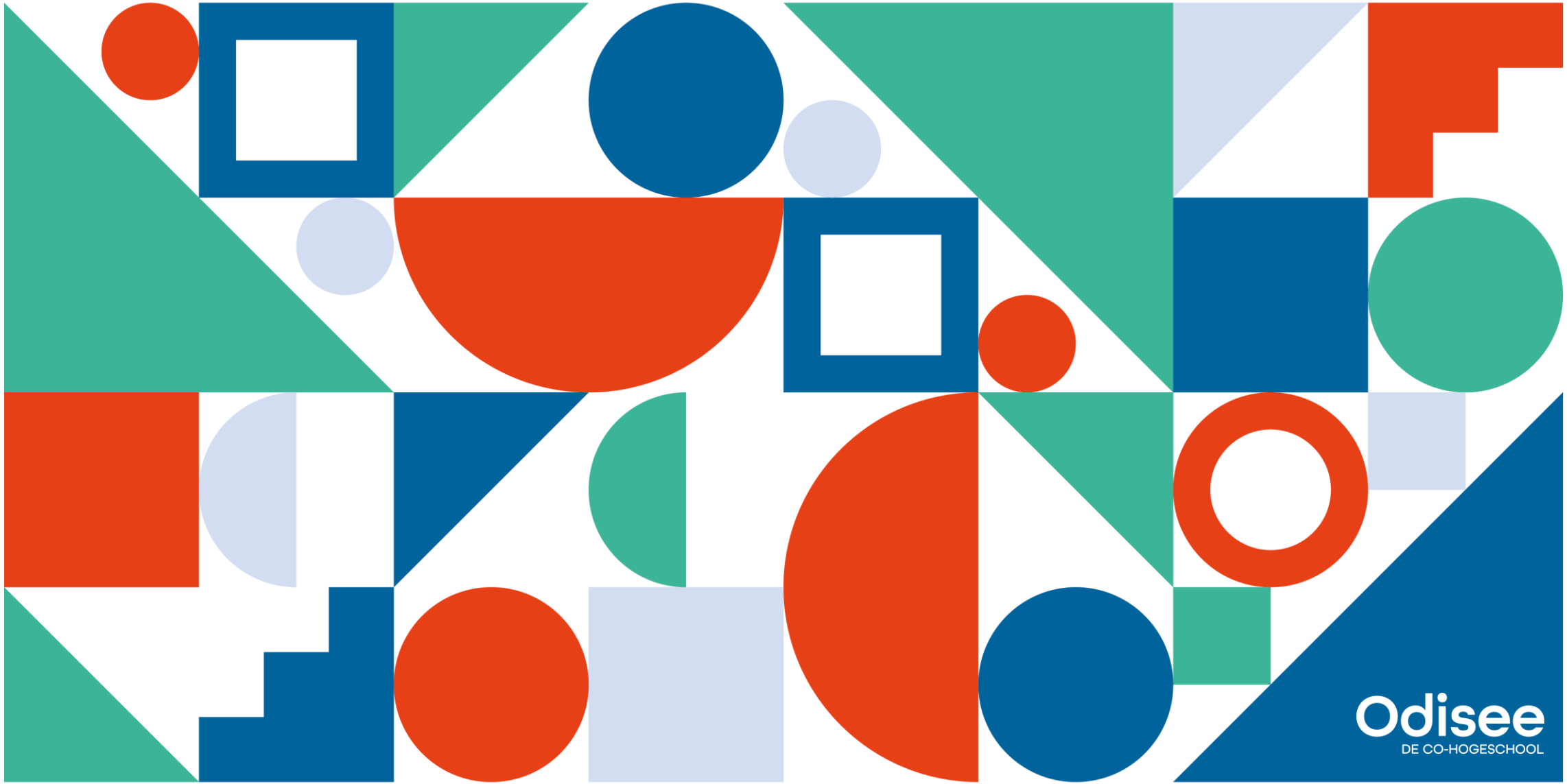

Odisee
DE CO-HOGESCHOOL

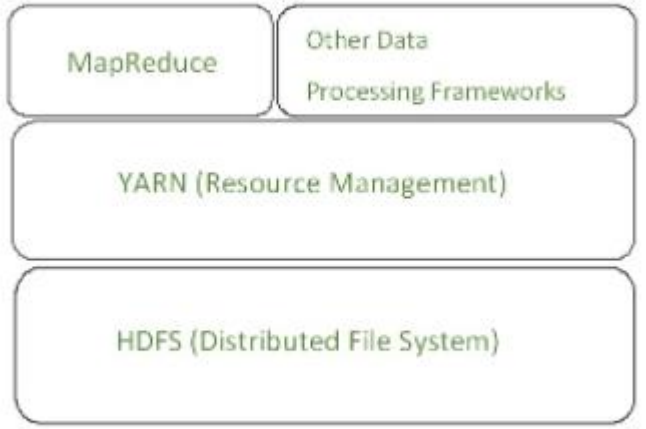
# YARN

Jens Baetens

Odisee
DE CO-HOGESCHOOL

# Wat is YARN?

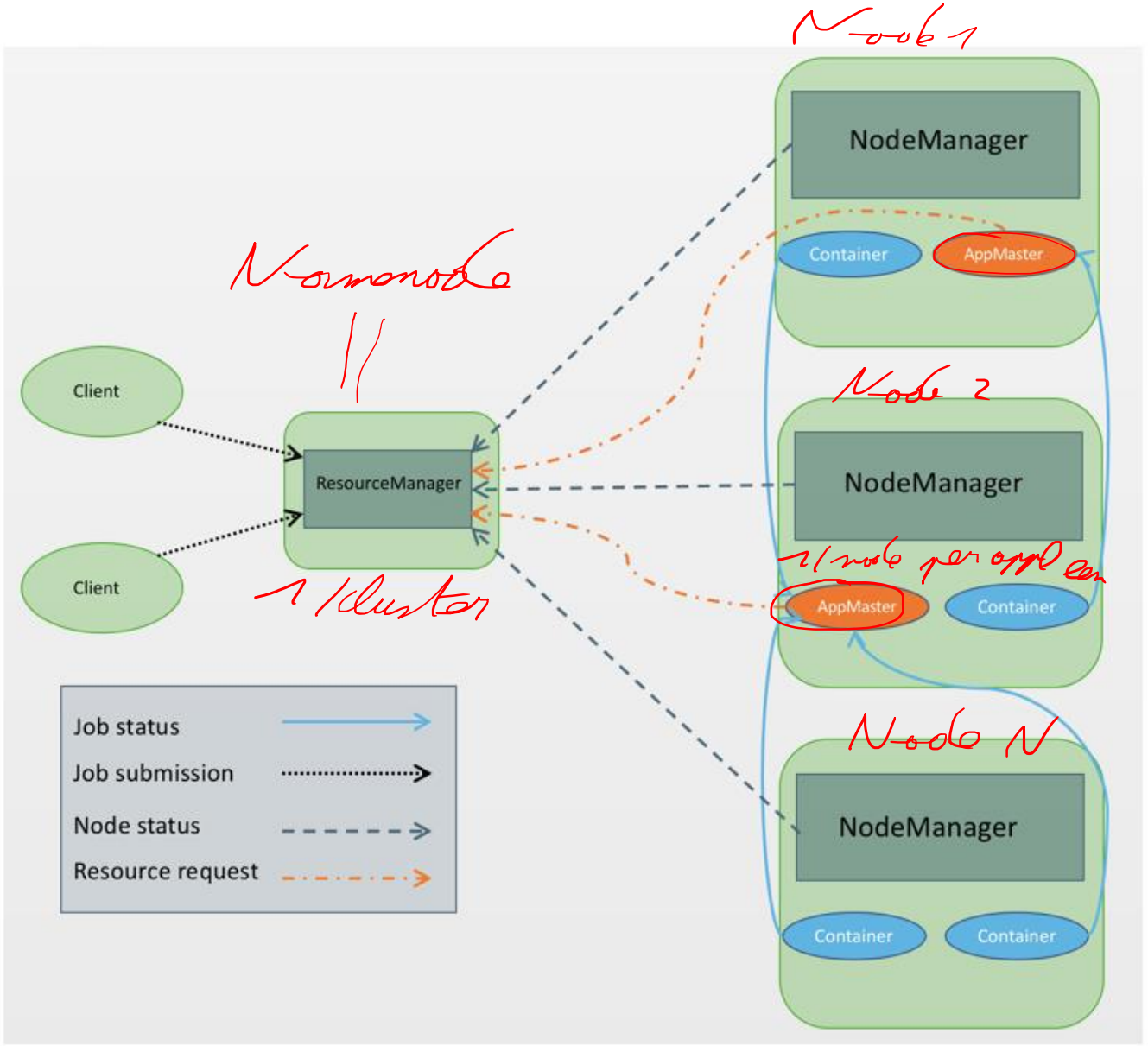
- Yet-Another-Resource-Manager
- Splitsen resource-management en rekenlaag
- Laat niet alleen map reduce toe maar ook:
  - Graph processing
  - Stream processing
  - Interactive processing
  - …



