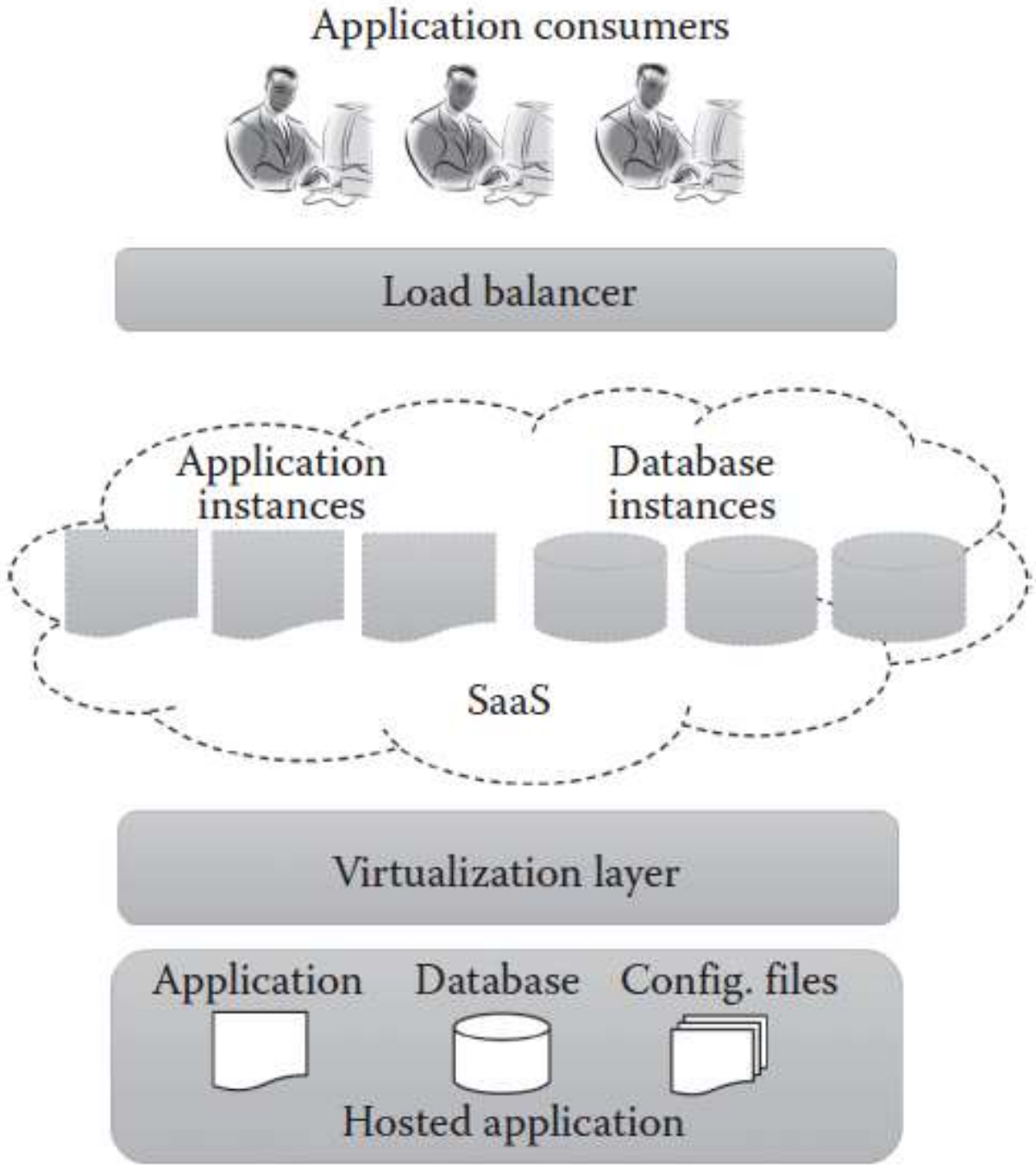
Software as a Service

Cloud Computing

Software as a service

When applications are run in the cloud, they are accessed as a service - this is known as Software as a Service (SaaS).

