HDFS:

\*Hadoop distributed File System.

\*It is a Big Data analytics tool which is used for storage .

\*The file store in HDFS provides scalable, fault-tolerant storage at low cost.

\* It stores files across the collection of servers in a cluster.

\* It is used for storage purpose of data before processing and result after processing.

\* It enables data transfer among the nodes.

\* The HDFS software detects and compensates for hardware issues, including disk problems and server failure.

\* Files are decomposed into blocks and each block is written to more than one of the servers.

\* HDFS applications is based on a write-once and read-many access model for files.

**\*** A file once created, written, and closed need not be changed.

Hadoop cluster:

\*A hadoop cluster is a special type of computational cluster designed specifically for storing and analyzing huge amounts of unstructured data in a distributed system environment.

\* Hadoop clusters are comprised of three different node types: master nodes, worker nodes, and client nodes.

\* After hadoop clusters, the data reading speed and the time taken to read of data is reduced.

\*The files are replicated and stored in different locations.

\*This replication provides both fault-tolerance and performance.

HDFS Block:

\*In hdfs files are stored in blocks.

\*A block is the smallest unit of data that can be stored or retrieved from the disk.

\*The blocks are stored in continuous format unlike the local file system where it is stored in random manner.

\*The default size of HDFS block is 128MB.

\*Since they are divided into a no. of blocks HDFS ensures fault tolerance and high availability.

\*The file can be stored only if the size of the file is less than that of the block.

\*It gets blocks of local file system contiguously to minimise head seek time.