Migrating Complex Data Aggregations from Hadoop to Spark

puneet.kumar@pubmatic.com

ashish.singh@pubmatic.com

PubMatic

Agenda

- Aggregation and Scale @PubMatic
- Problem Statement: Why Spark when we run Hadoop
- 3 Use Cases
- Configuration Tuning
- Challenges & Learnings

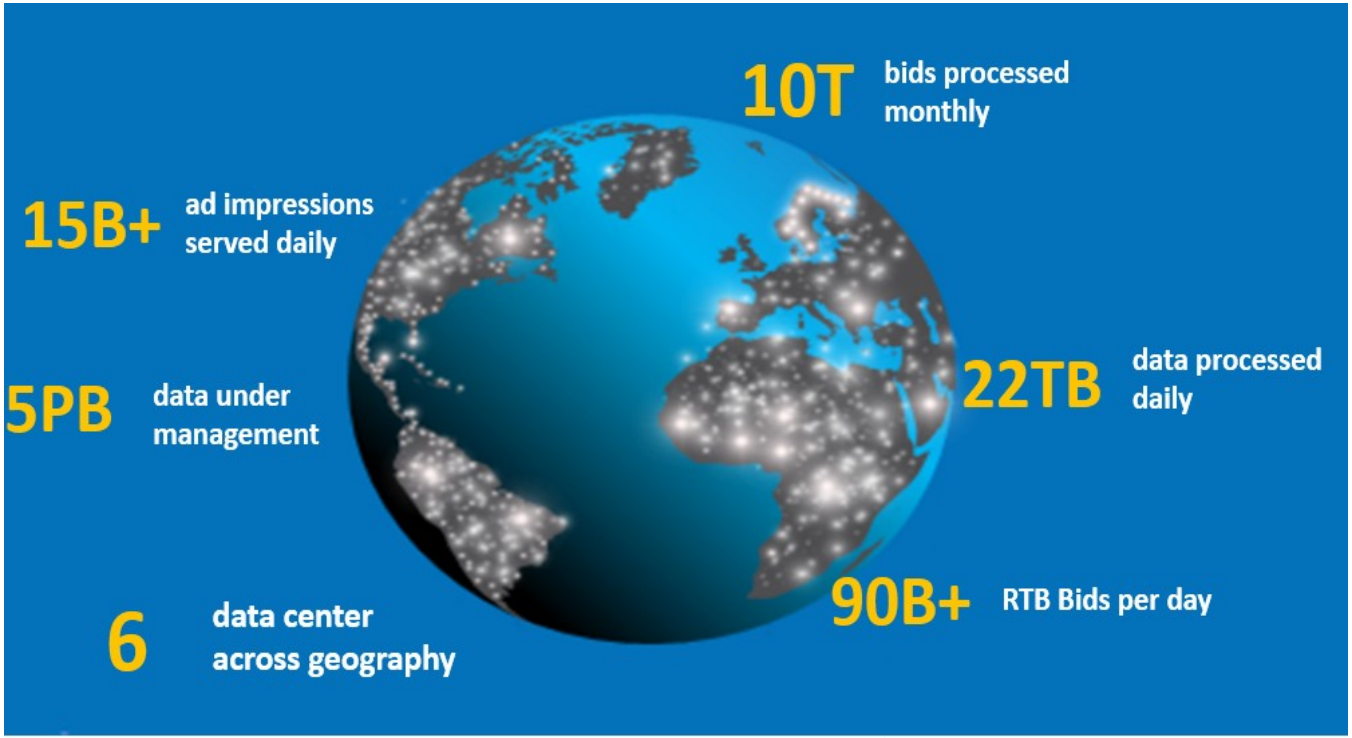
Who we are?