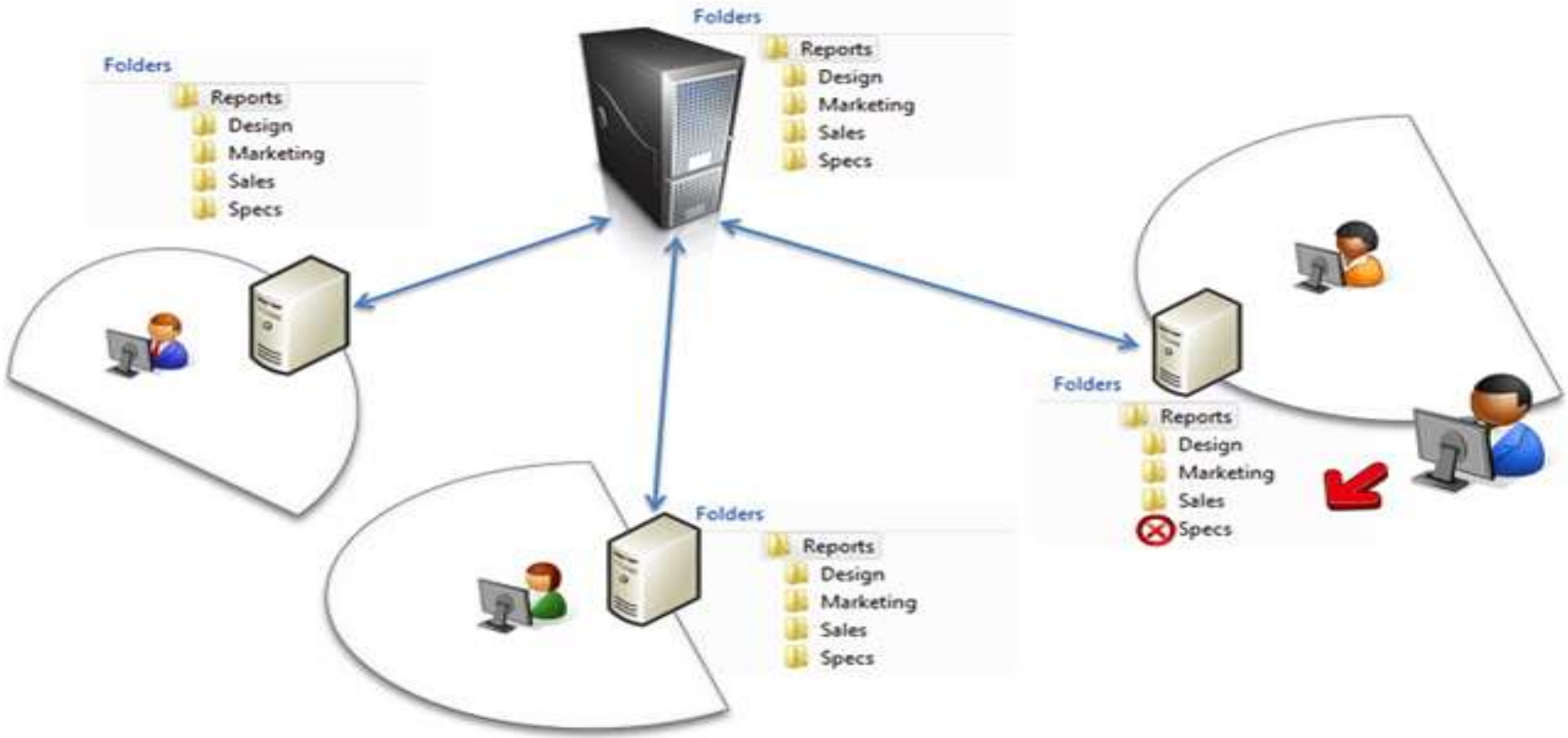
The software delivery model that allows the customers to access the software that is hosted in the service provider data center through the Internet is known as Software as a Service (SaaS).


Application consumers
Load balancer
Application instances
Database instances
SaaS
Virtualization layer
Application
Database
Config. files
Hosted application

