

恬泰

阿里云智能高级技术专家

阿里云Cassandra技术架构及最佳实践

恬泰

大纲



Apache
Cassandra 介
绍



Cassandra 架
构介绍



阿里云
Cassandra 特
性介绍

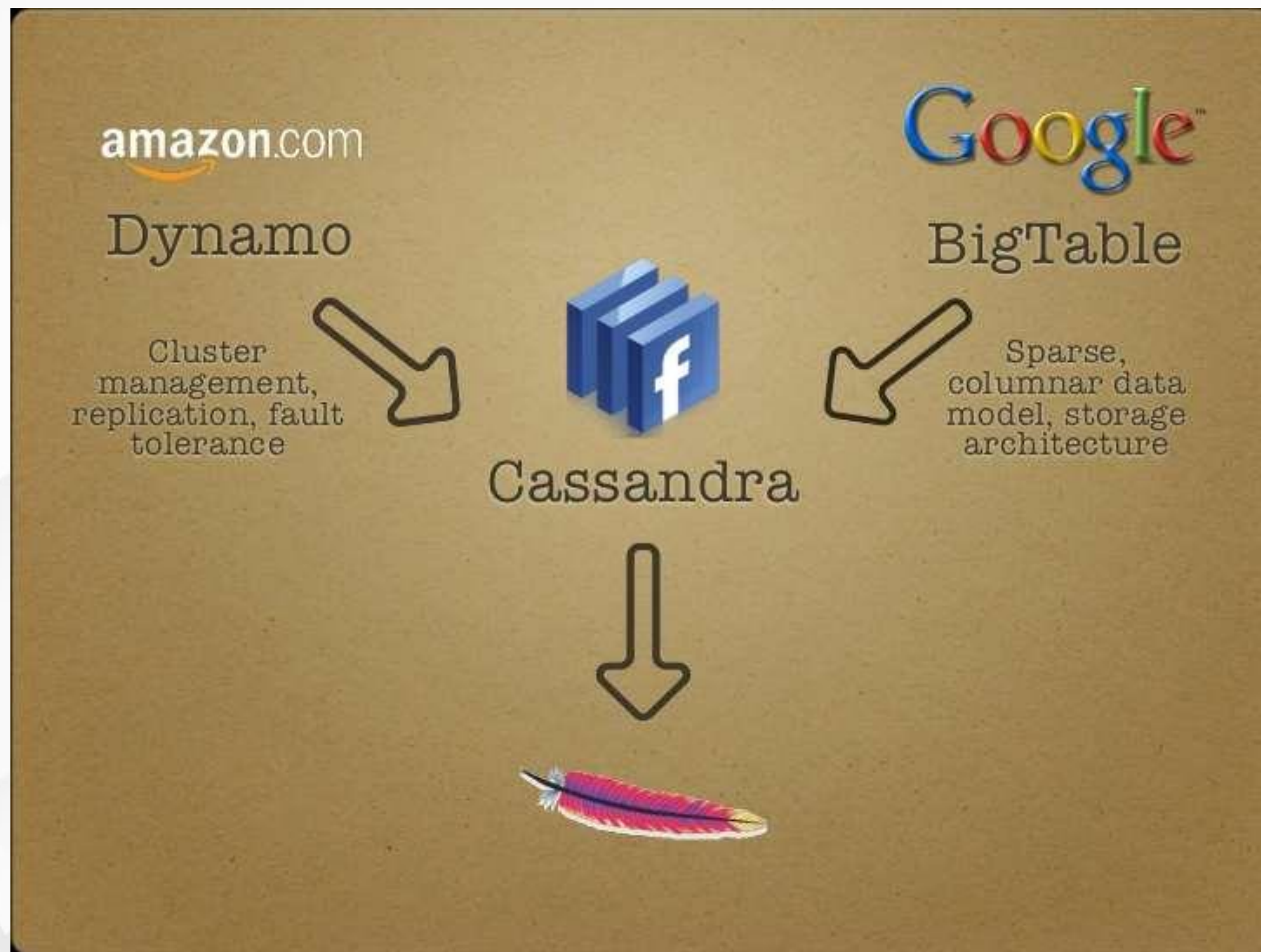


最佳实践及
案例

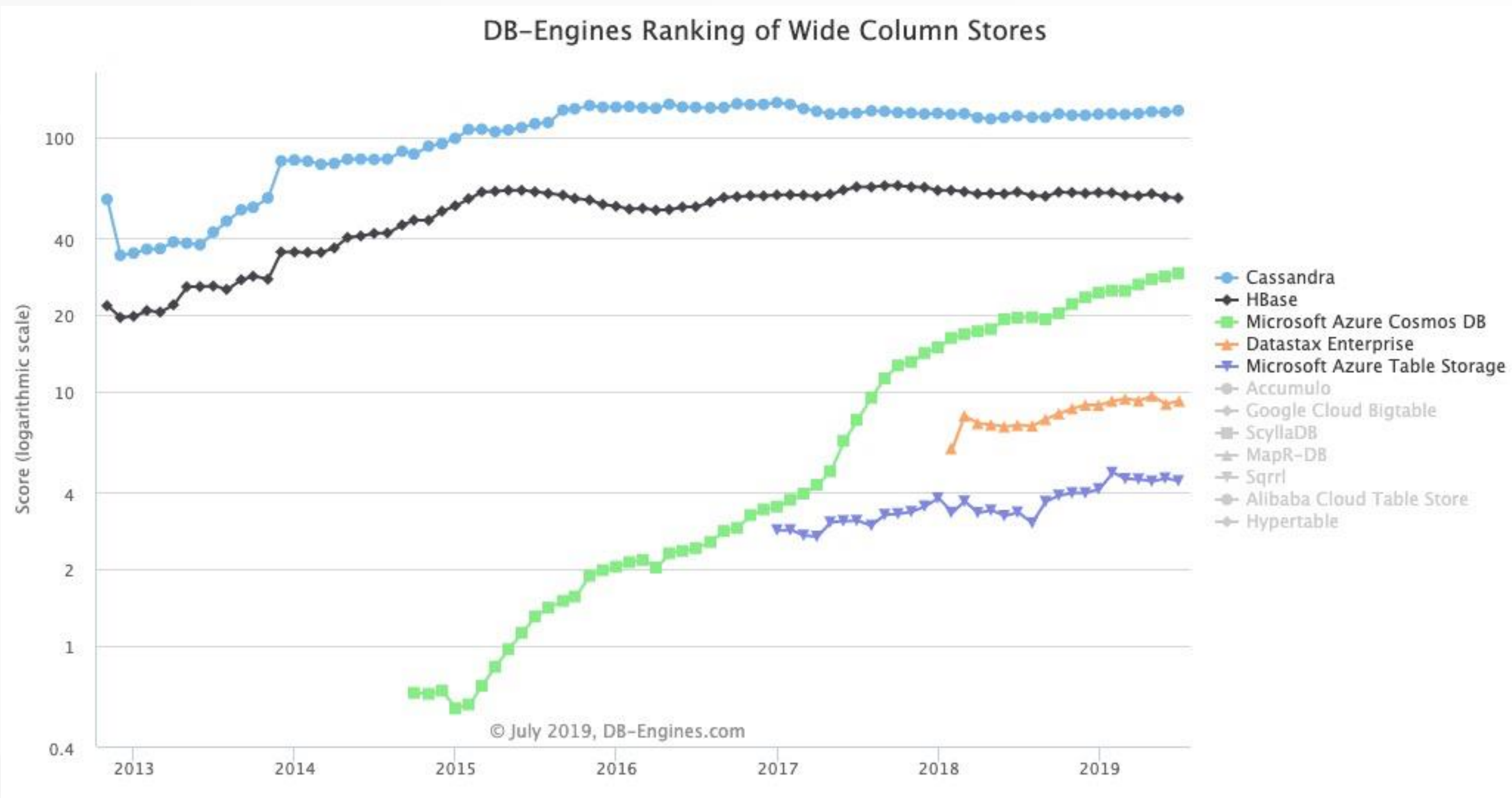
Why does Cassandra Exist?



