

Przechowywanie oraz przetwarzanie danych

Apache Hadoop

Agenda

1. Wprowadzenie
2. HDFS
 - a. Architektura
 - b. Rozproszone przechowywanie informacji
 - c. Bezpieczeństwo danych
 - d. Korzystanie z CLI
3. YARN
 - a. Czym jest YARN i powody powstania
 - b. Architektura klastrów obliczeniowych opartych o YARN
 - c. Zarządzanie zasobami
4. MapReduce
5. Tez
6. Uruchamiamy nasz pierwszy klaster (Część warsztatowa)
 - a. HDFS CLI
 - b. YARN CLI

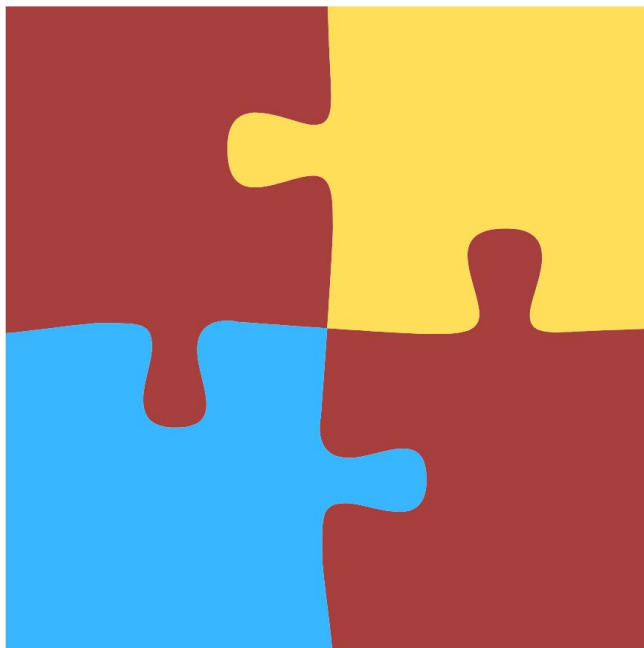
Apache Hadoop

TOOLS

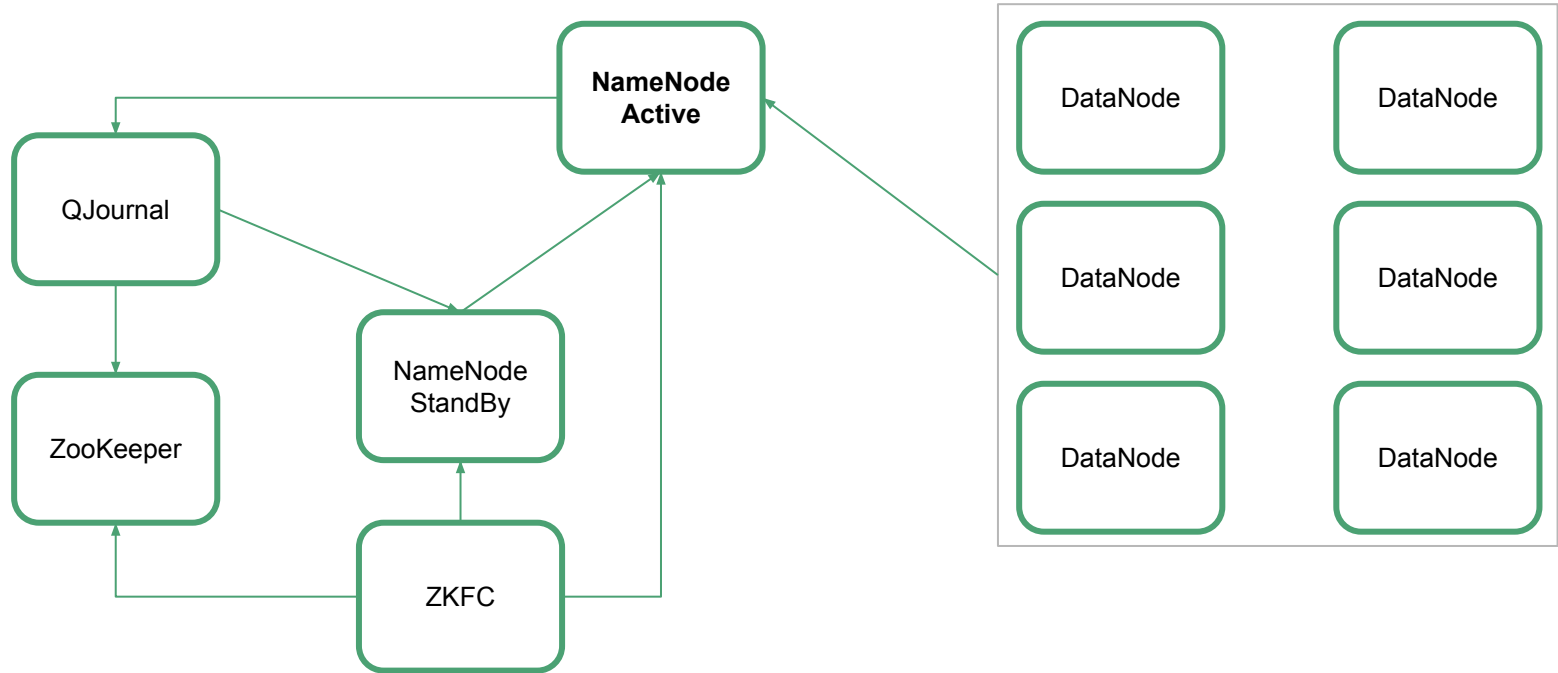
FRAMEWORKS