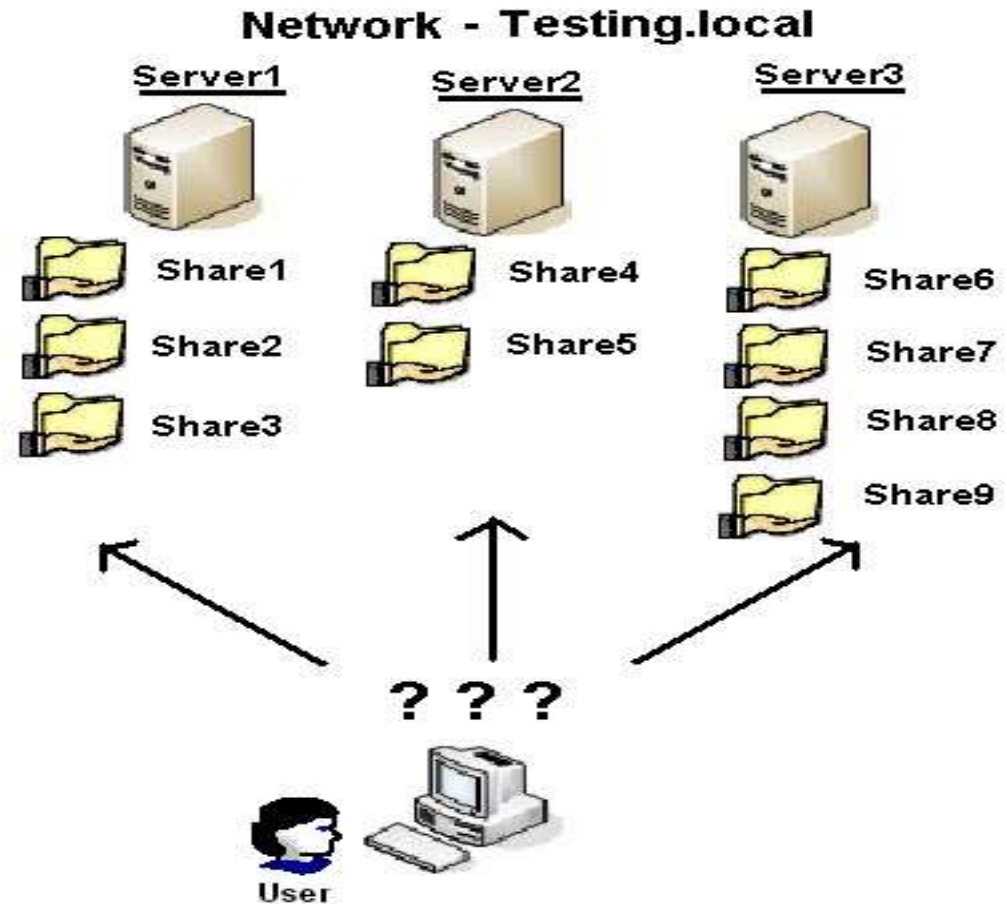
Distributed File System

- When data grows over storage capacity of single machine we need to distribute the data on a network.
- A distributed file system (DFS) is a file system with data stored on a server.
- The data is accessed and processed as if it was stored on the local client machine.
- The **DFS** makes it convenient to share information and files among users on a network in a **controlled** and authorized way

Distributed File System



- **Distributed File System** (**DFS**) allows administrators to group shared folders located on different servers by transparently connecting them to one or more **DFS** namespaces.
- A **DFS** namespace is a virtual view of shared folders in an organization.



List of DFS

- Apache Hadoop
- Parallel Virtual File Server
- IBM General Parallel File Server
- XtreemFS
- ObjectiveFS
- RedHat Storage Server
- BigCouch
- Cloudant
- Etc…

Hadoop

- Hadoop Distributed File System.
- Open source apache framework **to store and process** huge amount of data efficiently.
- It is not recommended for small amount of datasets.
- HDFS uses **clusters of commodity hardware**.
- It doesn’t requires high reliable hardware.
- Cluster : Machines within LAN.

