SparkCruise: Automatic Computation Reuse in Spark
Abhishek Roy, Priyanka Gomatam
Microsoft
Agenda

1. SparkCruise
   - Workload Optimization
   - Computation Reuse Problem
   - System Design
   - Demonstration and Results

2. Workload Insights Notebook
   - Motivation
   - Demonstration
The Overwhelmed DBA

On-premise databases

Database Administrators (DBA) tune the queries with sophisticated admin tools
Schedule maintenance tasks (e.g., collect statistics, create views) in offline cycles

Cloud database services

DBAs and cloud developers can tune but lack the tools to optimize massive workloads
No offline cycles in serverless systems

#Datateams #SparkAIsummit
Workload Optimization

**Why Workload Optimization**

Excellent query optimizers +CBO +AQE

Production Workloads

Many optimization opportunities

Reduce total cost for user

**Feedback Loop**

Learn from past query workloads

Adapt to changing workloads

Continuous feedback loop to optimizer

Feedback Loop in Spark:

#Datateams #SparkAIsummit
Computation Reuse Problem

Overlapping queries
- Parts of computation are duplicated across queries
- Overlapping queries
  - 95% in TPC-DS
  - 60% in production

Workloads are not fixed but recurring
- Queries are run repeatedly, e.g., hourly, or daily, over changing data possibly with different parameters
Reuse with SparkCruise

Workload on Day 1

Workload on Day 2

SparkCruise = ON

Materialized Output (Strict match)

#Datateams #SparkAIsummit
SparkCruise Design

Q₁, Q₂, ..., Qₙ

OPTIMIZER EXTENSIONS

ANNOTATED QUERY PLANS

FEEDBACK

ANALYSIS

INGESTION

Annotations → Selected Views → View Selection → Workload Table → Parsers → Application Logs

Catalyst Optimizer
Demo of SparkCruise
TPC-DS Results

Running Time Ratio on TPC-DS workload

Wall Clock Time ↓ 29%
CPU Time ↓ 31%
Geomean ↓ 33%
SparkCruise Summary

Hands-free computation reuse system for Spark

Workload-based feedback loop in Spark query engine
Select high utility common computations
Automatic materialization and reuse
Works with out-of-the-box Spark 2.3+
Workload Insights Notebook

Will SparkCruise benefit my workload?
Users can analyze individual queries, but not the complete workload

Inform users about
Redundancies in their workload
Potential savings from SparkCruise

Per Query View of Spark History Server

Details for Query 0
Submitted Time: 2020/06/02 18:21:44
Duration: 5 s
Succeeded Jobs: 0

HiveTableScan
number of output rows: 58,793

WholeStageCodegen
211 ms (22 ms, 159 ms, 159 ms)

HashAggregate
aggregate time total (min, med, max): 177 ms (25 ms, 142 ms, 142 ms)
number of output rows: 2

Exchange
data size total (min, med, max): 30.0 B (13.0 B, 13.0 B, 13.0 B)
Workload Insights Notebook

Different entities in Spark

Creates tabular representation of workload

Available for instant querying
Demo of Workload Insights Notebook
Resources

Software

SparkCruise available as an experimental feature on Azure HDInsight
To be released on Azure Synapse

Papers


Contact

Abhishek Roy (abhishek.roy@microsoft.com)
Priyanka Gomatam (priyanka.gomatam@microsoft.com)