HDFS

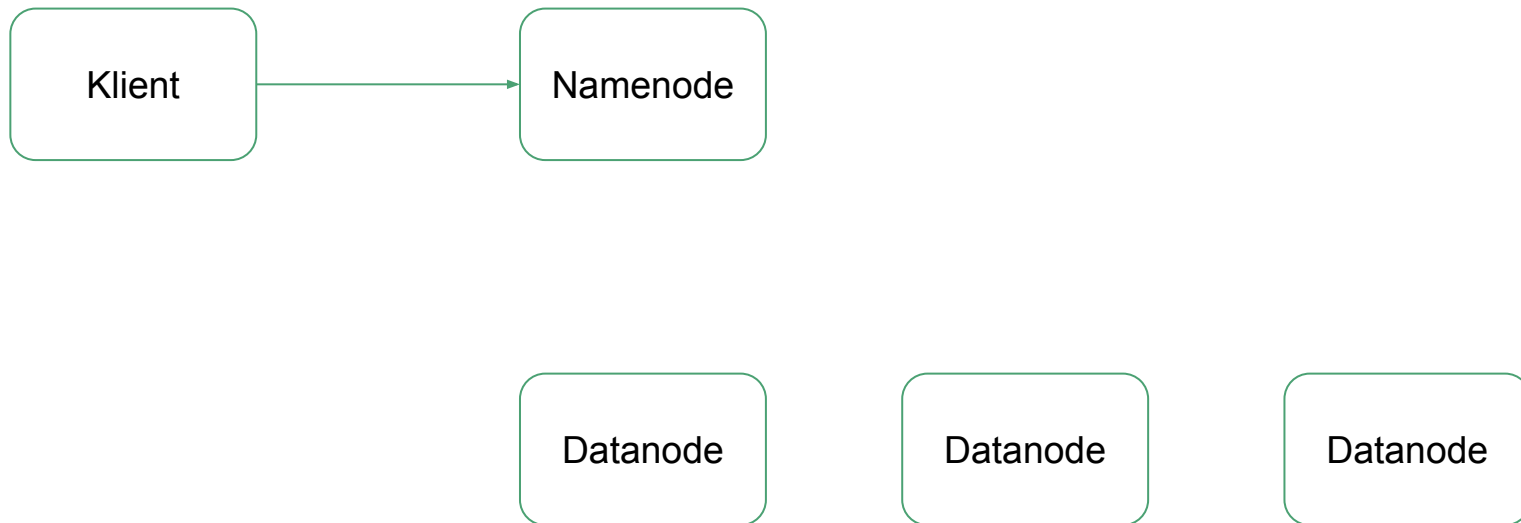
YARN



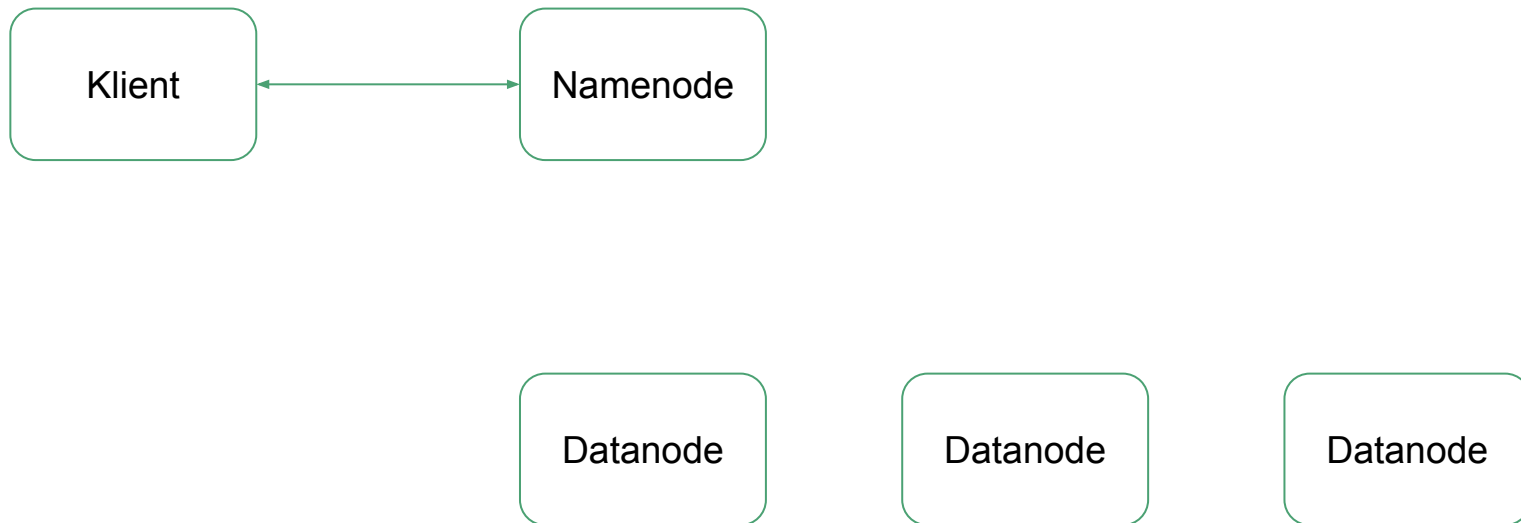
HDFS - Architektura



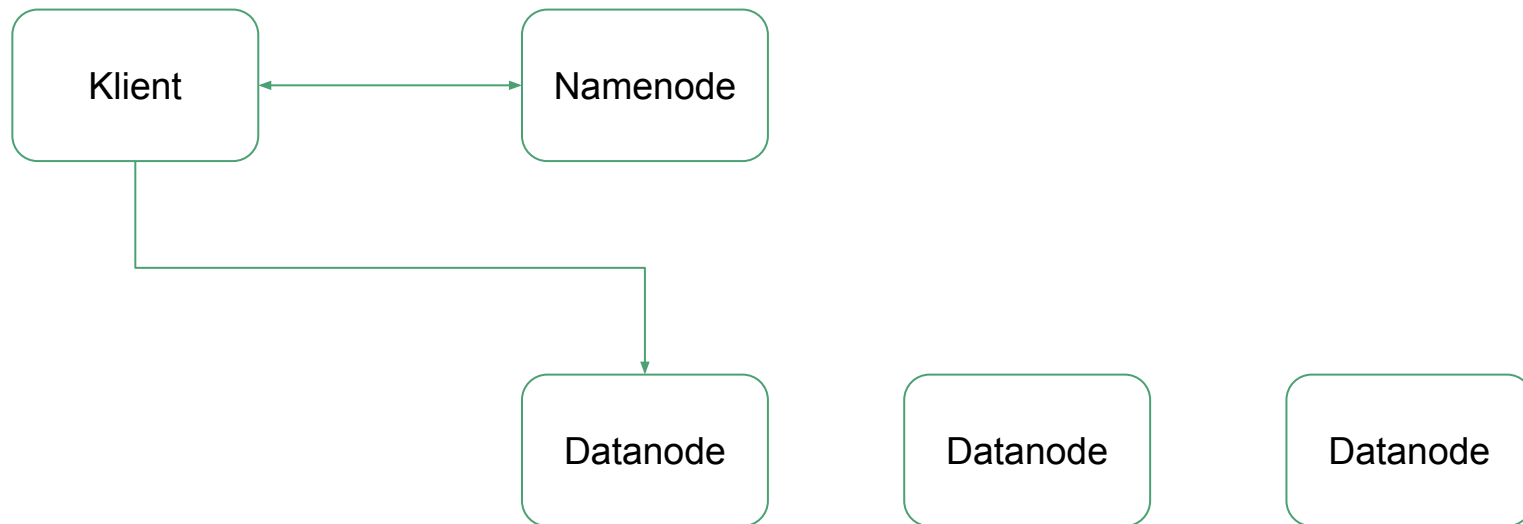
HDFS - Zapis pliku



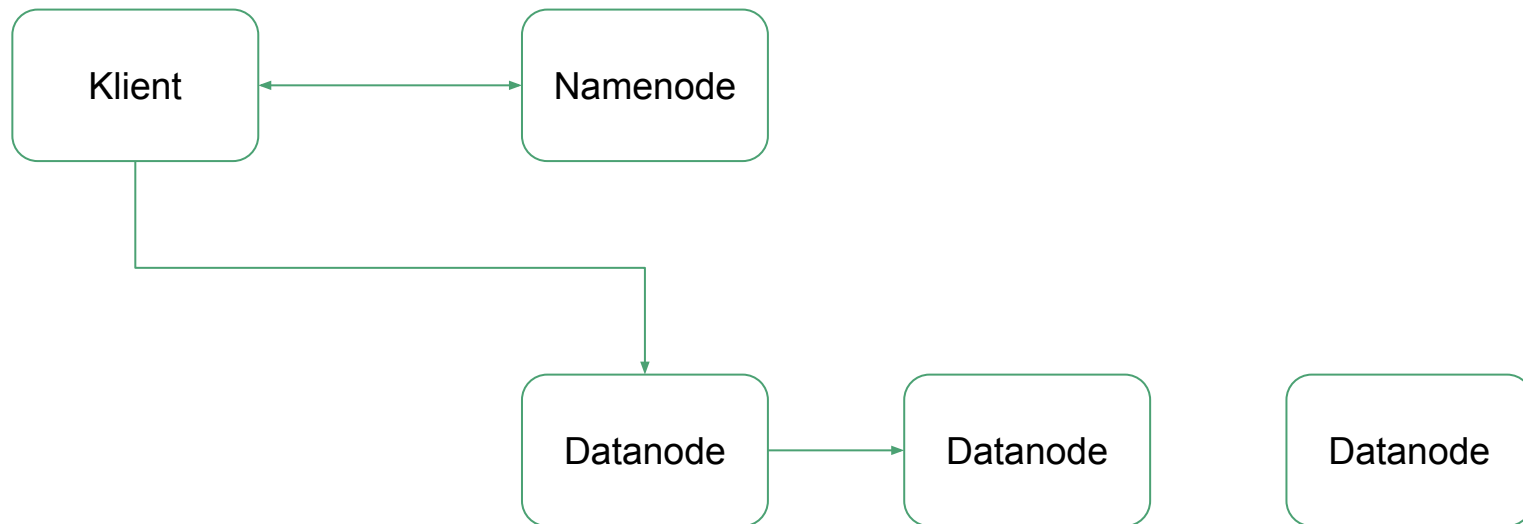
HDFS - Zapis pliku



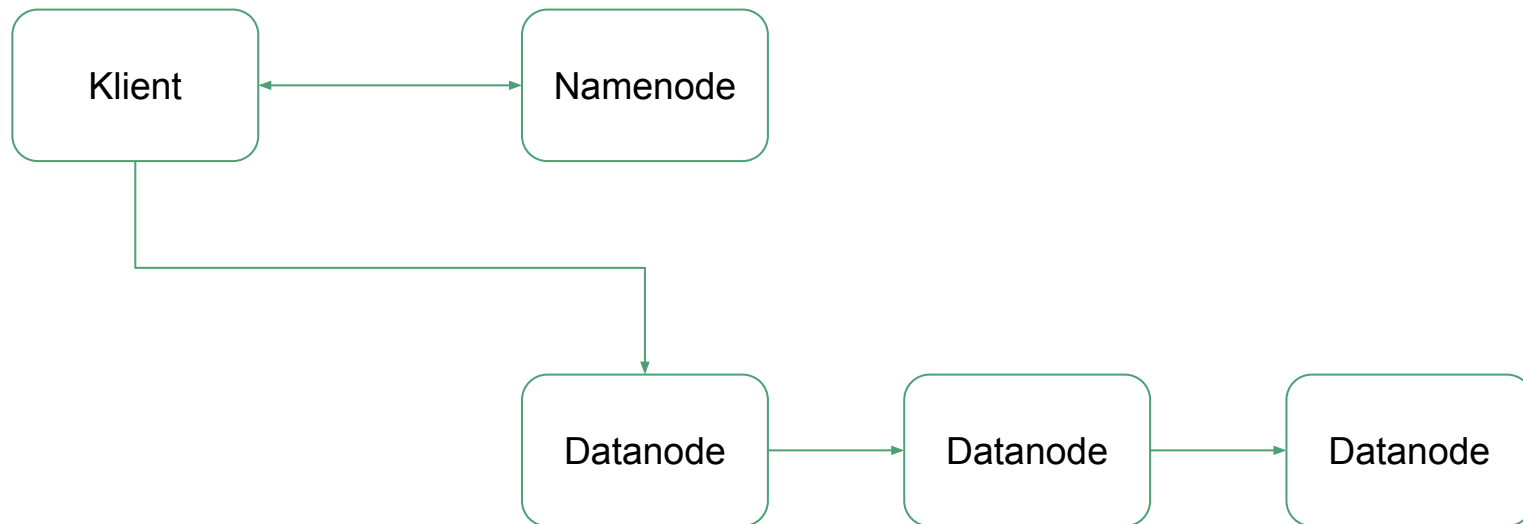
