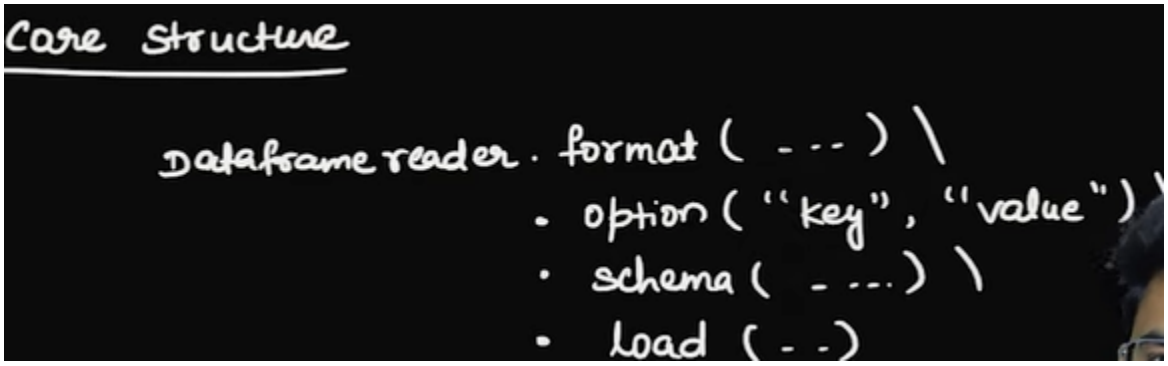
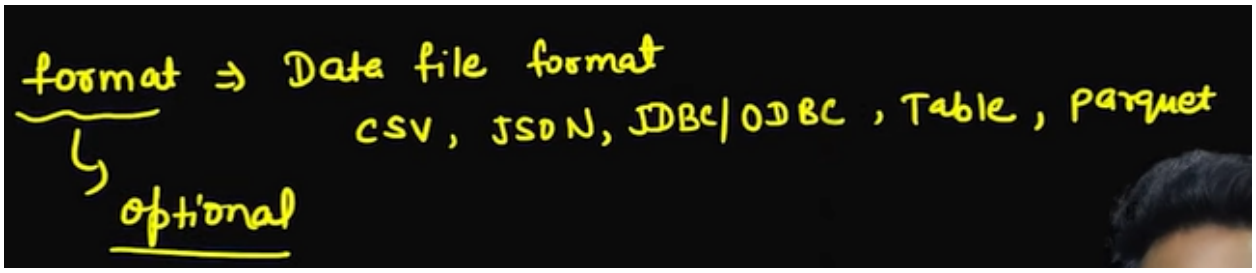
Read Data in SPARK

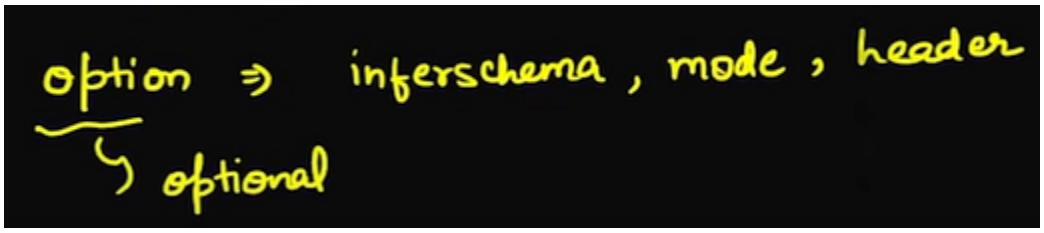
1. The core structure of code would be



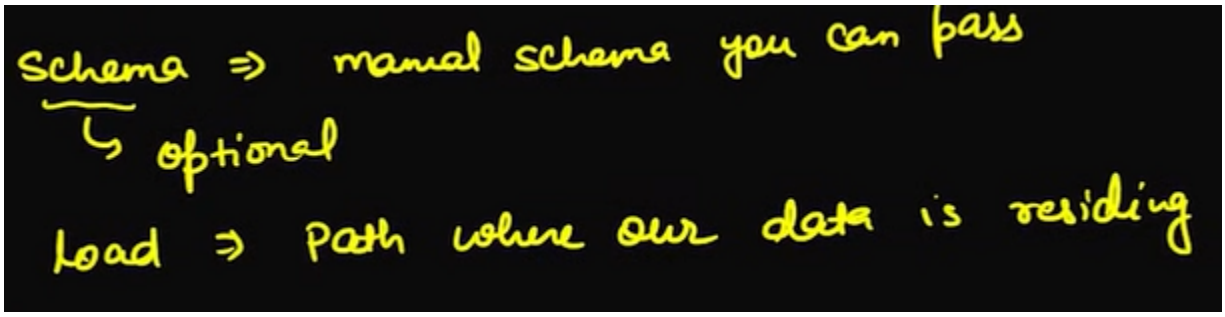
2. Here format is data file format(CSV, JSON, ODBC/JSON etc, PARQUET)..in spark the default file format is parquet