Hadoop 2.0

# Componenten

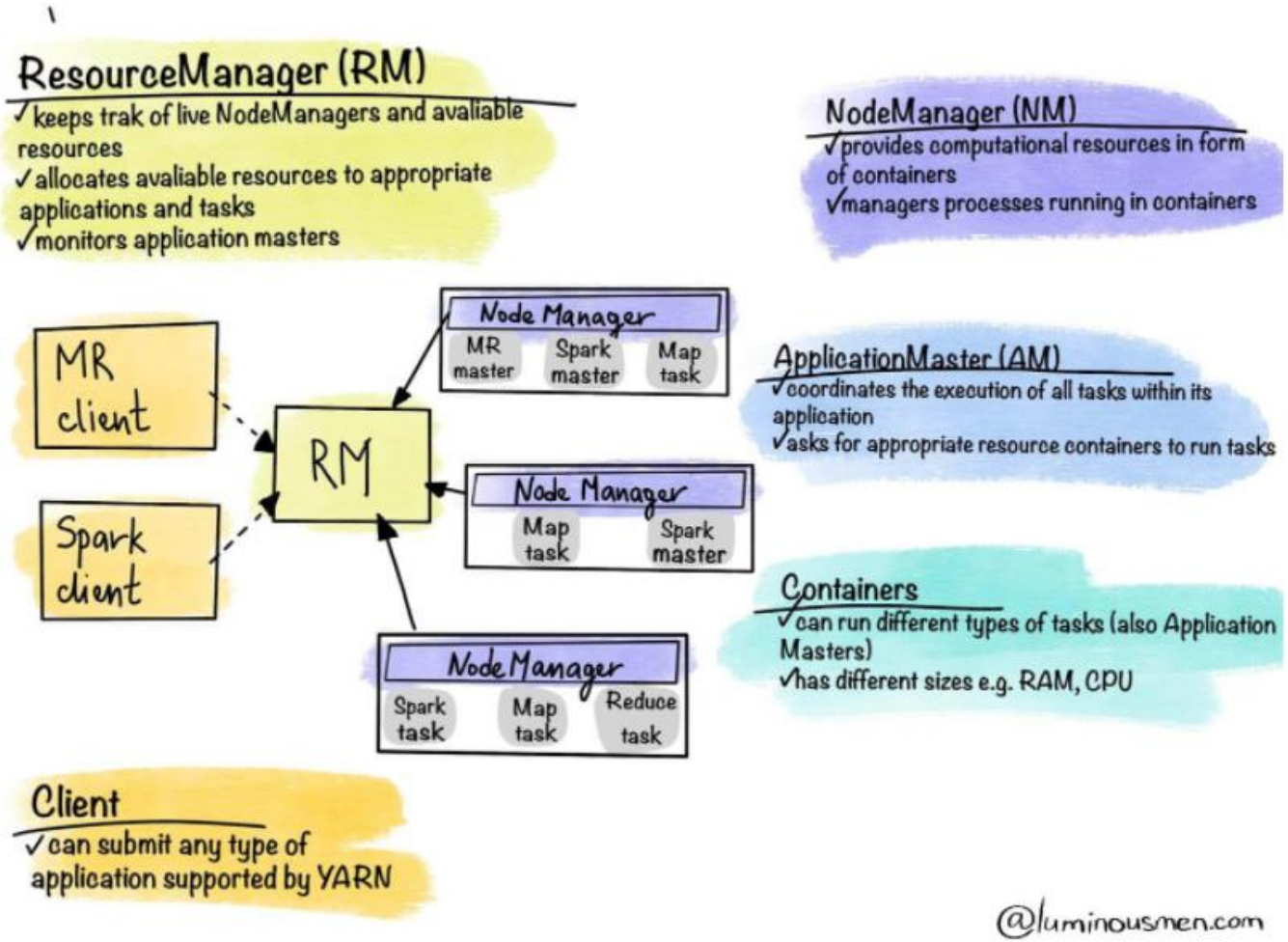
- Client
- Resource manager
  - Scheduler
  - Application manager
- Node manager
- Application master
- Container




