msup

分布式多模数据库设计与实现浅析

讲师简介

简怀兵
字节跳动 架构师

程序员，先后在YY、腾讯、唯品会工作，现在在字节跳动参与基础架构相关工作。分布式系统和数据库技术爱好者。

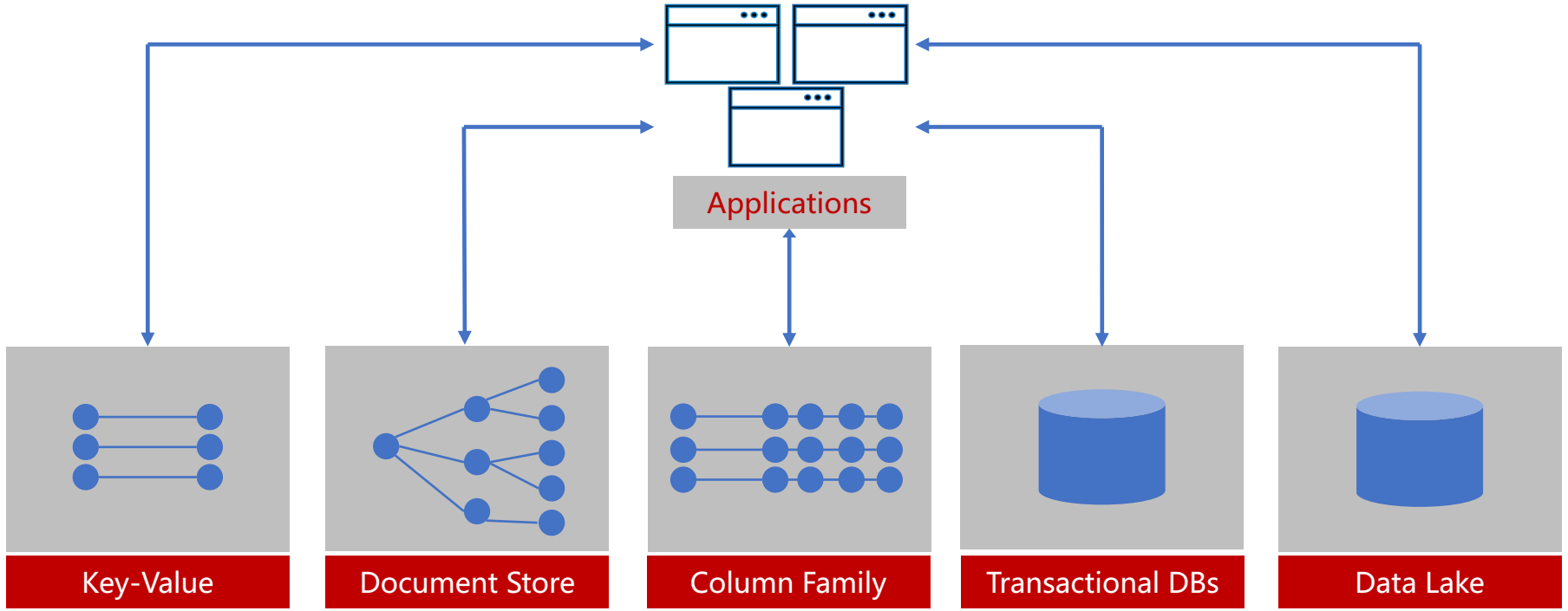
