

CockroachDB在京东物流的应用和实践

京东物流
史季强



目录



- 业务需求



- 系统架构



- 配置 & 性能



- 收益 & 问题



- 近期规划



- 期望



目录



- 业务需求

- 系统架构

- 配置 & 性能

- 收益 & 问题

- 近期规划

- 期望



线上历史数据结转平台DAP

数据库选型



可扩展性

- 数据量大
- 水平扩展



并发性能

- 生产库多
- 并发连接多
- 写入量大



支持事务

- 数据一致性
- RDBMS



标准SQL

- 便于查询
- 多表Join
- 学习成本



	可扩展性	并发支持	支持事务	SQL支持	其他
HBase	✓	✓	✗	✓	Rowkey设计 基于Hdfs
Greenplum	✓	✗	✓	✓	Mirror的稳定性 配置略繁琐
Tokudb	✗	✓	✓	✓	扩展性不满足
Cassandra	✓	✓	✗	✓	不支持Join





目录



- 业务需求



- **系统架构**



- 配置 & 性能



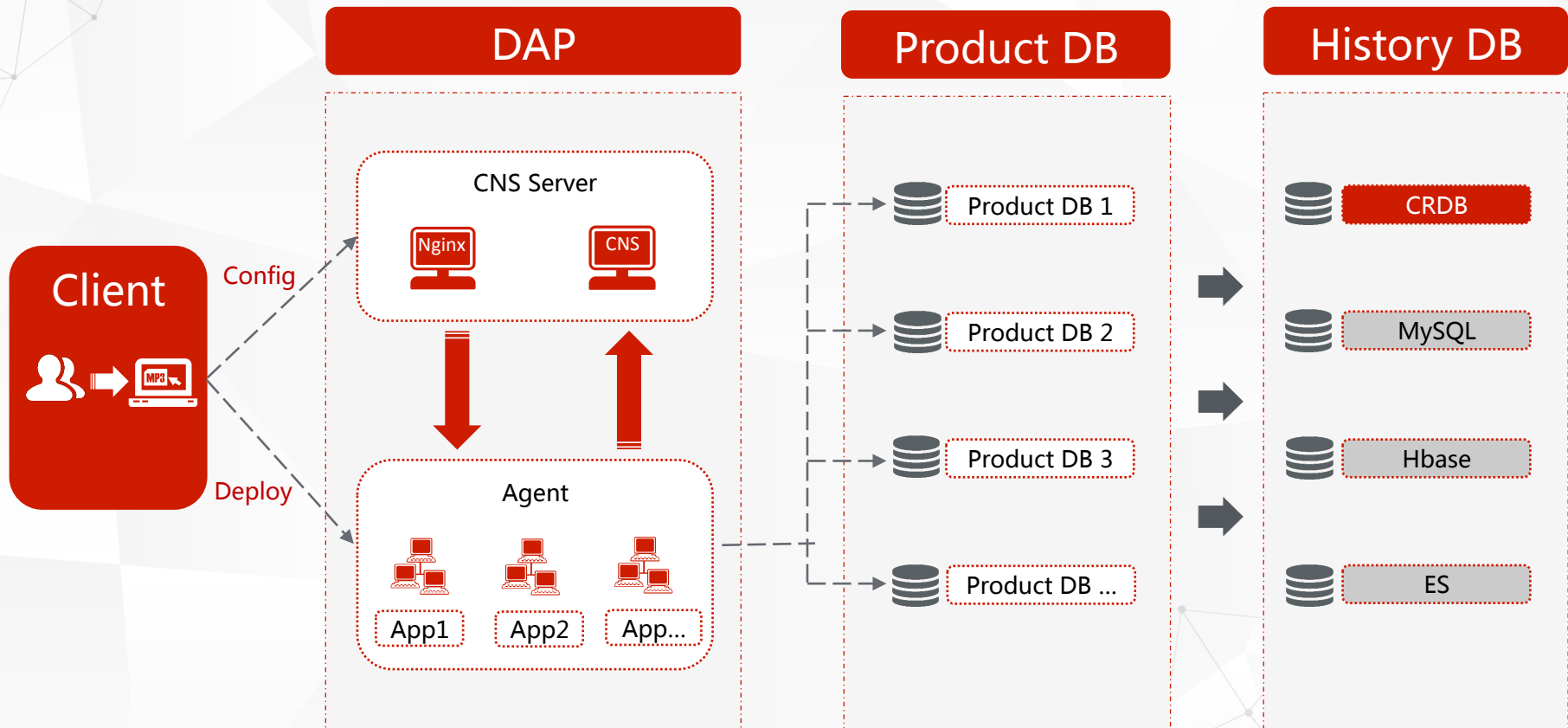
- 收益 & 问题



- 近期规划



- 期望





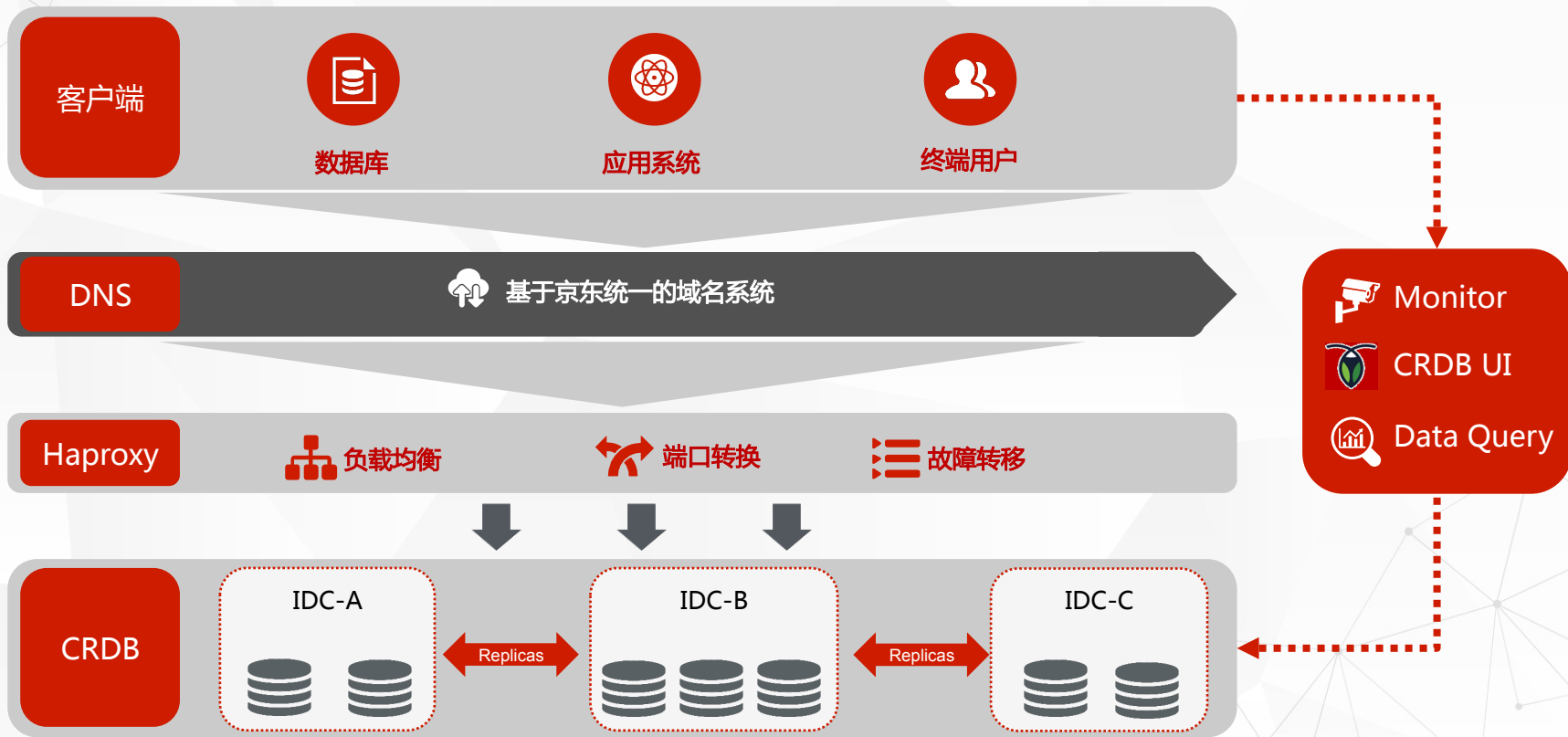
MySQL	CRDB
BOOLEAN	BOOLEAN
BIGINT	BIGINT
BINARY	BYTEA
BIT	BOOLEAN
CHAR	CHAR
CHARACTER	CHARACTER
DATE	DATE
DATETIME	TIMESTAMP WITH TIME ZONE
DECIMAL	DECIMAL
DEC	DEC
DOUBLE	DOUBLE PRECISION
FLOAT	REAL
INT	INT
INTEGER	INTEGER
SMALLINT	SMALLINT
TINYBLOB	BYTEA
BLOB	BYTEA
MEDIUMBLOB	BYTEA
LONGBLOB	BYTEA
TINYINT	SMALLINT
TINYTEXT	TEXT
TEXT	TEXT
MEDIUMTEXT	TEXT
LONGTEXT	TEXT