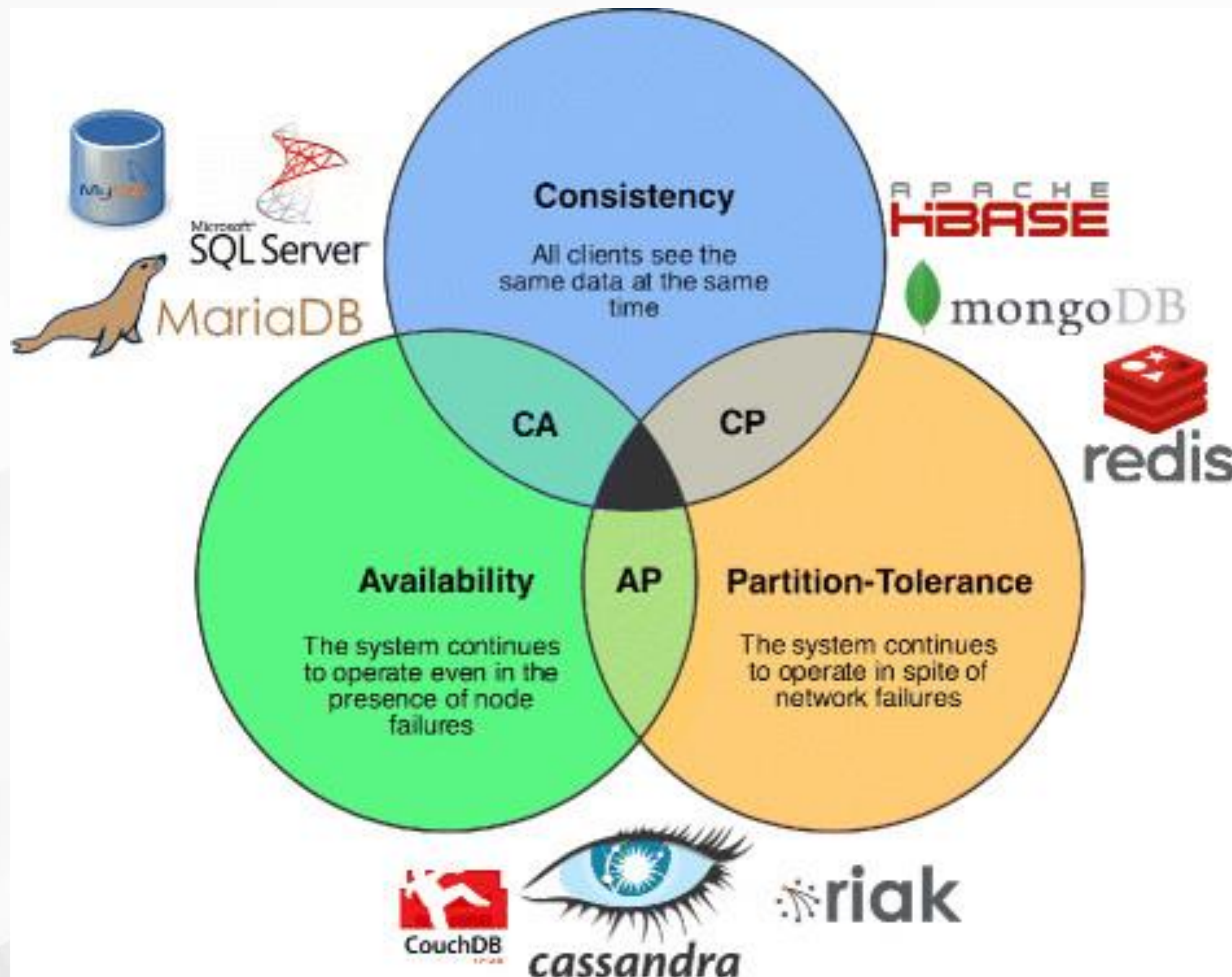
Cassandra 历史



DB-Engines Ranking



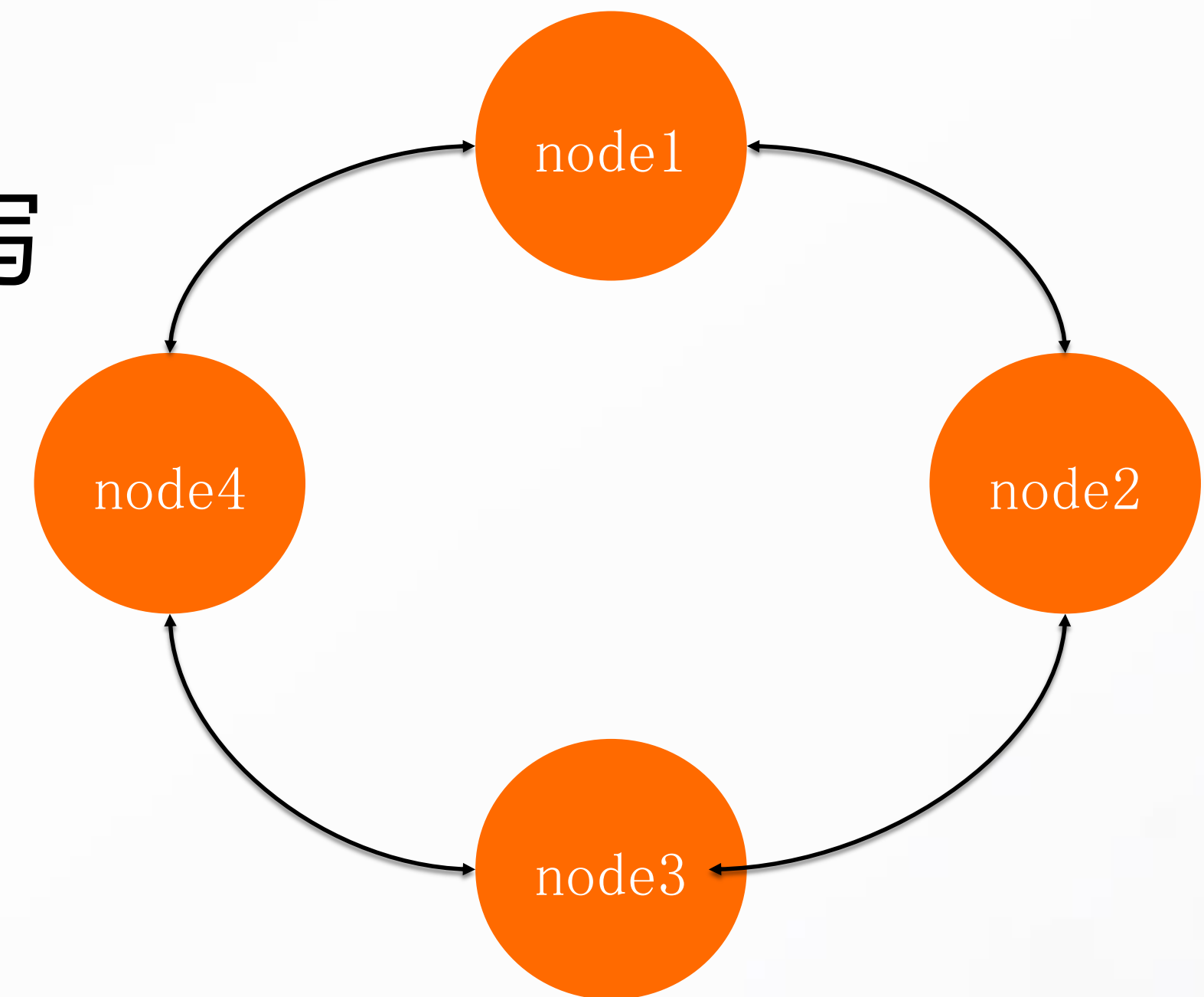
CAP 理论



Cassandra 架构介绍

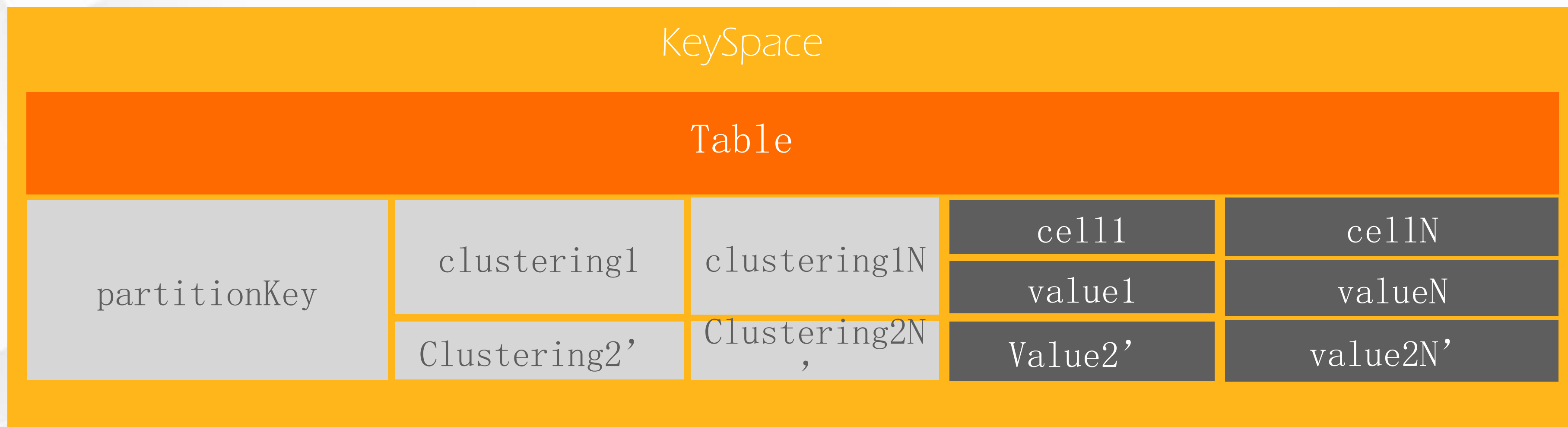
Cassandra 节点对等

- 集群中所有节点对等，角色相同,副本可同时写
- Shared nothing，竖井式架构
- 按需随意增加/移除节点
- 更多能力？加一台机器
- Gossip 交换状态，token 等信息



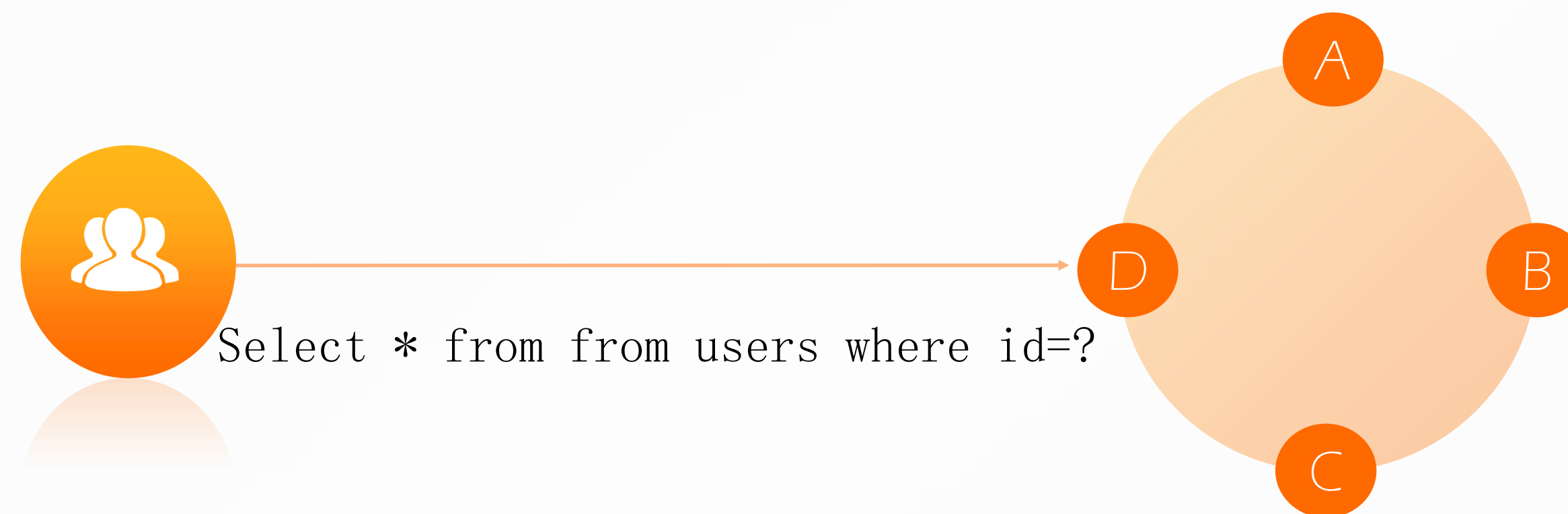
数据模型

```
CREATE TABLE timeline (  
  userid uuid,  
  posted_month int,  
  posted_time uuid,  
  body text,  
  posted_by text,  
  PRIMARY KEY (userid, posted_month, posted_time)  
) WITH compaction = { 'class' : 'LeveledCompactionStrategy' };
```