目录

业务价值

关键技术

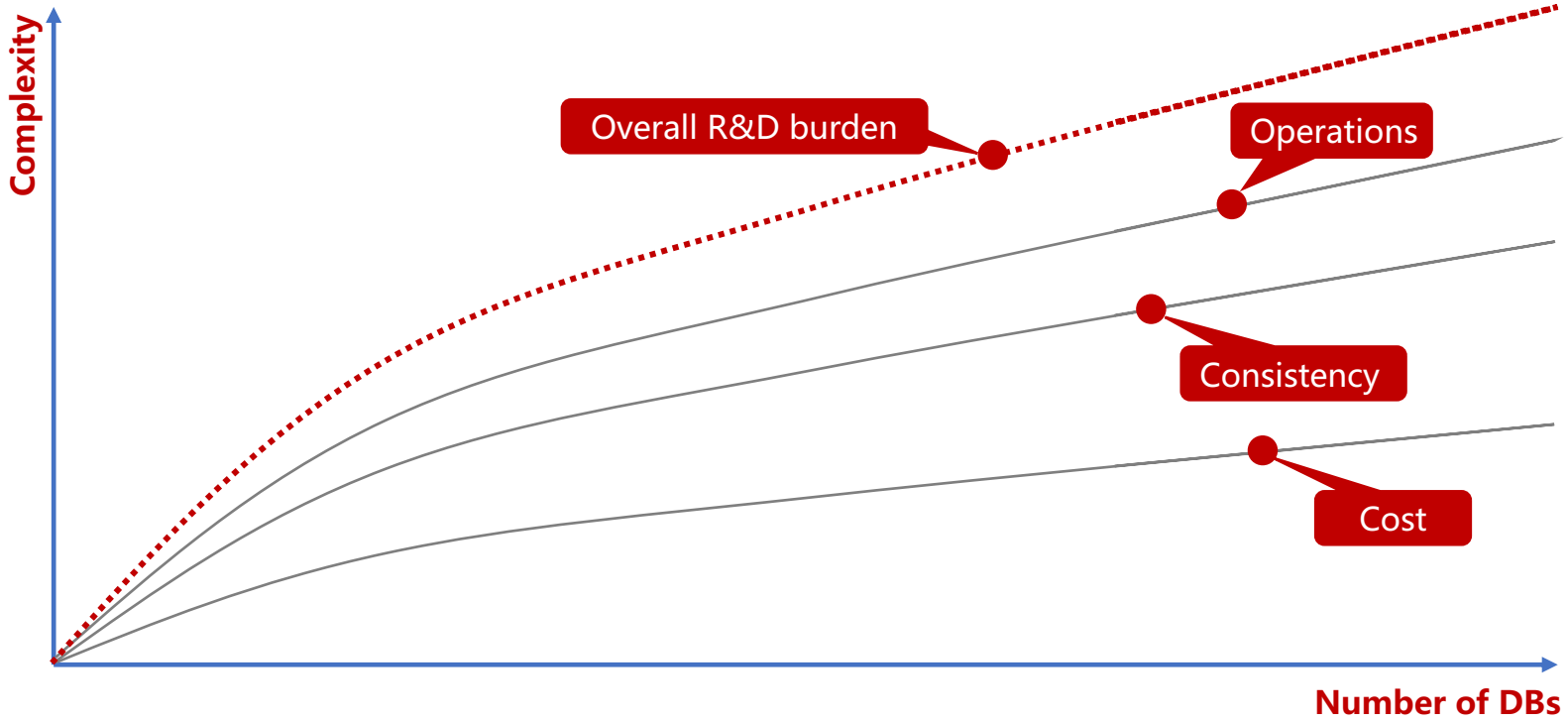
趋势展望

背景导读 – Polyglot Persistence

