Presto@Alibaba Group

AnalyticDB - High-Performance Real-Time OLAP Cloud Database Service

Jessica Zhang

Presto Meetup

10/17/2019

About Us – Alibaba Group

Alibaba Group

Digital Media & Entertainment

YOUKU · UC · 土豆 tudou.com · 阿里体育 Alisports.com · 阿里巴巴影业集团 Alibaba Pictures · 阿里音乐 Alibaba Music · 大麦 · 阿里游戏 · UC News · 微博 weibo.com

Core E-Commerce

阿里巴巴 1688.com · 淘宝 Taobao · TMALL天猫 · 天猫国际 TMALL GLOBAL · 聚划算 juhuasuan.com · 闲鱼 · Alibaba.com · 天猫超市 CHAOSHI.TMALL.COM · AliExpress Smarter Shopping, Better Living! · LAZADA Effortless Shopping · 农村淘宝 cun.taobao.com · 阿里拍卖

Local Services

高德地图 amap.com · 飞猪 Fliggy · 盒马 · 淘票票 · 口碑 Koubei · 饿了么

蚂蚁金服 ANT FINANCIAL — Financial Service

CAINIAO 菜鸟 — Logistics

alimama — Marketing

Alibaba Cloud — Cloud Computing

870 Million

Active Users of Alipay Around the Globe

Serving 15 million SMEs worldwide

100 Million

Packages Processed by Cai Niao Logistic Network Each Day

Covering 15 hundred cities and counties in mainland China

$768 Billion

GMV of FY2018

552 million monthly active users

AnalyticDB

online data analytics to explore business value of data for Alibaba cloud customers and Alibaba Group

Report

Delivers high performance through multi-dimensional query to support BI analysis

Interactive

Real-time interactive query from millions labeled customers to find high value targets (E-commerce, CRM…)

Intelligence

Internet company to build data analytics platform to manage & explore data value (Marketing, AD, O2O…)