PubMatic

- Marketing Automation Software Company
- Developed Industry's first Real Time Analytics Solution

Pubmatic Analytics

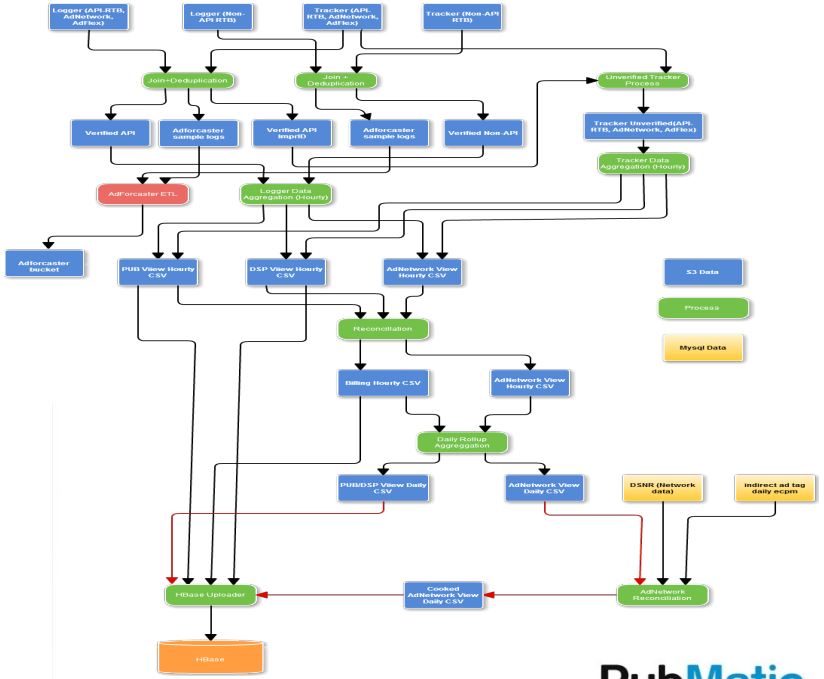
Data : Scale & Complexity



PubMatic

Challenges on Current Stack

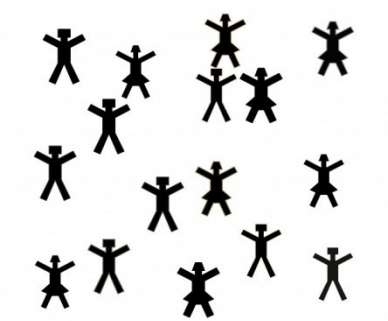
- Ever Increasing Hardware Costs
- Complex Data Flows
- Cardinality Estimation : Estimating Billion distinct users
- Multiple Grouping Sets
- Different flows for Real Time and Batch Analytics