https://docs.cloudera.com/HDPDocuments/HDP3/HDP-3.0.0/data-operating-system/content/apache_yarn.html
https://www.geeksforgeeks.org/hadoop-yarn-architecture/

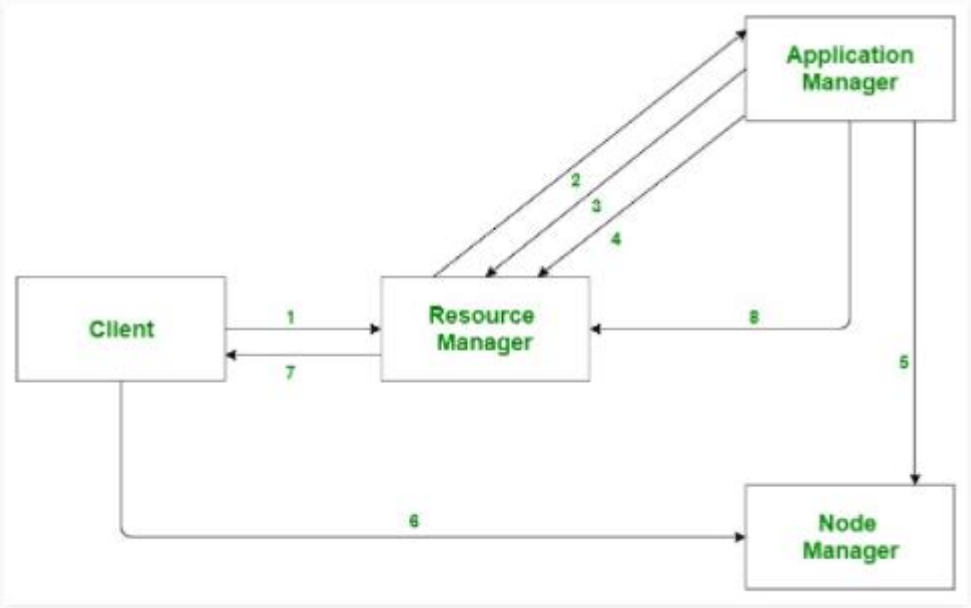
# Componenten

- Client
- Resource manager
  - Scheduler
  - Application manager
- Node manager
- Application master
- Container