3. Here in option we have things like

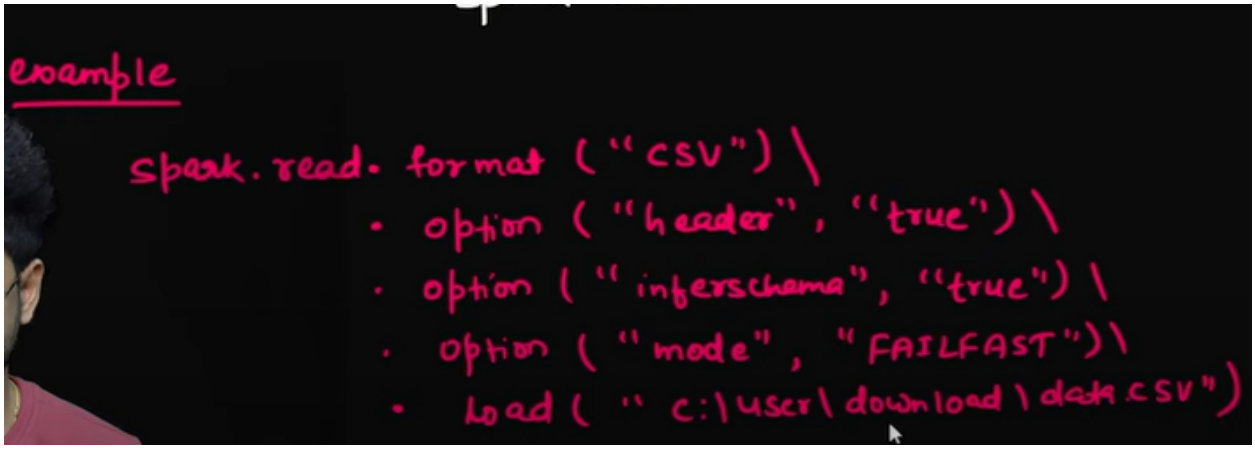


4. In Schema ..we can pass manual schema if we have one
5. And in load we have to give location of our data



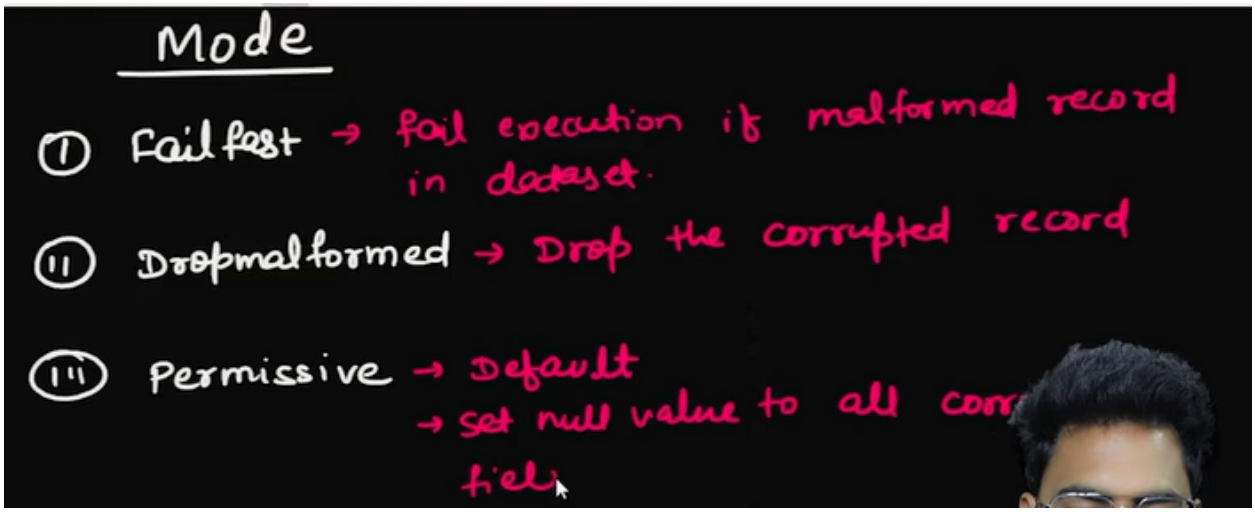
6. How to access dataframe reader API? ..for that we have to initialize the spark session and use spark.read

7. Example code would look like this



Here we didn't used .schema because we have used inferSchema

8. Lets learn about mode



9. We have 3 modes and Permissive mode is default mode..it sets null value to all corrupted fields
10. Now we'll do hands-on

11.
```
flight_df_header_schem=spark.read.format("csv")\
            .option("header","true")\
            .option("inferschema","true")\
            .option("mode","FAILFAST")\
            .load("/FileStore/tables/flight_data.csv")

flight_df_header.show(5)
```
12. Notes on inferschema : https://g.co/gemini/share/43be33546bfa
13.