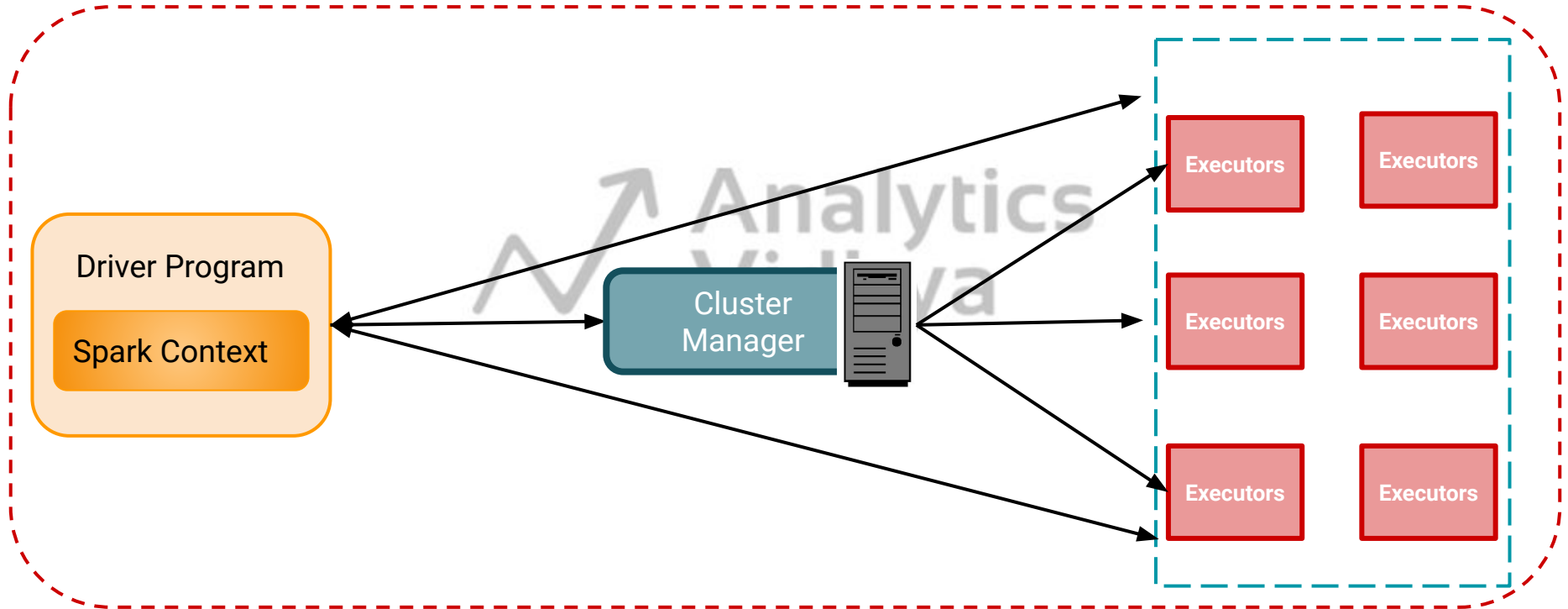
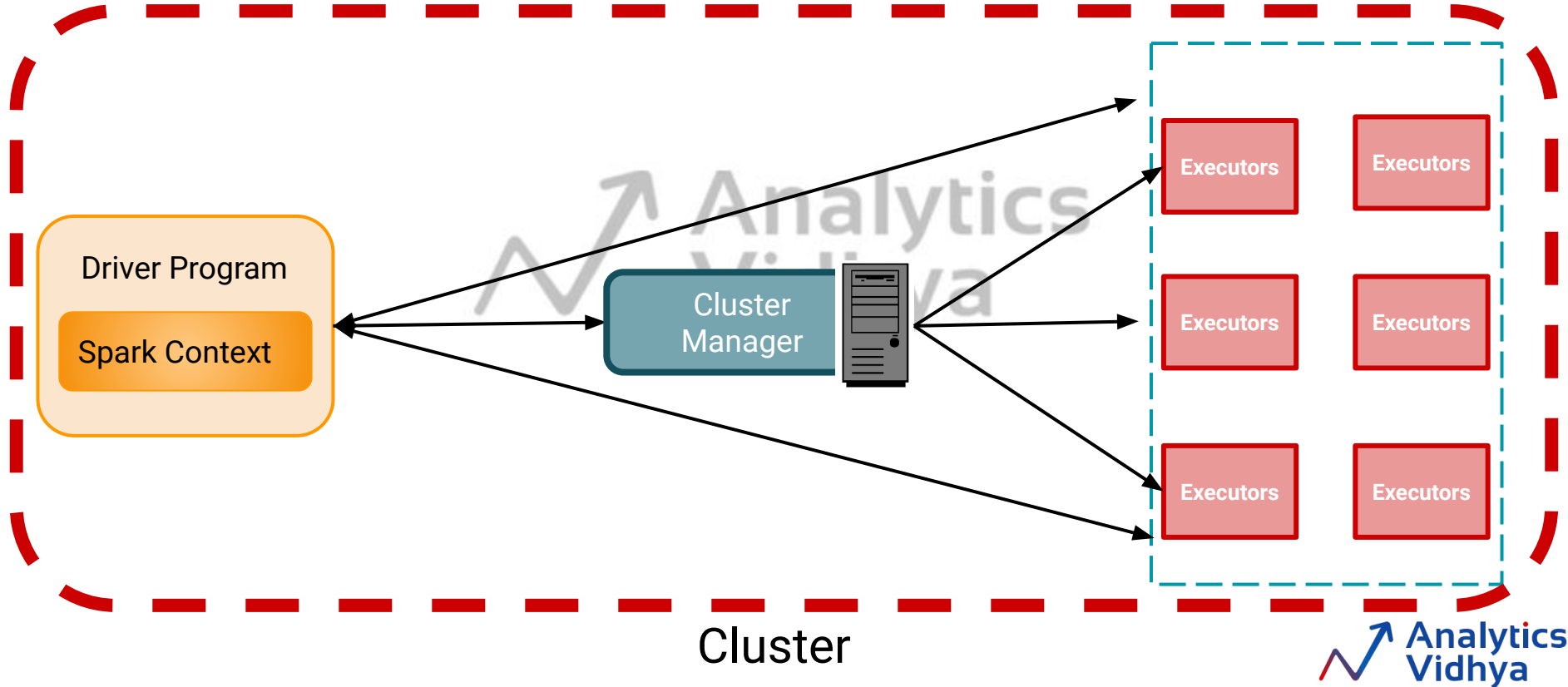