HDFS - Zapis pliku



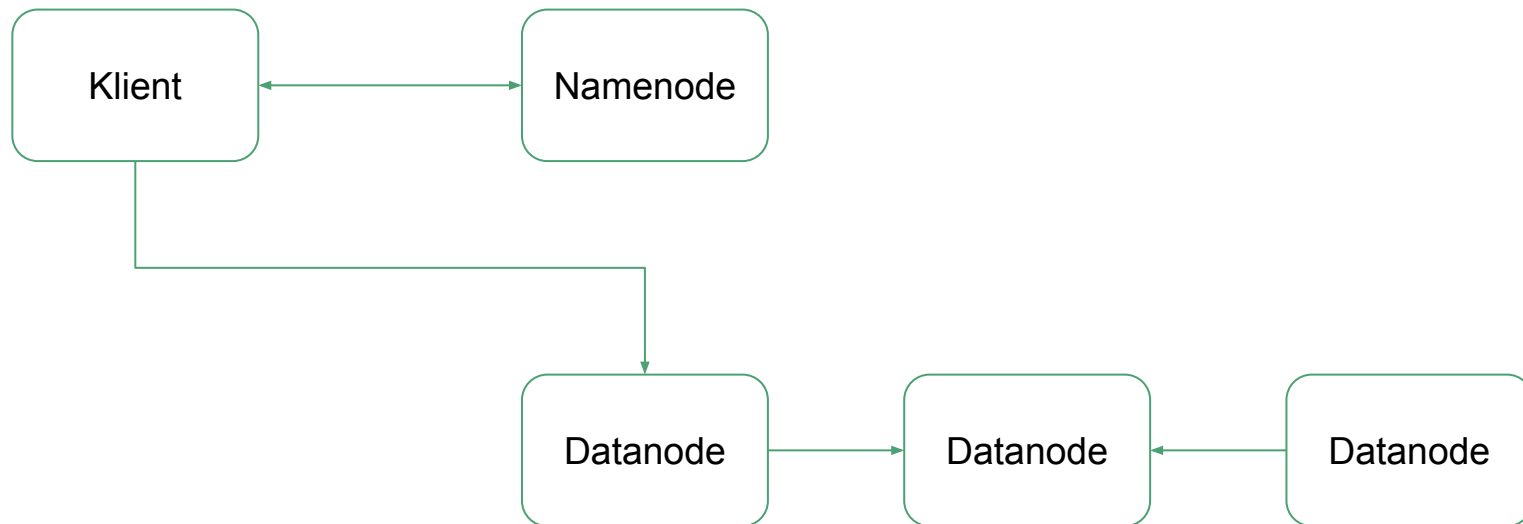
HDFS - Zapis pliku



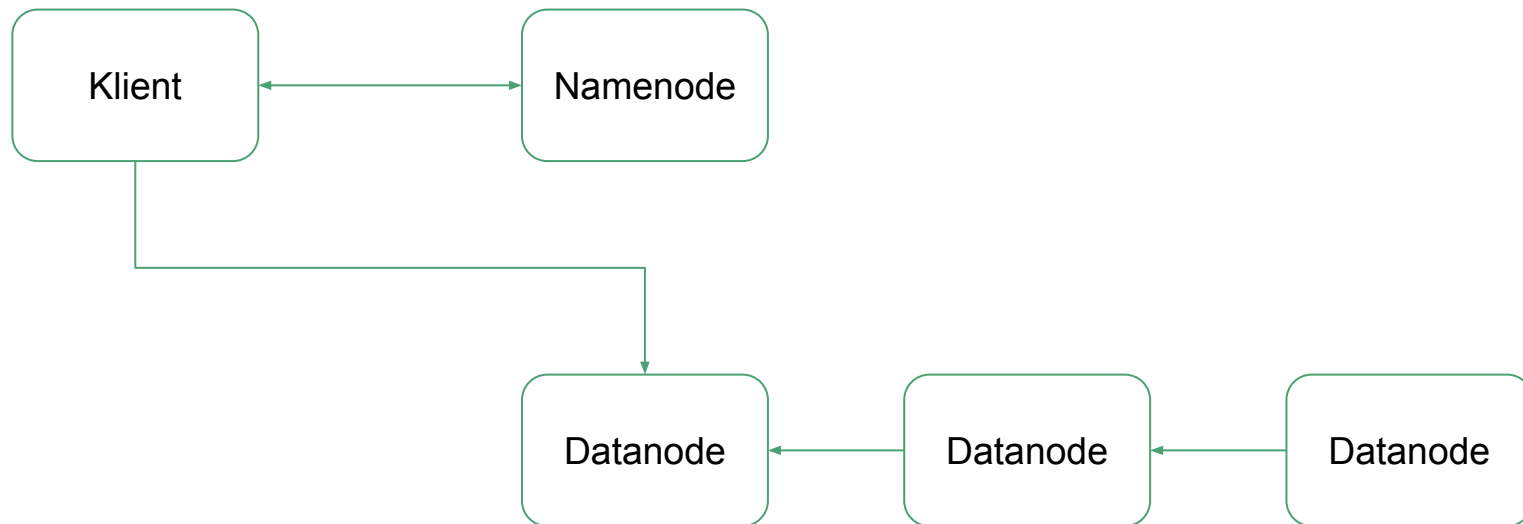
HDFS - Zapis pliku



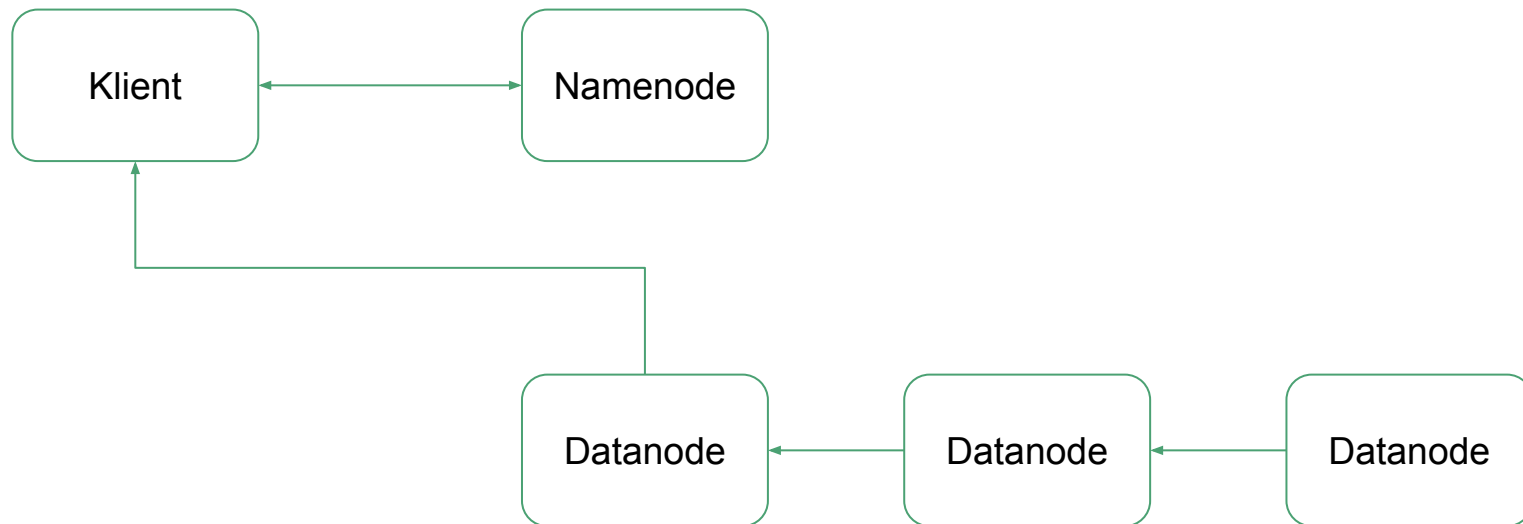
HDFS - Zapis pliku



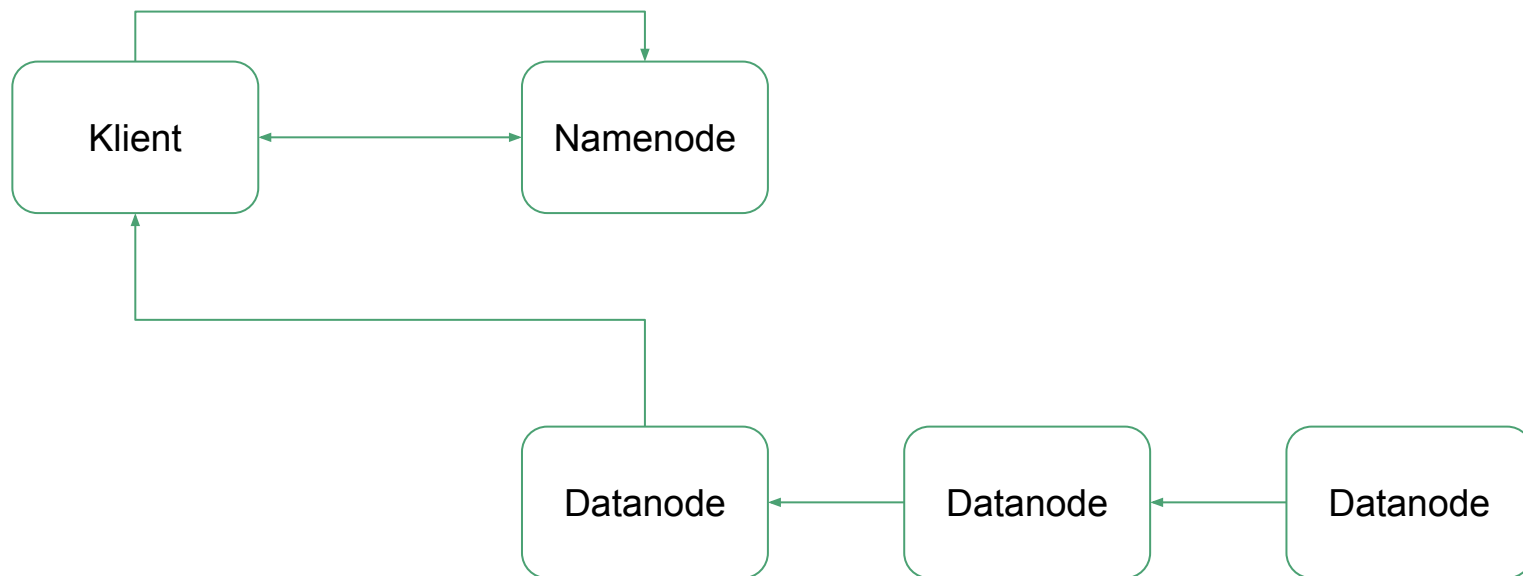
HDFS - Zapis pliku