Our Customers

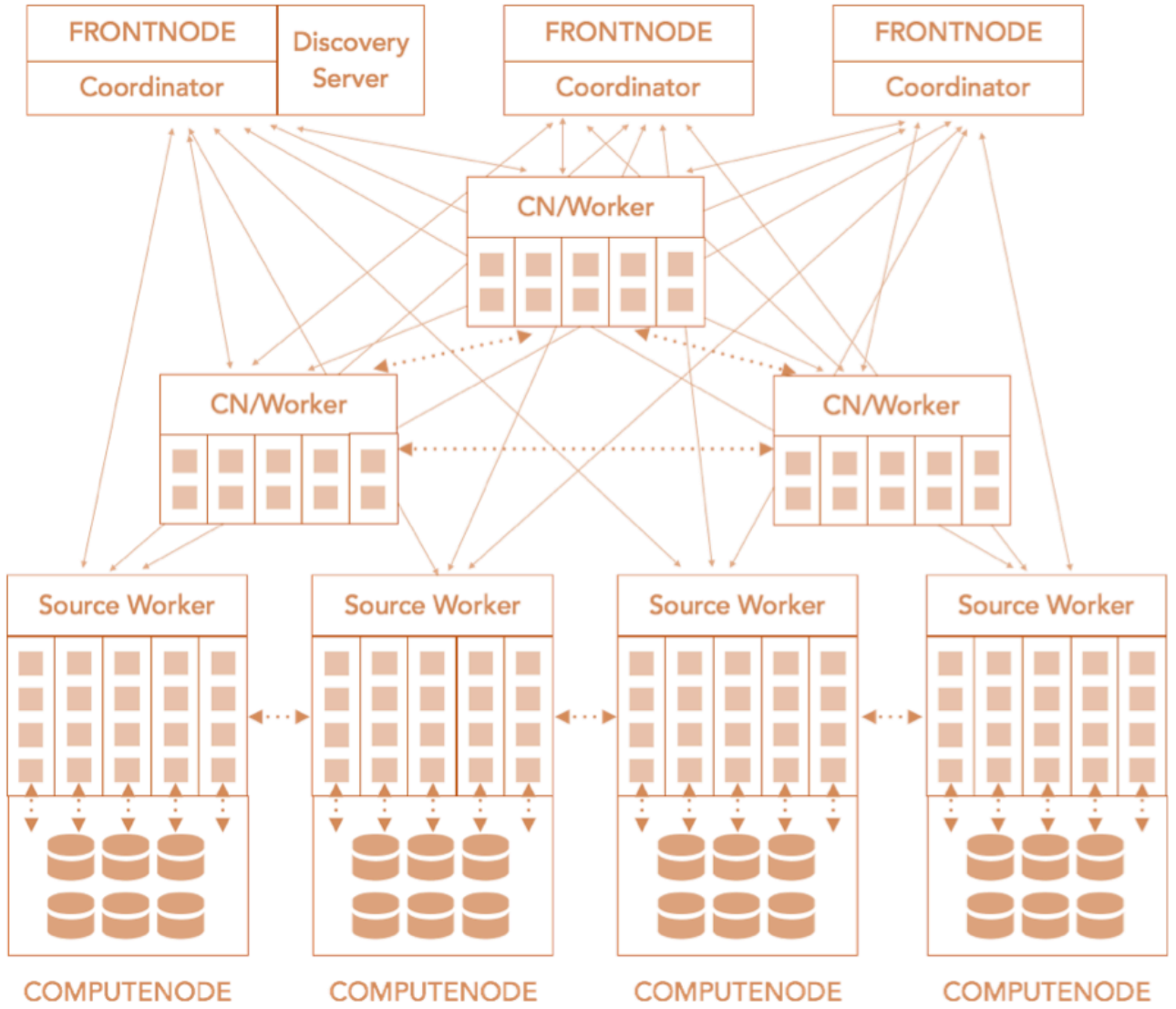
Alibaba Group

- Public Cloud & Private Cloud & Alibaba Group
  - Across 10+ industries

- Single DB
  - thousands nodes
  - thousands QPS
  - hundreds Concurrent Queries