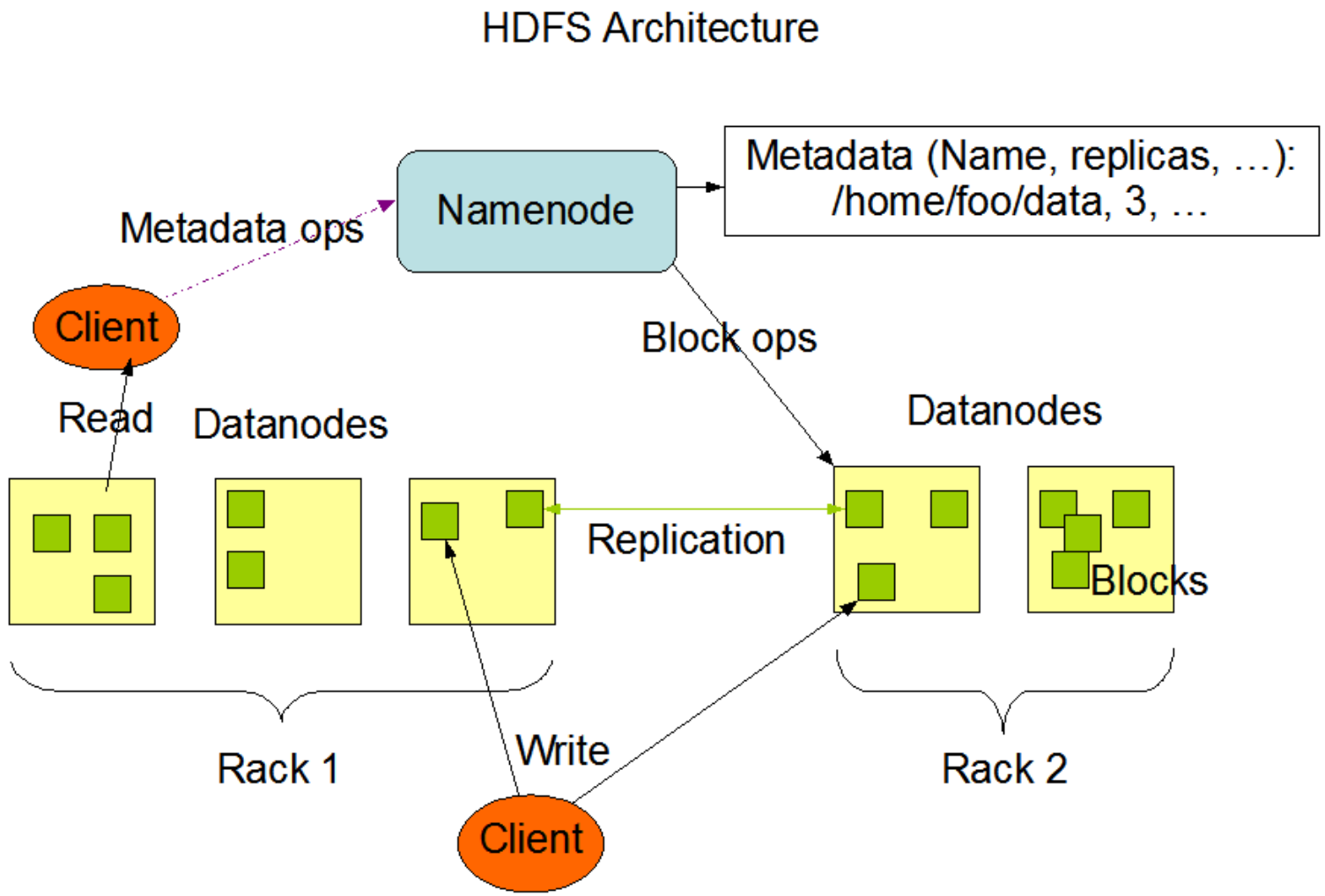
Hadoop

- Main elements of Hadoop.
  - HDFS
  - Map Reduce
- **HDFS** is used to **Storing** the data into Systems
- **MapReduce** is used to **Process** the data within the cluster.

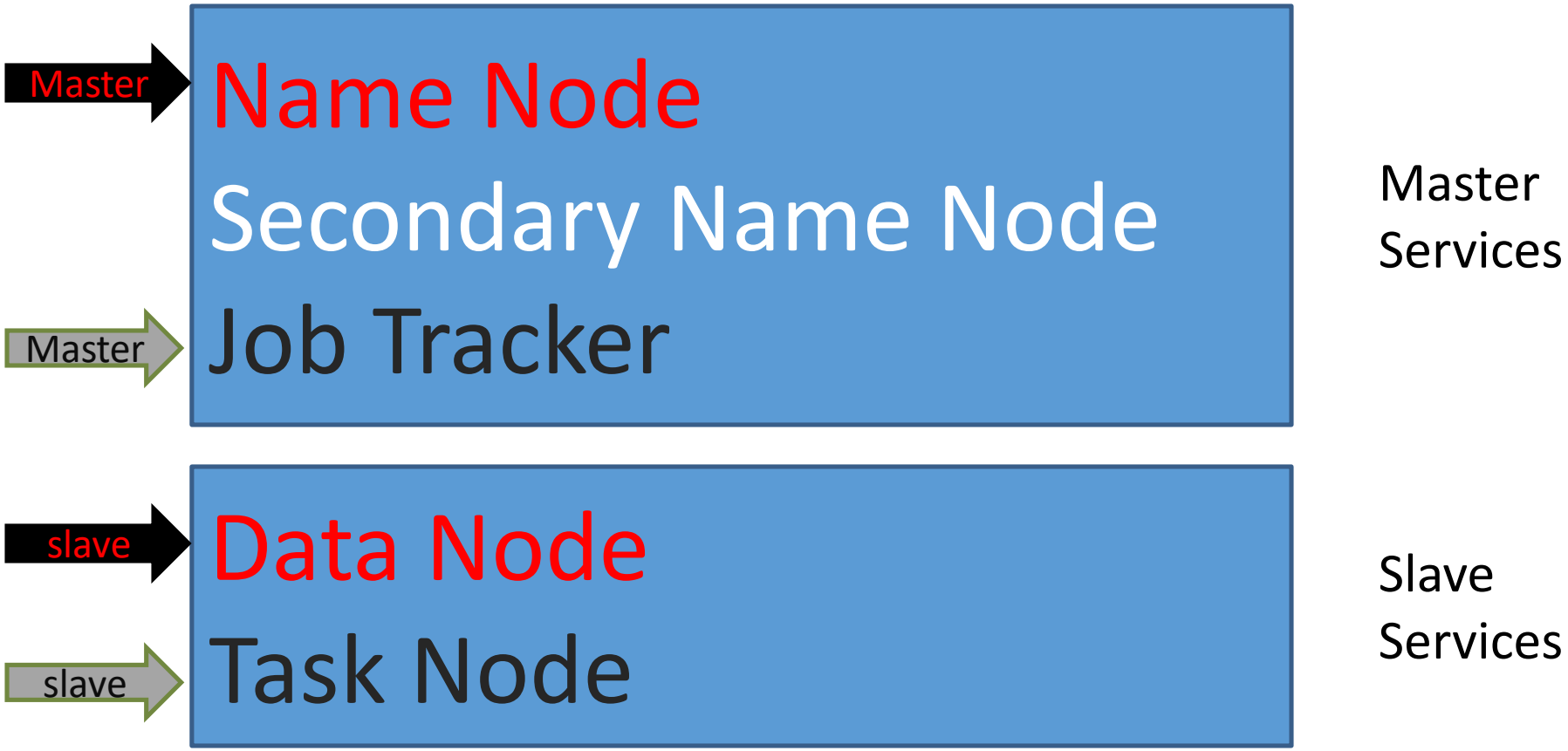
HDFS

- Hadoop Distributed File System.
- It is specially designed **file system** to storing huge dataset with **clusters** of **commodity hardware** with **streaming access pattern**.
- **streaming access pattern :** write once, read any number of time, but don’t try to change the content of that file once that data is kept into HDFS.

