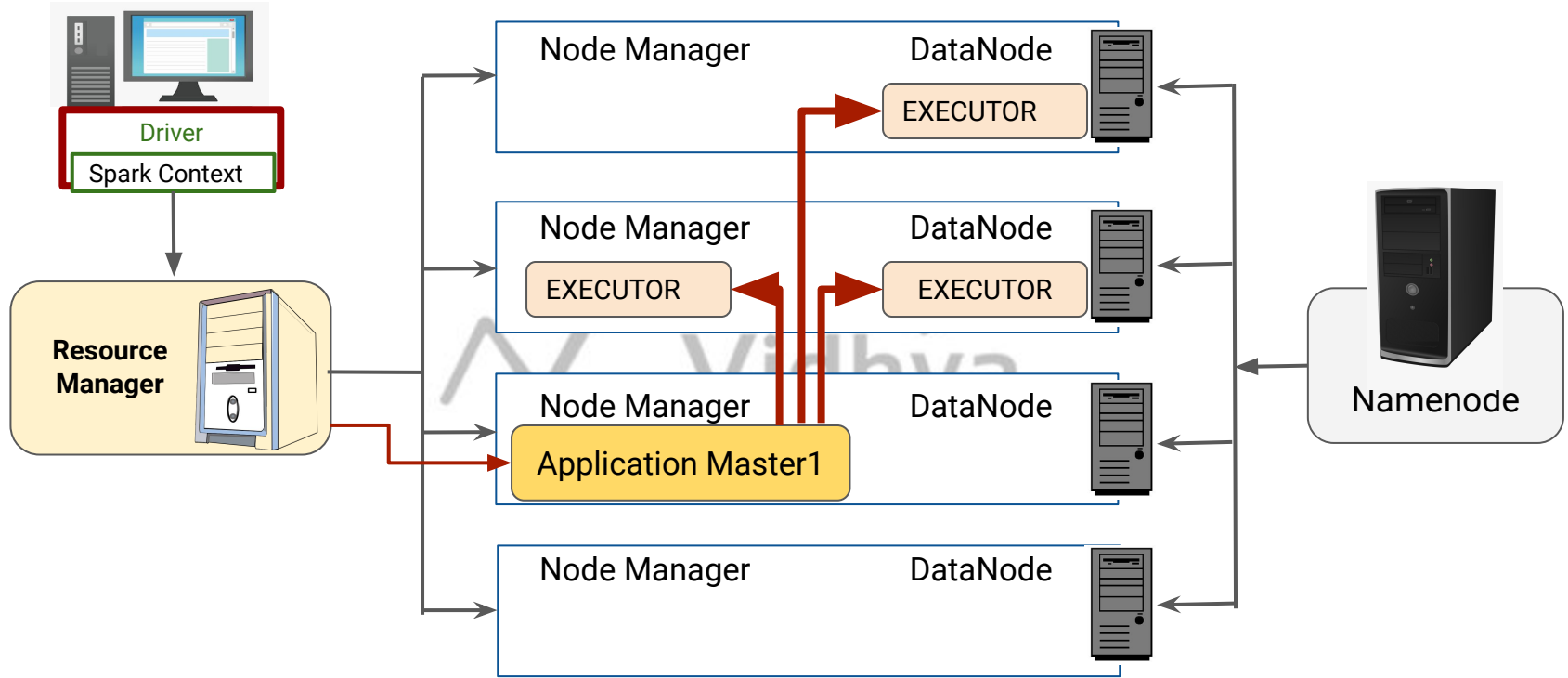
Running Spark Application on YARN

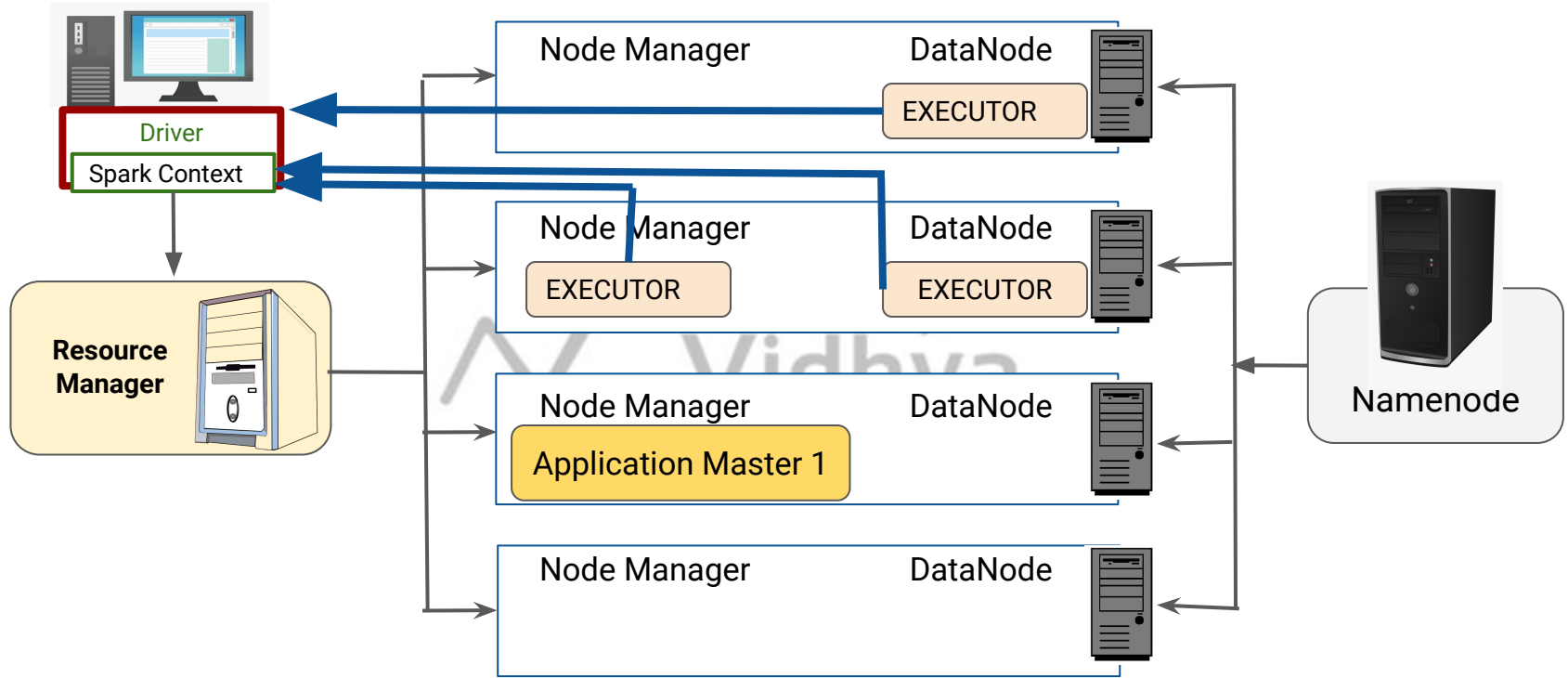
Deployment Modes in YARN

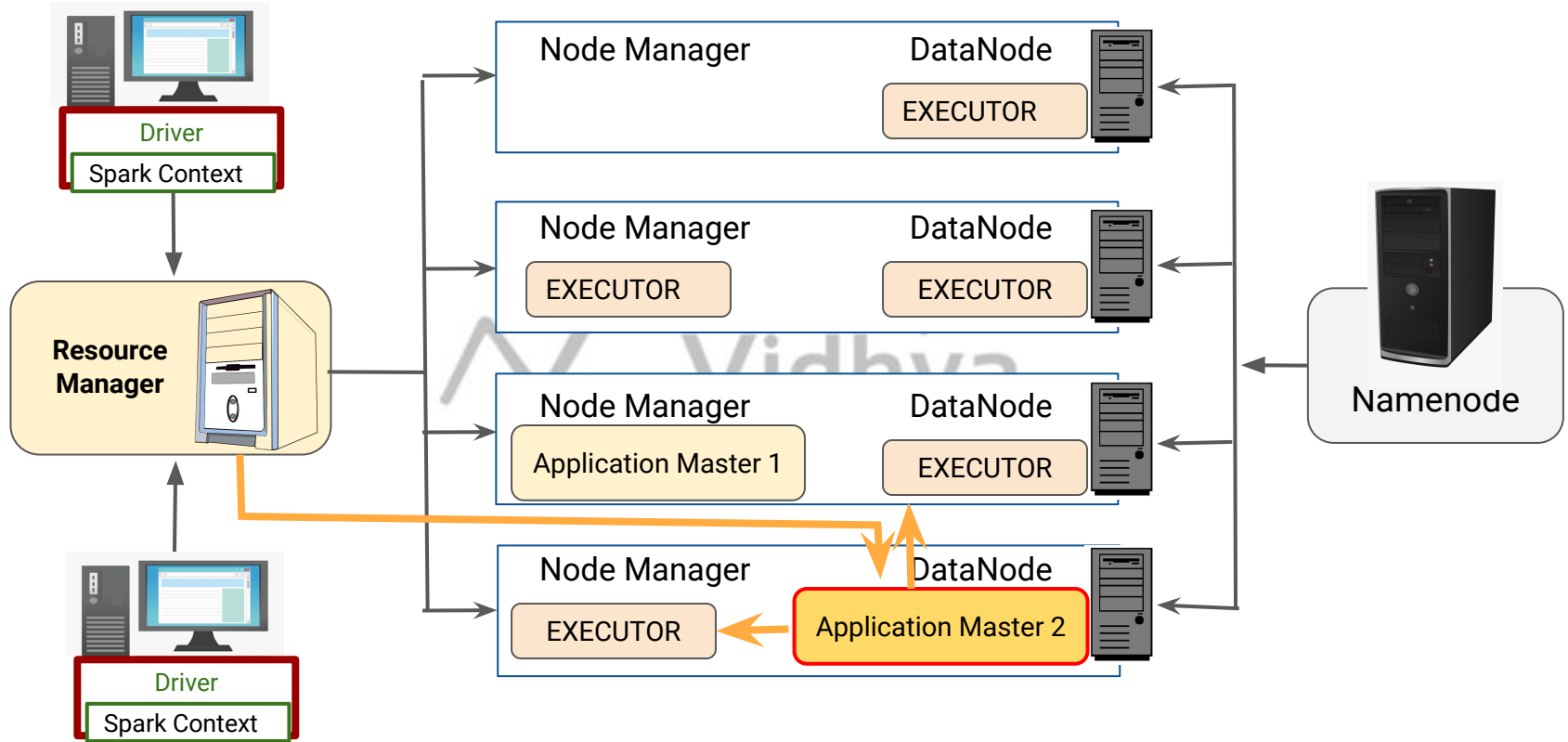


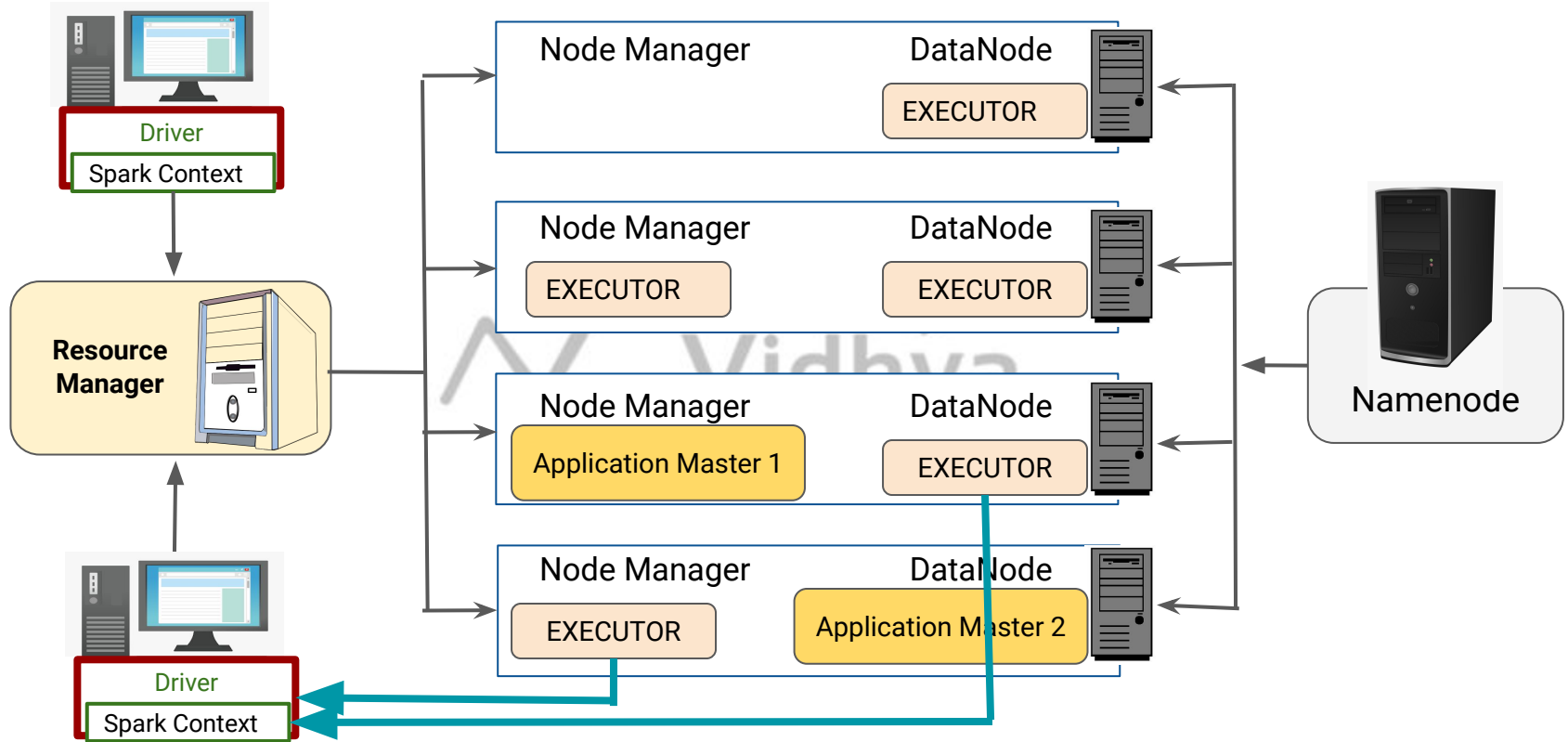


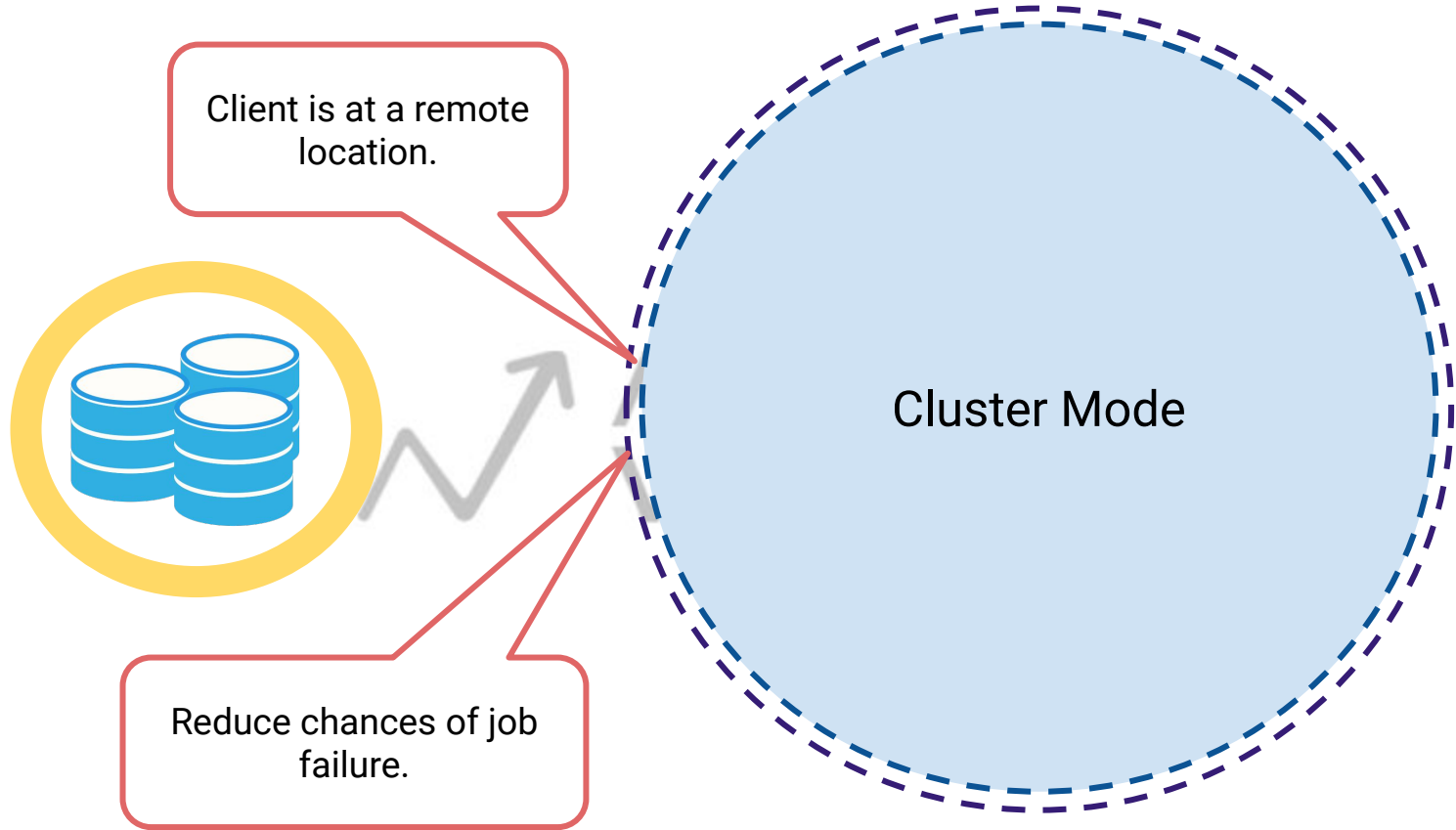


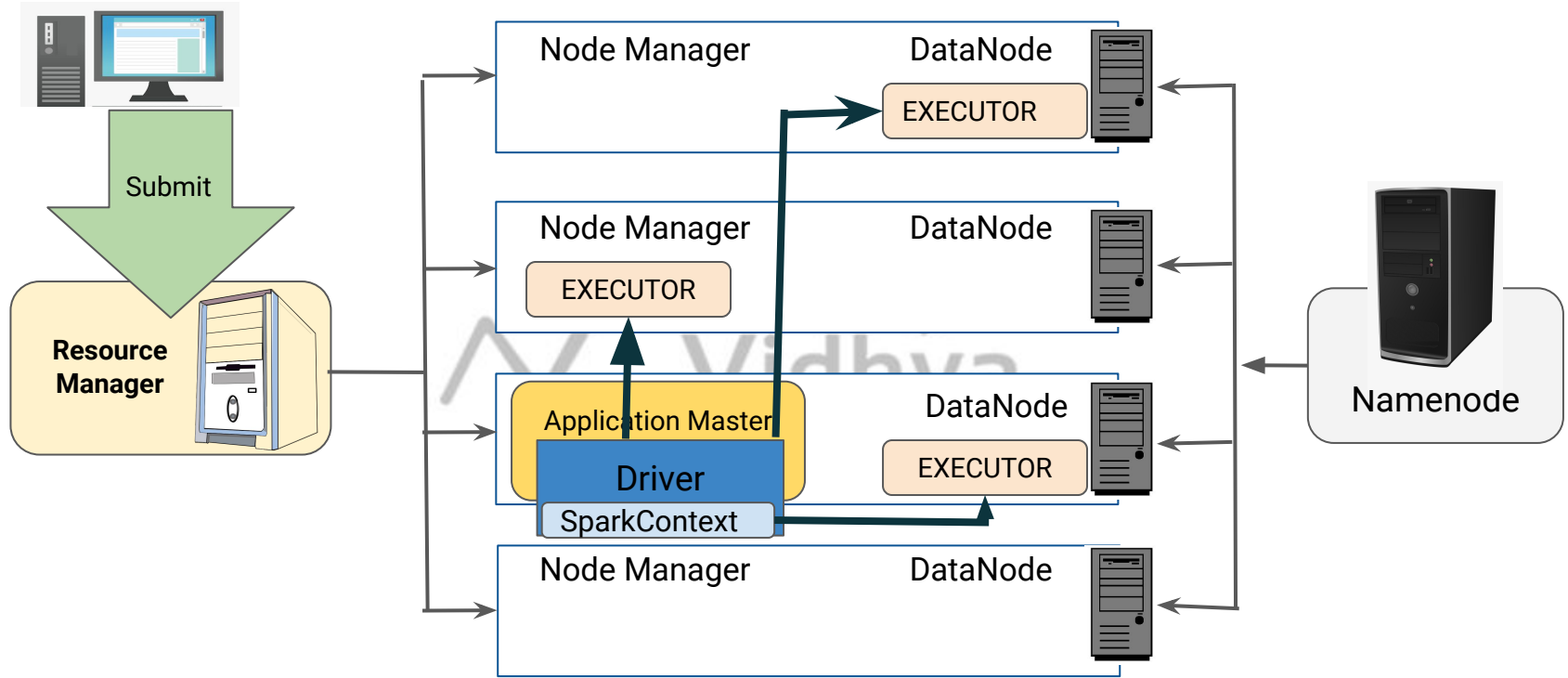


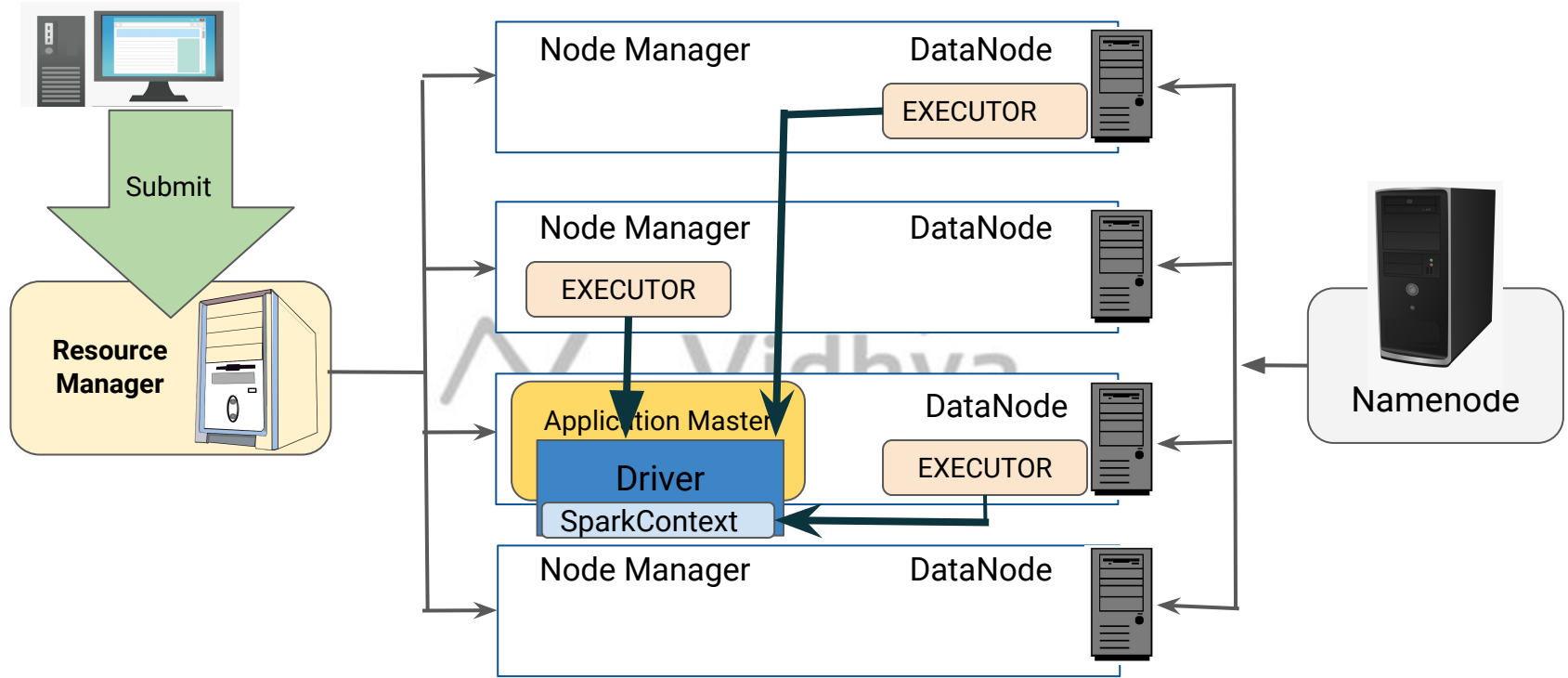












Running a Spark Application on a Cluster