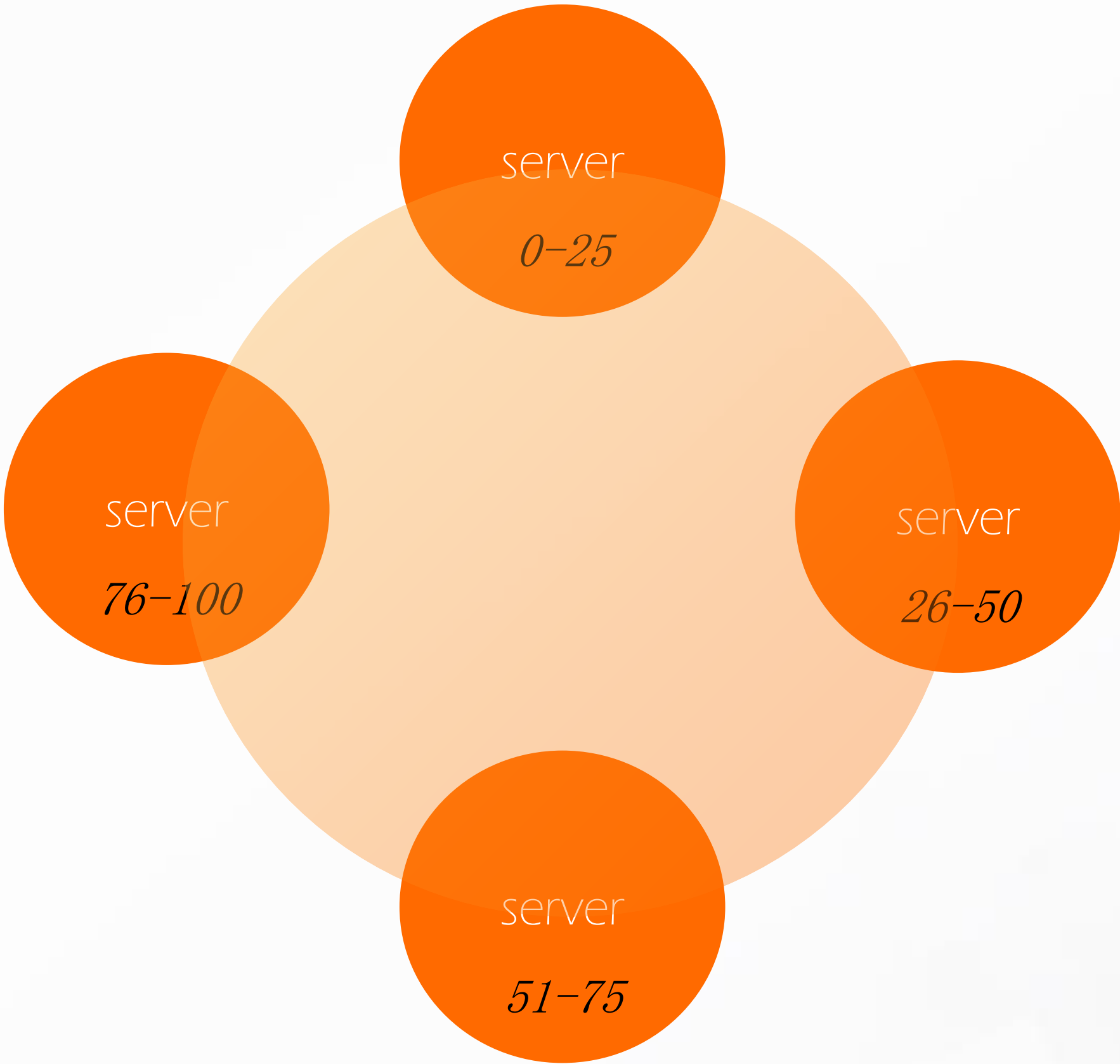
CQL 语言

- Cql 类似于 SQL
- DDL 操作表
- 支持 DML 操作 INSERT、UPDATE、DELETE 等等
- 查询数据通过 select



Token Ring

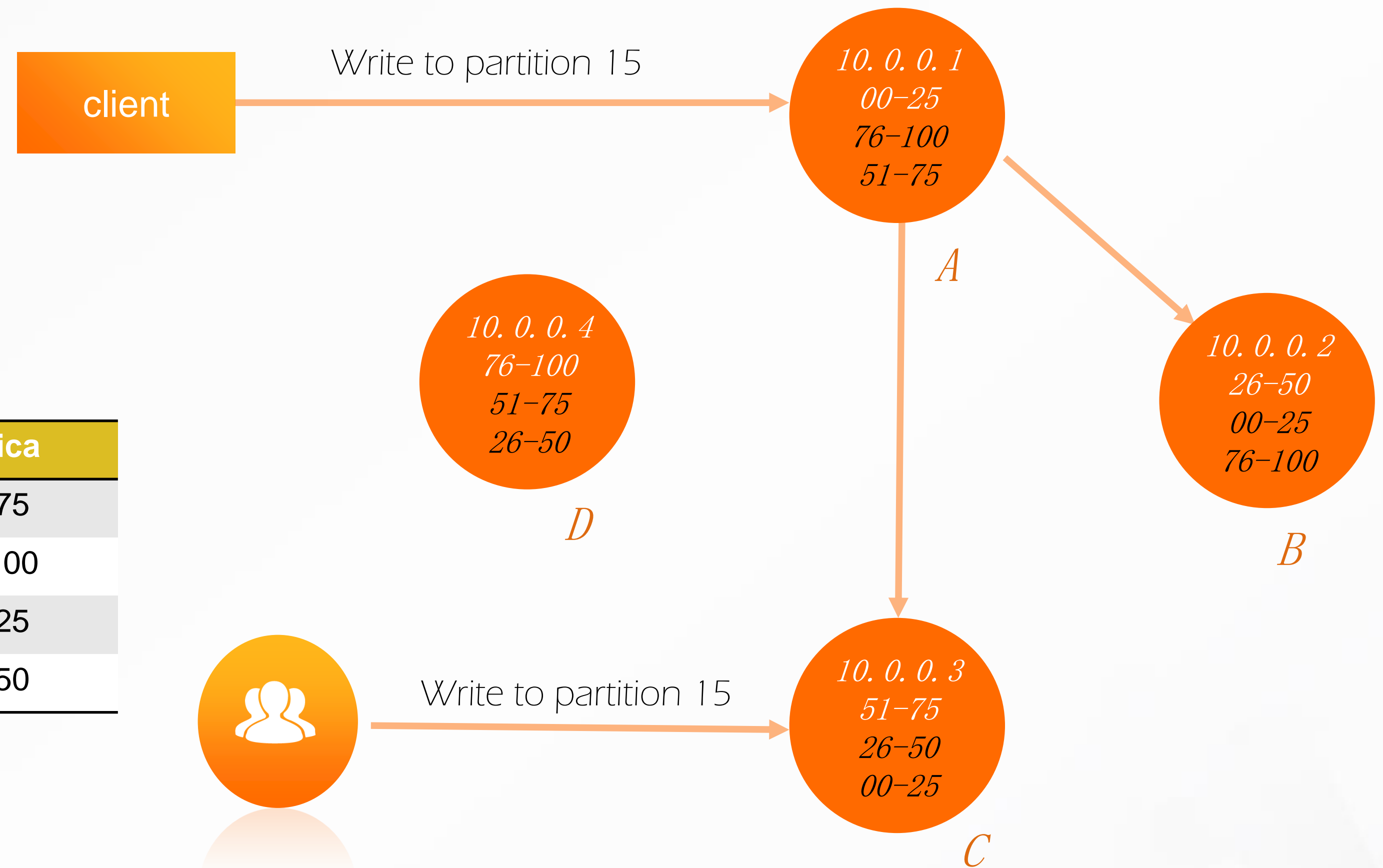
Token	Range
0	0-25
26	26-50
51	51-75
76	76-100