# Spark Cluster Managers

# Spark Cluster Overview

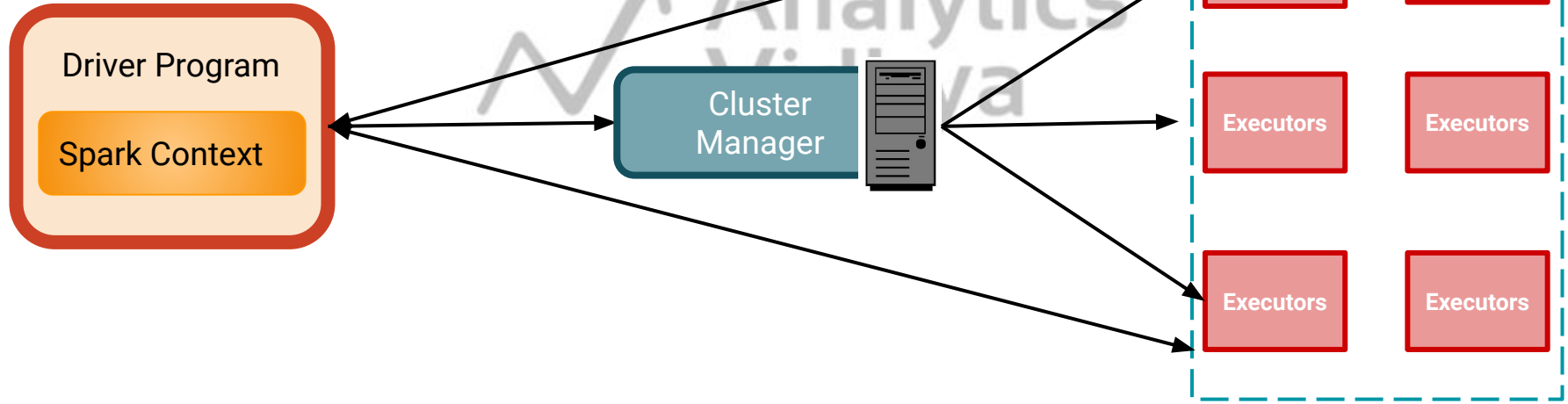


# Spark Cluster Overview

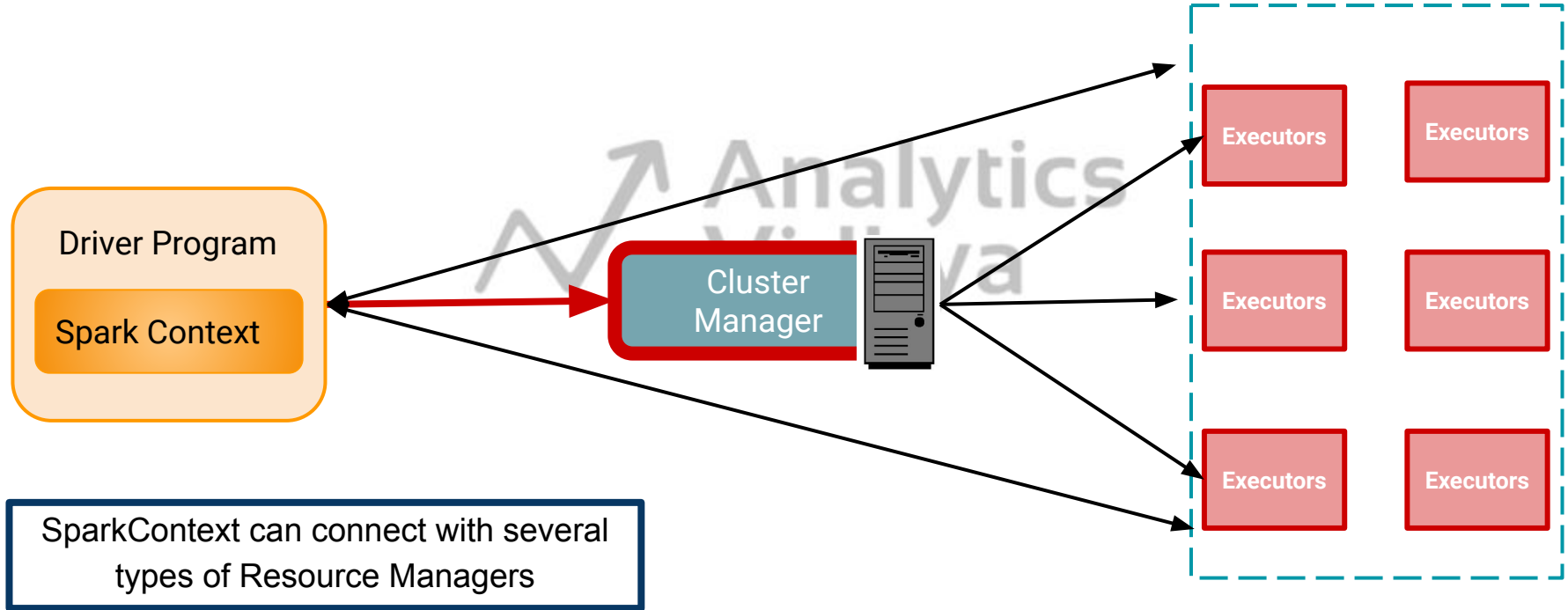


# Spark Cluster Overview