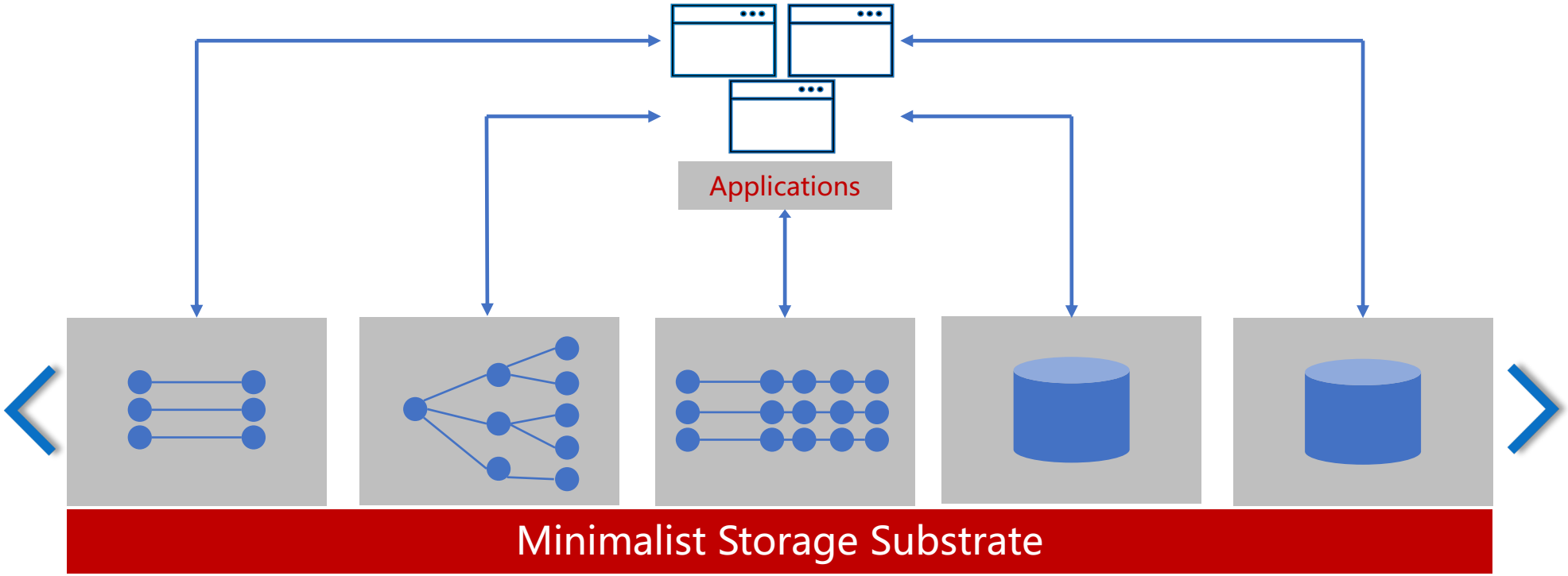


Polyglot Persistence

背景导读 – Overall R&D burden



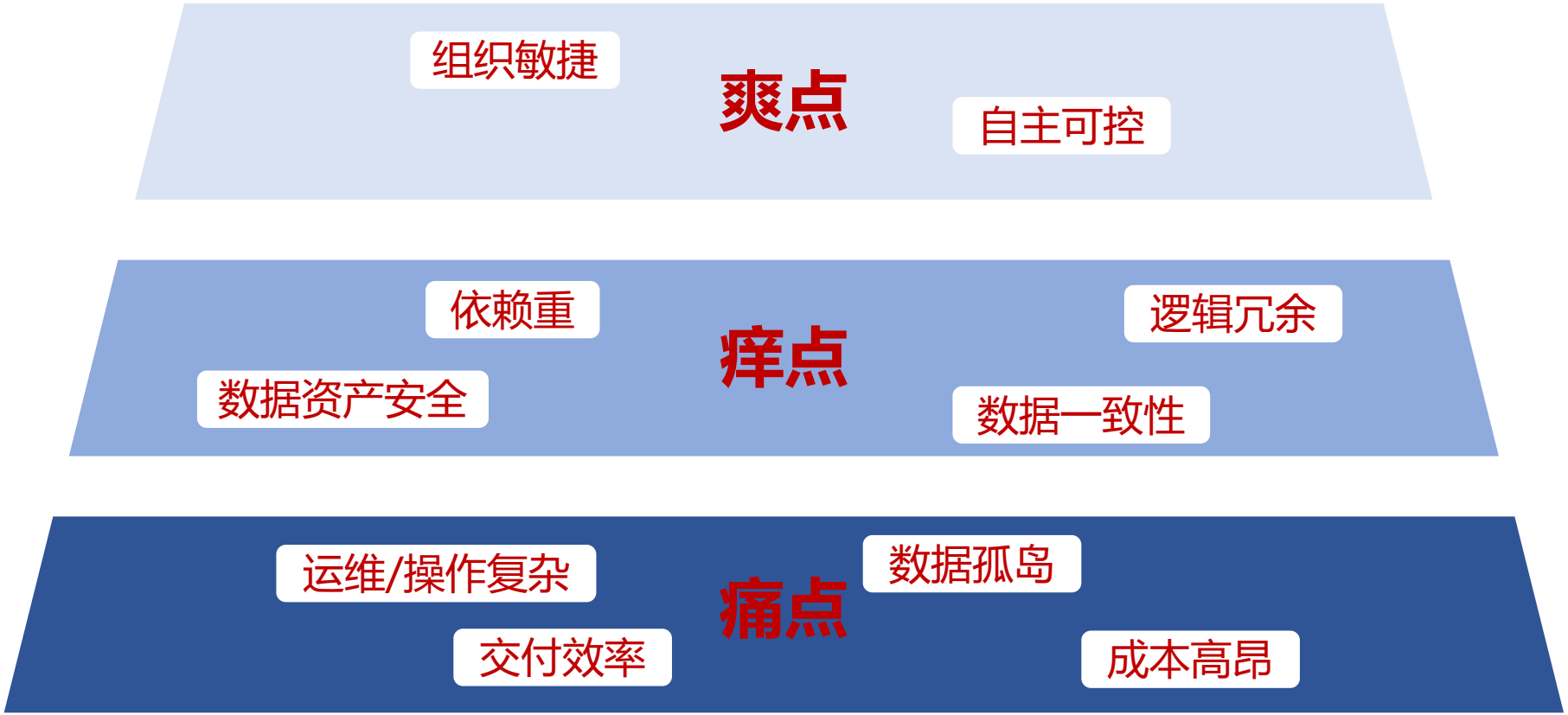
