

HBase Query Interface

Jiang XianLi
2012/8/29

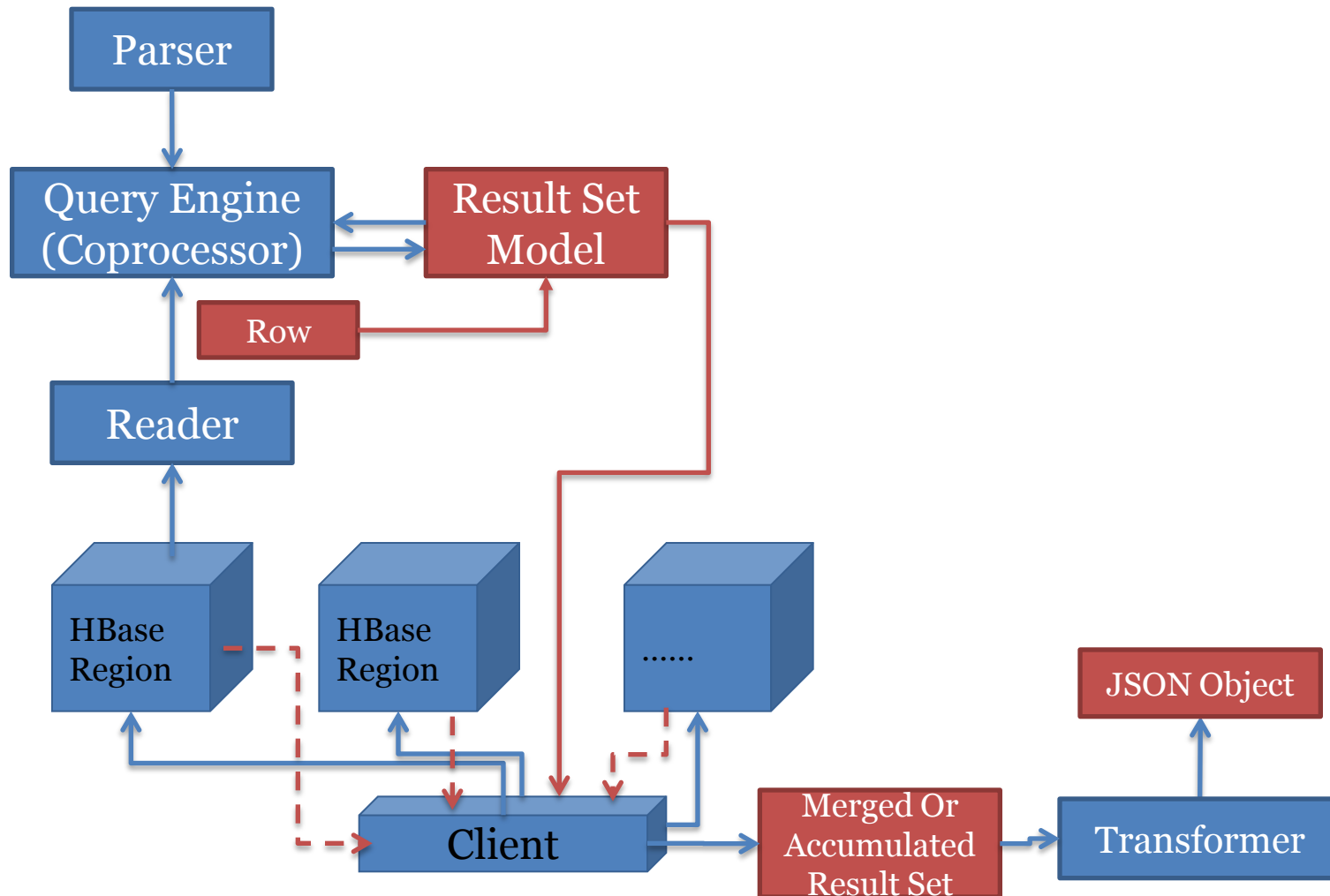
Agenda

- Background
- Design
- Implementation
- Assess

Background