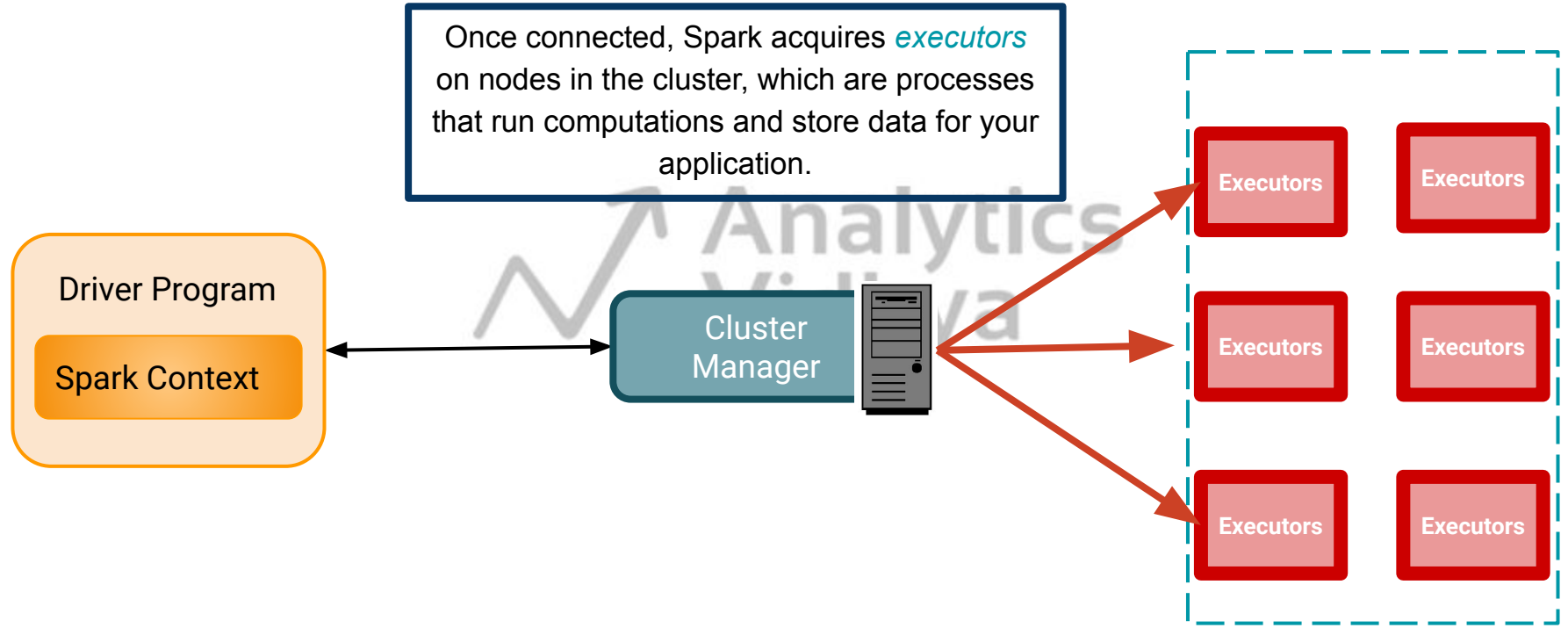
Spark applications run as independent sets of processes on a cluster, coordinated by the SparkContext object in your main program (called the *driver program*).



