AnalyticDB : Real-time OLAP Database System at Alibaba Cloud

Chaoqun Zhan, Maomeng Su, Chuangxian Wei, Xiaoqiang Peng, Liang Lin, Sheng Wang, Zhe Chen, Feifei Li, Yue Pan, Fang Zheng, Chengliang Chai
Alibaba

Abstract: In this paper, we introduce AnalyticDB, a real-time OLAP database system developed at Alibaba. AnalyticDB maintains all-column indexes in an asynchronous manner with acceptable overhead, which provides low latency for complex ad-hoc queries. Its storage engine extends hybrid row-column layout for fast retrieval of both structured data and data of complex types. To handle large-scale data with high query concurrency and write throughput, AnalyticDB decouples read and write access paths. To further reduce query latency, novel storage-aware SQL optimizer and execution engine are developed to fully utilize the advantages of the underlying storage and indexes. AnalyticDB has been successfully deployed on Alibaba Cloud to serve numerous customers (both large and small). It is capable of holding 100 trillion rows of records, i.e., 10PB+ in size. At the same time, it is able to serve 10m+ writes and 100k+ queries per second, while completing complex queries within hundreds of milliseconds.

Storage Architecture

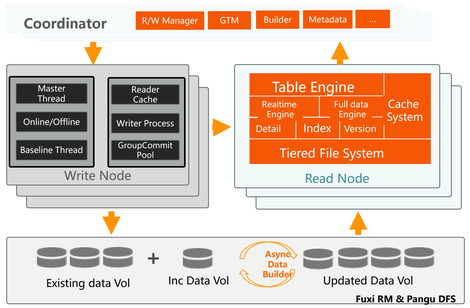
Read/Write decoupling
- High-throughput write
- High-throughput query

High Scalability
- Scale transparently
- up to 1024 nodes/DB

High Availability
- Fault-tolerant
- Self-healing
- All replicas are active

Strong Consistency
- Real-time read

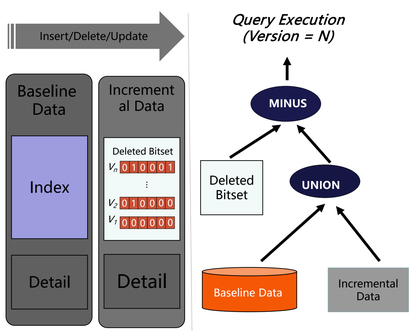
Async Data Builder
- All-Column Index Builder
- Re-partition Builder
- Re-Clustered

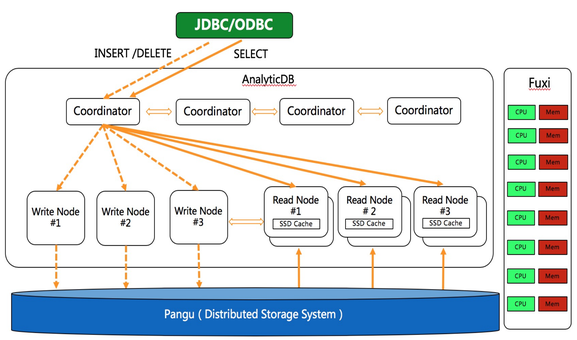


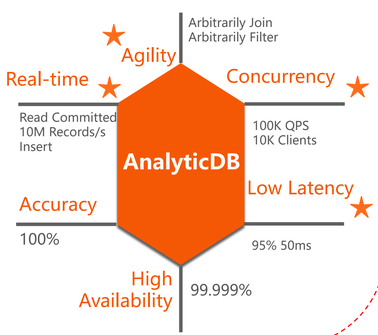
Lambda architecture
- Support fast insert
- Block index for Incremental data
- Column index for baseline data

Multi-Version
- Mark for delete with bitsets
- Copy-on-write for dedup
- Support snapshot read
- Support delete and update

Merge
- Incremental index build
- Time/size based merge
- Merge in background
- Data vacuum







