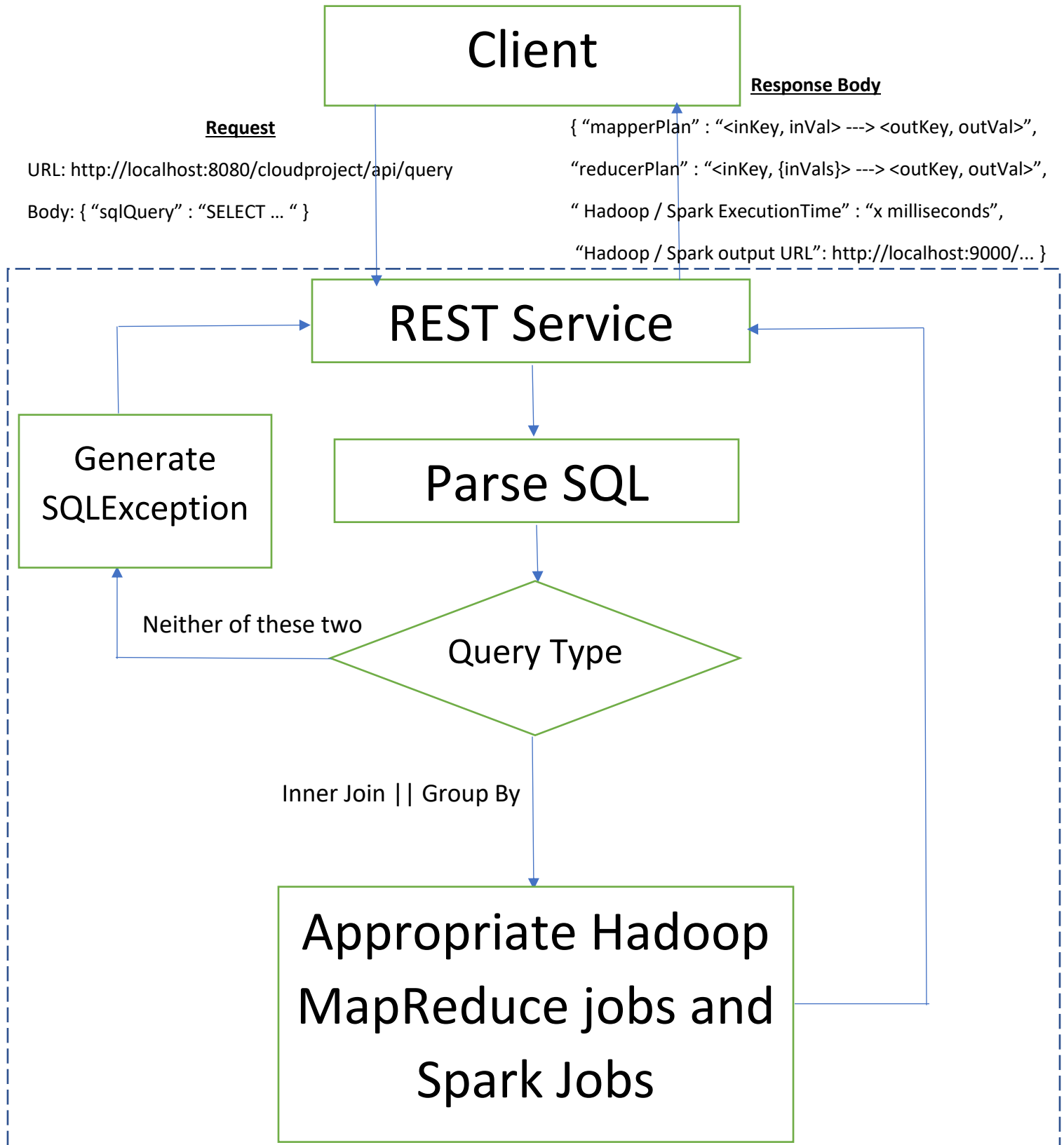
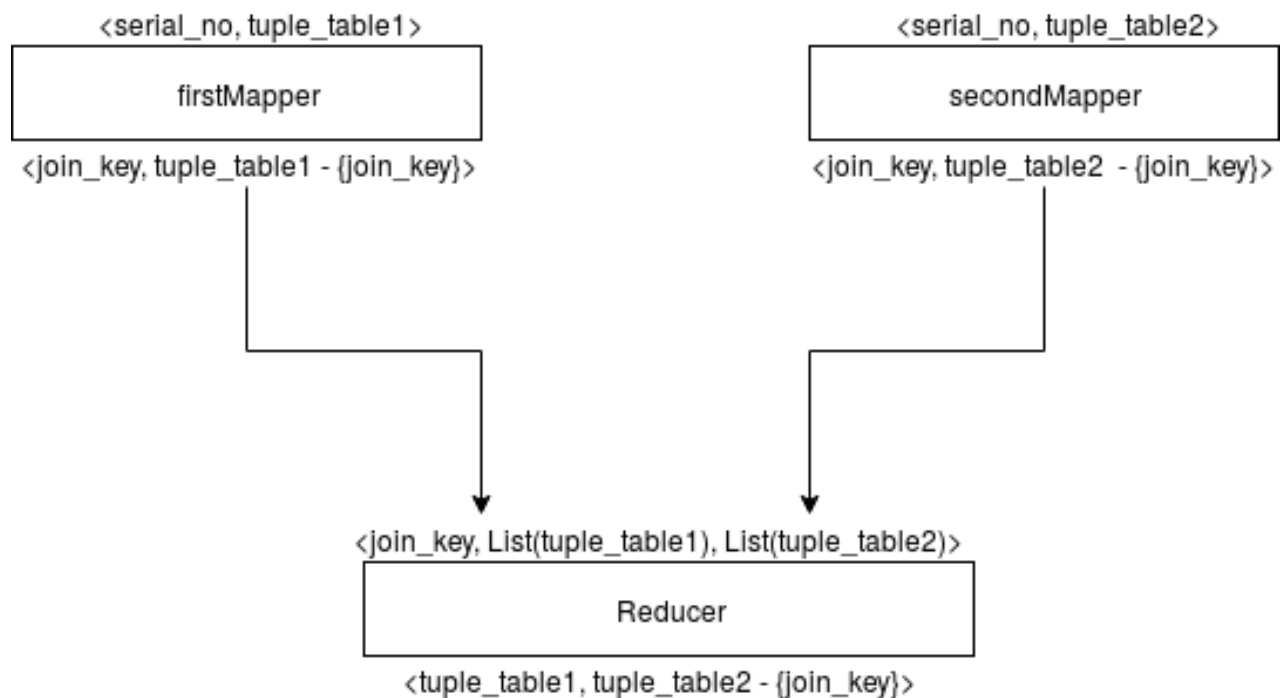


