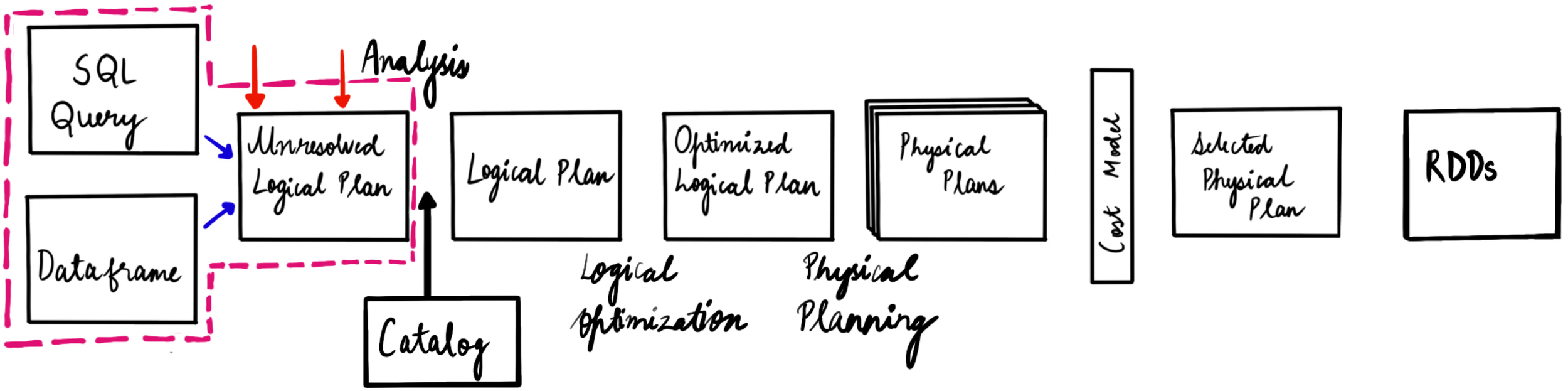
Logical & Physical Planning

in

Under 5 Seconds

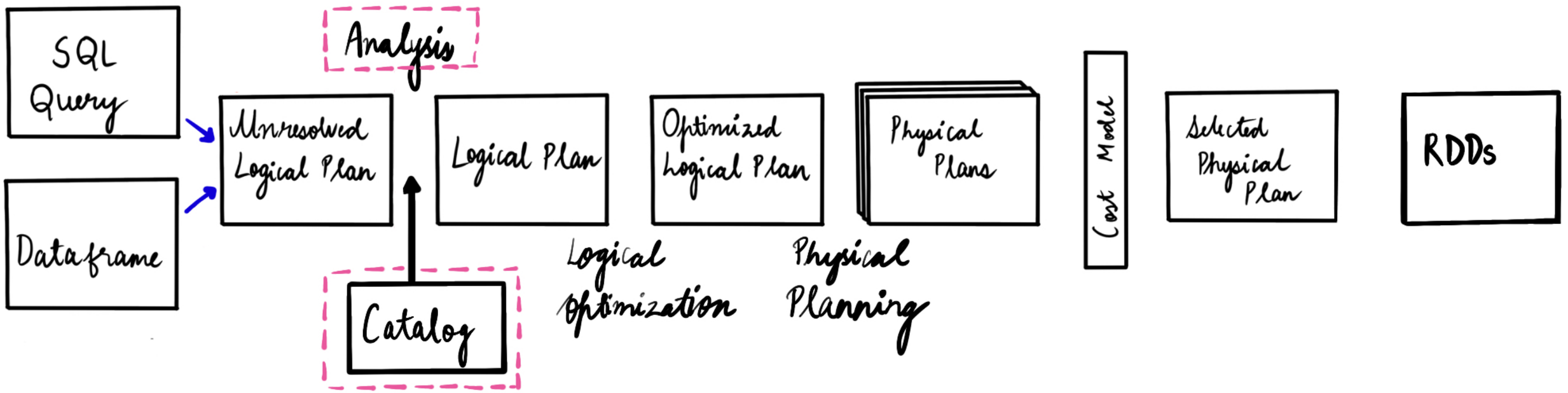


Spark Context Generates Unresolved Logical Plan

Unresolved Logical Plan

Plan Generated by SparkContext **without** Proper Checking of Schema.