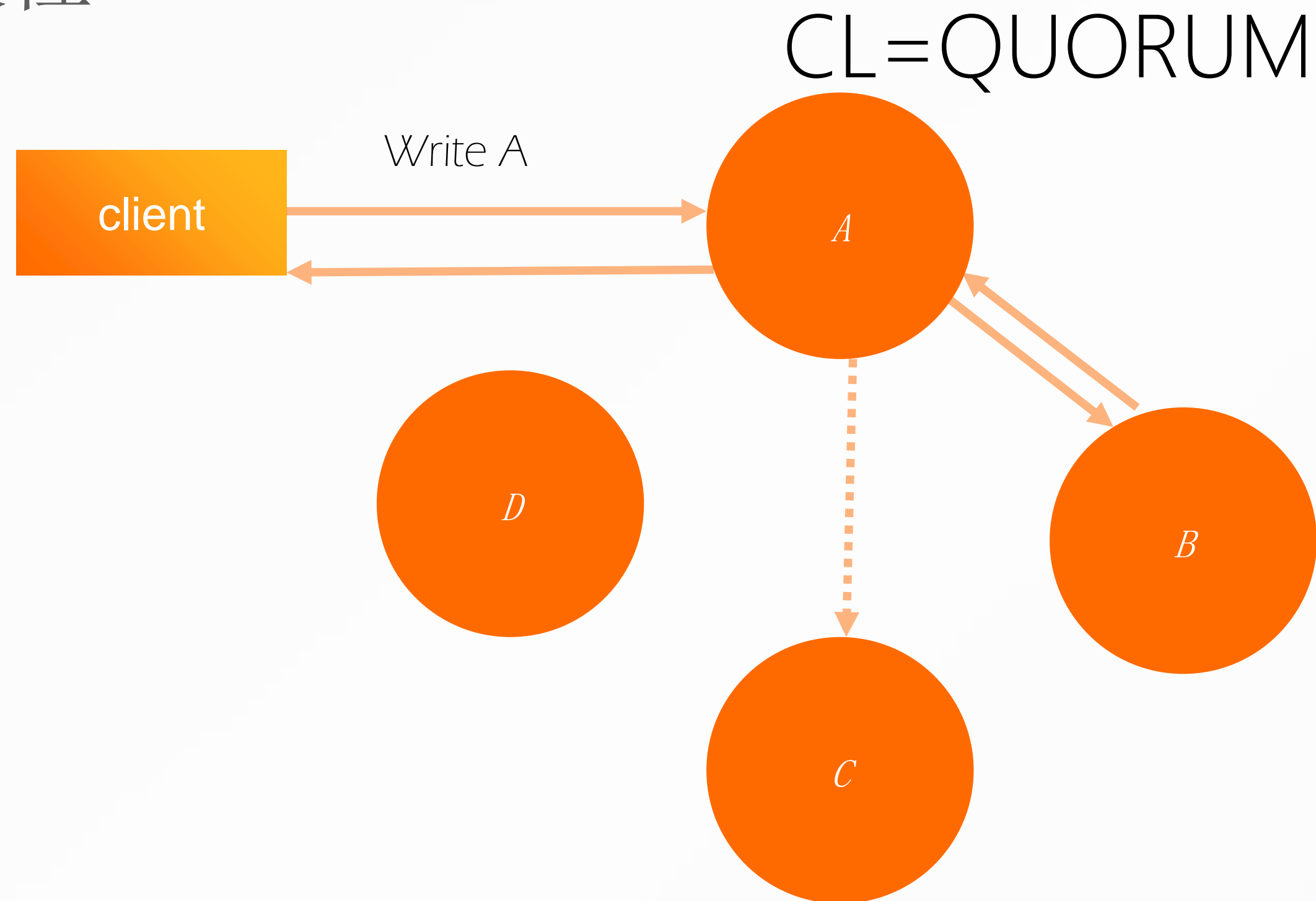
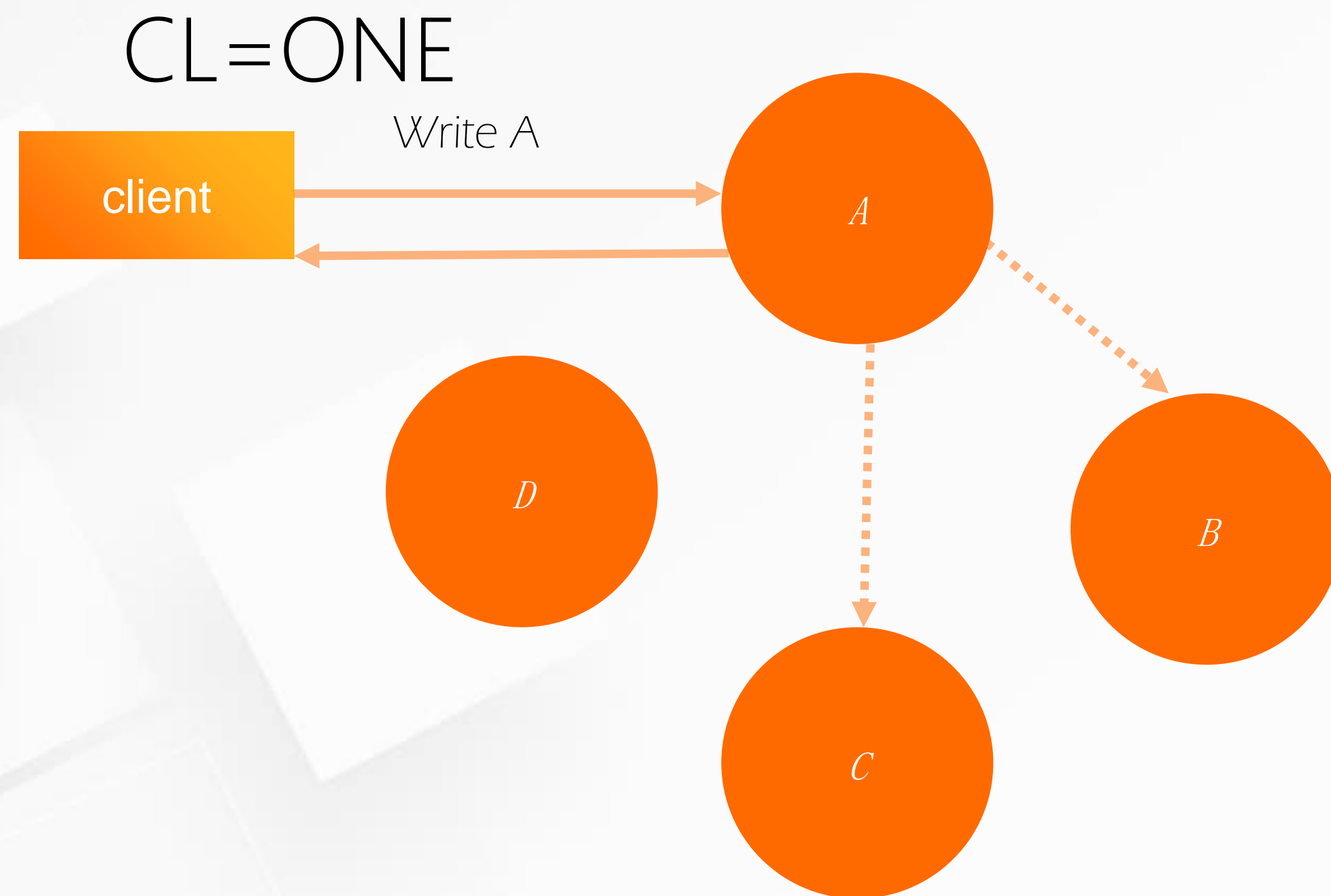
Replication

RF=3

Node	Primary	replica	replica
10.0.0.1	00-25	76-100	51-75
10.0.0.2	26-50	00-25	76-100
10.0.0.3	51-57	26-50	00-25
10.0.0.4	76-100	51-75	26-50



一致性



可调一致性

- 可选择从强一致到最终一致
- 可选: ANY, ONE, QUORUM, ALL, SERIAL
- 修复: read-repair/hint/repair

Index

- 二级索引
 - * 本地二级索引
 - * SASI
- MV

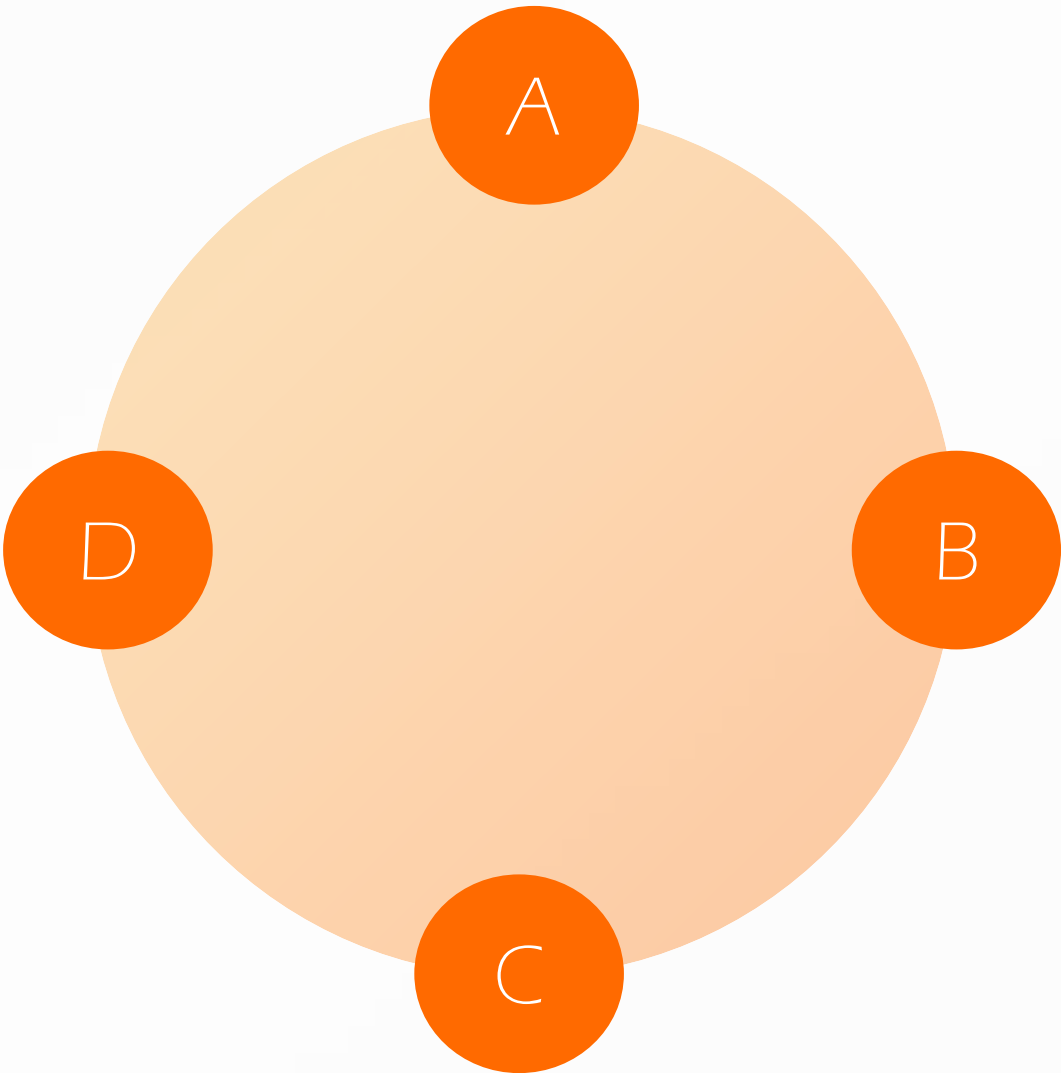
SELECT * FROM user WHERE country LIKE 'U%'