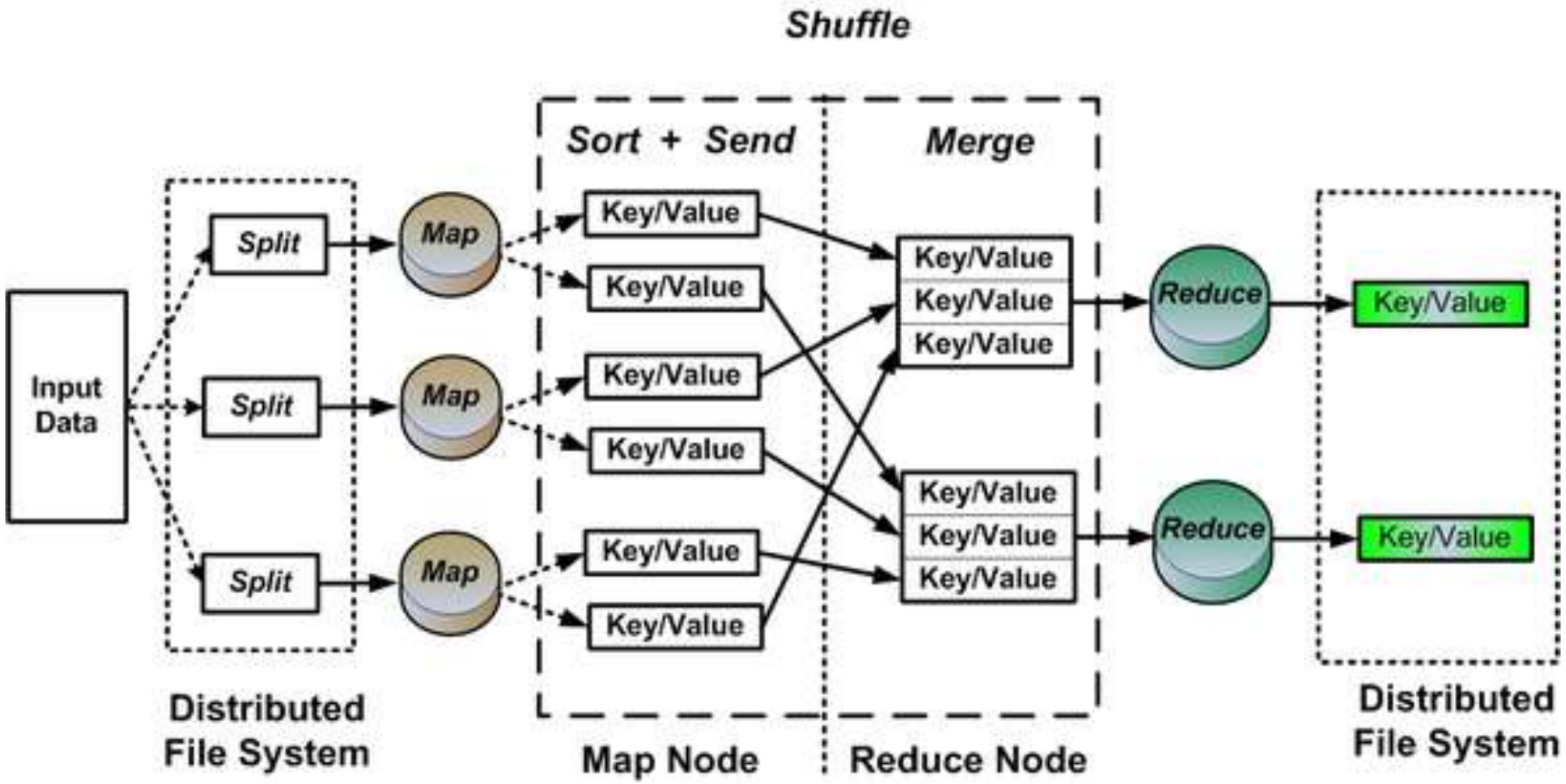
HDFS

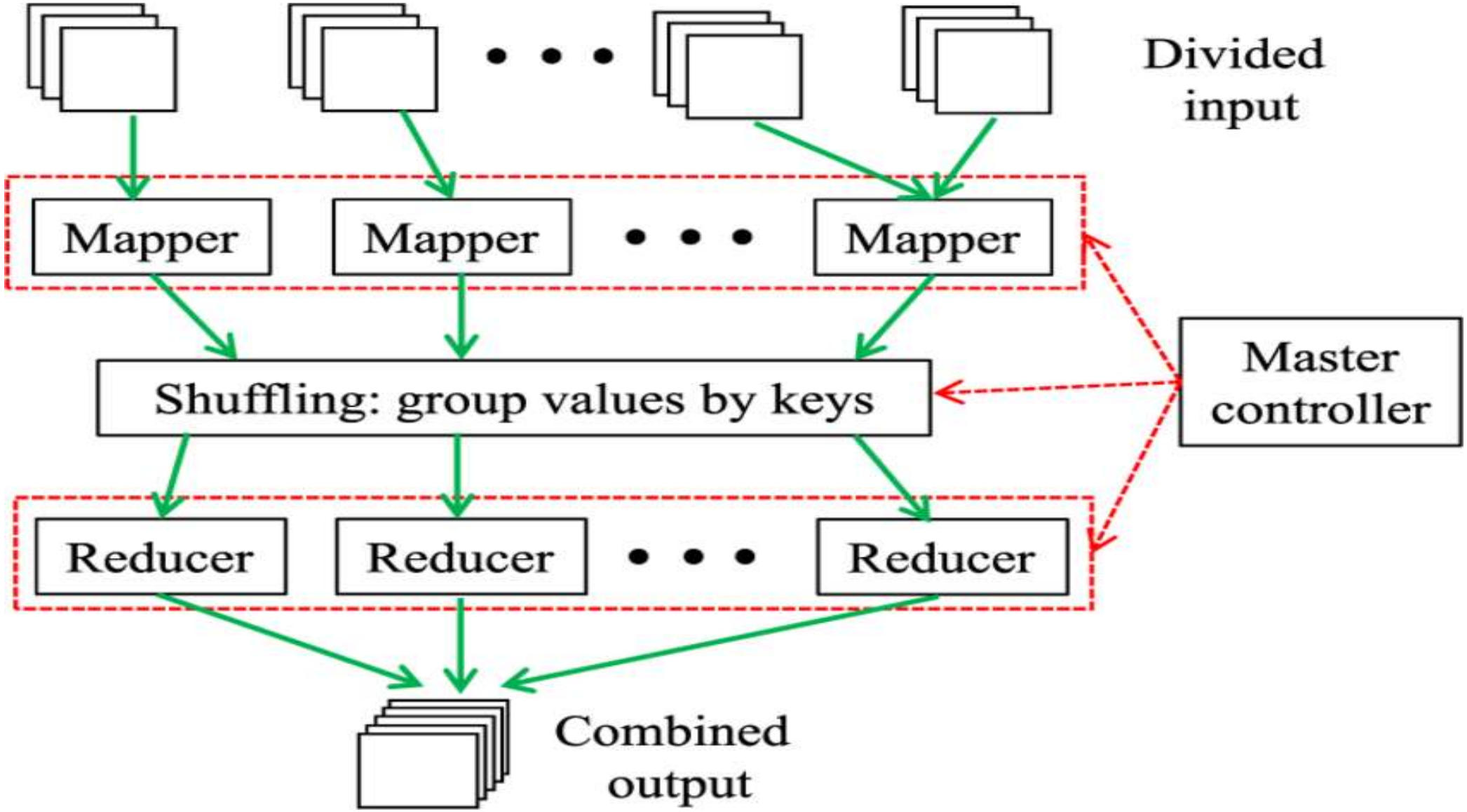

HDFS Architecture
Namenode
Metadata (Name, replicas, …):
/home/foo/data, 3, …
Metadata ops
Client
Block ops
Read
Datanodes
Datanodes
Replication
Blocks
Rack 1
Write
Rack 2
Client

HDFS Services.



Map Reduce Flow Chart




Divided input
Mapper
Mapper
Mapper
Shuffling: group values by keys
Master controller