背景导读 – Multi-Model



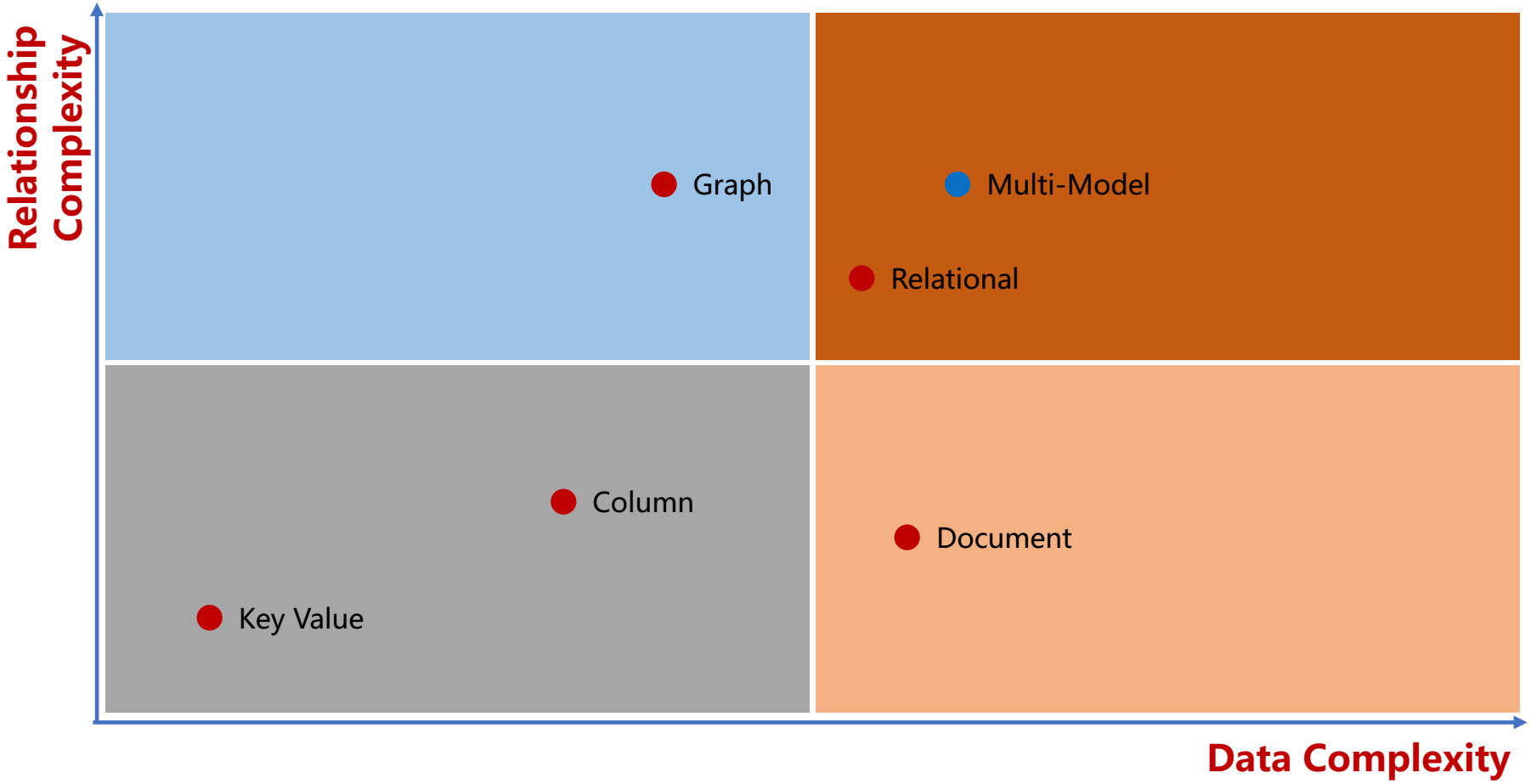
From Polyglot Persistence To Multi-Model