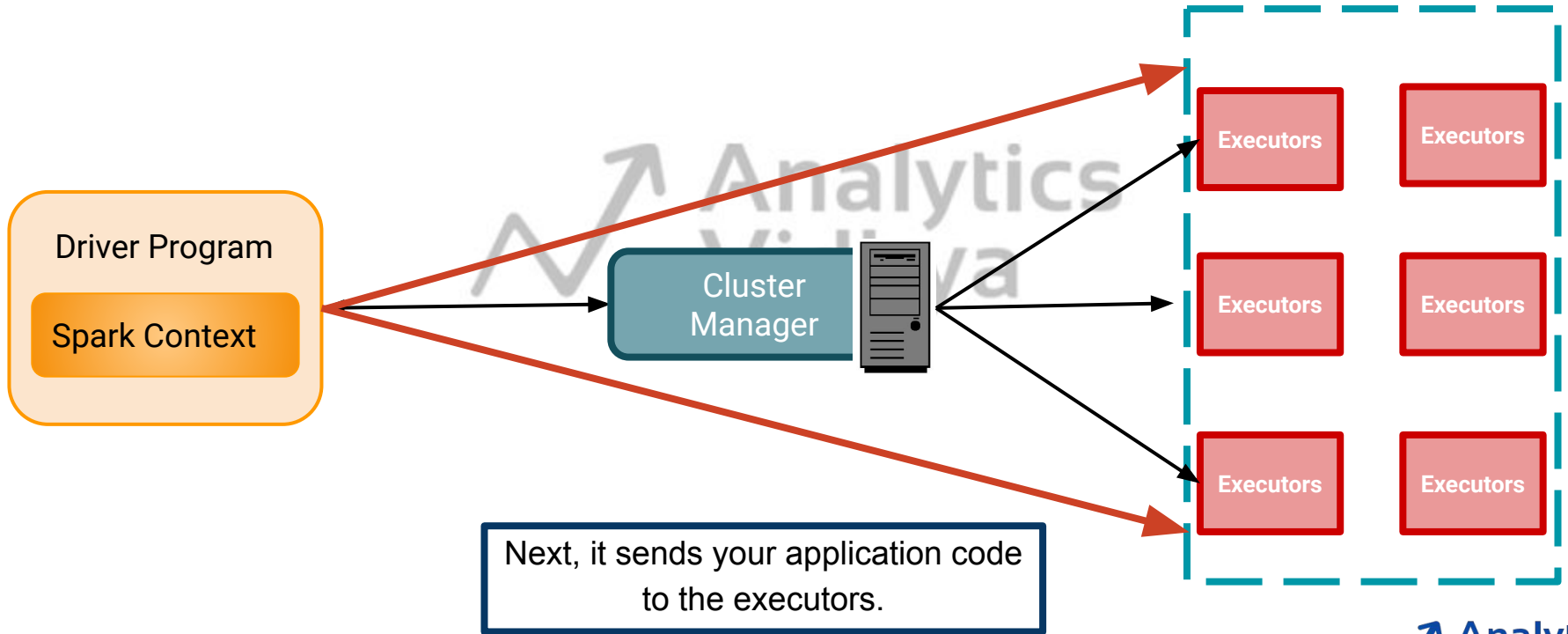
# Spark Cluster Overview



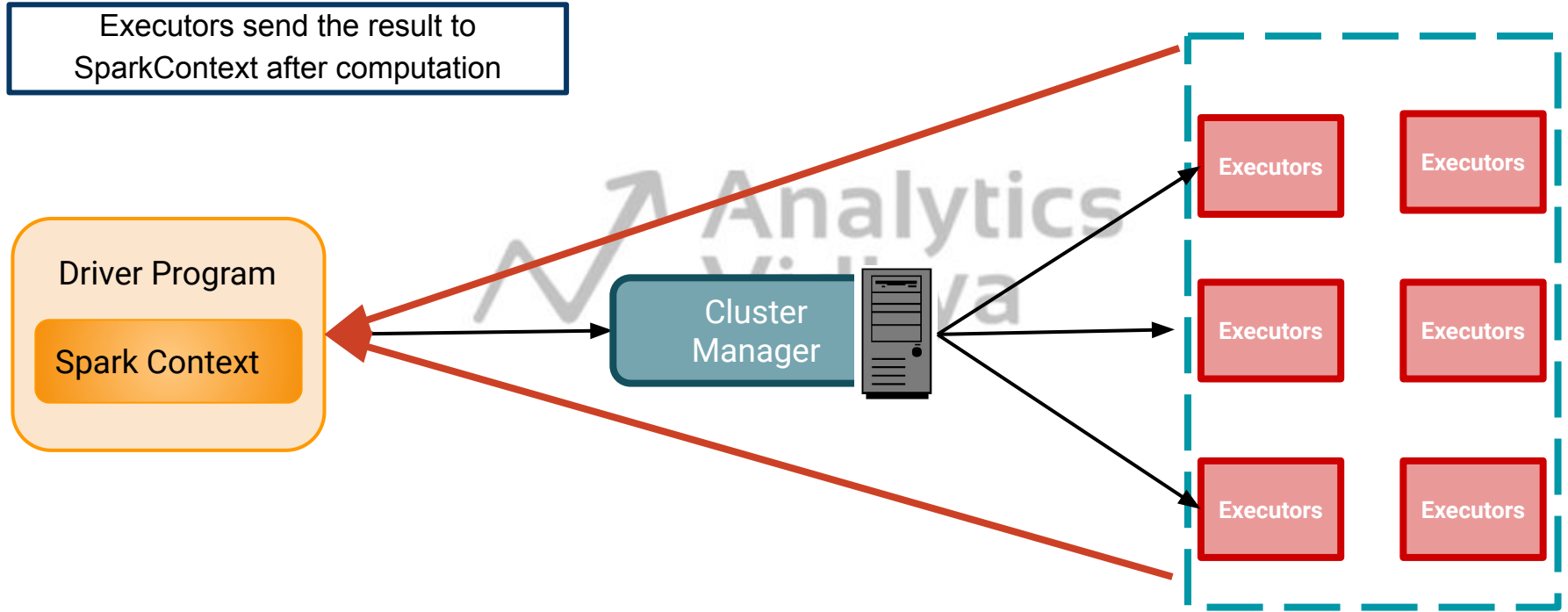
# Spark Cluster Overview



# Spark Cluster Overview



# Spark Cluster Overview







# How to run Spark Applications

# How to run Spark Applications

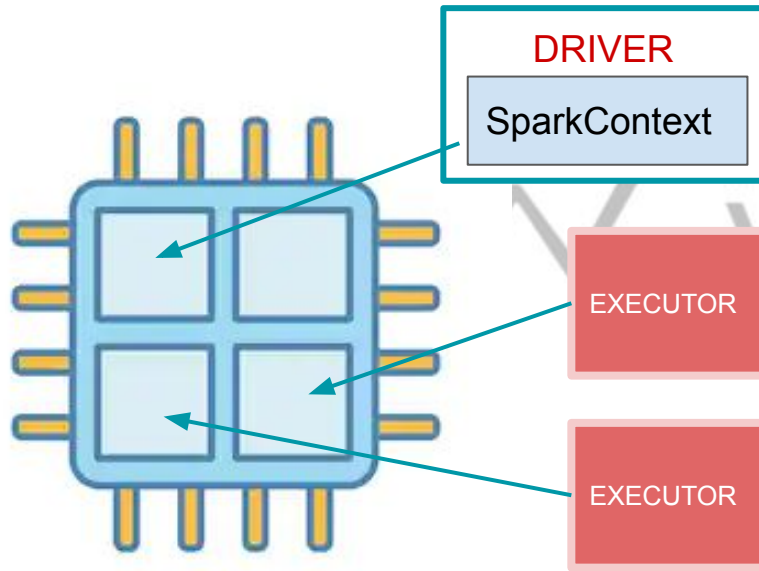
Spark Application that Run Locally

Spark Application that Run on a Cluster



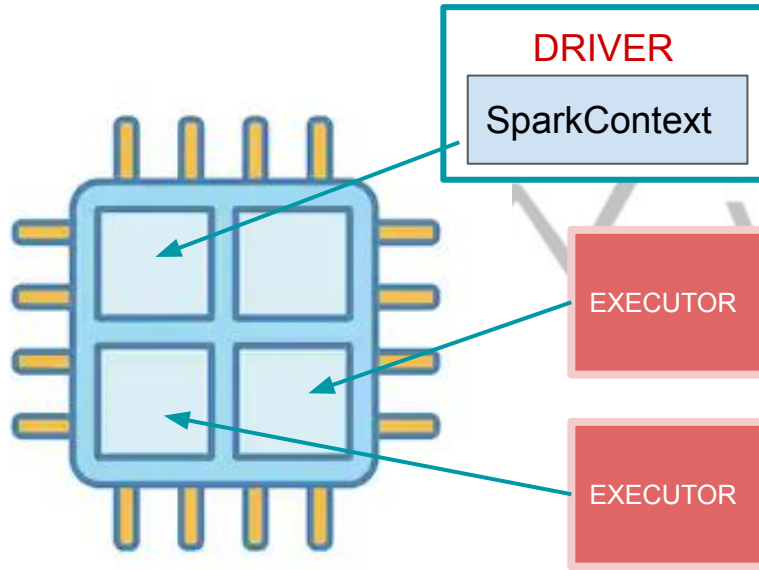
# How to run Spark Applications

## Spark Application that Run Locally

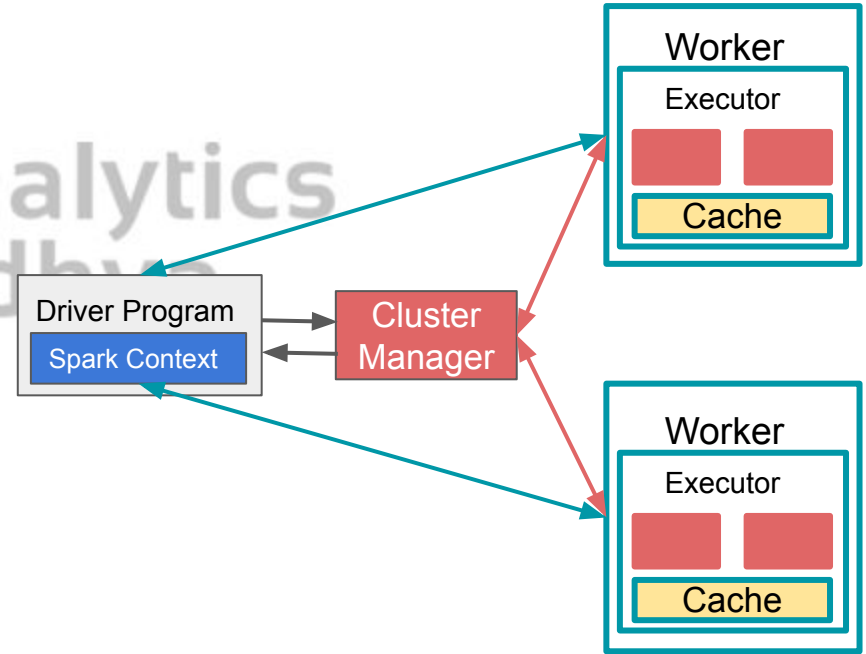