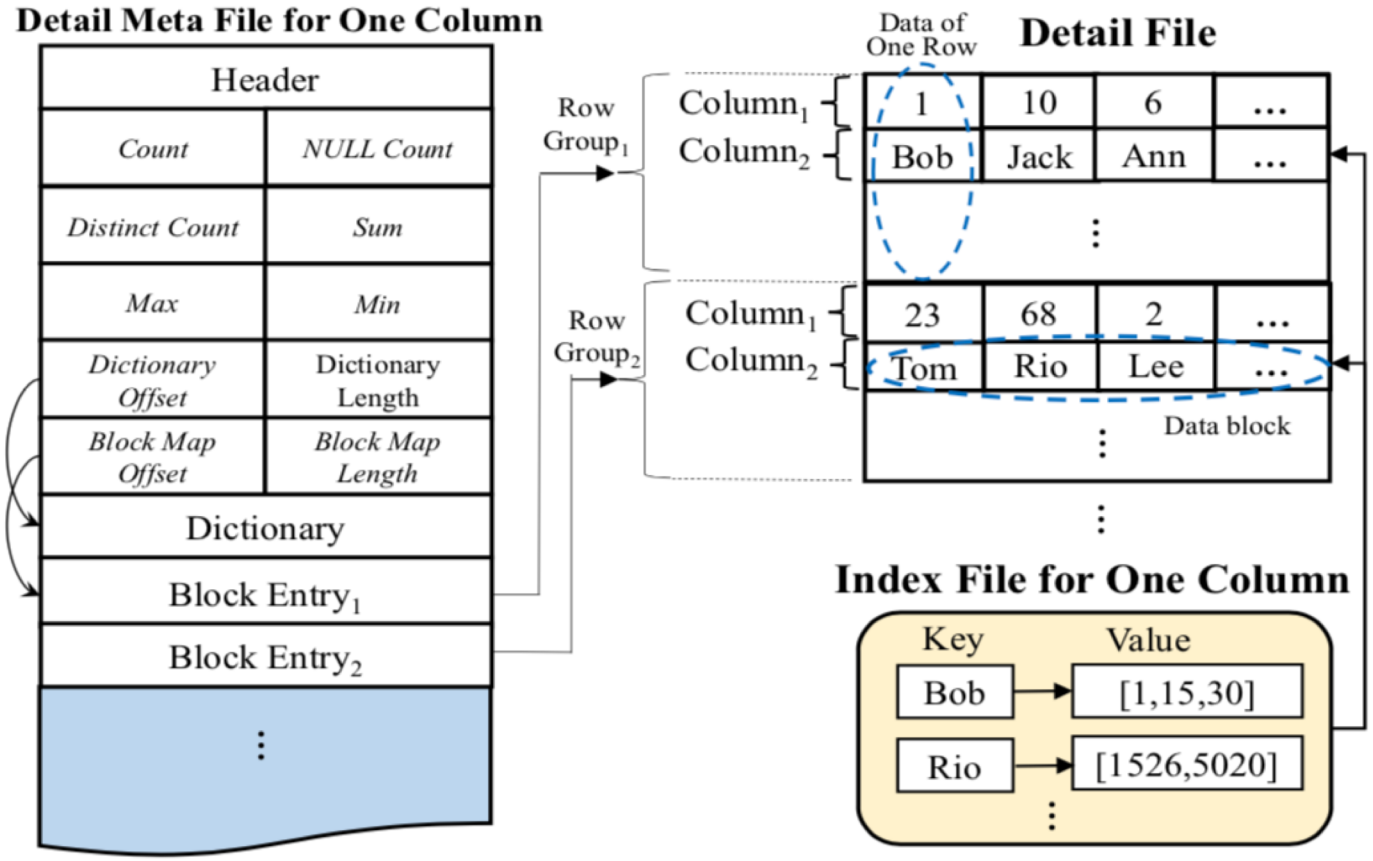
AnalyticDB Architecture

- Distributed processing system
- High availability
- Capable of scaling to thousands of nodes
- Extremely low latency, ~ms level



AnalyticDB Storage Engine

- File format
  - Hybrid row-column store
  - Fixed-size storage model for complex data type
    (JSON, Full Text, Vector data)
- Index engine
  - All-column index
  - Space-efficient async index building
  - Support Insert / Delete / Update baseline data + incremental data
- Distributed storage engine
  - Hash partition
  - High availability



Presto in AnalyticDB

Alibaba Group

- Connector to storage engine
- execution engine improvement

Connector – Storage-aware optimization

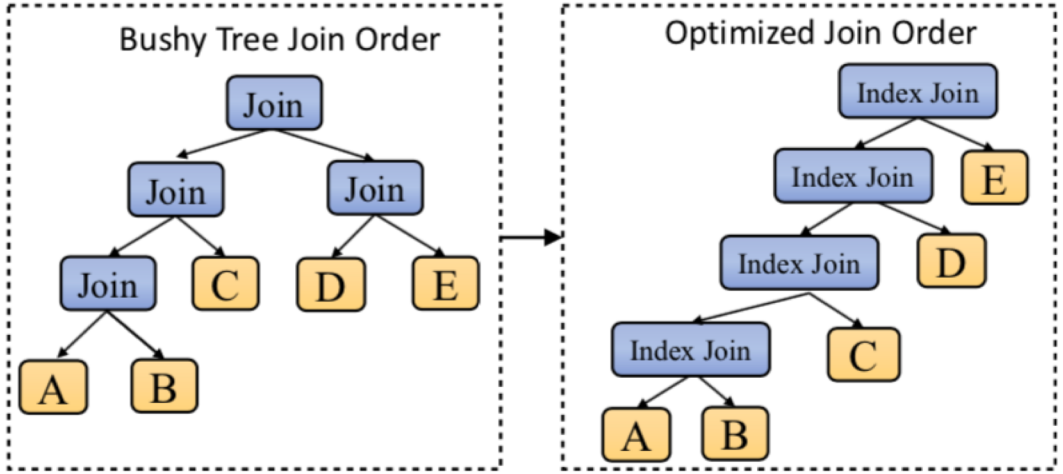
Alibaba Group

- Predicate pushdown
  - Filter condition on columns (comparison, LIKE, IN, etc.)
  - UDF (time function, vector function, etc.)
- Projection pushdown
- Colocated join / aggregation
  - When join keys or aggregation keys match hash partitioning keys
  - Reduce shuffle cost
- Bucket Pruning
  - When Equal or IN predicate on hash partitioning column
  - Avoid scheduling tasks on unnecessary nodes