https://luminousmen.com/post/hadoop-yarn-spark

# Yarn - Workflow

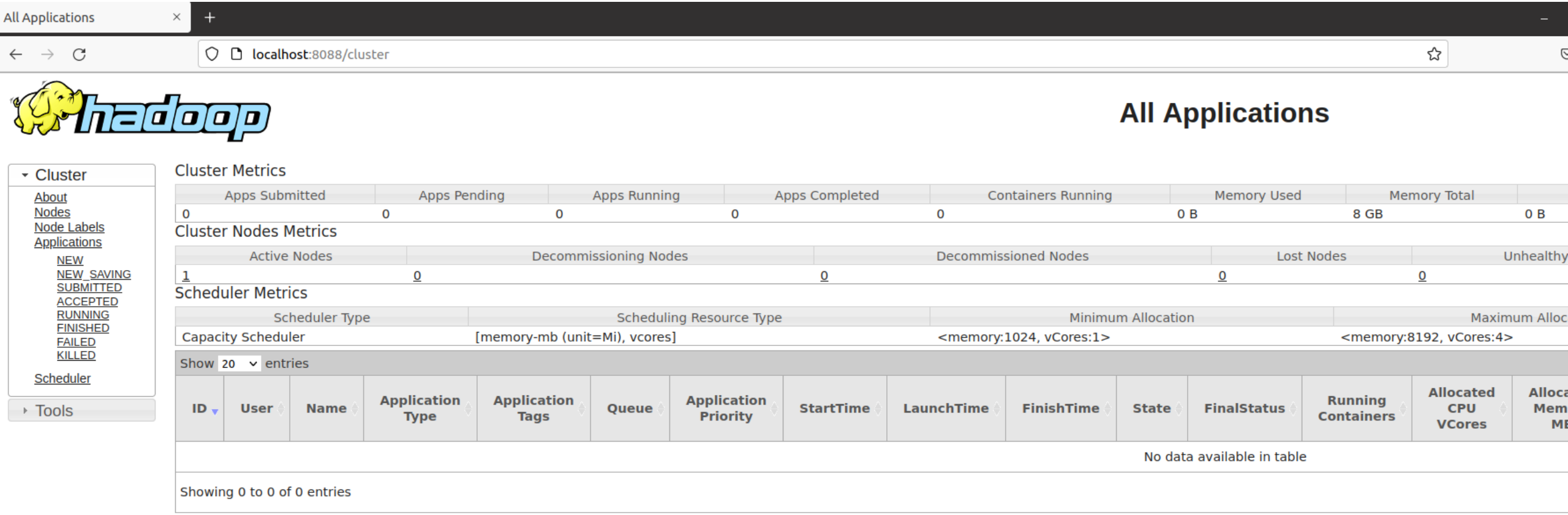



- De client voegt een applicatie toe
- De resource manager maakt een container voor de application manager
- De application manager registreert zich bij de resource manager
- De application manager vraagt containers aan de resource manager
- De application manager vraagt aan de node manager om de containers op te starten
- De applicatie wordt uitgevoerd in de containers
- De client contacteert de resource manager om de status te monitoren
- Wanneer de applicatie is uitgevoerd, meld de application manager dit aan de resource manager om alles vrij te geven

# Features

- Goede (horizontale) schaalbaarheid
- Compatibel met oudere versies
- Meerdere resource types kunnen ingesteld worden (CPU, GPU, Ram)
- Benutting van de cluster geoptimaliseerd door dynamische allocatie
- Containers om meerdere versies van dezelfde applicatie tegelijkertijd te starten
- Multi-tenancy: Alternatieven voor map-reduce voor simultane, batch, interactieve en real-time toegang tot een dataset