# How to run Spark Applications

## Spark Application that Run Locally

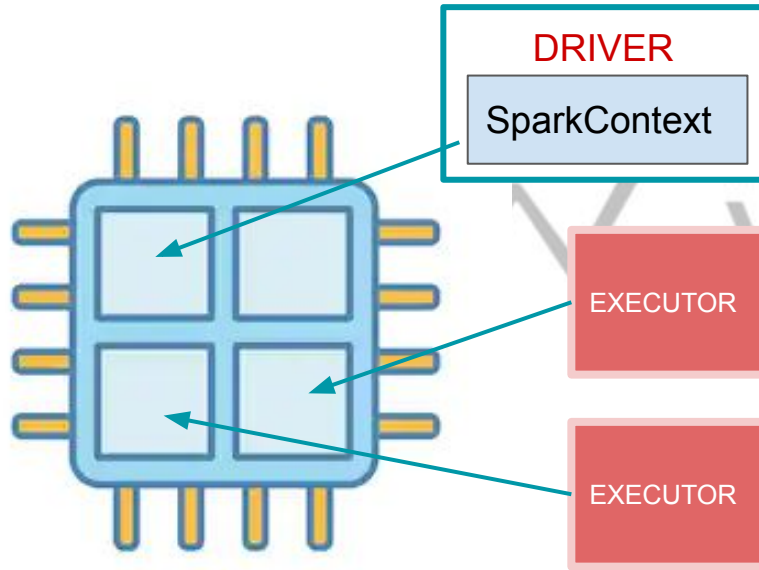


## Spark Application that Run on a Cluster



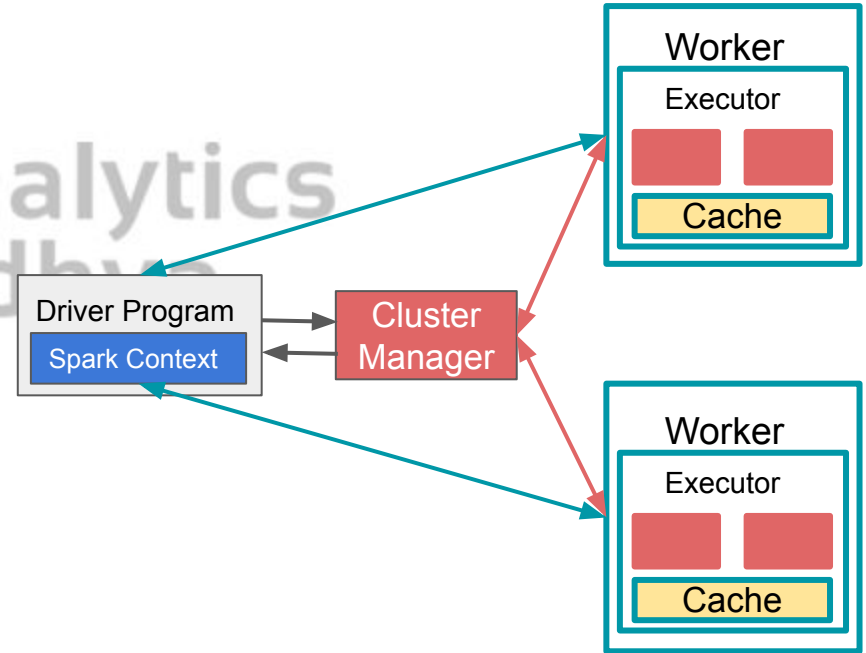
# How to run Spark Applications

## Spark Application that Run Locally



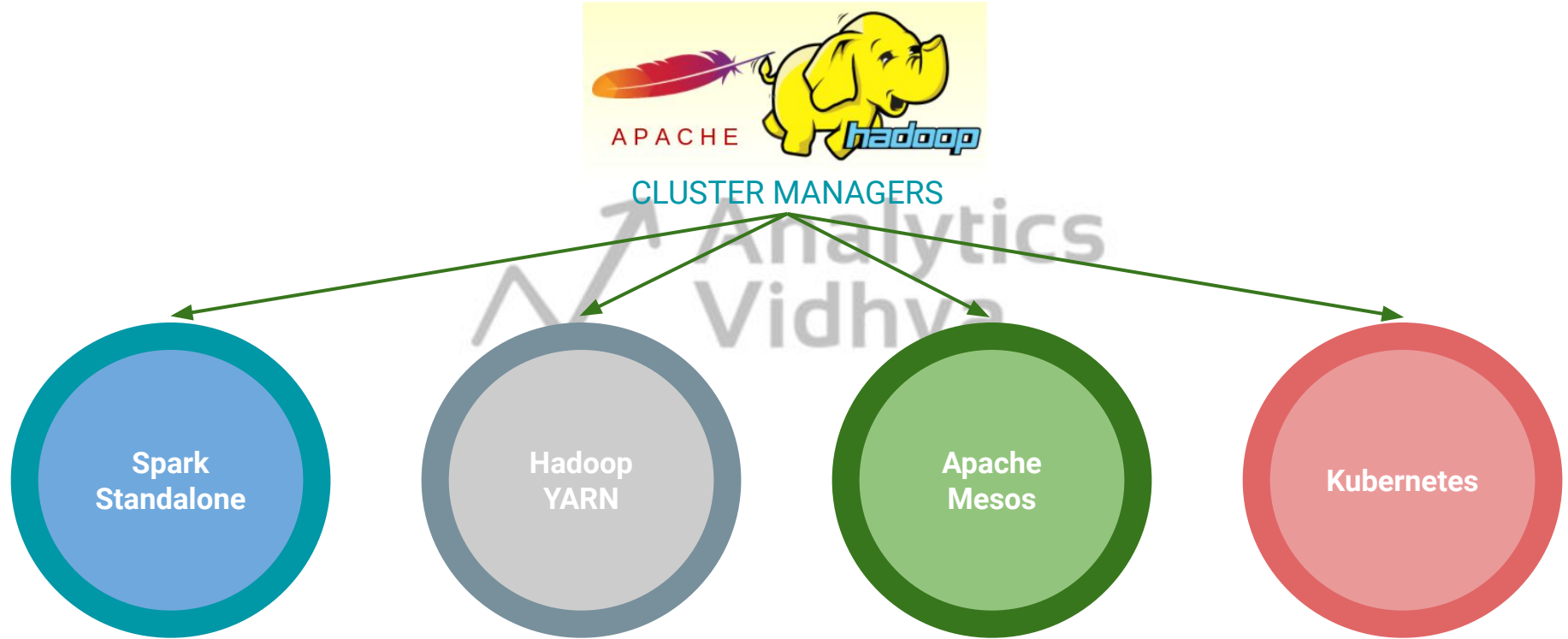
Local Mode: Useful for development and Testing

## Spark Application that Run on a Cluster

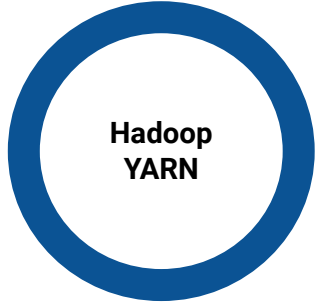


Cluster Mode: Useful for Production

# Supported Cluster Resource Managers



# Supported Cluster Resource Managers



- For production sites or Live applications
- Allows sharing cluster resources with other applications running in the same cluster



# Supported Cluster Resource Managers



## Hadoop YARN

- For production sites or Live applications
- Allows sharing cluster resources with other applications running in the same cluster



## Apache Mesos

- First platform to be supported by Apache Spark
- Not as famous as Apache YARN



# Supported Cluster Resource Managers

## Hadoop YARN

- For production sites or Live applications
- Allows sharing cluster resources with other applications running in the same cluster

## Apache Mesos

- First platform to be supported by Apache Spark
- Not as famous as Apache YARN

## Spark Standalone

- Included with Spark distribution
- Easy to install and run
- Useful for Testing, Development and Learning
- But its has limited scalability

# Supported Cluster Resource Managers

## Hadoop YARN

- For production sites or Live applications
- Allows sharing cluster resources with other applications running in the same cluster

## Apache Mesos

- First platform to be supported by Apache Spark
- Not as famous as Apache YARN

## Spark Standalone

- Included with Spark distribution
- Easy to install and run
- Useful for Testing, Development and Learning
- But its has limited scalability

## Kubernetes

- an open-source system for automating deployment, scaling, and management of containerized applications



Thank You