Data Flow Diagram : Batch

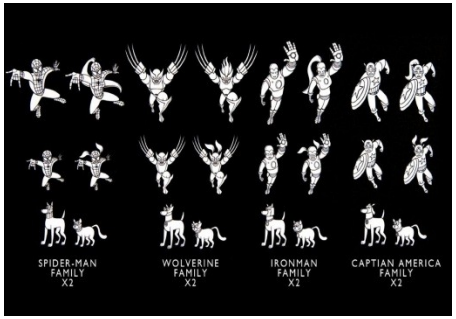
PubMatic

3 Use cases



Cardinality Estimation

Multi Stage Workflows



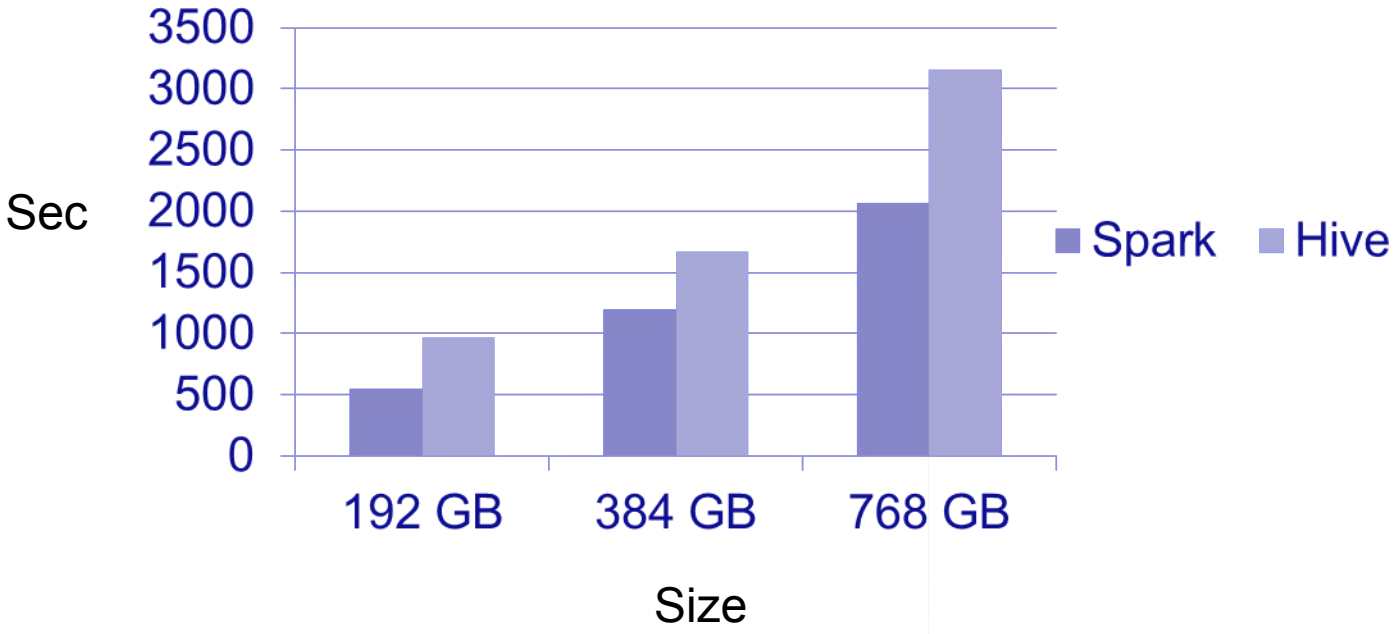
Grouping sets

Data Flow Diagram : Batch

Why Spark ?

- Efficient for dependent Data Flows
- Memory : Cheaper (Moore's Law)
- Optimized Hardware Usage
- Unified stack for Real Time & Batch
- Awesome Scala API's