Reducer
Reducer
Reducer
Combined output

Map-Reduce Flow

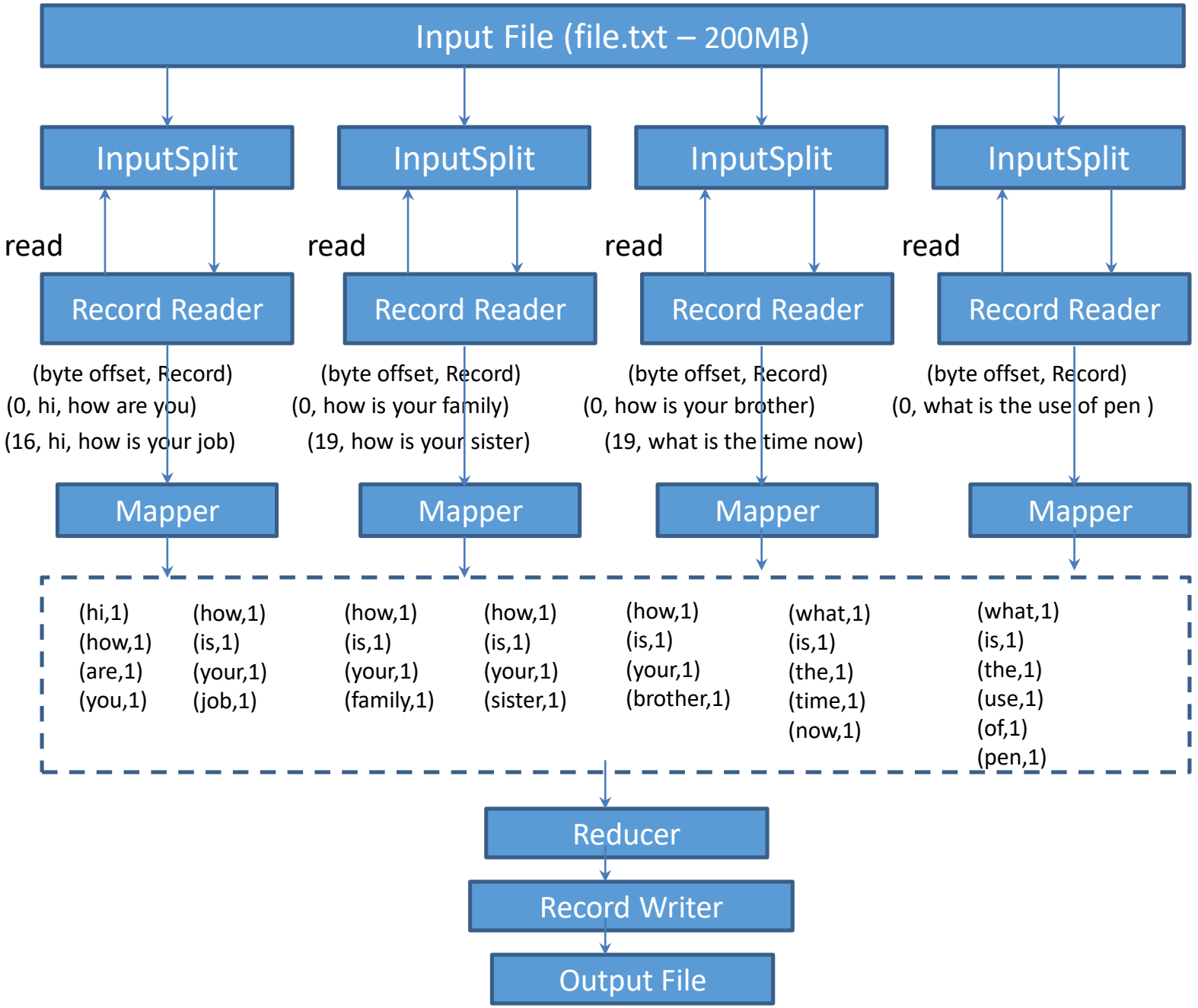
- Mapper and Reducer work only using (Key, Value) pair.
- InputSplits will be sended to Mappers with the help of Record Reader.
- Record Reader converts text into records with byte offset and text line.
- Every input splits will have a RecordReader interface to convert records into key value pair.