HDFS - Zapis pliku



HDFS - Zapis pliku



HDFS - Odczyt pliku



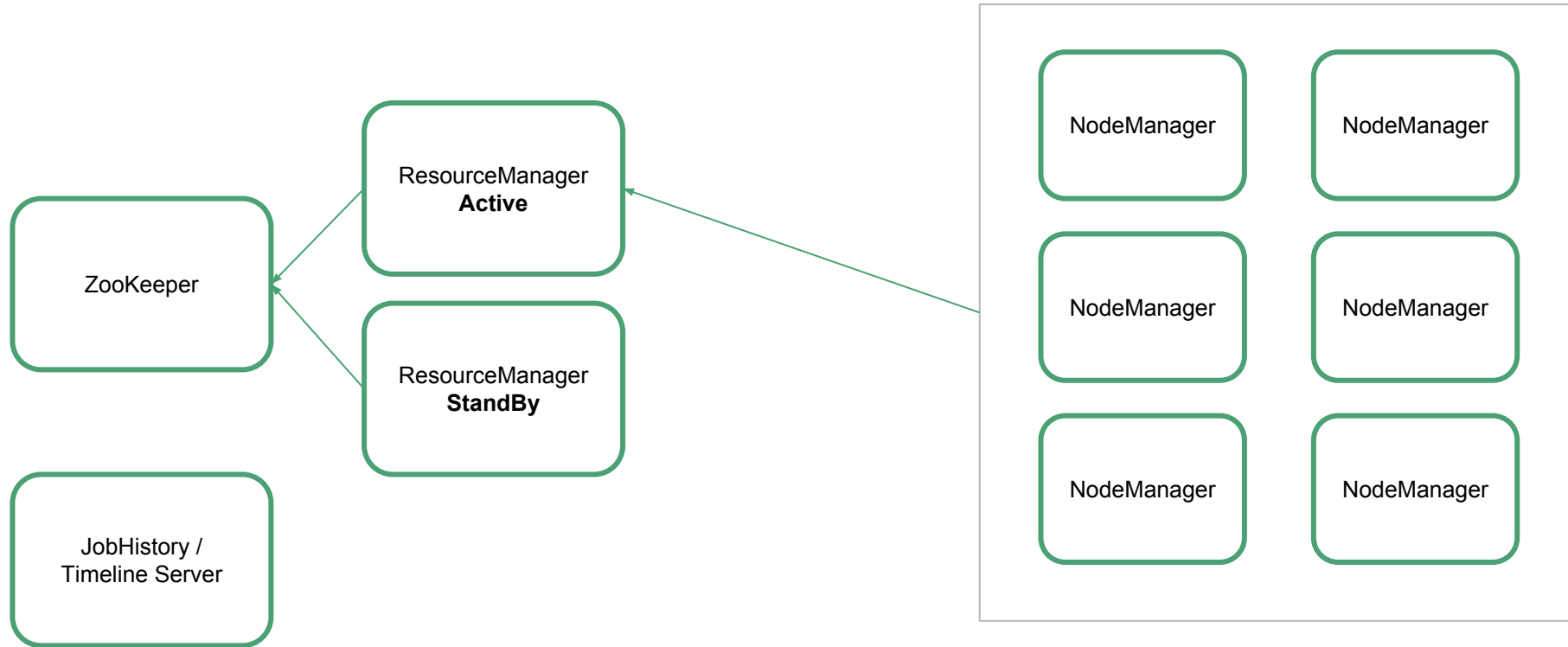
HDFS - Uprawnienia

```
> hdfs dfs -ls hdfs:///tmp/sparkPiExample-all.jar  
  
-rw-r--r--    3 hdfs  hadoop    11670716 2016-09-22 14:49  
hdfs:///tmp/sparkPiExample-all.jar
```