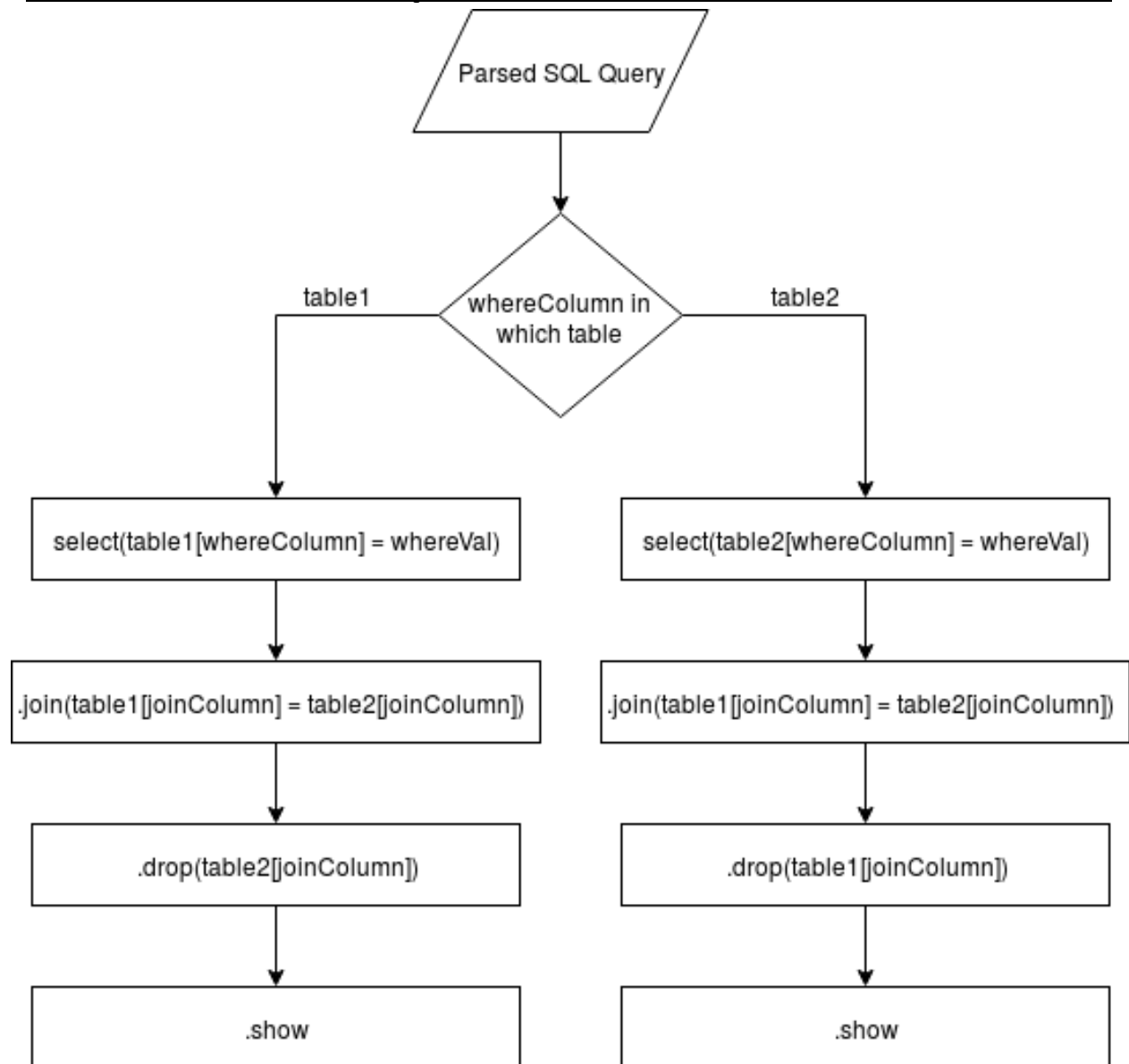
Overall Plan



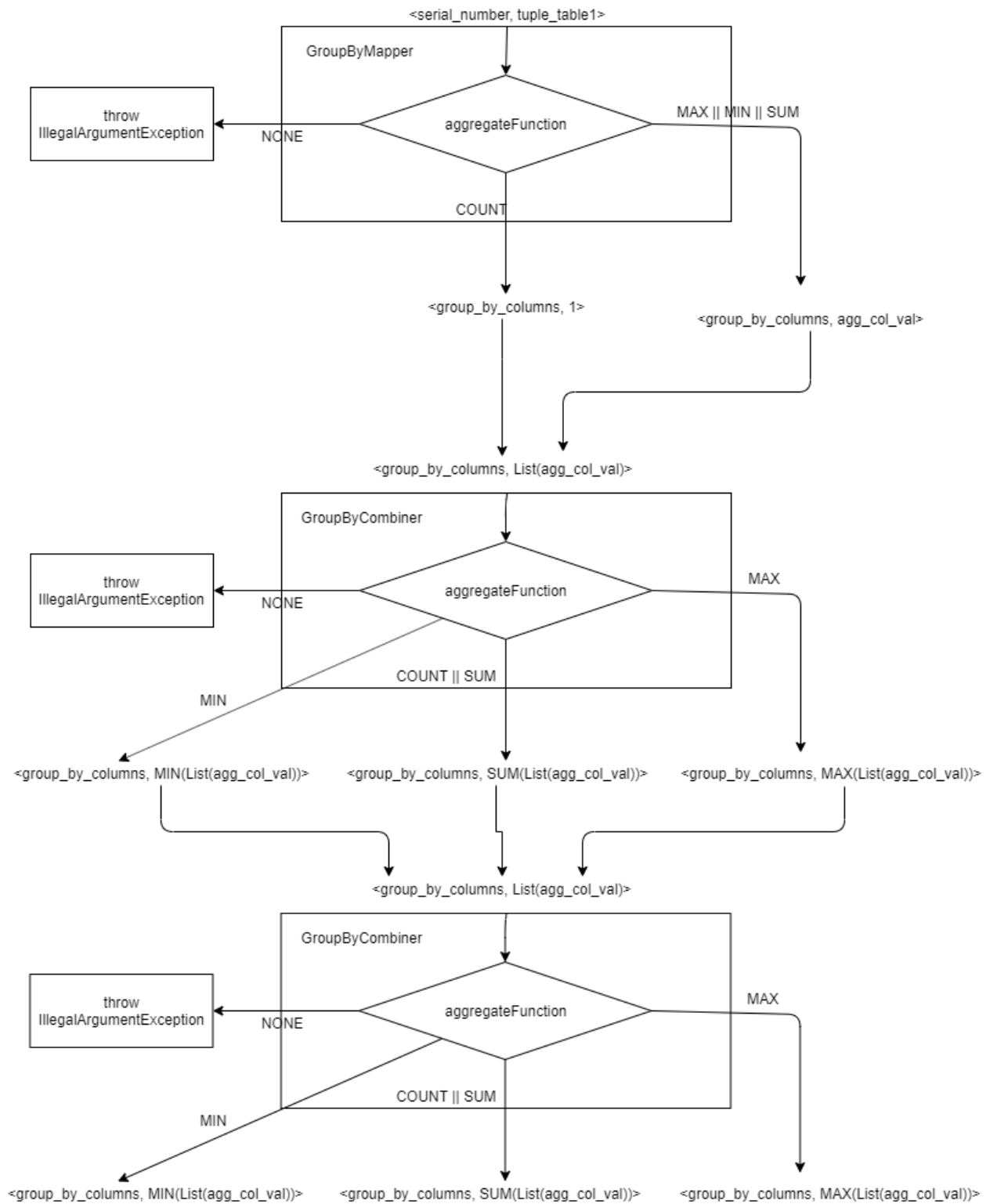
Inner Join Hadoop MapReduce



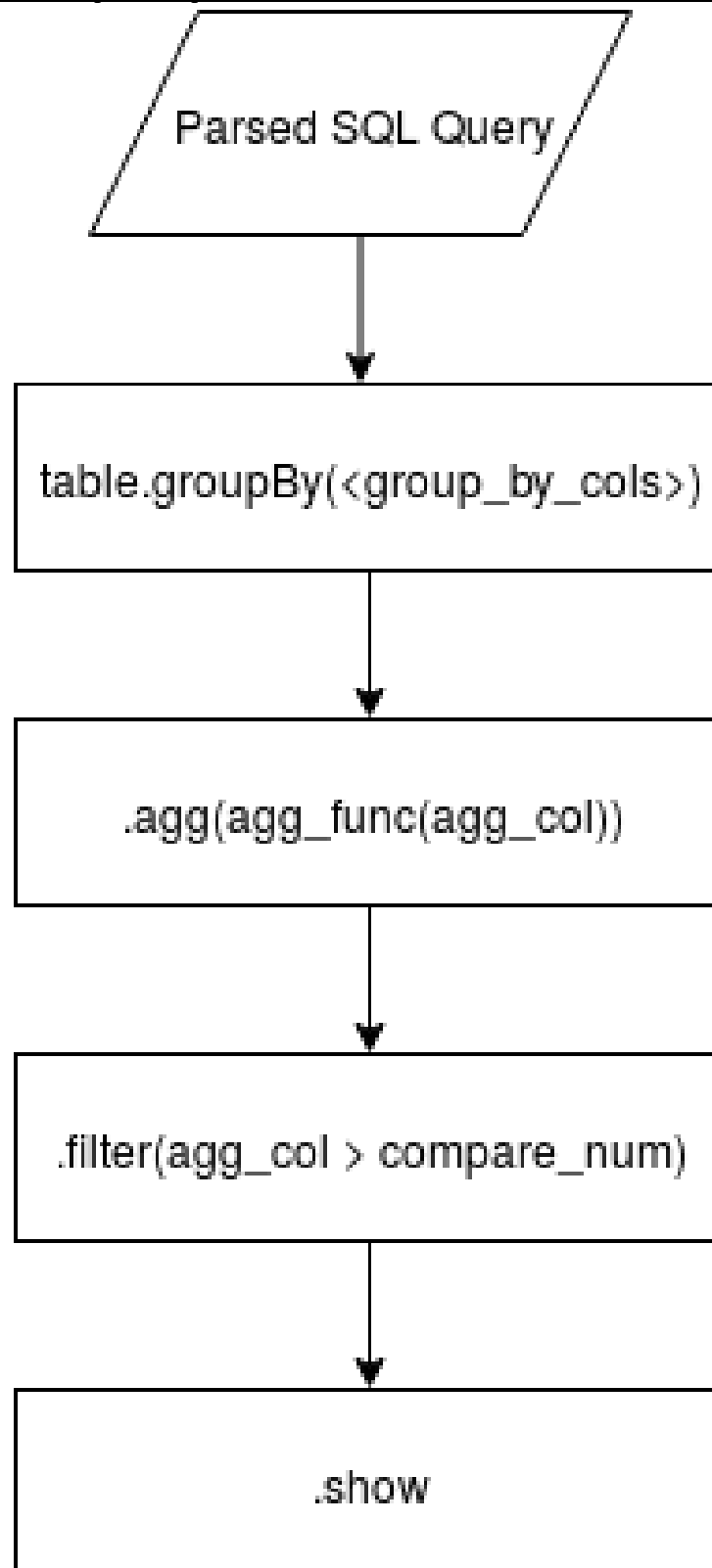
Inner Join Spark Transformations



Group By Hadoop MapReduce



Group By Spark Transformations



Design Rationale

- Reduce-Side join has been implemented. This is efficient when both the join tables are large. Also, Reduce-Side join requires less strict format than Map-side join, hence easier to implement.
- Used contract classes for index to column mapping and column to index mapping.
- Used genson for POJO to JSON mapping in the REST API.