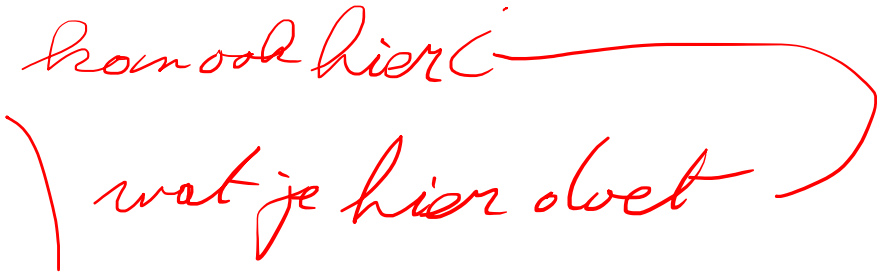
# YARN - UI


All Applications
localhost:8088/cluster
hadoop
All Applications
Cluster
About
Nodes
Node Labels
Applications
NEW
NEW_SAVING
SUBMITTED
ACCEPTED
RUNNING
FINISHED
FAILED
KILLED
Scheduler
Tools
Cluster Metrics
Apps Submitted
Apps Pending
Apps Running
Apps Completed
Containers Running
Memory Used
Memory Total
0
0
0
0
0
0 B
8 GB
0 B
Cluster Nodes Metrics
Active Nodes
Decommissioning Nodes
Decommissioned Nodes
Lost Nodes
Unhealthy
1
0
0
0
0
Scheduler Metrics
Scheduler Type
Scheduling Resource Type
Minimum Allocation
Maximum Alloc
Capacity Scheduler
[memory-mb (unit=Mi), vcores]
<memory:1024, vCores:1>
<memory:8192, vCores:4>
Show 20 entries
ID
User
Name
Application Type
Application Tags
Queue
Application Priority
StartTime
LaunchTime
FinishTime
State
FinalStatus
Running Containers
Allocated CPU VCores
No data available in table
Showing 0 to 0 of 0 entries

# Configuration

# Configuration

- Drie files bevatten hier informatie over
  - hadoop/etc/hadoop/yarn-site.xml
  - hadoop/etc/hadoop/resource-types.xml
  - hadoop/etc/hadoop/node-resources.xml




- https://hadoop.apache.org/docs/stable/hadoop-yarn/hadoop-yarn-site/ResourceModel.html

# resource-types.xml

- Defineer types van resources
- Ram en memory standard
- Bvb GPU, licenties, …

- Kan ook in yarn-site.xml

```
<configuration>
  <property>
    <name>yarn.resource-types</name>
    <value>resource1, resource2</value>
  </property>

  <property>
    <name>yarn.resource-types.resource1.units</name>
    <value>G</value>
  </property>

  <property>
    <name>yarn.resource-types.resource2.minimum-allocation</name>
    <value>1</value>
  </property>

  <property>
    <name>yarn.resource-types.resource2.maximum-allocation</name>
    <value>1024</value>
  </property>
</configuration>
```

| Configuration Property | Description |
|---|---|
| yarn.resource-types | Comma-separated list of additional resources. May not include memory, memory-mb, or vcores |
| yarn.resource-types.<resource>.units | Default unit for the specified resource type |
| yarn.resource-types.<resource>.minimum-allocation | The minimum request for the specified resource type |
| yarn.resource-types.<resource>.maximum-allocation | The maximum request for the specified resource type |

# node-resources.xml

- Geeft weer wat elke node aanbiedt
- Kan ook in de yarn-site.xml

```
<configuration>
 <property>
   <name>yarn.nodemanager.resource-type.resource1</name>
   <value>5G</value>
 </property>

 <property>
   <name>yarn.nodemanager.resource-type.resource2</name>
   <value>2m</value>
 </property>

</configuration>
```

# yarn-site.xml

- Aantal map reduce settings (met default waarden)
- Laat ook toe om resource profiles aan te maken
  - Eenvoudig bepaalde configuratie aanvragen voor een applicatie

→ niet mogelijk in pseudo-distributed

```
{
    "small": {
        "memory-mb" : 1024,
        "vcores" : 1
    },
    "default" : {
        "memory-mb" : 2048,
        "vcores" : 2
    },
    "large" : {
        "memory-mb": 4096,
        "vcores" : 4
    },
    "compute" : {
        "memory-mb" : 2048,
        "vcores" : 2,
        "gpu" : 1
    }
}
```