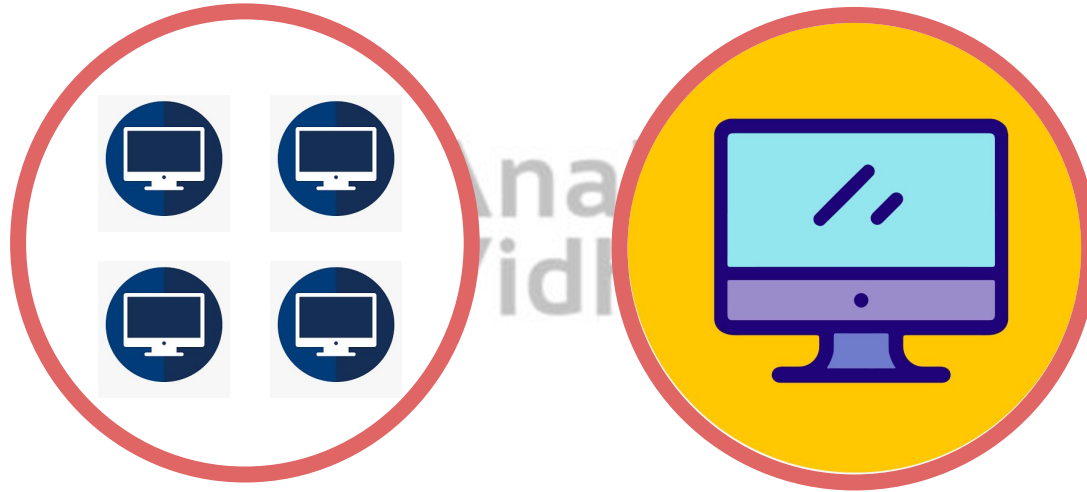
Run on a YARN client

```
./bin/spark-submit \  
--class org.apache.spark.examples.SparkPi \  
--master yarn \  
--deploy-mode client
```

Run on a YARN cluster

```
./bin/spark-submit \  
--class org.apache.spark.examples.SparkPi \  
--master yarn \  
--deploy-mode cluster
```

Cluster Mode vs Client Mode



CLUSTER MODE	CLIENT MODE
