

Elastic Map Reduce

- Hadoop in AWS
- When creating you can specify:
 - instance type of the nodes
 - number of nodes
 - version of hadoop
 - Additional Tools
 - Hive
 - Pig
 - Spak
 - Presto

File systems

```
## HDFS
- use for persistent clusters
- lowest latency
- you can use instance store (much faster)
## EMRFS
- Stored in S3
- can be used for transient clusters
```

- You can use combination of both HDFS and EMRFS for your cluster

Use cases

- Log processing
- Clicksteram Analytics
- Genomis and life sciences