Index and Data Store

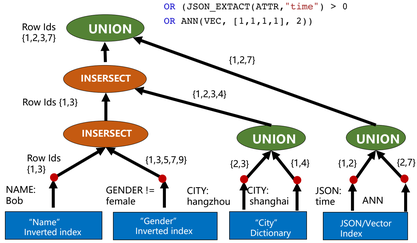
All-columns Indexing
- Indices built for all columns (automatic/optional)
- Runtime index selection

High performance ad-hoc
- Index Computing
- K-way merge for indexing results

Various Data Type
- int/varchar/time/date/...
- Full text and JSON

Example:

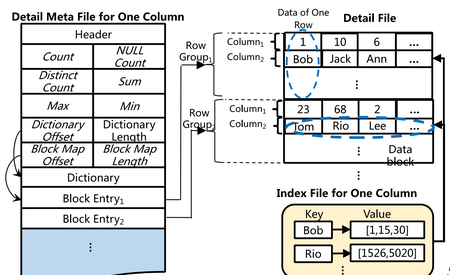
```
SELECT ... From t WHERE (name = "Bob"
    AND gender != "female"
    AND (CITY = "Hangzhou"  OR CITY = "Shanghai"))
      OR (JSON_EXTACT(ATTR,"time") > 0
      OR ANN(VEC, [1,1,1,1], 2))
```



Multi-dimensional Analysis
- Any column join
- Complex long computing tasks, ETL

Complex Query
- 1000+ columns extremely wide table
- Semi-structured, large fields

Real-time Read/Write
- Live updates
- 10 million TPS
- 10K+ QPS



Execution and Optimization

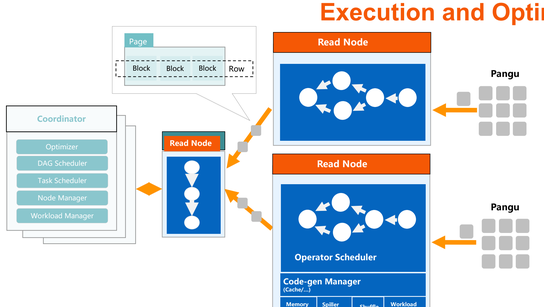
Pipelining

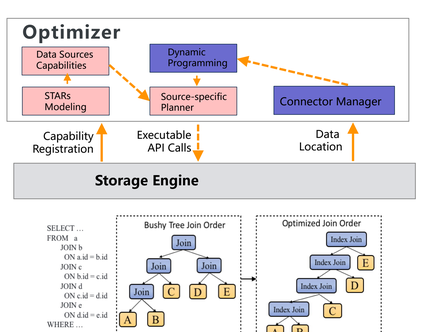
Codegen

Mixed workload

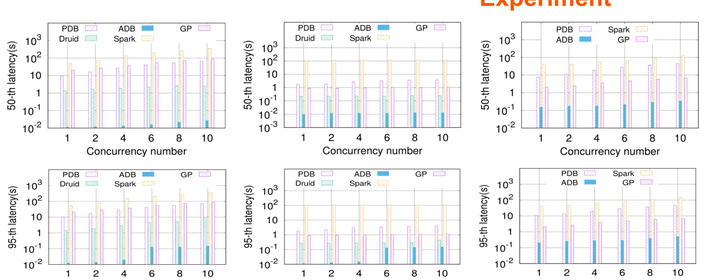
Vectorized execution

Memory pool/cache





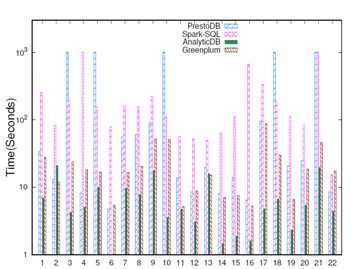
Experiment



Full Scan Query

Point Lookup Query

Multi-Join Query



TPC-H 22 Queries