Network latency is less.	Network latency is relatively higher.



CLUSTER MODE	CLIENT MODE
Network latency is less.	Network latency is relatively higher.
The yarn-cluster mode is not well suited to using Spark interactively	But the yarn-client mode is.

Vidhaya

CLUSTER MODE	CLIENT MODE
Network latency is less.	Network latency is relatively higher.
The yarn-cluster mode is not well suited to using Spark interactively	But the yarn-client mode is.
The Driver can disconnect after submitting the job.	The client needs to stay connected.

CLUSTER MODE	CLIENT MODE
Network latency is less.	Network latency is relatively higher.
The yarn-cluster mode is not well suited to using Spark interactively	But the yarn-client mode is.
The Driver can disconnect after submitting the job.	The client needs to stay connected.
Driver and Workers run on the same infrastructure.	Driver and Workers may run not on the same infrastructure.

Running a Spark Application Locally

Running a Spark Application Locally - Example

spark-submit is a utility to submit your spark program (or job) to Spark clusters.

```
siddharth@siddharth:~$ spark-submit --master local[2] --deploy-mode client --executor-memory 1G --num-executors 2 /home/siddharth/Documents/BIGDATA/WordCount.py
```

Running a Spark Application Locally - Example

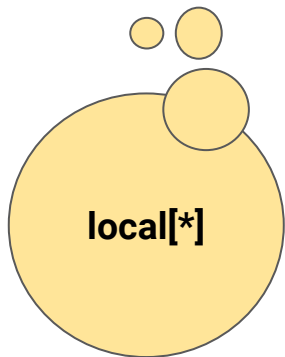
We specify the master as local with 2 cores.

```
siddharth@siddharth:~$ spark-submit --master local[2] --deploy-mode client --executor-memory 1G --num-executors 2 /home/siddharth/Documents/BIGDATA/WordCount.py
```

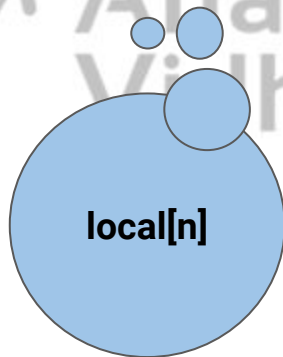
Running a Spark Application Locally

To run a Spark application locally, use

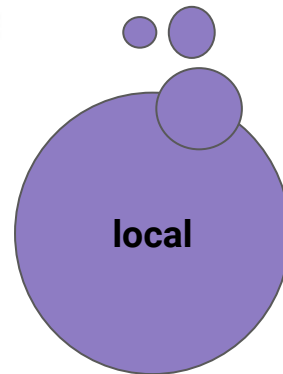
- `spark-submit --master local[*]`
- `spark-submit --master local[n]`
- `spark-submit --master local` to specify the cluster option



Use `local[*]` to run application locally with as many threads as the cores



Use `local[n]` to run application locally with n threads



Use `local` to run application locally with a single thread



Thank You