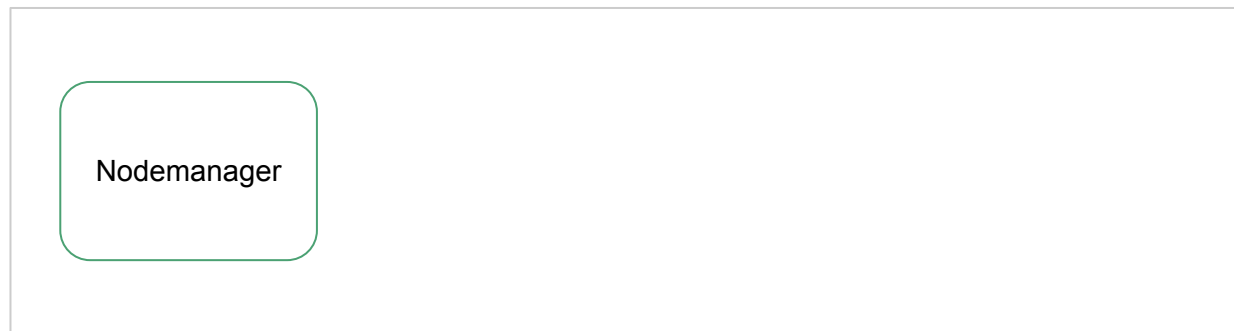
HDFS - CLI

- `hdfs dfs`
- `hdfs fsck`
- `hdfs dfsadmin`
- `hdfs haadmin`

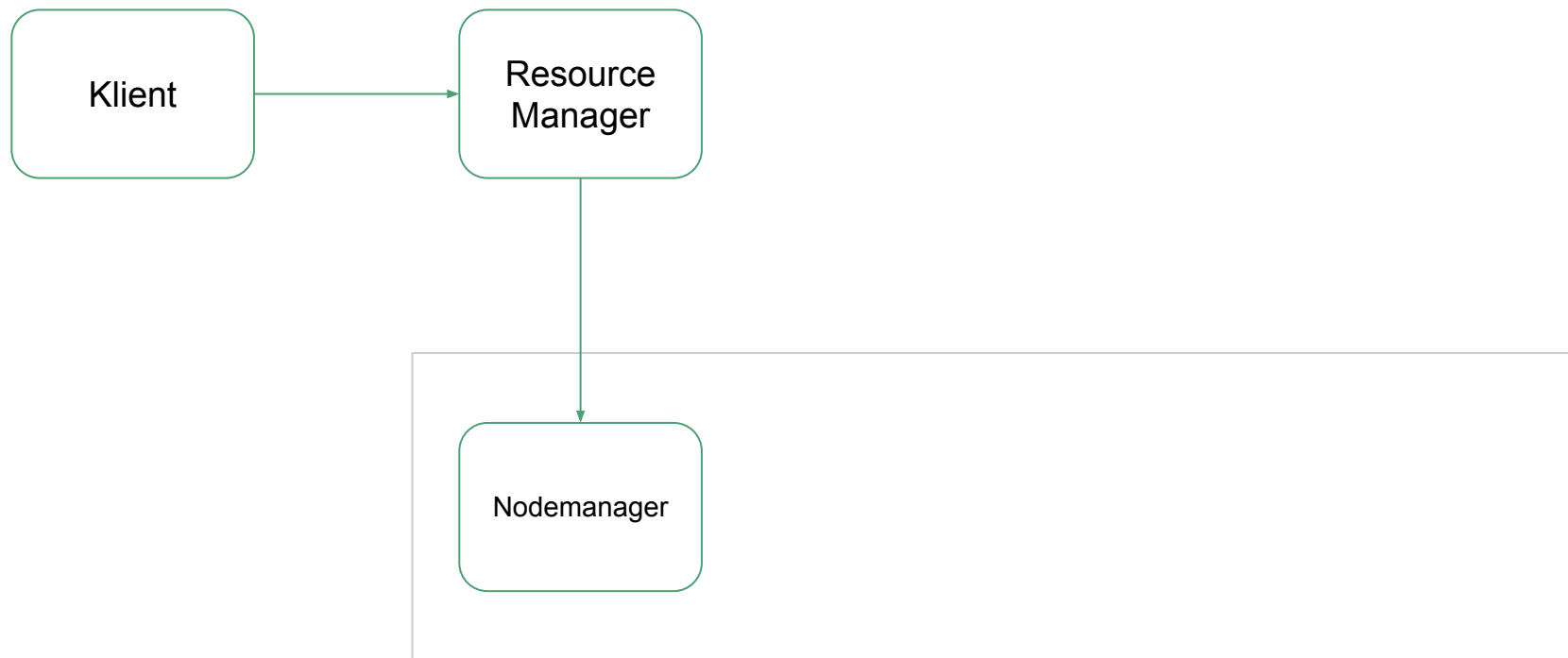
YARN - Architektura



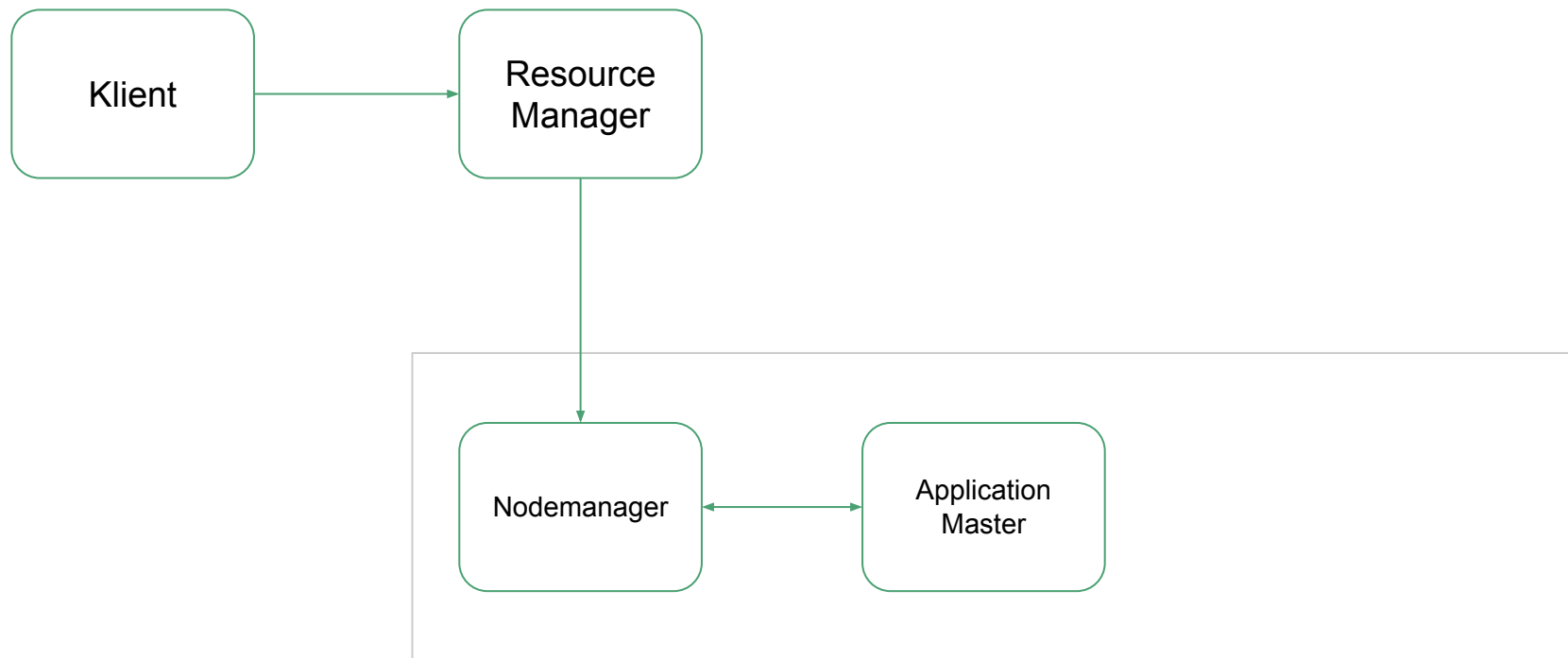