MySQL	CRDB
TIME	TIME // CockroachDB不支持TIME
TIMESTAMP	TIMESTAMP WITH TIME ZONE
VARBINARY	BYTEA
VARCHAR	VARCHAR
BIGINT AUTO INCREMENT	BIGSERIAL
INT AUTO INCREMENT	SERIAL
INTEGER AUTO INCREMENT	SERIAL
SMALLINT AUTO INCREMENT	SMALLSERIAL
TINYINT AUTO INCREMENT	SMALLSERIAL
BIGINT UNSIGNED AUTO INCREMENT	BIGSERIAL
BIGINT UNSIGNED AUTO INCREMENT	BIGSERIAL
INT UNSIGNED AUTO INCREMENT	SERIAL
INTEGER UNSIGNED AUTO INCREMENT	SERIAL
SMALLINT UNSIGNED AUTO INCREMENT	SMALLSERIAL
TINYINT UNSIGNED AUTO INCREMENT	SMALLSERIAL
BIGINT UNSIGNED	NUMERIC
FLOAT UNSIGNED	REAL
INT UNSIGNED	BIGINT
MEDIUMINT UNSIGNED	INTEGER
SMALLINT UNSIGNED	INTEGER
TINYINT UNSIGNED	SMALLINT
YEAR	DATE
MEDIUMINT	INTEGER
NUMERIC	NUMERIC

CRDB部署架构



京东物流
JD Logistics





目录



- 业务需求



- 系统架构



- **配置 & 性能**



- 收益 & 问题



- 近期规划



- 期望



硬件层面



推荐使用SSD硬盘

- 性能
- 稳定性



NTP时间同步

- MVCC
- 数据一致性
- Crash风险



跨机房部署

- 本地机房，网络延迟低
- 异地机房，网络延迟高，资源要求



软件层面



必须使用加密模式

- 数据安全
- 合规审计



使用代理访问

- 避免直接暴露IP和端口
- Load balance
- Failover



配套监控

- 服务器资源监控
- 数据库性能监控



及时升级

- 性能
- Bug



数据安全性



数据隔离性



集群高可用

Cluster

Database

Table

Index

Row

策略优先级

◆调整系统范围的复制

.default .meta .liveness .system .timeseries

◆跨数据中心进行复制

3个数据中心，每个中心2节点即可

◆数据写在特定的数据中心

基于constraints 和 lease_preferences进行配置

◆多个应用程序写入不同的数据库

针对database和table的constraints配置

◆Index和row级的复制



节点数	CPU	磁盘	内存	网卡
9	Xeon E5-2683V4 32C	SSD 12*1T	256G	万兆

OVERVIEW DASHBOARD

GRAPH: CLUSTER ▾

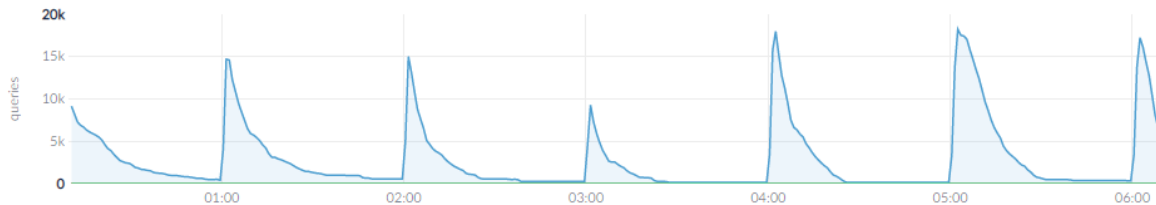
DASHBOARD: OVERVIEW ▾



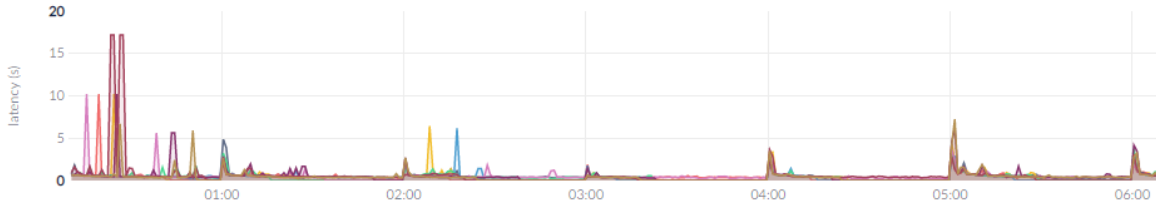
LAST 6 HOURS



SQL Queries ⓘ



Service Latency: SQL, 99th percentile ⓘ



Summary

Total Nodes [View nodes list](#) 9

Capacity Used **42.72%**
You are using 21.5 TiB of 50.3 TiB usable storage capacity across all nodes.

