Modeling Analytics for Computational Storage Veronica Lagrange, Harry Li, Anahita Shayesteh

A Contraction of the second se

Memory Solutions Lab Samsung Semiconductor, Inc.

07 April 2020 Version 1.2



Modeling Analytics for Computational Storage

- Motivation
- Near storage opportunities
- Deconstruction of "big data" queries
- Push down to Near Storage
- Workload:TPC-DS
- Modeling Methodologies and Results



Motivation





SAMSUNG

Motivation SSD

SAMSUNG

SSD

Server



Motivation: Near storage OLAP



SAMSUNG



Motivation: Near storage OLAP

Read IN just needle.

Where IS that needle ???

SmartSSD

Where IS that needle ???

SmartSSD

Server







Near storage opportunities

- Compression/Decompression;
- Encoding/Decoding;
- Filter;
- Projection;
- Some aggregates (SUM, COUNT);
 SORT;
- Some JOINs.

Deconstruction of "big data" queries

TPC-DS Q44:

"List the best and worst performing products measured by net profit." For a specific store.



```
select asceding.rnk, il.i_product_name best_performing,
        i2.i_product_name worst_performing
from(select *
   from (select item_sk,rank() over (order by rank_col asc) rnk
       from (select ss_item_sk item_sk,avg(ss_net_profit) rank_col
           from store_sales ss l
           where ss store sk = 2
           group by ss_item_sk
           having avg(ss_net_profit) > 0.9*(select avg(ss_net_profit) rank_col
                                   from store sales
                                   where ss_store_sk = 2 and ss_hdemo_sk is null
                                   group by ss_store_sk))VI)VII
   where rnk < 11 asceding,
  (select *
   from (select item_sk,rank() over (order by rank_col desc) rnk
       from (select ss_item_sk item_sk,avg(ss_net_profit) rank_col
           from store_sales ss l
           where ss_store_sk = 2
           group by ss_item_sk
           having avg(ss_net_profit) > 0.9*(select avg(ss_net_profit) rank_col
                                   from store sales
                                   where ss store sk = 2 and ss hdemo sk is null
                                   group by ss_store_sk))V2)V21
   where rnk < 11 descending,
item i1, item i2
where asceding.rnk = descending.rnk
 and il.i_item_sk=asceding.item_sk
 and i2.i_item_sk=descending.item_sk
order by asceding.rnk limit 100;
```





Push down to Near Storage

Operations pushed down:

- SCAN: I/O plus data transformation
- FILTER: row selection
- PROJECTION: column selection





Workload: TPC-DS

Two clusters:

- SPARK-SQL
- Presto

TPC-DS sf10,000 (10TB dataset)





99 TPC-DS queries have different characteristics and performance behavior.



Parquet File Format

Two 8-node Hadoop clusters:

- SPARK-SQL
- Presto

One file format – PARQUET:

- Columnar
- Designed for OLAP applications
- READ optimized
- Self-contained METADATA
- Existing Parquet Readers can FILTER/PROJECT certain datatypes using statistics in METADATA







SAMSUNG

SPARK-SQL modeling:

Timestamp	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Note
Stage-0																			Read dimension table: Scan, Filter, Project, Aggregate
Stage-1																			Read fact table: Scan, Filter, Project, Aggregate
Stage-2																			Read fact table: Scan, Filter, Project, Aggregate
Stage-3																			Sort, Aggregate
Stage-4																-			Sort, Aggregate
Stage-5																			Join



SPARK-SQL modeling:

Timestamp	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Note
Stage-0																			Read dimension table: Scan, Filter, Project, Aggregate
Stage-1																			Read fact table: Scan, Filter, Project, Aggregate
Stage-2																			Read fact table: Scan, Filter, Project, Aggregate
Stage-3																			Sort, Aggregate
Stage-4																			Sort, Aggregate
Stage-5																			Join





Presto modeling:

Origin	al Table:	Projection: selec	ct colur	nns A,B,C					
	R			n	E	6			
1	San Jose	23 Back Ave	Sn	nith	234567	2	>		
2	Seattle	45 Ocean Dr ,Ap 102	D	oe	345678	3	K	Filtering: select where G >	
 3	Reno	345 Yosemite Ln	1	IA	567890	0	,		
4	York	Magenta St	R	ay	2345	2			
5	San Diego	12 Main Ln	A	nn	65432	4	. Y		
6	Dallas	437 First Ave	Asi	mov	45214134				
7	Jacksonville	2 Broadway	K	ing	43211	-1	L		
8	Los Angeles	Santa Monica Blvd	L	ee	98756	0)		
9	San Diego	Palm Ln) CL	irry	9010280	1			
10	Sacramento	Capitol	Bro	ooks	834261000	2			
Mode	l Table fo	r Query "SELECT	T A ,B,C	WHERE	G > 2;"				
			А	В	С			ì	
			2	Seattle	45 Ocean Dr ,	Ap 102		;	
			5	San Diego	12 Main	4	Ļ		

- Run query with original tables. Repeat query with model tables.
- Presto generates same query plan in both cases.

Modeling Results





Presto Model Speedup



SAMSUNG

Modeling Results

Presto Q44 at sf10T is the best speed up observed.

- Total bytes READ much smaller with Model – must use LOG SCALE
- Avg CPU utilization 4x smaller
- Response time decreases from 18+ minutes to 19 seconds
- Presto plan for Q44 does not scale



Q44: Average CPU busy per Server



Q44: Mean Read KB per Server



Modeling Results



SAMSUNG



Conclusion

Modeling Analytics for Computational Storage

- Near Storage optimizations for OLAP NOT universal
- Some queries see significant speedup from Near Storage opportunities
- We covered only basic operations ("low hanging fruit")
- Other Operations also amenable to Push down to Near Storage