目录

业务价值 – 显性解决问题

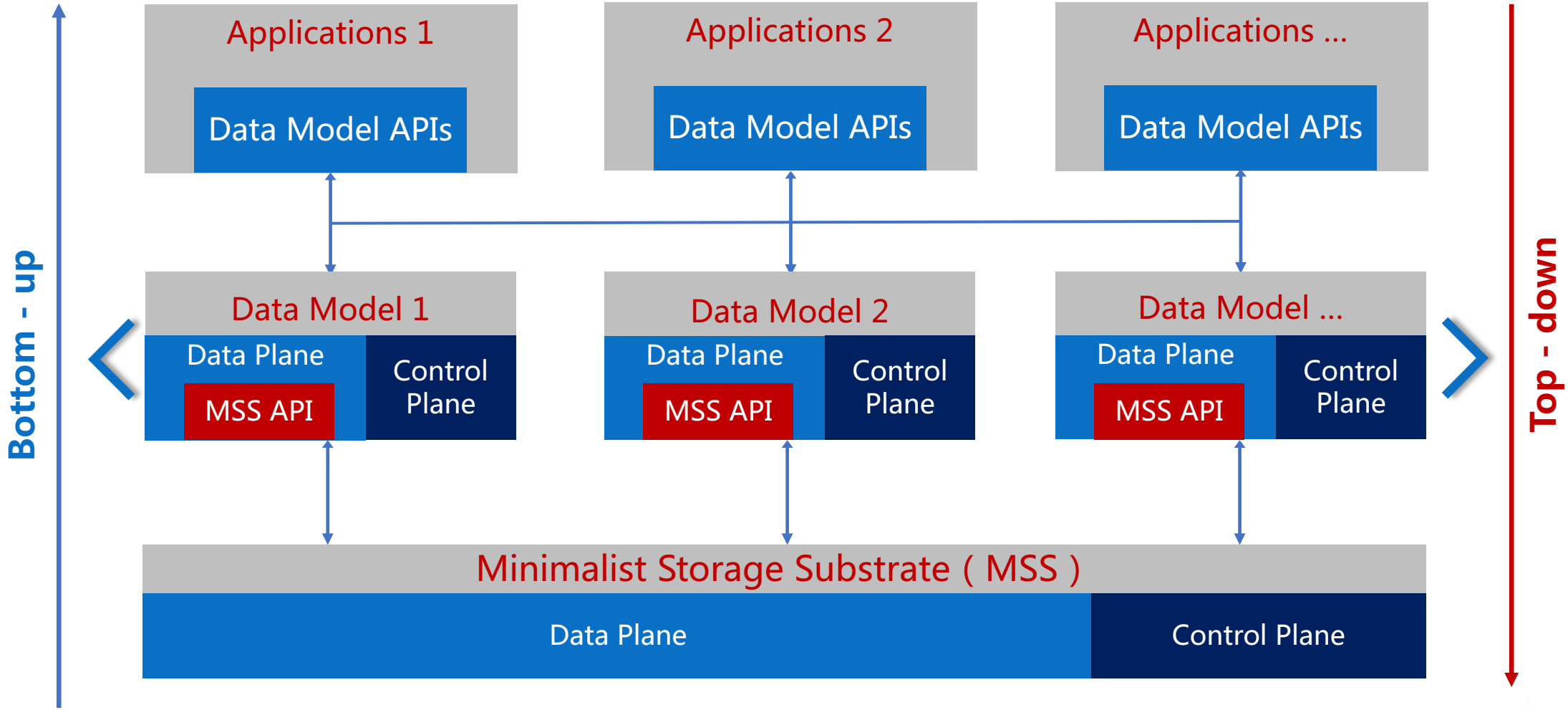

组织敏捷
爽点
自主可控
依赖重
逻辑冗余
痒点
数据资产安全
数据一致性
运维/操作复杂
数据孤岛
痛点
交付效率
成本高昂

