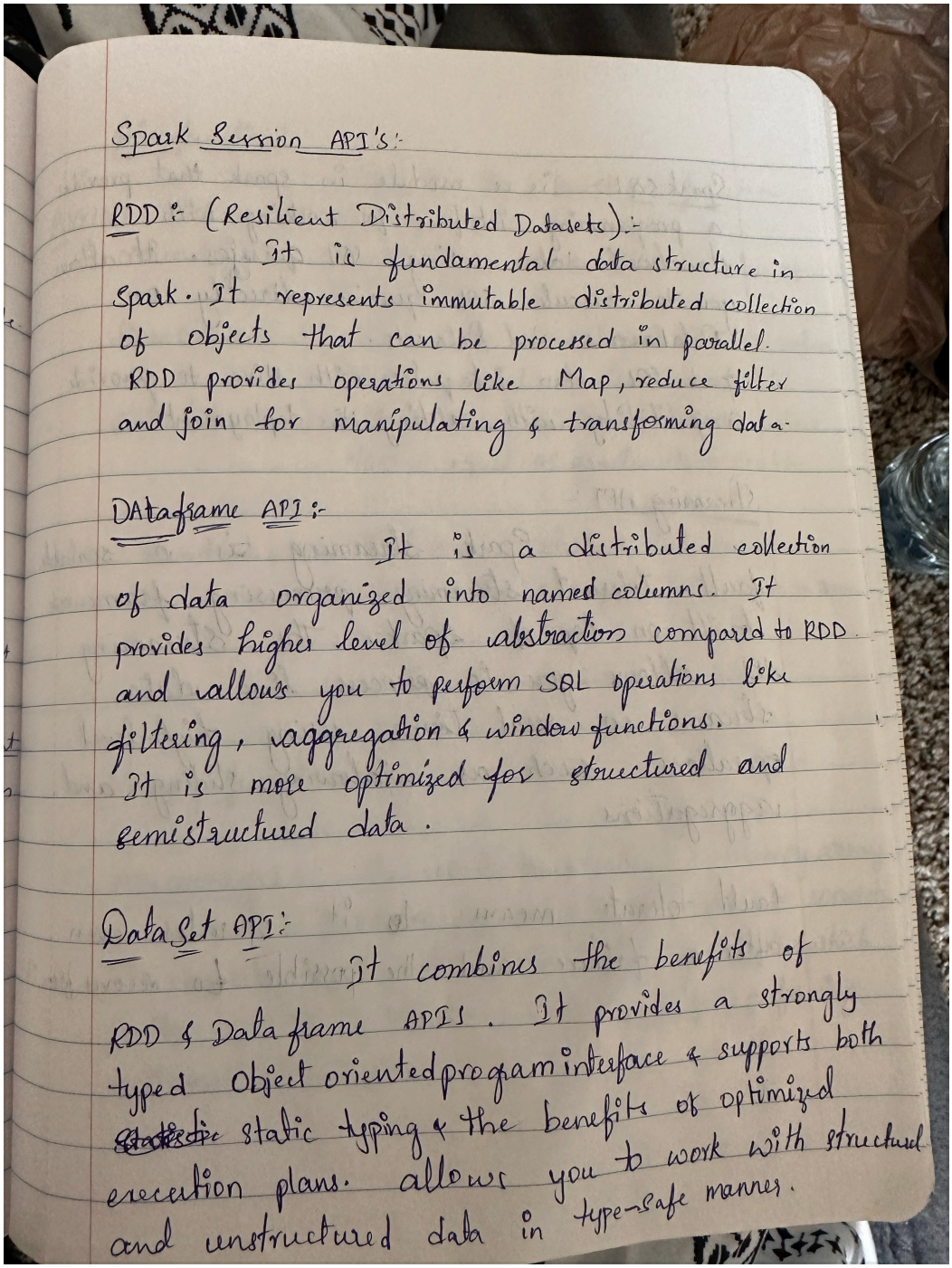
SPARK SESSION



Spark Session API's:-

RDD :- (Resilient Distributed Datasets):-
It is fundamental data structure in Spark. It represents immutable distributed collection of objects that can be processed in parallel. RDD provides operations like Map, reduce filter and join for manipulating & transforming data.

DAtaframe API :-
It is a distributed collection of data organized into named columns. It provides higher level of abstraction compared to RDD and allows you to perform SQL operations like filtering, aggregation & window functions. It is more optimized for structured and semistructured data.

Data Set API :-
It combines the benefits of RDD & Data frame APIs. It provides a strongly typed object oriented program interface & supports both ~~Statistic~~ static typing & the benefits of optimized execution plans. allows you to work with structured and unstructured data in type-safe manner.