On what basis Record Reader convert file into key value pair.

- Based on Format of a File.
  - 4 types.
    - Text Input Format – (by default)
    - Key Value Text Input Format
    - Sequence File Input Format
    - Sequence File As Text Input Format
- Mapper can only work with key value pair & Mapper will give only key value pair.

file.txt (200MB)

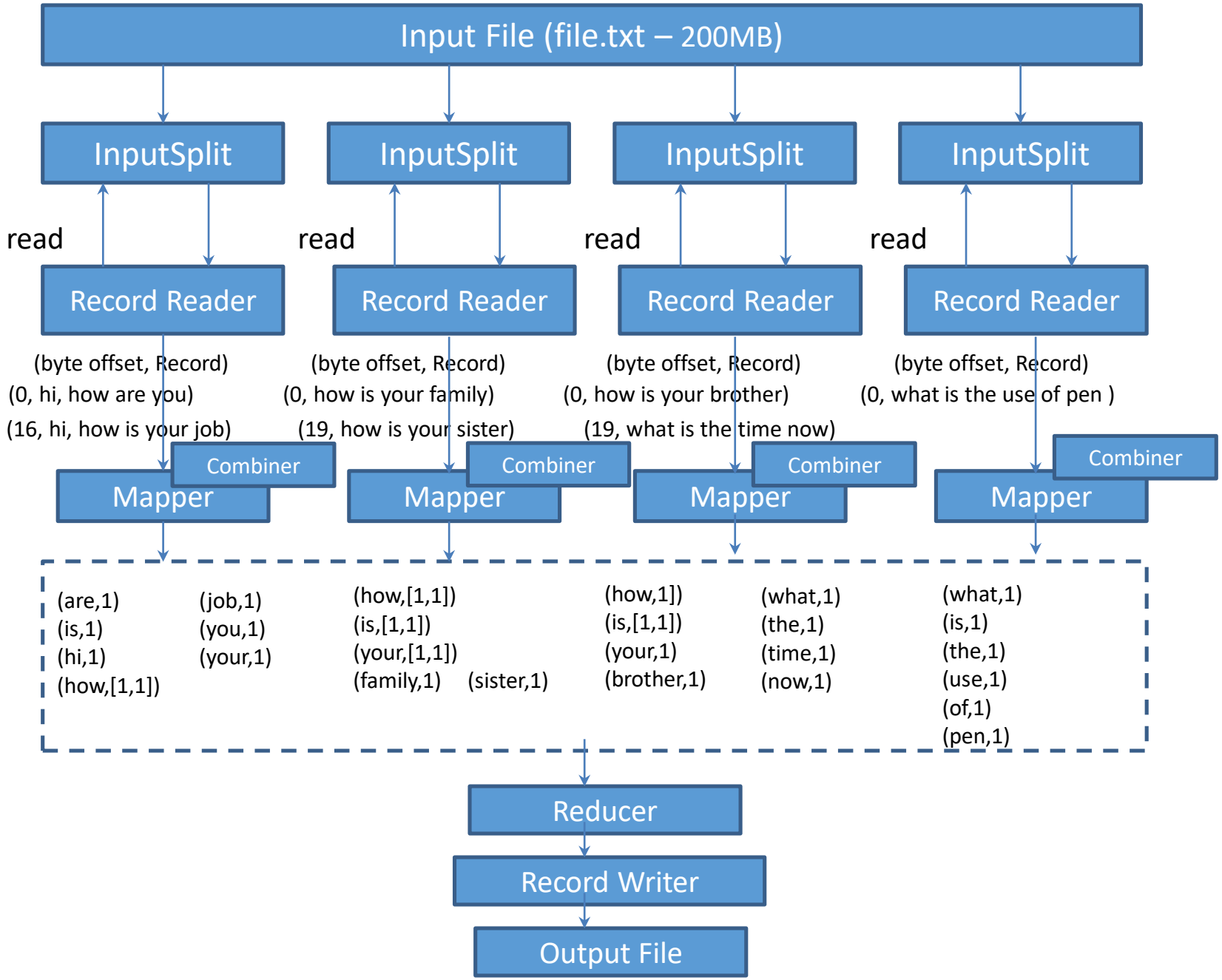
hi, how are you how is your job	64MB
how is your family how is your sister	64MB
how is your brother what is the time now	64MB
what is the use of pen	8MB