Unavailable ranges **0**Queries per second **4694.5**P50 latency **117.4 ms**P99 latency **637.5 ms**

Events

Table Created: User report_... 22 days ago

Schema Change Completed... 23 days ago

Schema Change Completed...



节点数	CPU	磁盘	内存	网卡
5	Xeon E5-2683V4 32C	SSD 12*1T	256G	万兆

OVERVIEW DASHBOARD

GRAPH: CLUSTER ▾

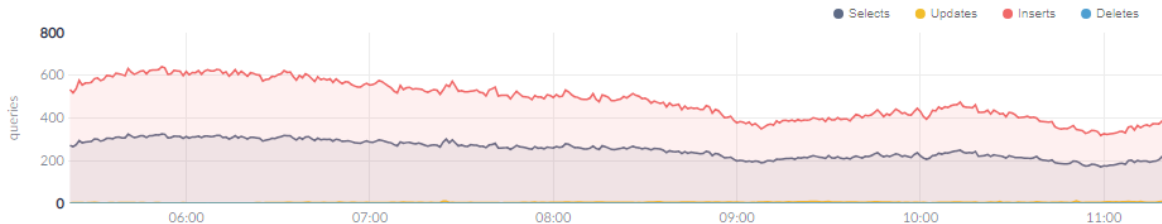
DASHBOARD: OVERVIEW ▾



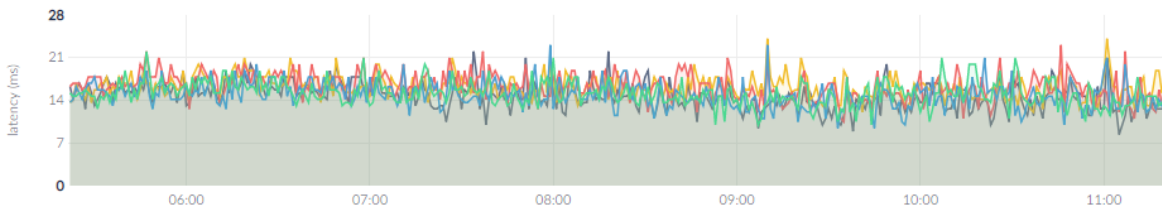
LAST 6 HOURS



SQL Queries ⓘ



Service Latency: SQL, 99th percentile ⓘ



Summary

Total Nodes [View nodes list](#) 5Capacity Used 5.86%
You are using 2.0 TiB of 34.9 TiB usable storage capacity across all nodes.

Unavailable ranges 0

Queries per second 611.9

P50 latency 2.0 ms

P99 latency 18.9 ms

Events

Schema Change: User root ... 4 days ago

Cluster Setting Changed: U... a month ago

Node Rejoined: Node 1 rei a month ago



目录

- 业务需求
- 系统架构
- 配置 & 性能
- **收益 & 问题**
- 近期规划
- 期望



稳定性高

虽然个别版本曾经出现Bug，导致节点宕机，但是集群整体来看稳定性极高，没有出现重大故障

数据扩展性问题

不用依赖中间件或者其他工具，可以自动进行数据水平扩展，提高了系统的伸缩性

性能满足要求

性能基本可以满足我们目前业务的需求，并且可以看到随着版本的升级，性能也在逐渐提升

新的DB解决方案

提供了一种新的数据库解决方案，可以根据业务的特点进行选择



- 字符大小写敏感
- 2.0版本下没有DOUBLE PRECISION类型，会导致DOUBLE字段精度丢失
- 2.1版本下无法修改字段长度
- 2.0.2版本会有OutOfMemory和CPU使用率100%的问题，导致个别节点宕机或者不可用
- 使用HDD硬盘也会有节点宕机的问题
- 并发量激增的时候延迟会明显增大



目录



- 业务需求



- 系统架构



- 配置 & 性能



- 收益 & 问题



- **近期规划**



- 期望



OLTP
并发量可控的在线交易业务



OLAP
轻量级的数据分析业务



历史数据
大容量的数据存储业务



容器化部署
基于K8S的虚拟化部署





目录



- 业务需求



- 系统架构



- 配置 & 性能



- 收益 & 问题



- 近期规划



- **期望**



兼容MySQL协议

更有利于CRDB的推广和使用，
减少学习成本和迁移成本



数据压缩功能

可以对于大数据量存储的
业务进行空间优化

表分区功能

便于对数据的分类和生命周期
进行管理



一款非常棒的分布式数据库，希望持续完善功能，提高性能！

Thanks !