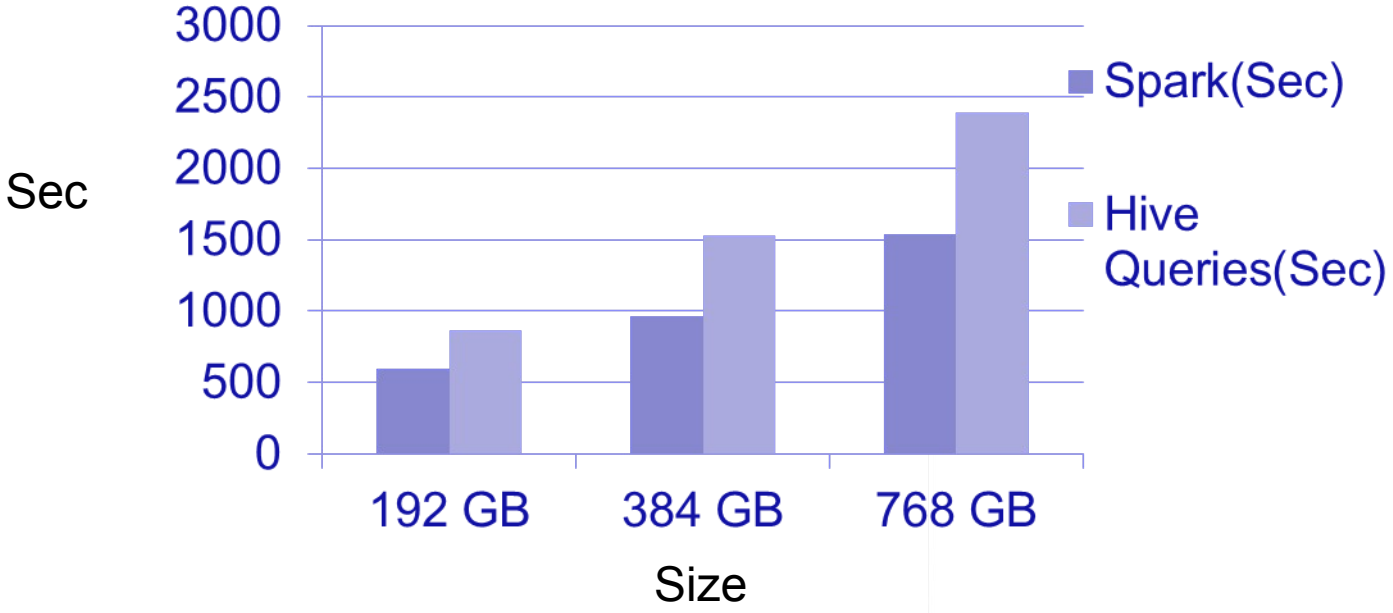
Case 1: Cardinality Estimation



Spark is ~ 25-30 % faster than Hive on MR

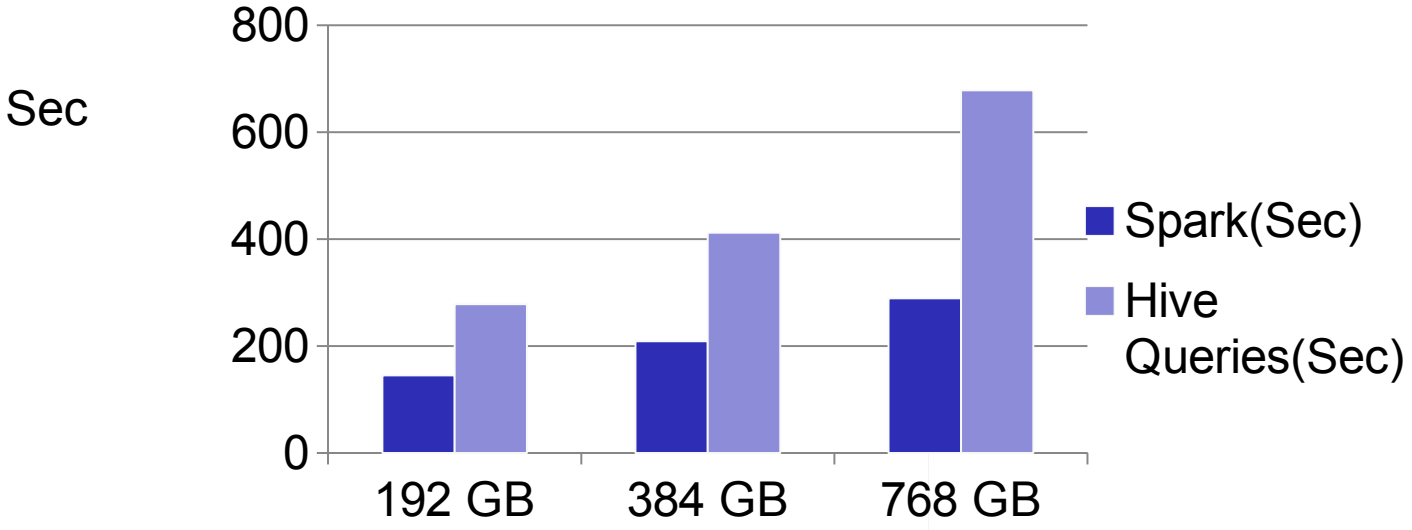
PubMatic

Case 2 :Multi Stage Data Flow



Spark is ~ 85 % faster than Hive on MR

Case 3 : Grouping Sets



Spark is ~ 150 % faster than Hive on MR

PubMatic

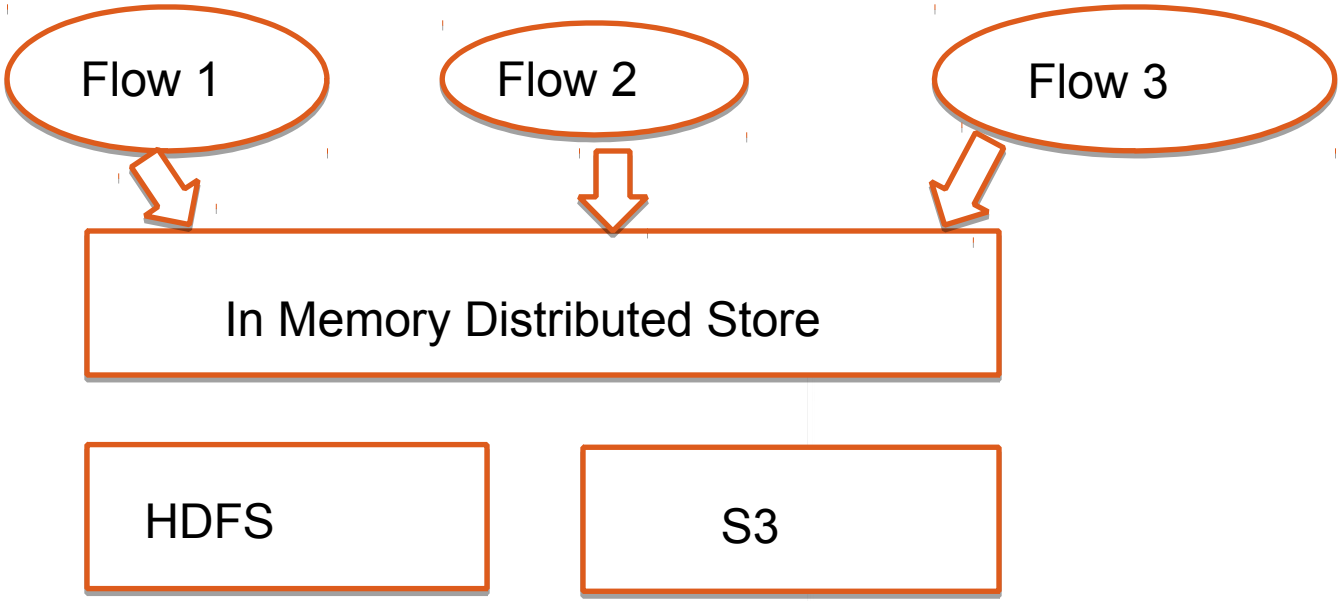
Challenges faced

- Spark on YARN : executors did not use full memory
- Reading Nested Avro Schemas until Spark 1.2 was tedious
- Had to rewrite code to leverage Spark-Avro with Spark 1.3(DataFrames)
- Join and Deduplication was slow for Spark vs Hive

Important Performance Params

- SET spark.default.parallelism;
- SET spark.serializer : Kyro Serialization improved the runtime.
- SET spark.sql.inMemoryColumnarStorage.compressed : Snappy compassion set to true
- SET spark.sql.inMemoryColumnarStorage.batchSize : Increasing it to a higher optimum value.
- SET spark.shuffle.memorySize

Memory Based Architecture



Conclusions :

- Spark Multi Stage workflows were faster by 85 % over Hive on MR
- Single stage workflows did not see huge benefits
- HLL mask generation and heavy jobs finished 20-30% faster
- Use In Memory Distributed Storage with Spark for multiple jobs on same Input
- Overall Hardware cost is expected to decrease by ~35% due to Spark usage(more memory , less nodes)

THANK YOU!