- Shuffling and Sorting
  - In Reducer, duplicate values is allowed but duplicate keys is not allowed.
  - Shuffling : Combines all values associated with single identical key.
  - Sorting : Sort the keys in ascending order.
- After Sorting and Shuffling Data will be sent to Reducer.
- Reducer will combine the output and forward the data to Record Writer.

file.txt (200MB)

hi, how are you how is your job	64MB
how is your family how is your sister	64MB
how is your brother what is the time now	64MB
what is the use of pen	8MB



Web 2.0

- Web 2.0 is the popular term given to the advanced Internet technology and applications that include blogs, wikis, really simple syndication (RSS), and social bookmarking.
- The two major contributors of Web 2.0 are the technological advances enabled by Ajax and other applications such as RSS and Eclipse that support the user interaction and their empowerment in dealing with the web.
- The term and web concepts and issues held by O'Reilly Media and MediaLive International in 2004.

- Differences between Web 2.0 and the traditional World Wide Web (referred to as Web 1.0) is that Web 2.0 facilitates greater collaboration and information sharing among Internet users, content providers, and enterprises.
- Web 2.0 is a migration from the read-only web to a read/write web.
- Web 2.0 provides the web users the ability to share and distribute information online with other users and sites.
- It refers to the transition from the static HTML web pages to a more dynamic web for serving web applications to users effectively.
- A Web 2.0 allows users to interact with each other in a social media dialogue.
- Examples of Web 2.0 include social networking sites, blogs, wikis, video-sharing sites, and any other hosted services or web applications that allow dynamic sharing of information among users.

- Web 2.0 technologies can be used as interactive tools to provide feedback on contents or information provided in the web page such as the best practices and recent updates.
- Cloud computing is closely related to the SOA of Web 2.0 and virtualization of hardware and software resources.
- Cloud computing makes it possible to build applications and services that can run/execute utilizing the resources (hardware and software) provided by the service providers, without restricting the application developers or consumers to the resources available on premise.

Characteristic features of Web 2.0

1. *Blogging*: Blogging allows a user to make a post to a web log or a blog.
2. *Usage of Ajax and other new technologies*: Ajax is a way of developing web applications that combines XHTML and CSS standards–based presentation.
3. *RSS-generated syndication*: RSS is a format for syndicating web content. It allows to *feed* the freshly published web content to the users through the RSS reader/aggregator.
4. *Social bookmarking*: Social bookmarking is a user-defined taxonomy system for storing tags to web contents. Instead of storing bookmarks in a folder on the user's computer, tagged pages are stored on the web increasing the accessibility from any computer connected to the Internet.
5. Mash-ups: A mash-up is a web page or an application that can integrate information from two or more sources.

Applications of Web 2.0

- Social Media
  - Way people communicate and share information.
- Marketing
  - offers excellent opportunities for marketing by engaging customers in various stages of the product development cycle.
- Education
  - education scenario by providing students and faculty with more opportunities to interact and collaborate with their peers.

Web 2.0 and Cloud Computing

- In Web 2.0, the metadata describing the web content is written in languages such as XML, which can be read and processed by the computers automatically.
- Various XML-based web protocols such as SOAP, WSDL, and UDDI help to integrate applications developed using different programming languages utilizing heterogeneous computing platforms and OSs.
- The applications can be hosted on the web and accessed by geographically separated clients over the Internet.
- **Web services** are such interoperable applications or services hosted on the web for remote use by multiple clients with heterogeneous platforms, and they can even be discovered dynamically on the fly with no prior knowledge of their existence.

- In Cloud computing, the application development infrastructures such as processors, storage, memory, OS, and application development tools and software can be accessed by the clients as services over the Internet in a pay-per-use model.
- In this model of service delivery, a huge pool of physical resources hosted on the web by the service providers will be shared by multiple clients as and when required.
- **Cloud computing is based on the SOA of Web 2.0 and virtualization of hardware and software resources stored hosted by the service providers.**
- Web 2.0 functionalities encouraged many SaaS applications to offer features that let its users work together, and distribute and share data and information.

Thank you.