# AWS

Provides:

Infrastructure-as-a-service

* EC2: Purchase of virtual machines of different capabilities, with different OS’s.
* S3: Purchase of storage that is accessed through a simple file system interface.

Platform-as-a-service

* EMR (Elastic Map Reduce): The ability to run programs written using the map-reduce programming language model over EC2 and S3 (data for running is usually stored in S3) infrastructure.

# Map Reduce

* Programming model
* Map and reduce are two functions used in the model.
* Infrastructure supports the running of Map Reduce applications in parallel.
* Hadoop is a widely used implementation of the map-reduce model.

## How does it work?

* Break the document/s up into parts – This is done by the infrastructure
* The map function does a simple job: Maps key to a value. (Generated the data)
* Infrastructure between Map and reduce organizes the key-value pairs
* The reduce function processes the data.

### Map

* Map(key1, value1) (e.g. document identifier, document)🡪 [(key2, value2)] (term, partial count)

### Reduce

* Reduce(key2, [value2]) (term, partial occurance) 🡪 [(key3, value3)] (e.g. term, total occurrence)