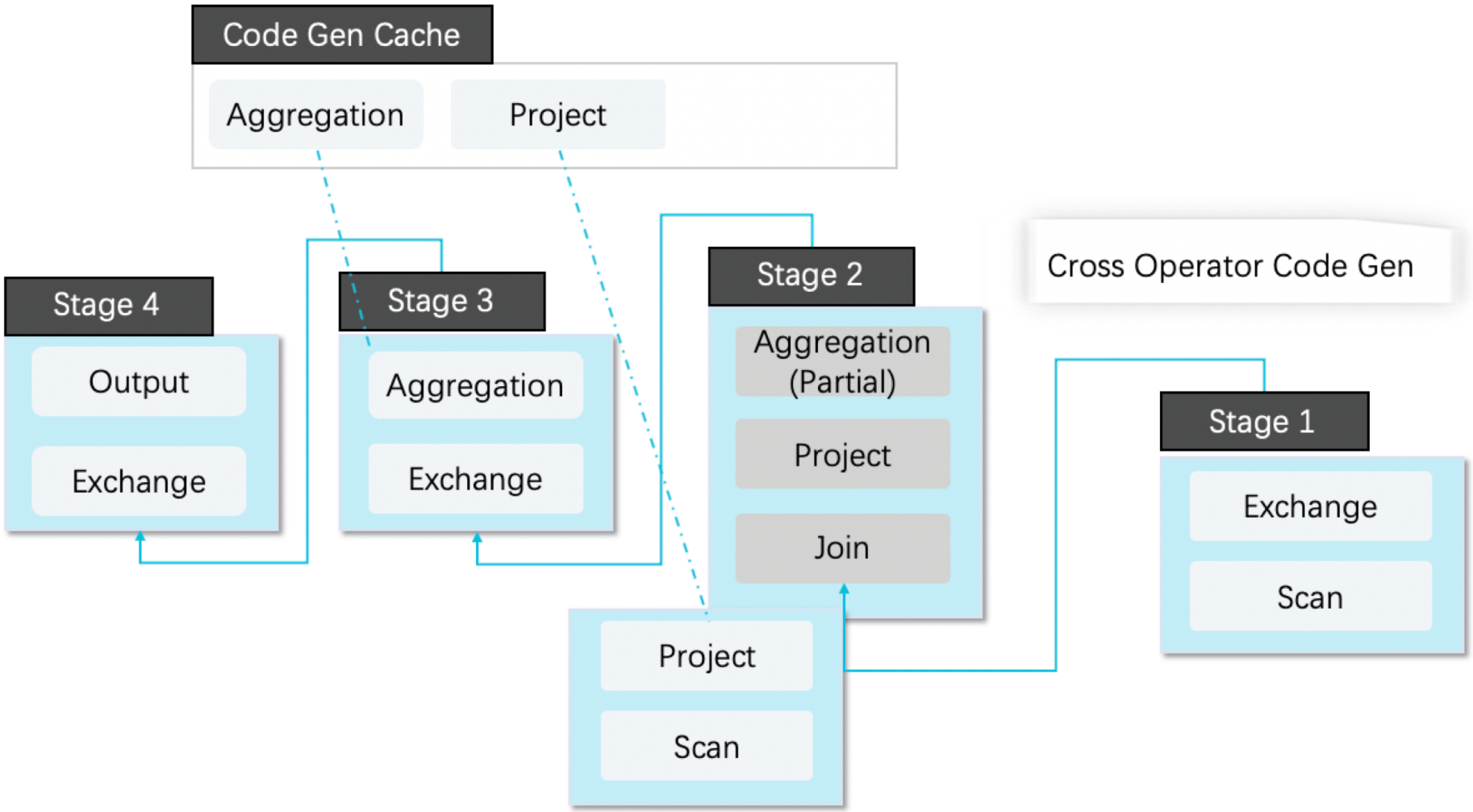
Connector – Nested loop index join

- Fit high-filtering lookup join case
- Preserve left side data distribution

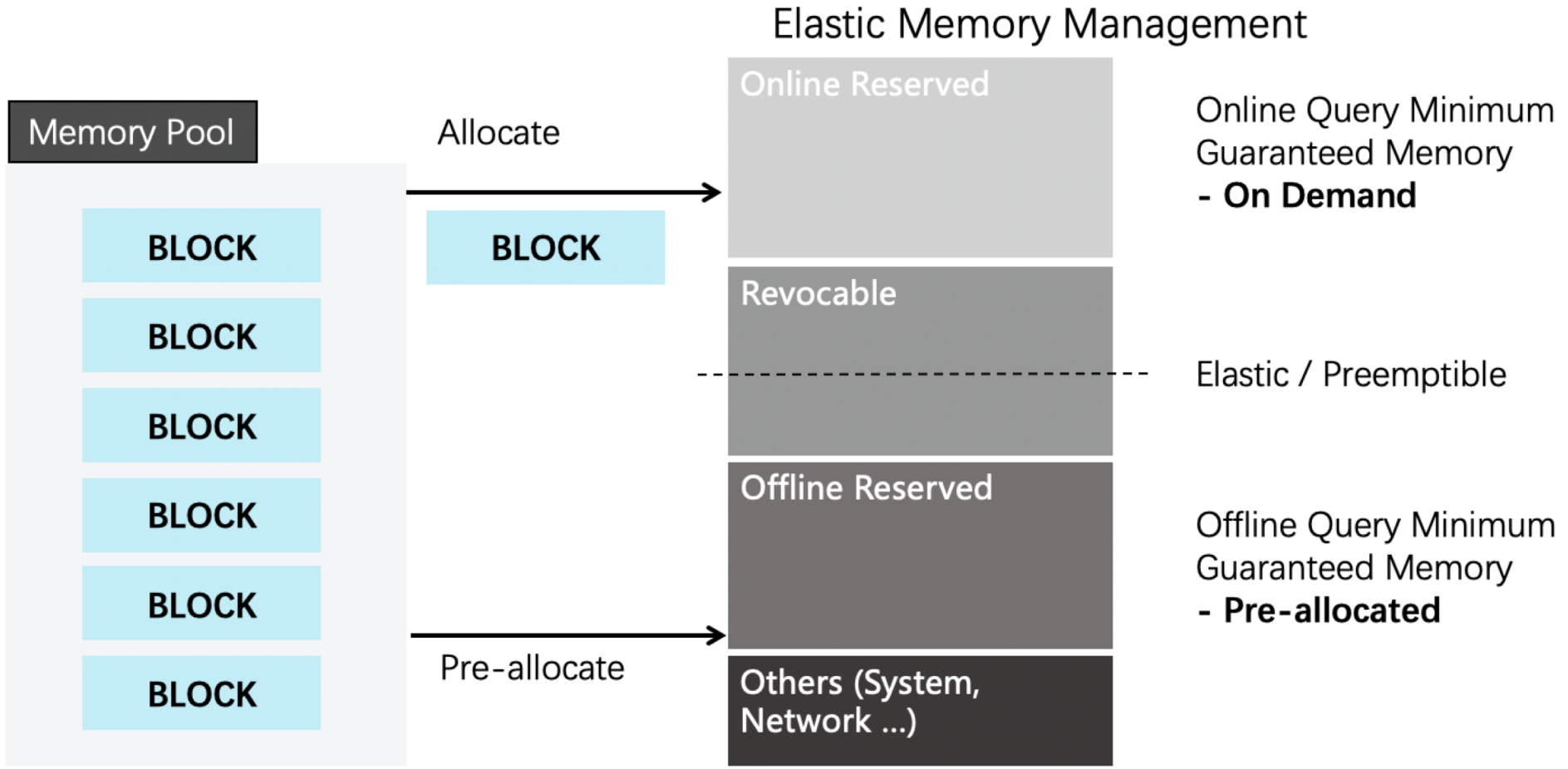
```
SELECT …
FROM   a
       JOIN b
         ON a.id = b.id
       JOIN c
         ON b.id = c.id
       JOIN d
         ON c.id = d.id
       JOIN e
         ON d.id = e.id
WHERE …
```



Compute Engine – Code Gen



Compute Engine - Memory



Compute Engine – Binary Block

Alibaba Group

- Pros
  - Eliminate Serde overhead
  - Reduce data shuffle cost
  - Reduce memory fragmentation
  - Cooperate with memory pool
- Cons
  - Degrade read performance

Optimizer – Statistics Collection

Alibaba Group

- Analyze
  - Table cardinality
  - Column cardinality
  - Column histogram
- Real-time sampling statistics
  - Efficient light weight
  - Utilize storage engine index

Optimizer -Tools

Alibaba Group

- Advanced plan explain
- Plan hint
- Planning time analysis
- Query pattern analysis

Optimizer - CBO

Alibaba Group

- Cost model
  - Filter factor
  - Derived column cardinality
  - Join cardinality
- Cost-based transformation
  - Join reorder
  - Join distribution type
  - Property enforcement
  - Partial aggregation pushdown
  - Filter join pushdown across aggregation

Thanks

Alibaba Group