YARN - Uruchamianie zadań



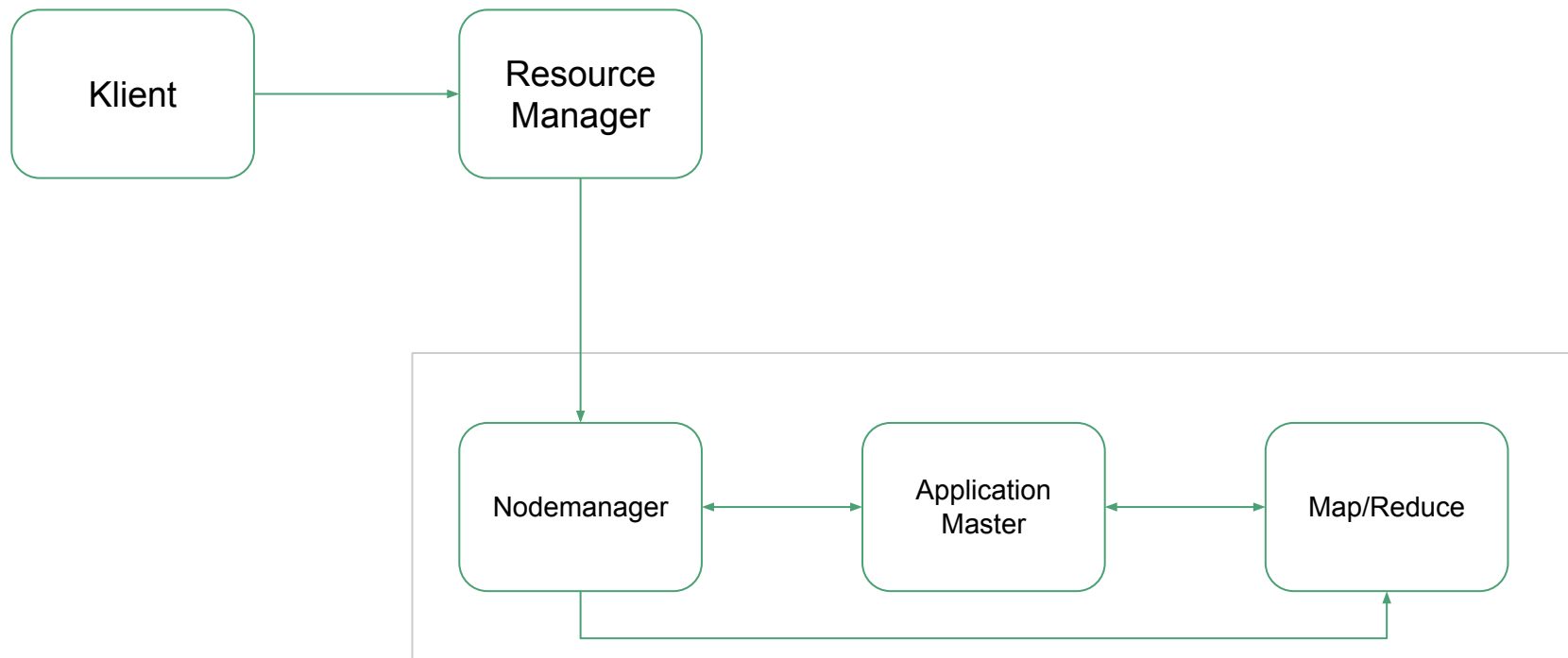
YARN - Uruchamianie zadań



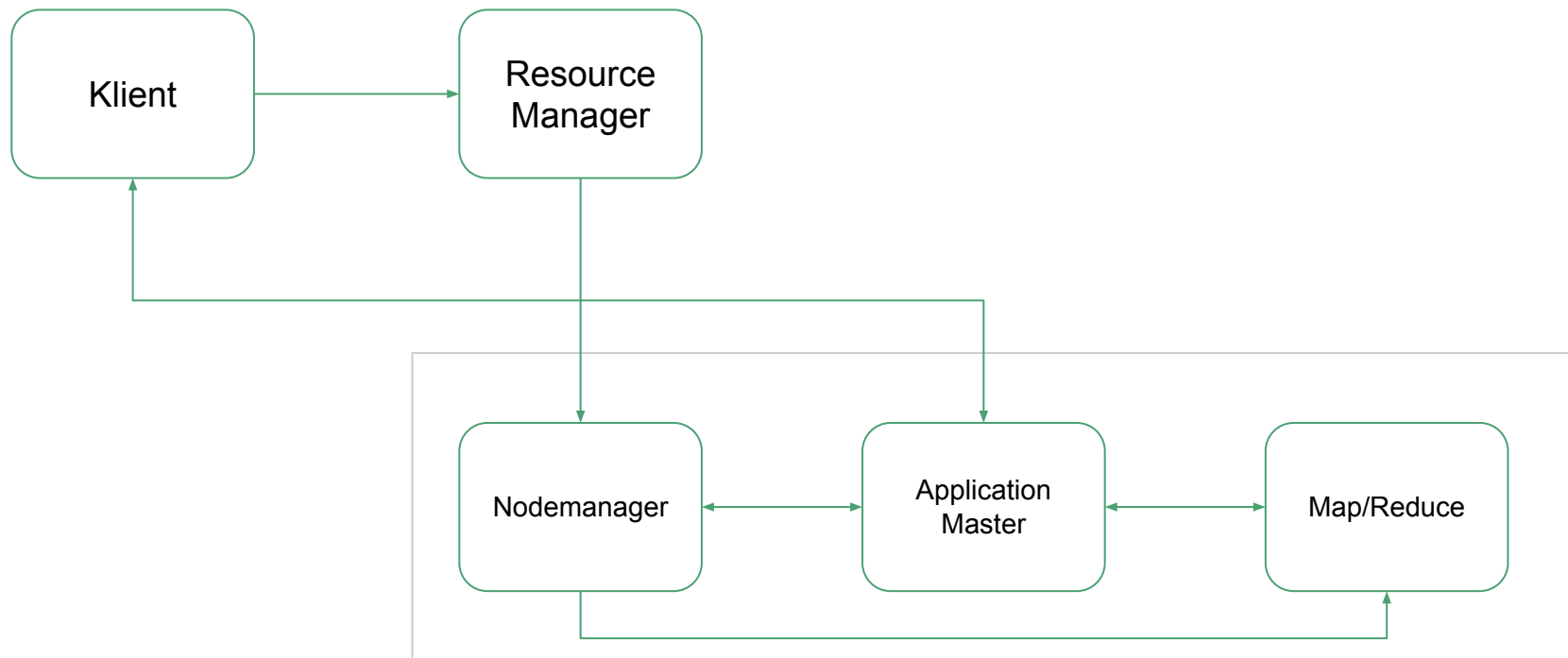
YARN - Uruchamianie zadań



YARN - Uruchamianie zadań



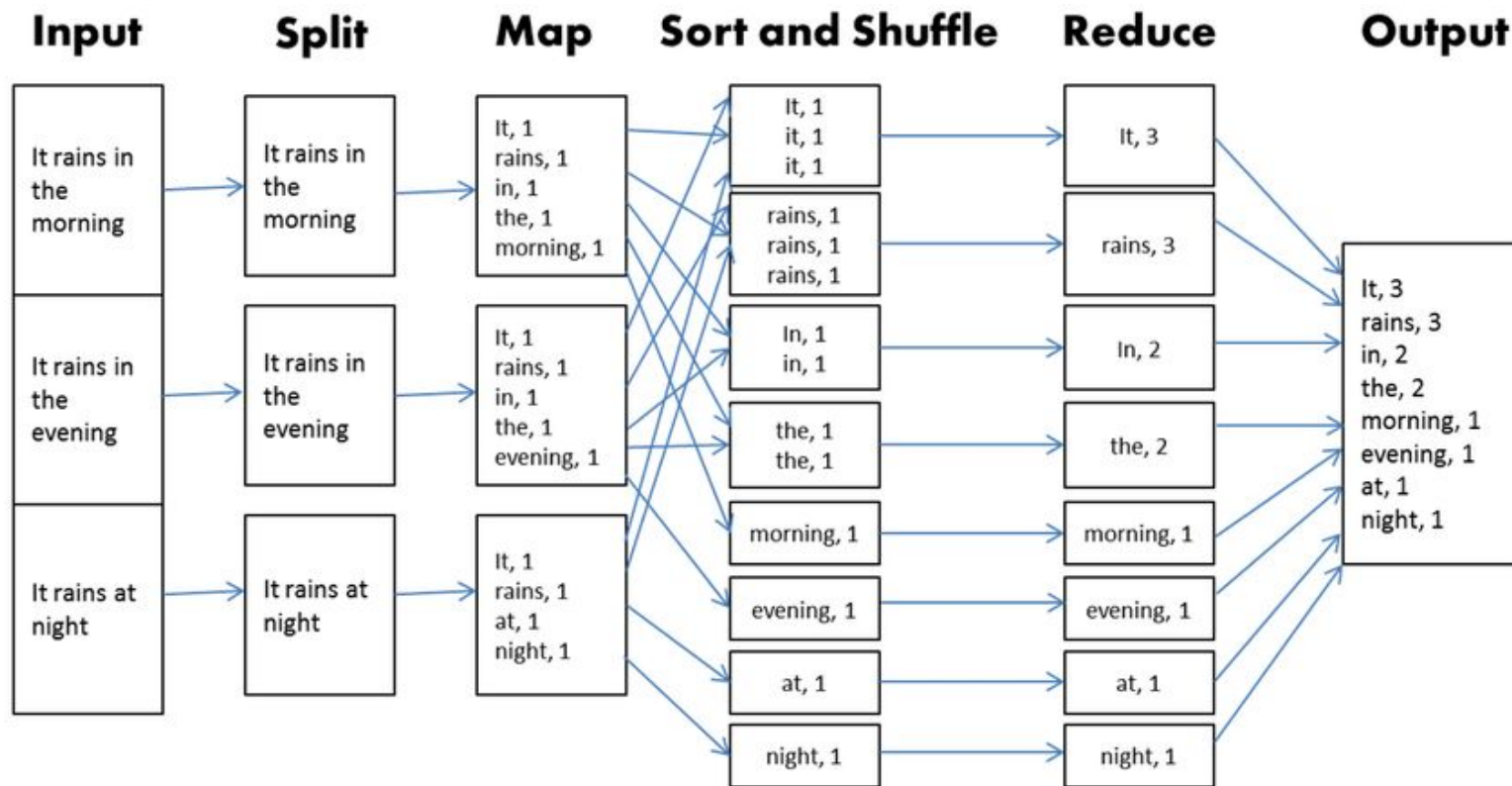
YARN - Uruchamianie zadań



YARN - CLI

- yarn application
- yarn jar
- yarn logs
- yarn queue
- yarn rmadmin

MapReduce



MapReduce in Bash

```
cat /var/log/syslog | cut -d" " -f1 | sort | uniq -c
```

TEZ