Data on node D

User_id1	FR
User_id2	US
User_id7	FR

Index on node D

FR	User_id1	User_id7	...
US	User_id2



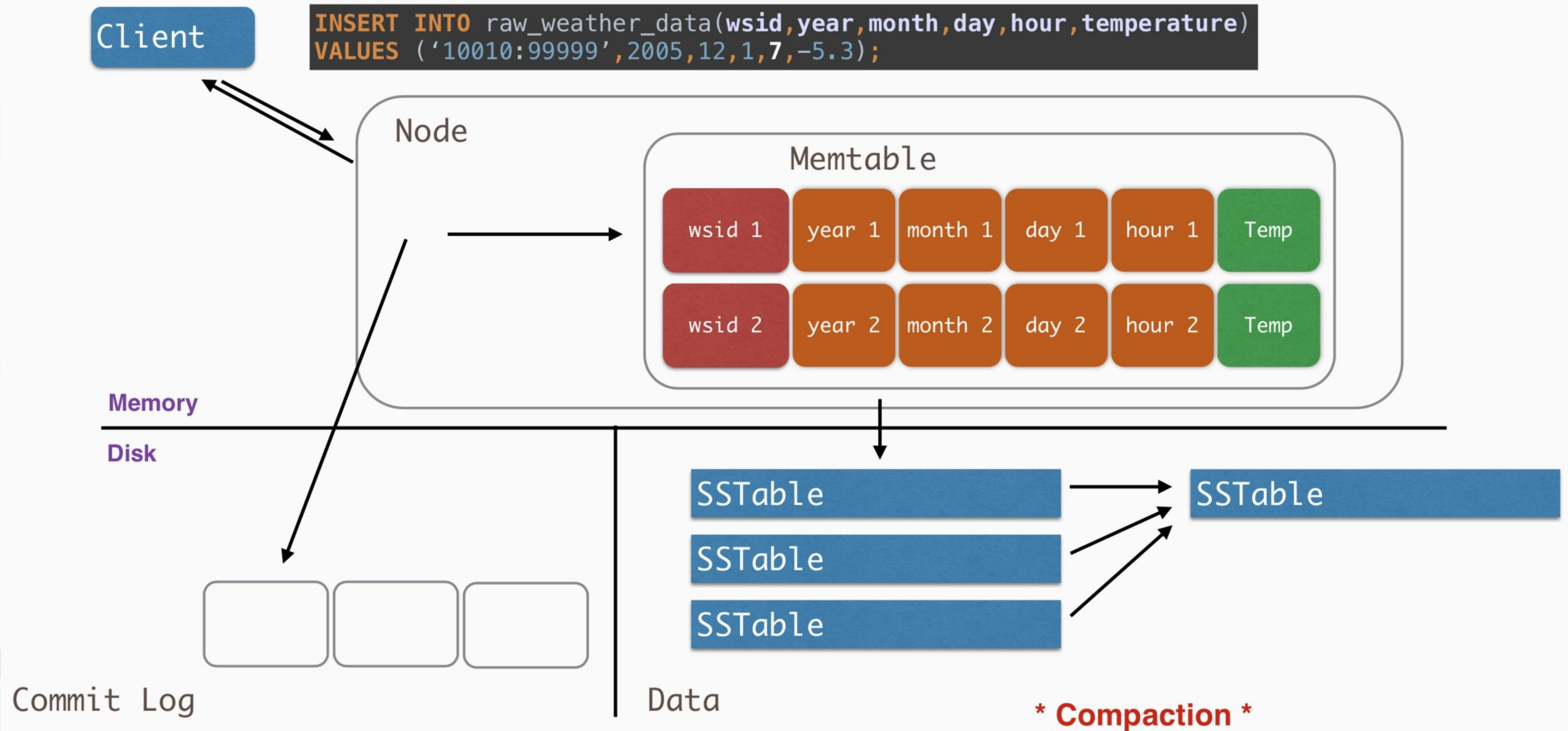
Data on node B

User_id4	US
User_id5	US

Index on node B

US	User_id4	User_id5	...
----	----------	----------	-----

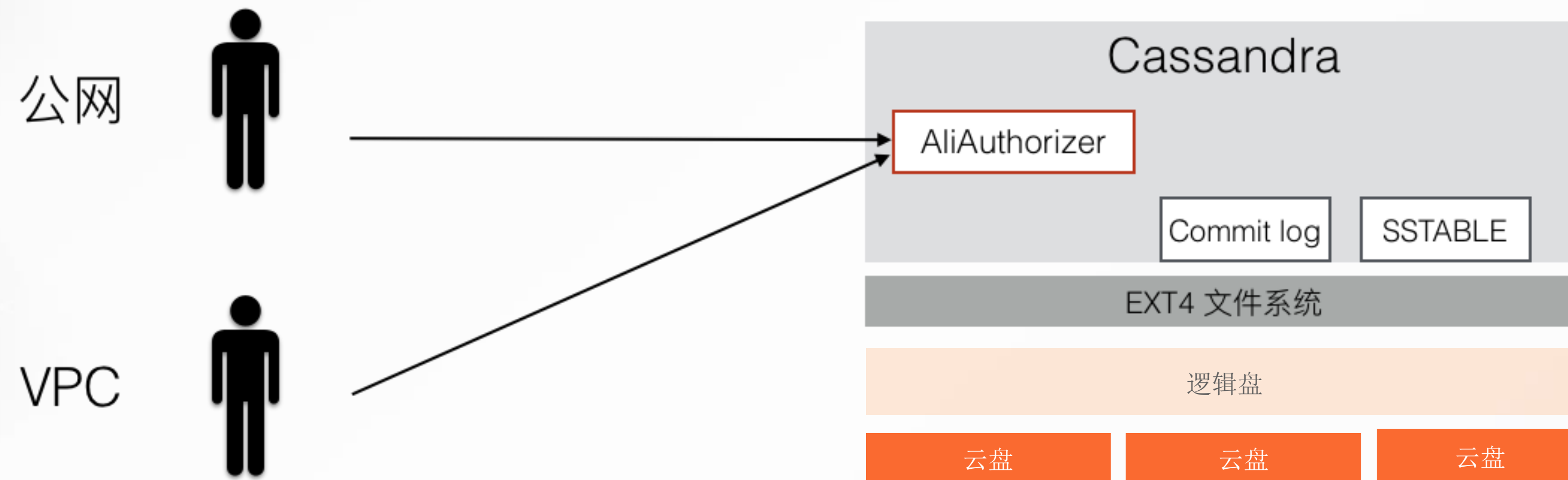
Write Path



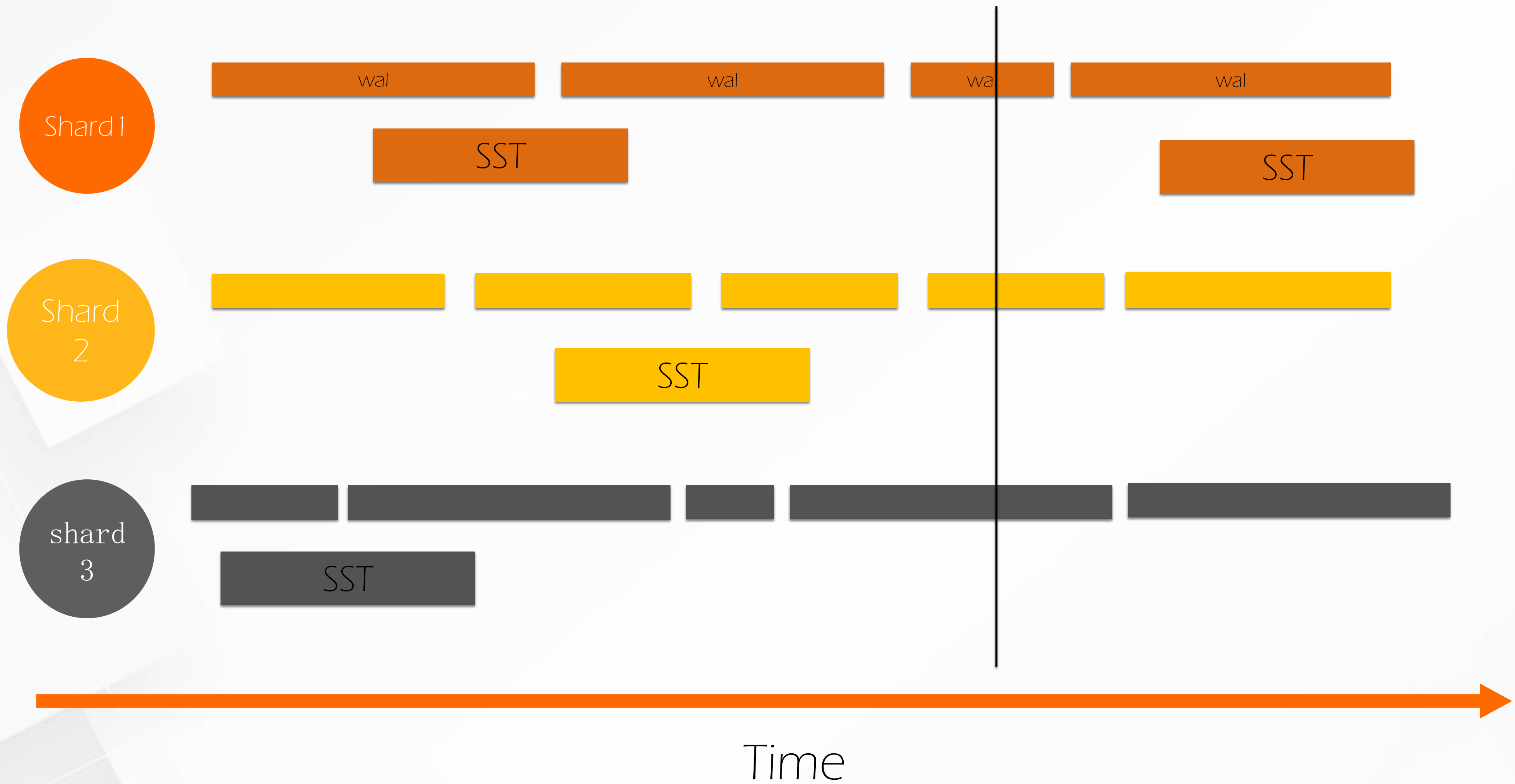
阿里云Cassandra特性

阿里云Cassandra-云化点

- 条带化管理磁盘：联合复用多盘能力，吞吐最高可提升22%
- 自研公网访问：多种网络环境vpc, 公网, Server 自动切换路由链路，开源客户端无需改动
- 自研分级安全插件：在原有super user上，多一级system user，保护云上核心资源不被篡改