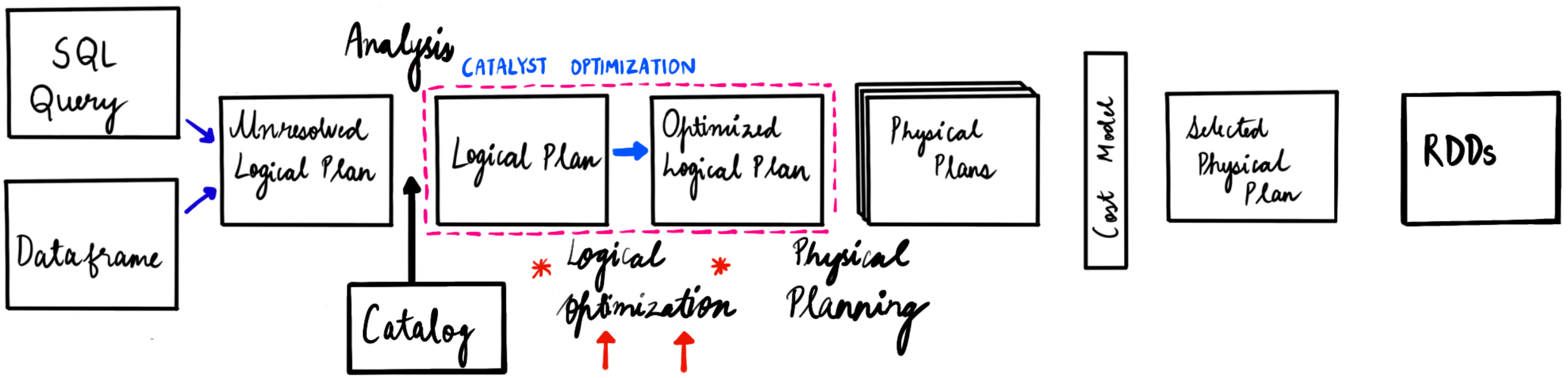
@chiragsehra



USE OF CATALOG & ANALYSER

CATALOG: ACTS AS METASTORE. IT CONTAINS INFO ABOUT DataFrame & DataSet API

ANALYSER: HELPS IN VERYFYING SEMANTICS & TABLE NAMES FROM CATALOG.



LOGICAL PLAN TO OPTIMIZED LOGICAL PLAN

CATALYST OPTIMIZER HELPS IN CREATING A OPTIMIZED LOGICAL PLAN.

IT CONTAINS RULE SETS FOR ANALYSIS, QUERY OPTIMIZATION, PHYSICAL PLANNING etc.

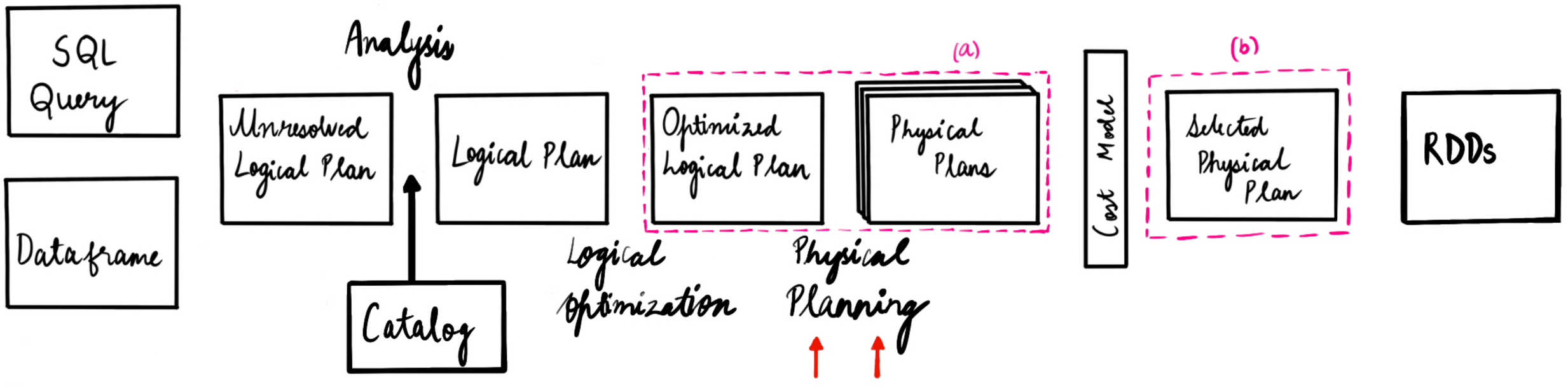
THIS INCLUDES

(a) CONSTANT FOLDING

(b) PREDICTIVE PUSHDOWN

(c) PROJECTION PRUNING.

@chiragsehra



PHYSICAL PLANNING

(a)

- PHYSICAL OPERATORS ARE USED FOR COST BASED MODELLING OF LOGICAL PLAN
- IT MEASURES COST FOR RUNNING QUERIES WITH DIFFERENT OPERATIONS. EX:- BROADCAST JOIN vs SORT-MERGE JOIN
- FINDS BEST OPTIMAL PHYSICAL PLAN

(b)

- SELECTED PHYSICAL PLAN WITH THE LOWEST COST PERFORMS CODEGEN I.E. CONVERT SPARK SQL TO LOW LEVEL CODE THAT IS DISTRIBUTED AMONGST THE EXECUTORS.