业务价值 – 隐性对抗 Complexity



目录

关键技术 – Open Layer Architecture



关键技术 – 架构原则

msup

Open Layer Architecture

按技术域封闭，按访问场景开放

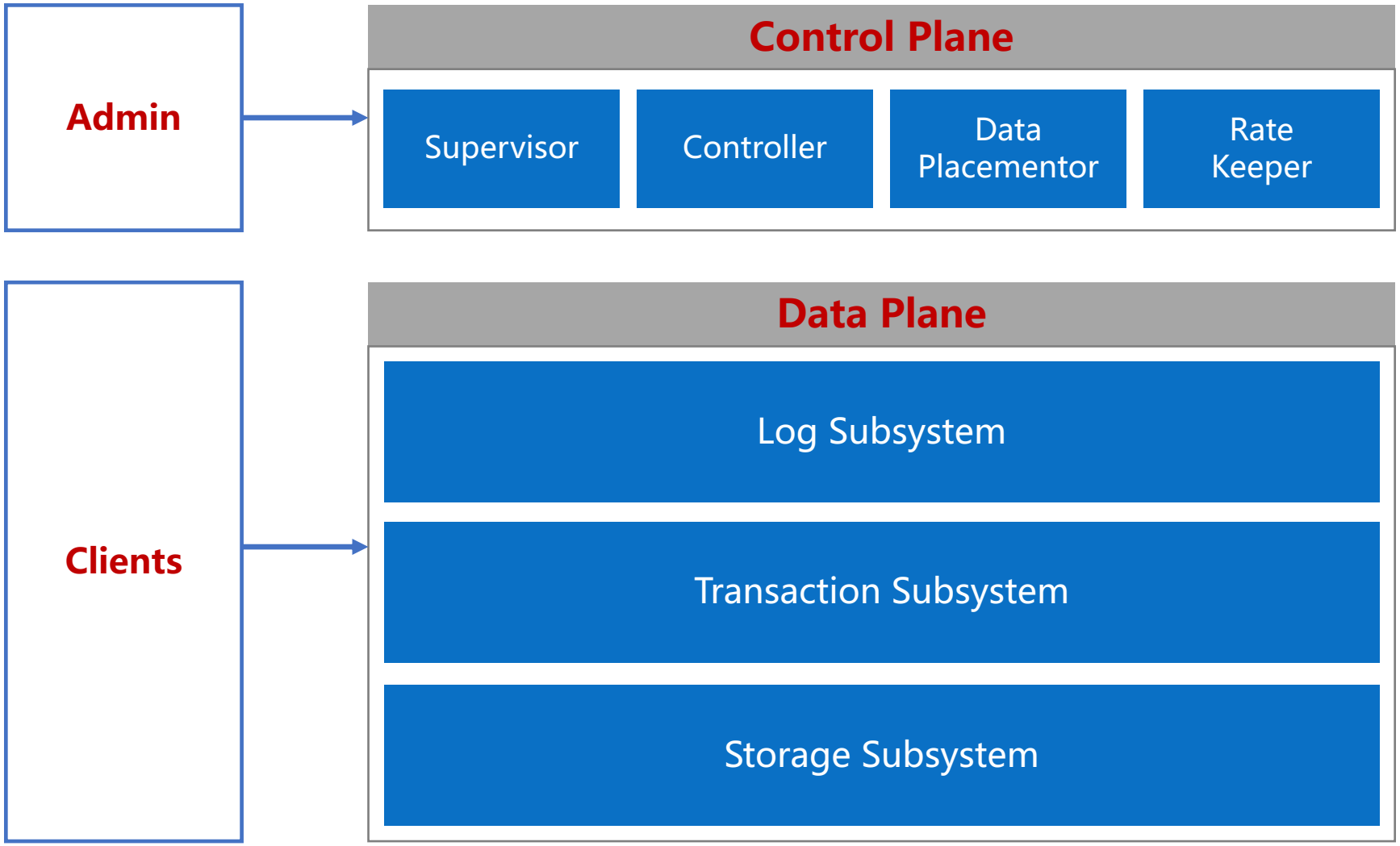
Control / Data Plane Separation

控制面和数据面分离，系统弹性、开放性与服务能力同等对待

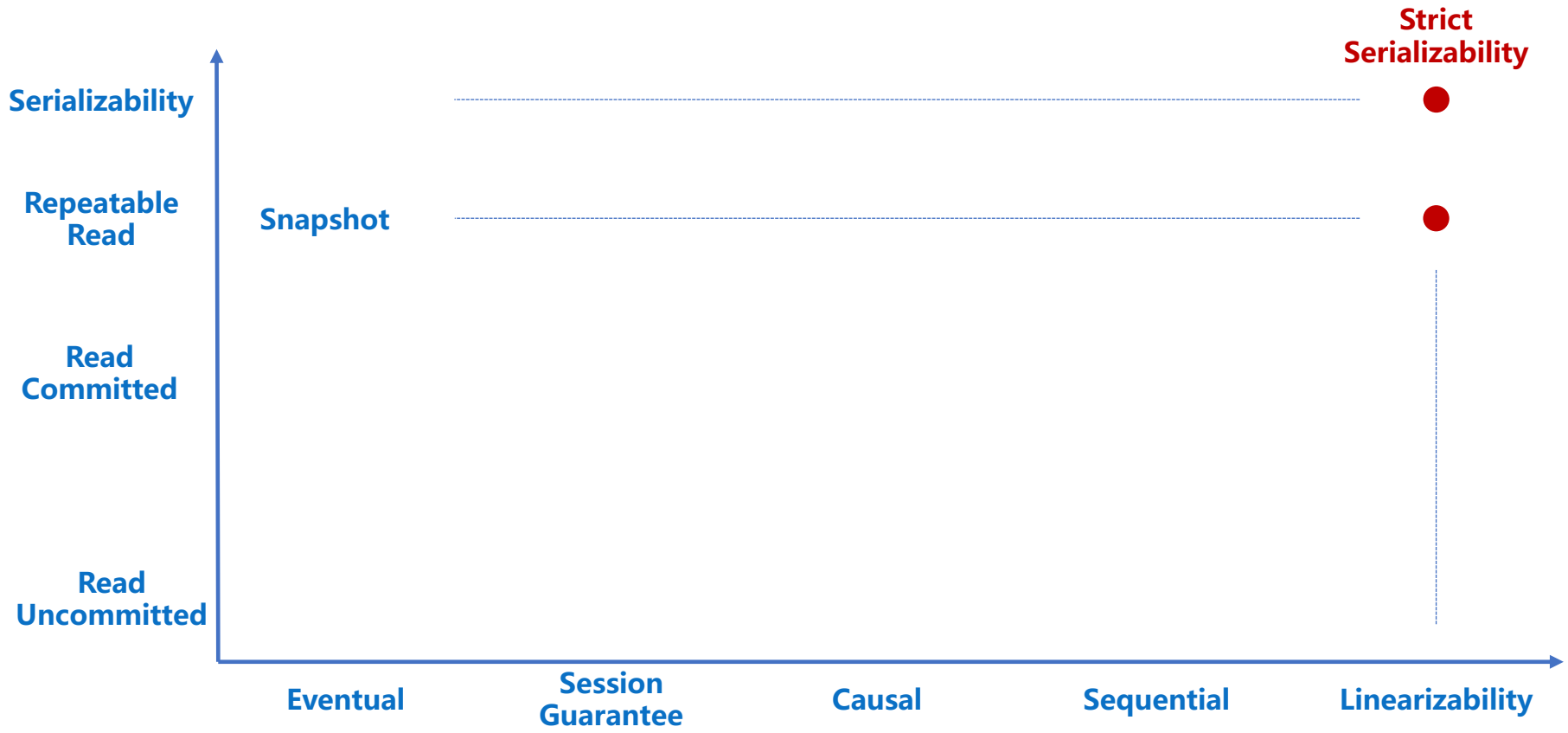