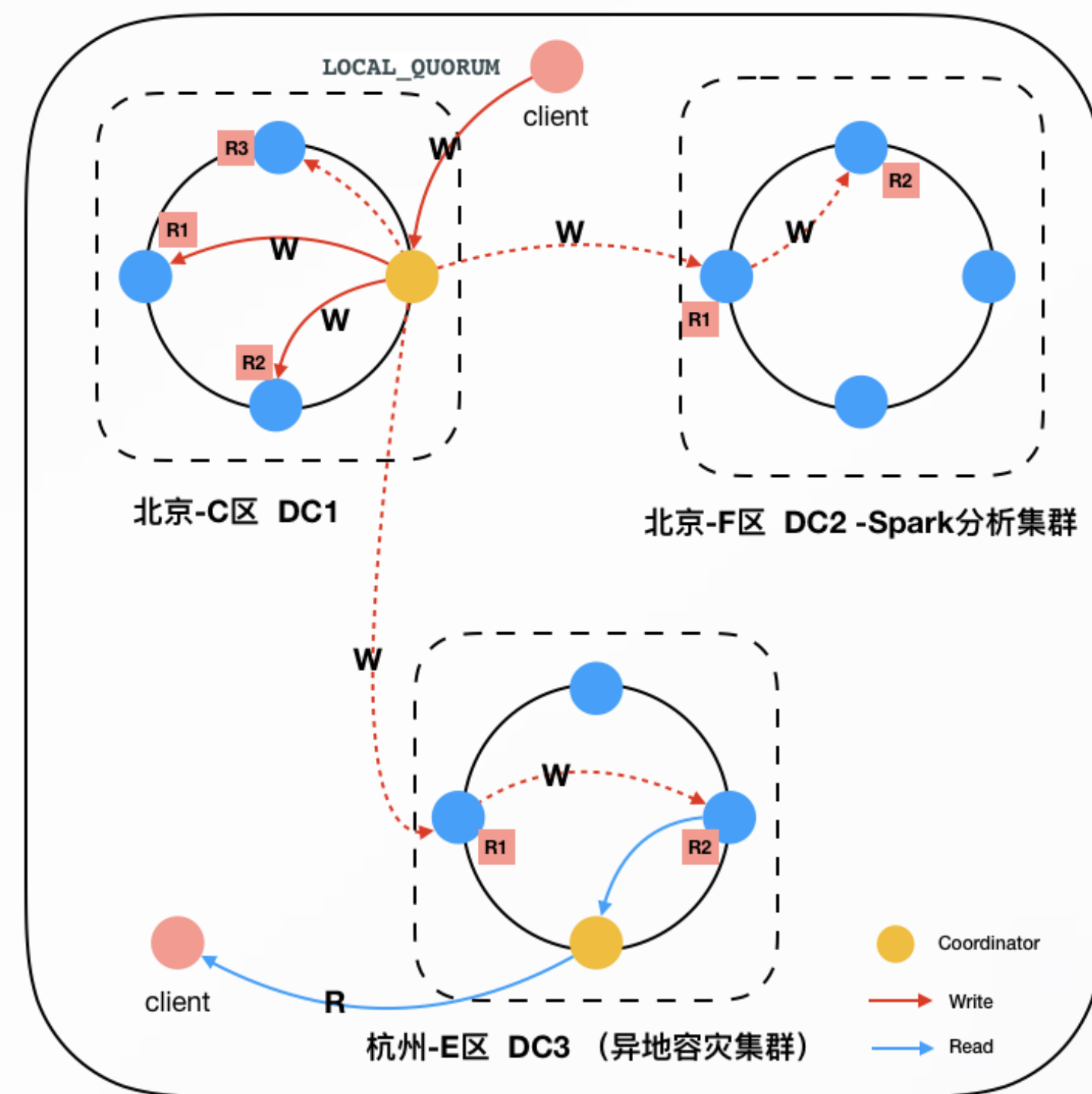


备份恢复



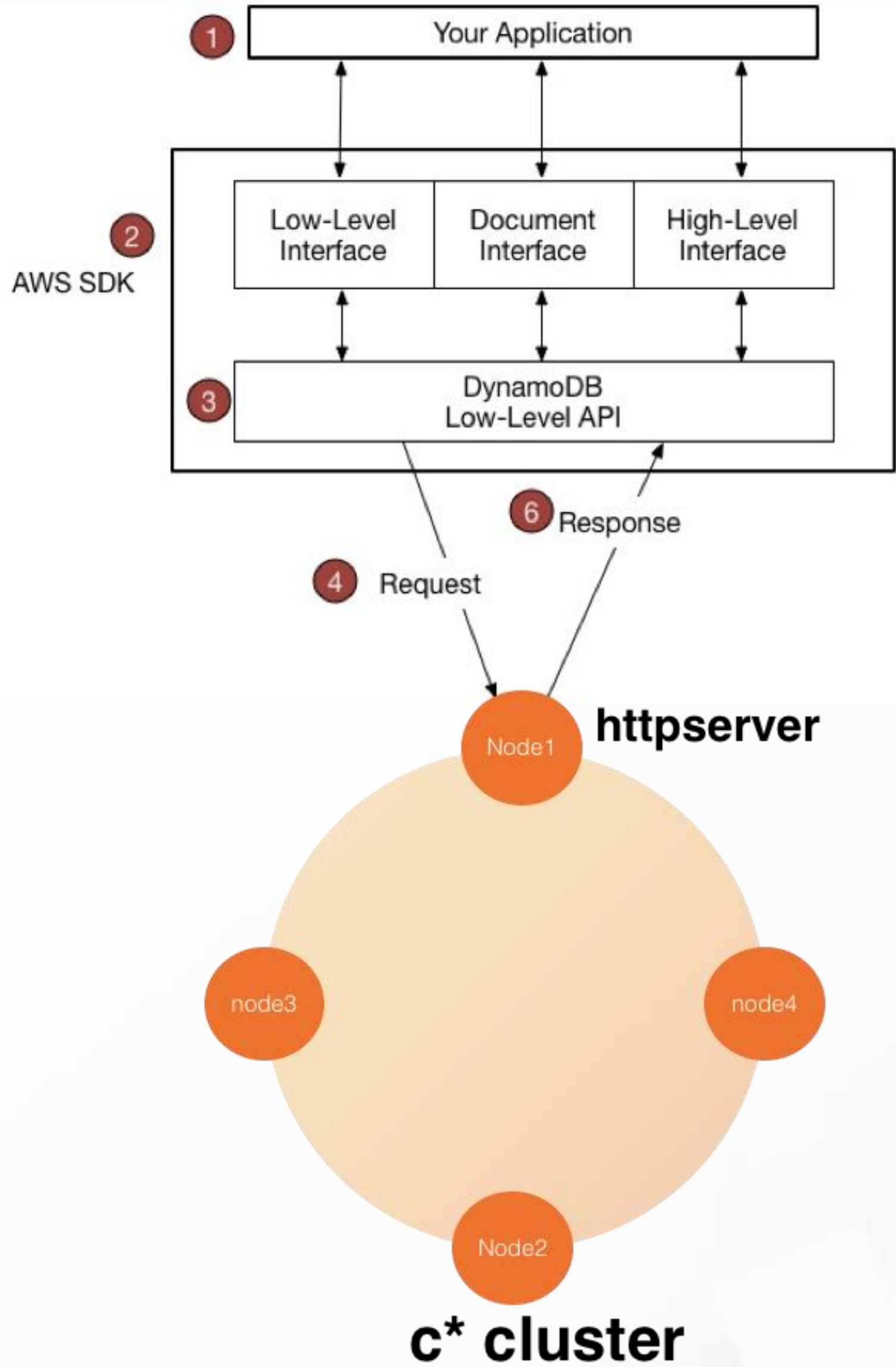
多地多活（待上线）

- 构建全地域集群：
一键购买 - 自由搭配
引擎内在支持异地及同城同步
- CREATE KeySpace IF NOT EXISTS test WITH REPLICATION =
{ 'class': 'NetworkTopologyStrategy', 'beijingC' : 3, 'beijingF':2,
'hangzhouE':2};
- 另一个DC，可用于spark分析

