1. What is HDFS?

It is a highly distributed, fault tolerant file system built on top of hadoop that can store data on thousands of off-the-shelf servers, with no special requirements

for hardware configuration.

It is rack aware, which means each replica of a block is placed on a different rack.

1. What are the kind of servers on HDFS?

Name nodes and data nodes.

1. What do name nodes store?

The name node maps every file to the list of blocks that the file consists of.

The name node also holds information about each block’s location—which data nodes the block is stored on and where on the data node it is.

1. Classification of hdfs data formats

Splittable and unsplittable

1. Why are binary formats better in general?

Binary formats have the ability to detect faulty/incomplete records in case there is a system failure while writing data and records are not loaded properly. Binary records help detect such records and ignore.