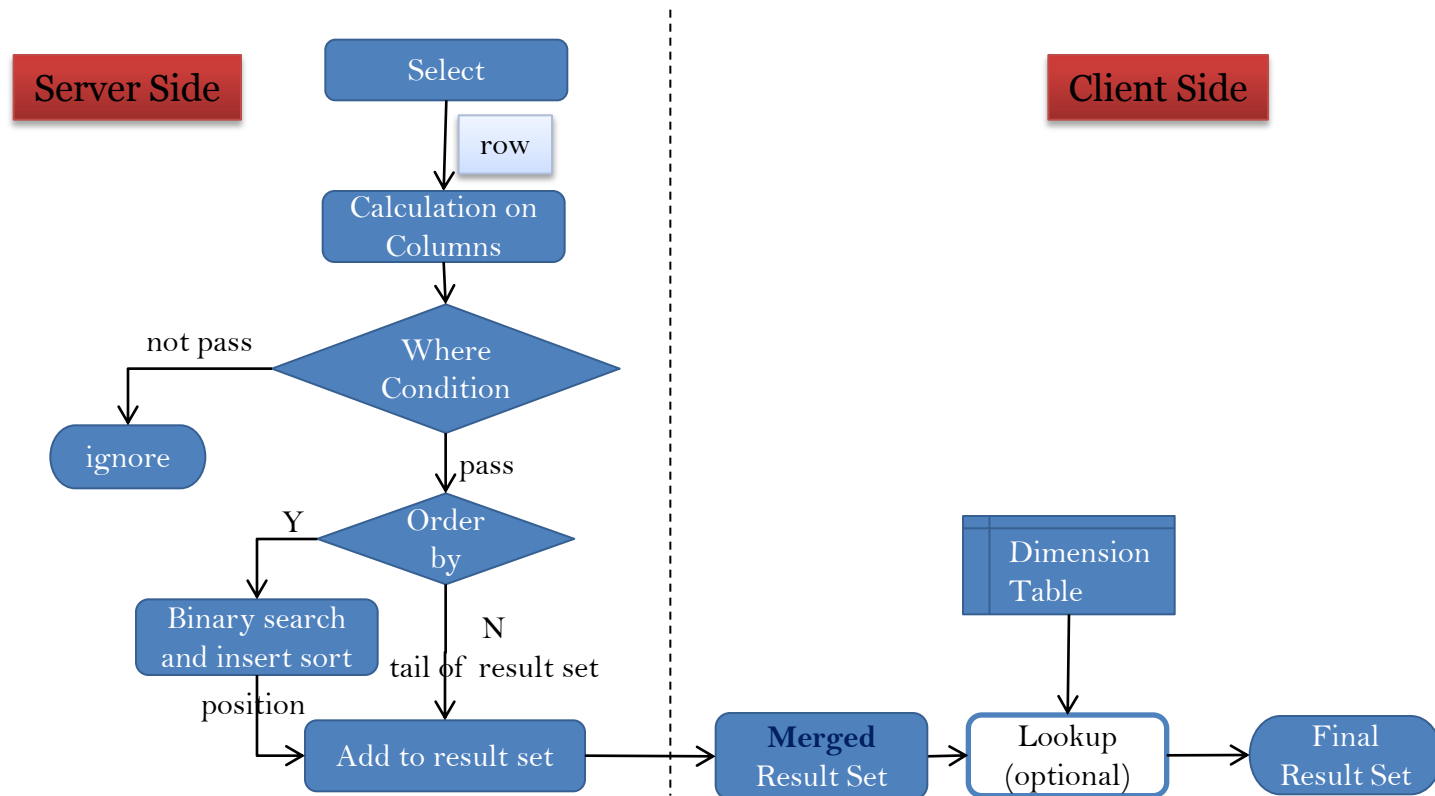
- What HBase has is simple
 - Get, Scan
- But we need more!
 - Aggregation
 - sum, avg, count, max, min
 - group by
 - having
 - Sort
 - Where
 - Join