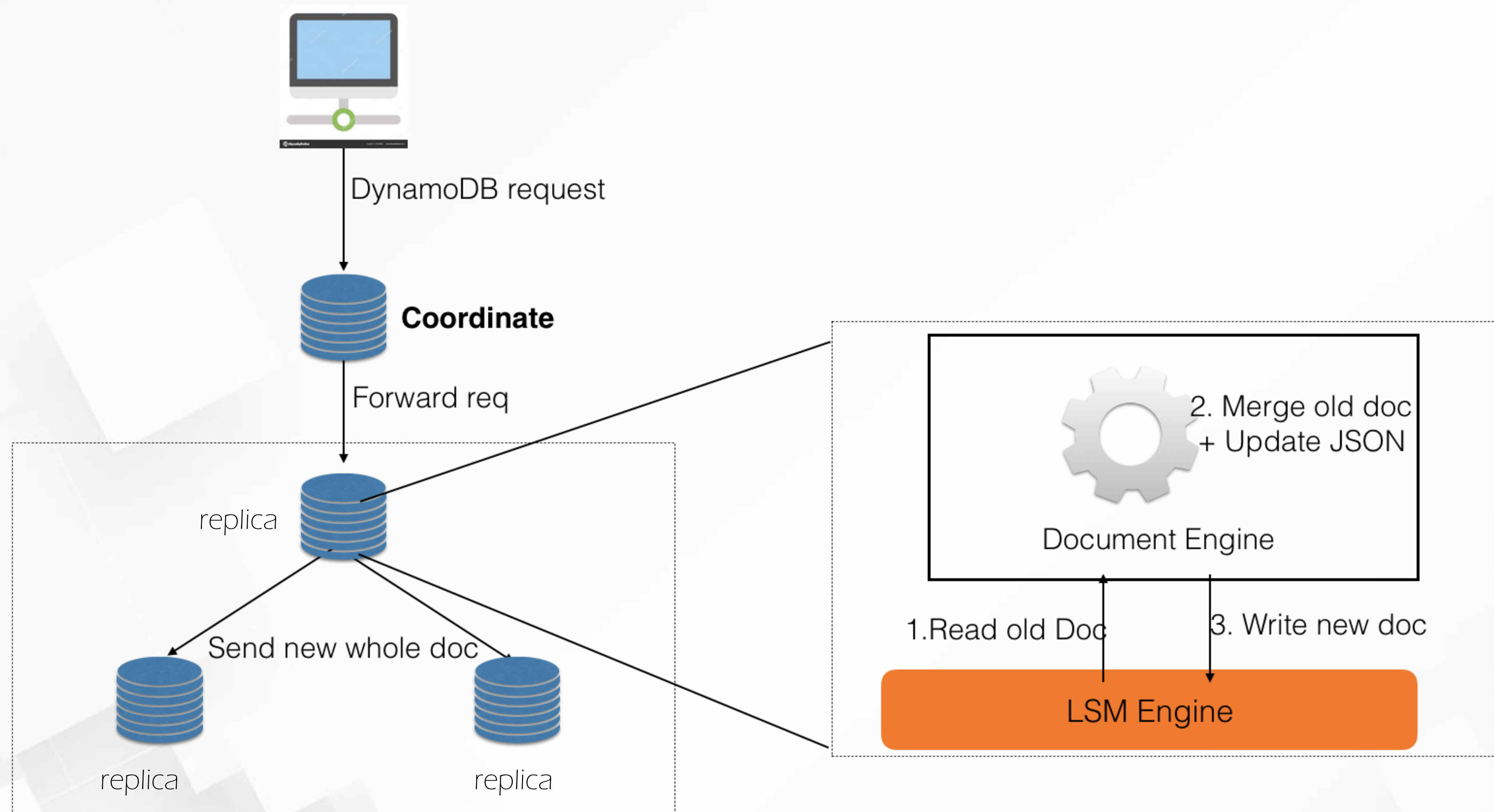


兼容dynamodb

事项	cassandra	说明
key	hashkey+sort key	表模型对应
简单类型	直接映射	S、N、B、BOOL、NULL SS、NS、BS
复杂嵌套类型	自定义DocNode	支持嵌套子文档
本地索引	SASI	cassandra自带
全局索引	MV	cassandra自带
单行事务	Paxos	cassandra自带
迁移	-	迁移工具一键迁移 DynamoDB

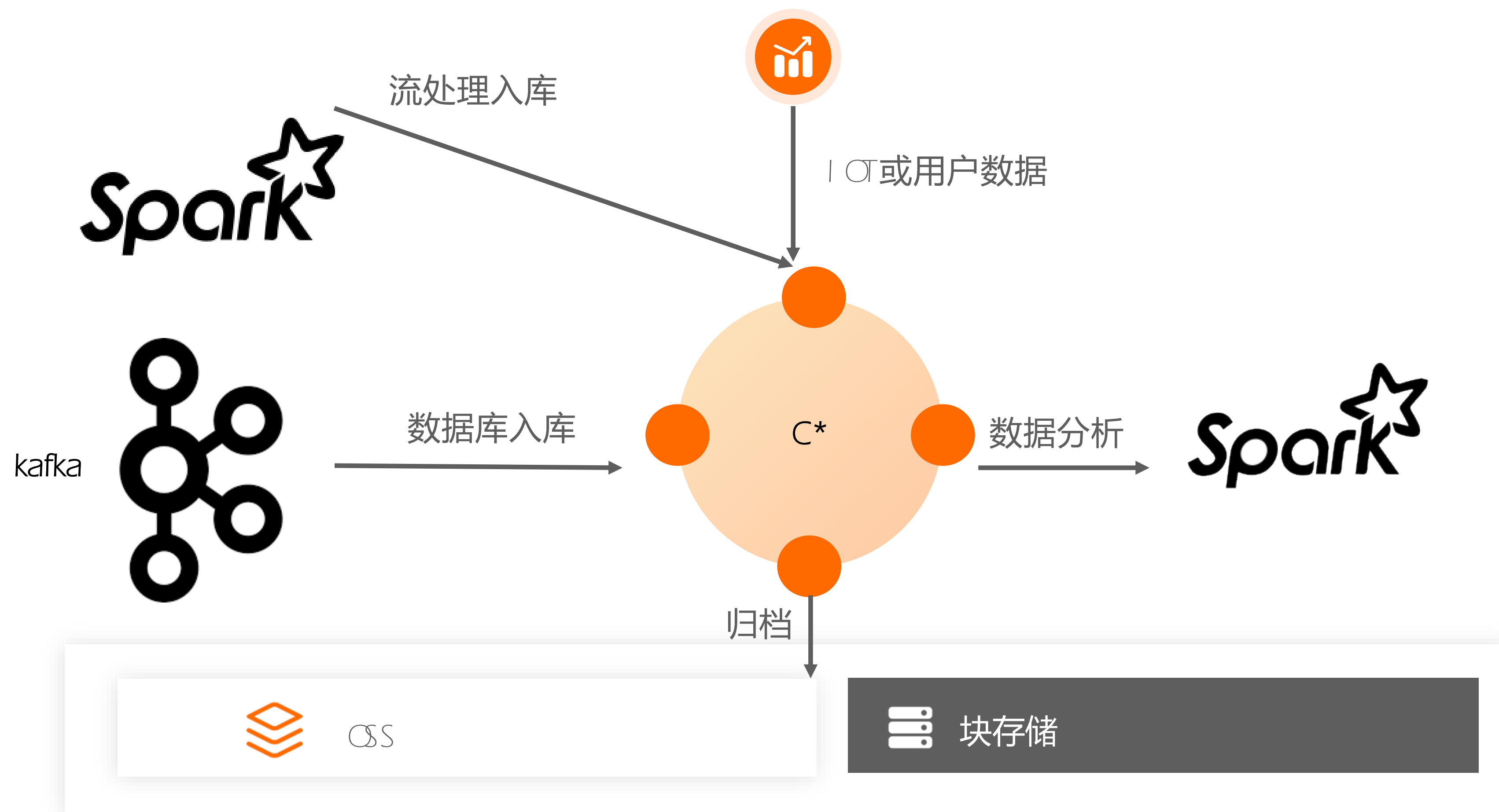


兼容dynamodb-I0路径



对接大数据生态

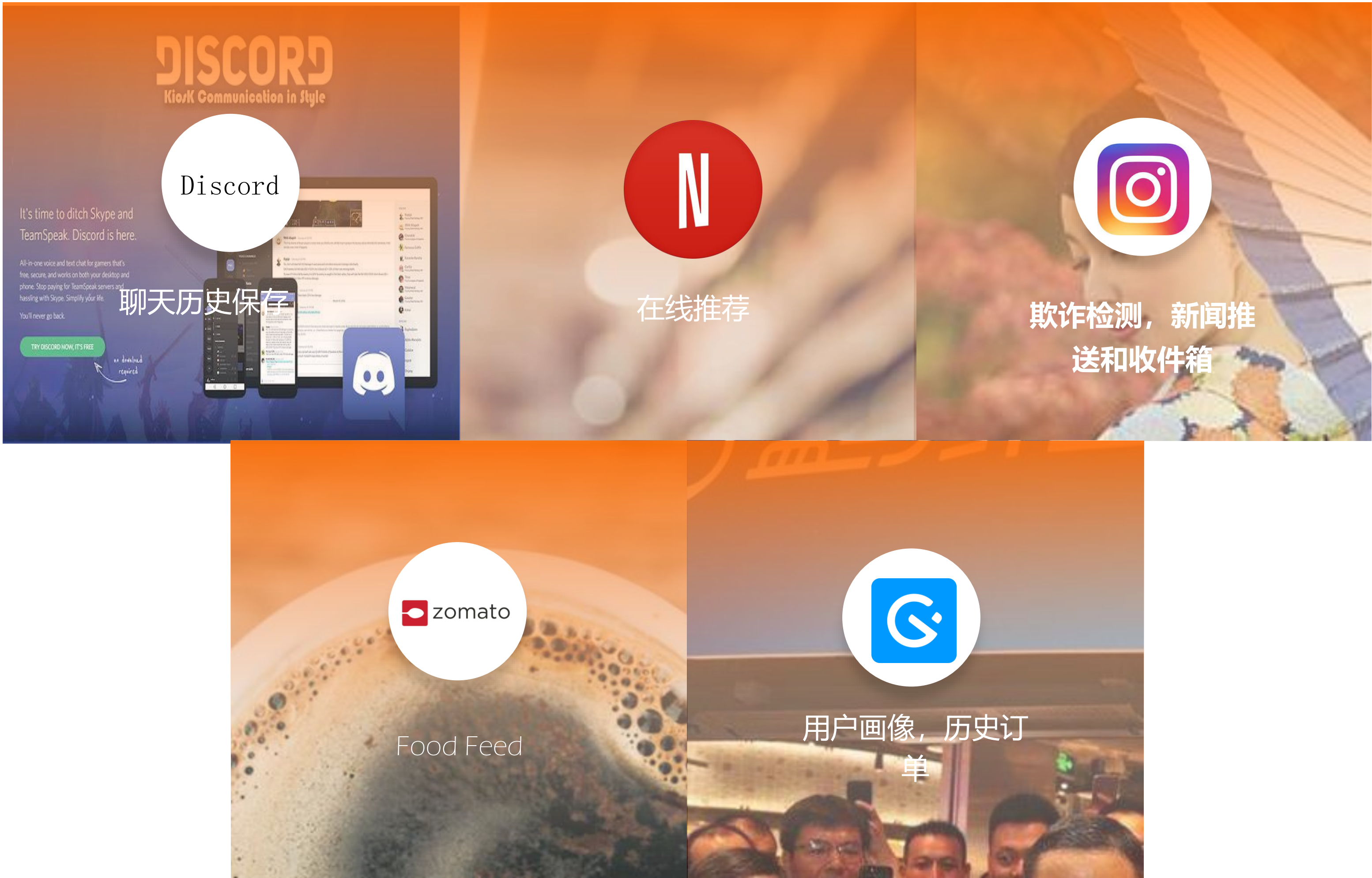
打通大数据生态上下游：kafka - cassandra - spark



后续规划

- 分级存储
- SearchIndex, 支持全文搜索
- 支持社区 4.0 & Incremental Repair
- 支持rocksdb引擎
- 添加事务

