

#### \* Namenode Failure Cause :

Zookeeper helps in switching the active & the standby namenodes.

The namenode maintains the reference to every file & block in the memory.

If we have too many files or folders in HDFS, the metadata could be huge. Therefore the memory of namenode may become insufficient.

#### \* Operations on a list

1. Append
2. Sort
3. update
4. Delete
5. POP
6. Max, Min
7. Length
8. Comma index

#### \* Spark Session

SparkSession.builder.master

SparkSession.newSession

SparkSession.getOrCreate

• Config

• master

• appName

#### \* Spark. Create Data Frame()

Spark.Sql

• Table

• read

• stop()

#### Pandas Functions :

read\_csv

head()

describe()

memory\_usage()

value\_count

drop\_duplicates()

value - count  
group-by()

numpy

numpy.linspace (spread numbers over specific intervals)

numpy.repeat()

numpy.random

numpy.polyfit

numpy.nan (contains nan)

Spark

count

avg

mean

max

✓ String Fun

✓ Date Fun

Spark session objects

Functions in RDD

String ! SQL

\* LEN \* LOWER

\* UPPER

\* ASCII

\* CHAR

\* (ONCAT)

Date Functions

DATEDIFF() diff b/w 2 dates

NOW()

EXTRACT() Returns single part of day/time

CURDATE()

CURTIME()

Functions in RDD

\* Count()

\* Union()

\* groupByKey()

\* First()