Design



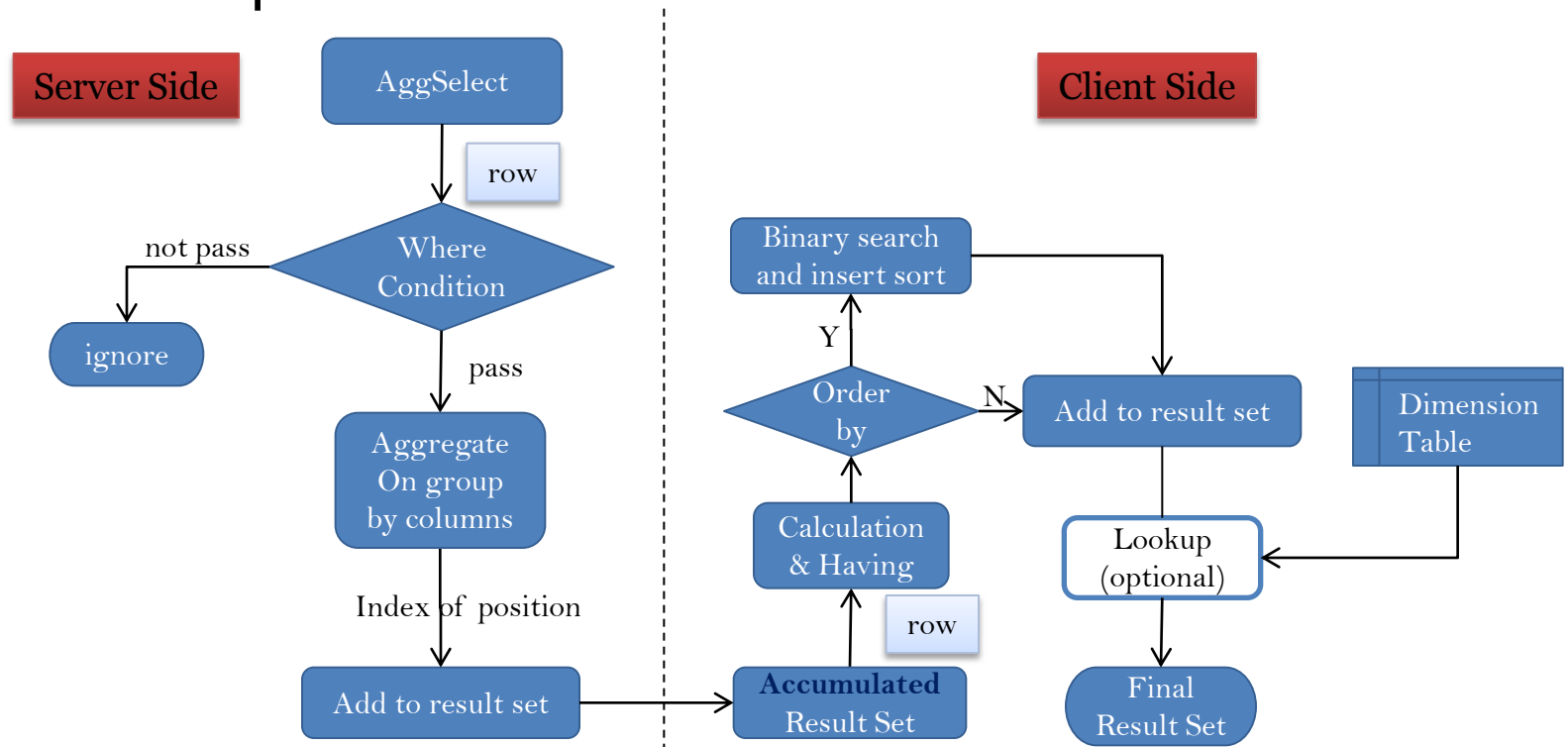
Implementation

- Simple query process
 - Select, Where, Order by, Lookup



Implementation

- Aggregate query process
 - AggSelect, Where, Group by, Having, Order by, Lookup



Assess

- Effect
 - Rich analysis
 - Multi-dimension analysis
 - Time-frame query
 - Ease ETL workload
 - 7 days, 15 days data can be calculated online
 - Complex measures can be calculated online
 - Different ratios
 - Server utilization
 - CPU, memory

Assess

- Benchmark

rows	orderby=null, top= -1	orderBy=null top=100	OrderBy=d:salary top=1000	OrderBy=d:salary, top=2000, Where d:name=xianli
10000	19533 ms	323 ms	2751 ms	3459 ms
50000	97783 ms	1191 ms	15089 ms	17762 ms
100000	182043 ms	2087 ms	21702 ms	34787 ms

rows	Groupby=null Having(null)	Groupby=d:name Having(null)	Groupby=d:name Having(d:salary>30000)
10000	276 ms	284 ms	279 ms
50000	721 ms	422 ms	548 ms
100000	648 ms	635 ms	591 ms

Q&A

- Thanks 😊