业界使用案例



Discord 背景 & 需求



高性能

强类型, 非raw二进制
读写比 1: 1



低维护成本

技术成熟, 经过验证



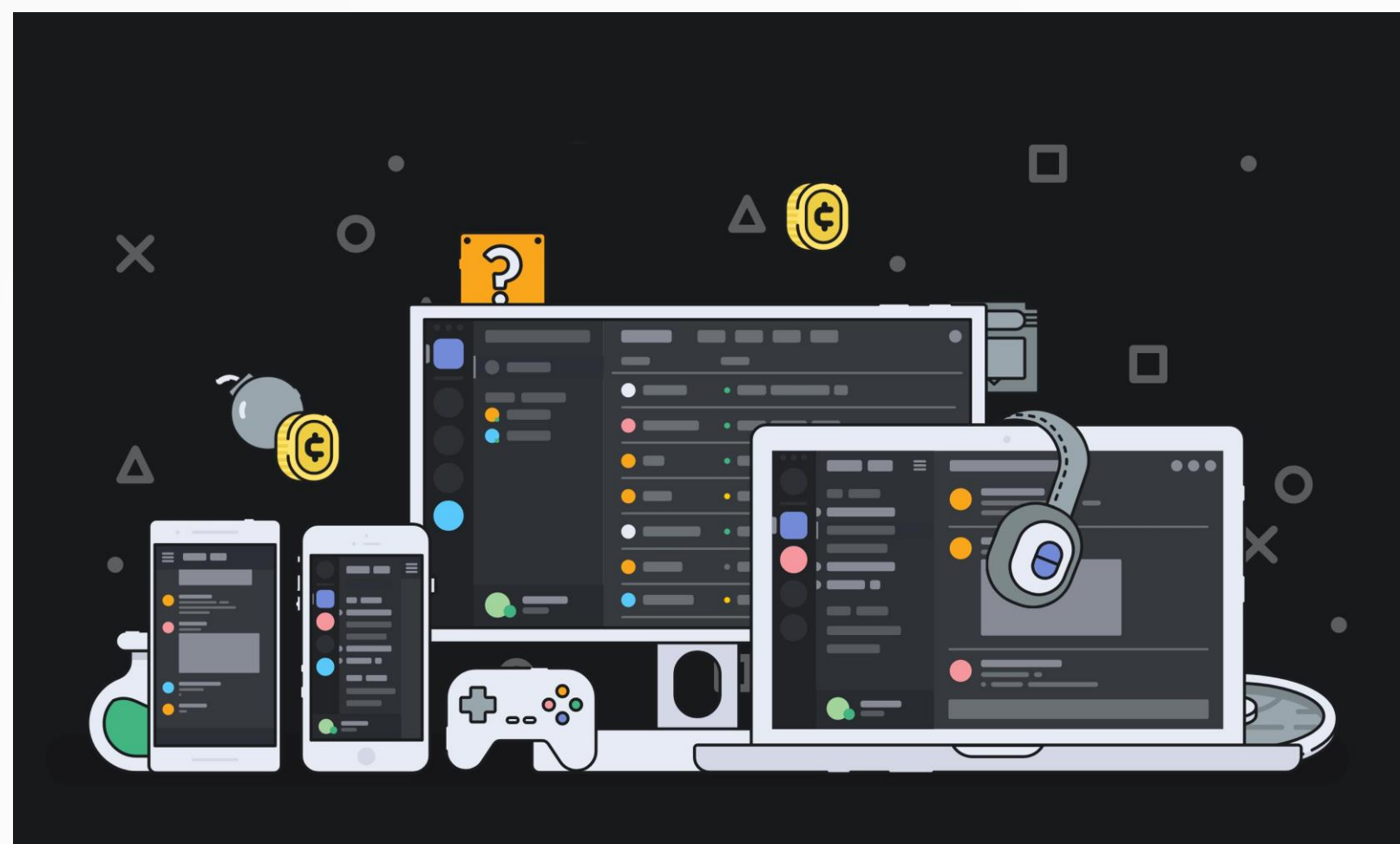
线性扩展

自动故障转移 failover



开源可控

方案及效果



一周内

C* 替换mongodb



RT稳定

保持着良好的性能
和稳定性

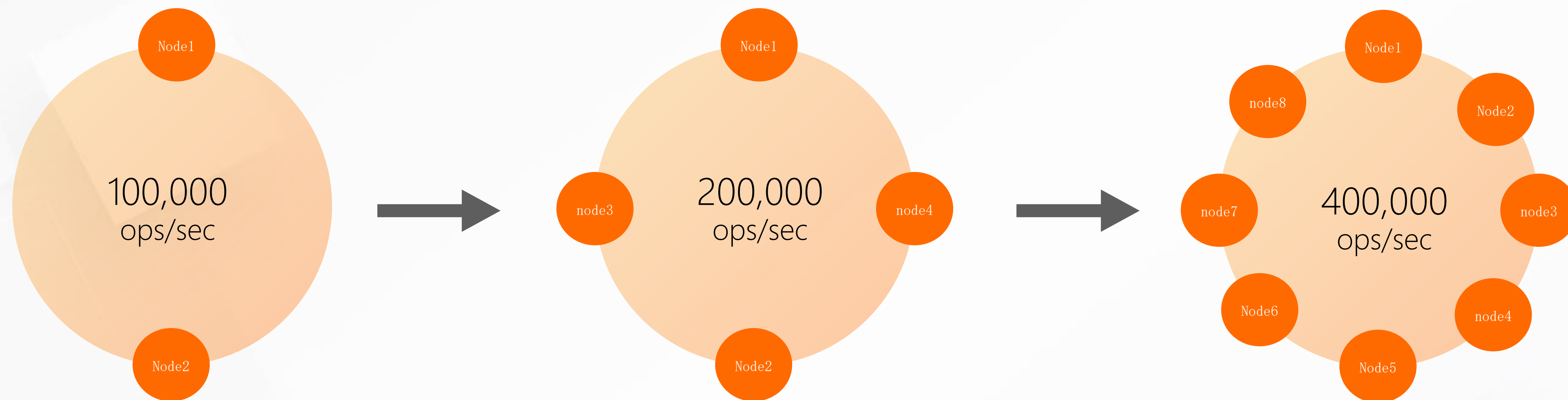


一亿增长至 >1.2亿

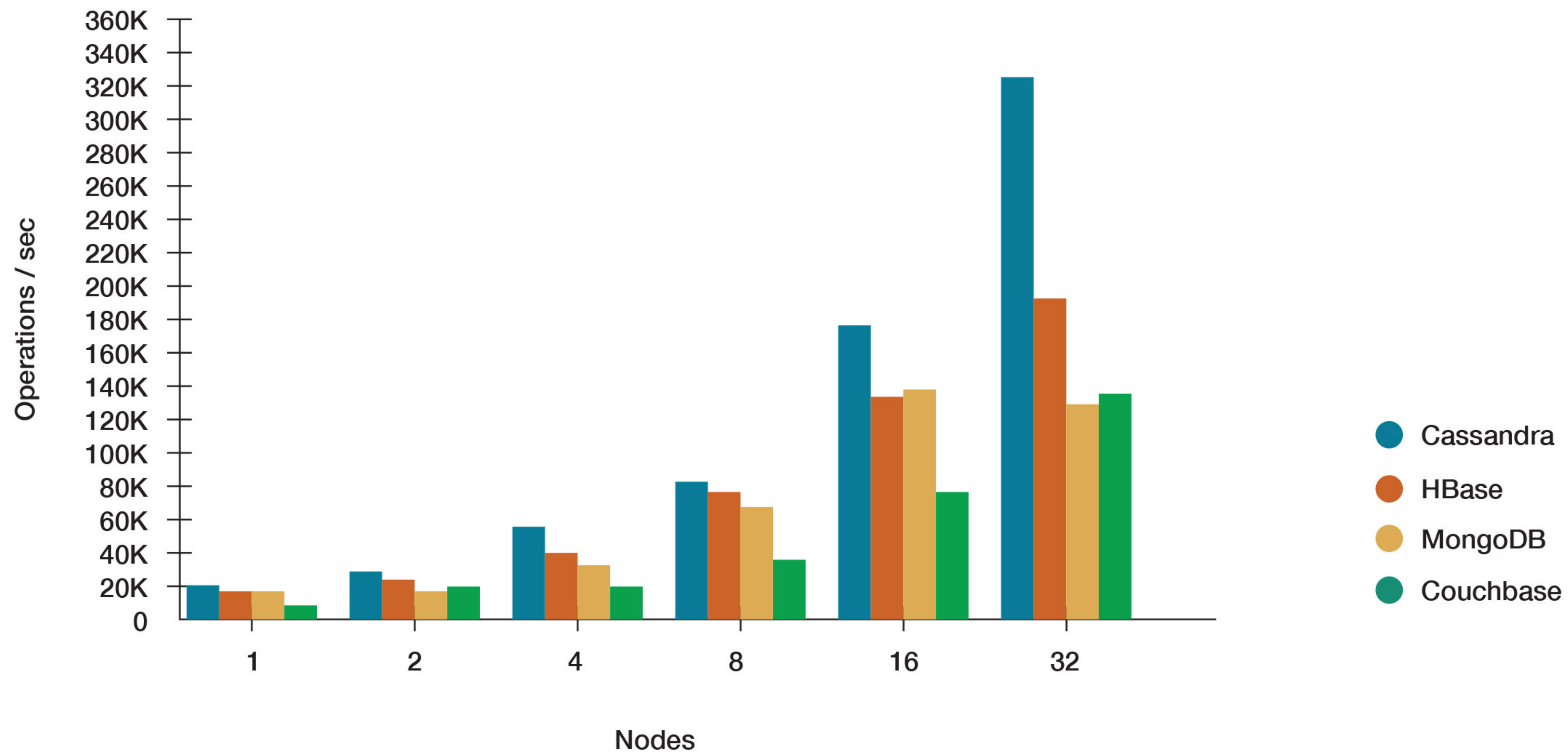
1天内聊天消息

总结-为什么选择Cassandra

线性扩展 (1...N)



性能对比



摘自: <https://www.datastax.com/nosql-databases/benchmarks-cassandra-vs-mongodb-vs-hbase>



功能强大

Cql/强大二级索引
/MV/LWT/UDF/CDC
，更像真正的数据库



开发成本

类sql的cql语言，多
语言driver



低成本

部署简单，运维成本
低，架构简洁，无浪
费



开源可控

开源成熟，可控，商
业公司支持



阿里云开发者社区

扫码加入社群
与志同道合的码友一起
Code Up



Cassandra钉钉技术社区

谢谢！