Cloud Native Design

从计算/存储分离到计算/状态分离，轻松构建/迁移至三方云

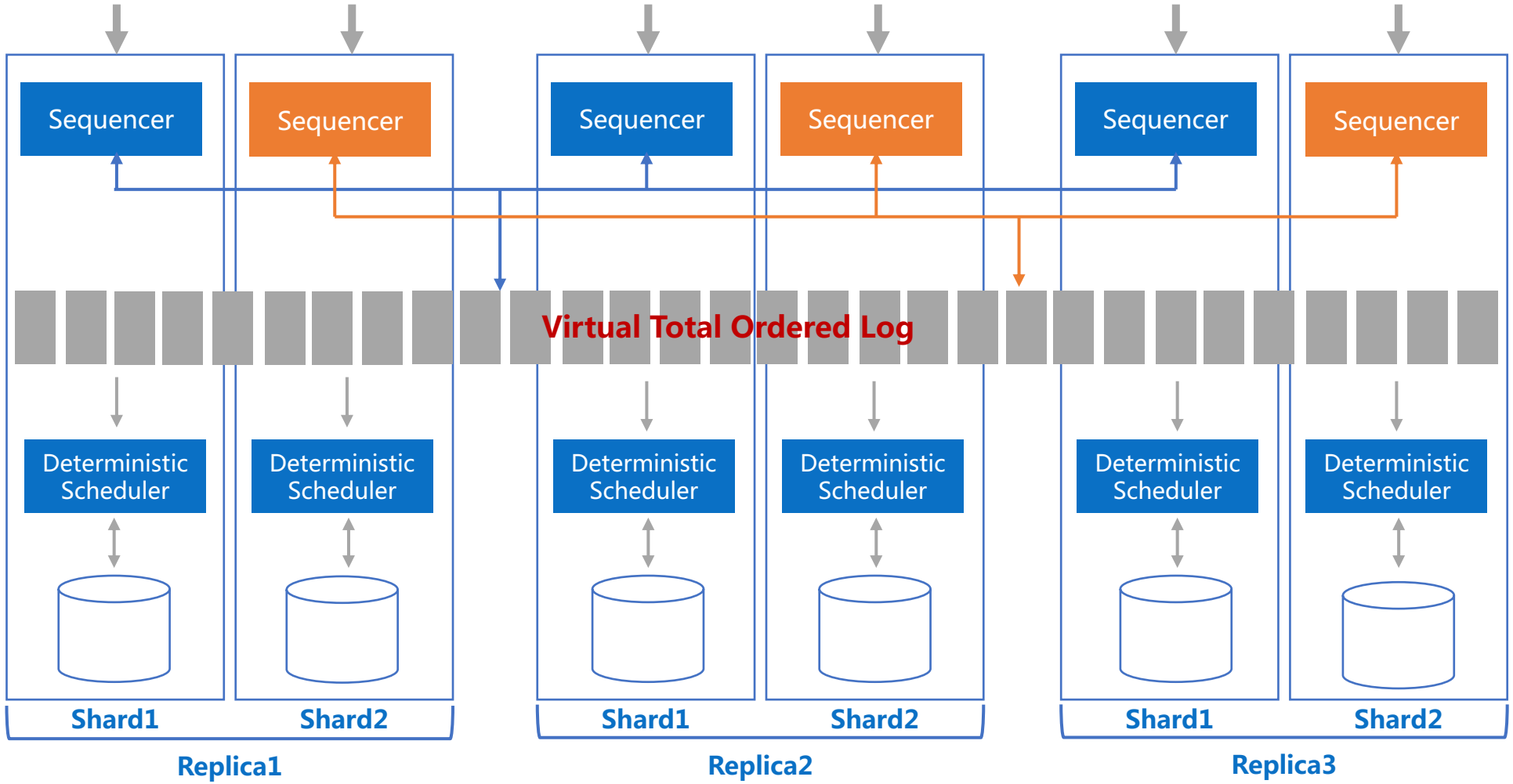
关键技术 – Minimalist Storage Substrate



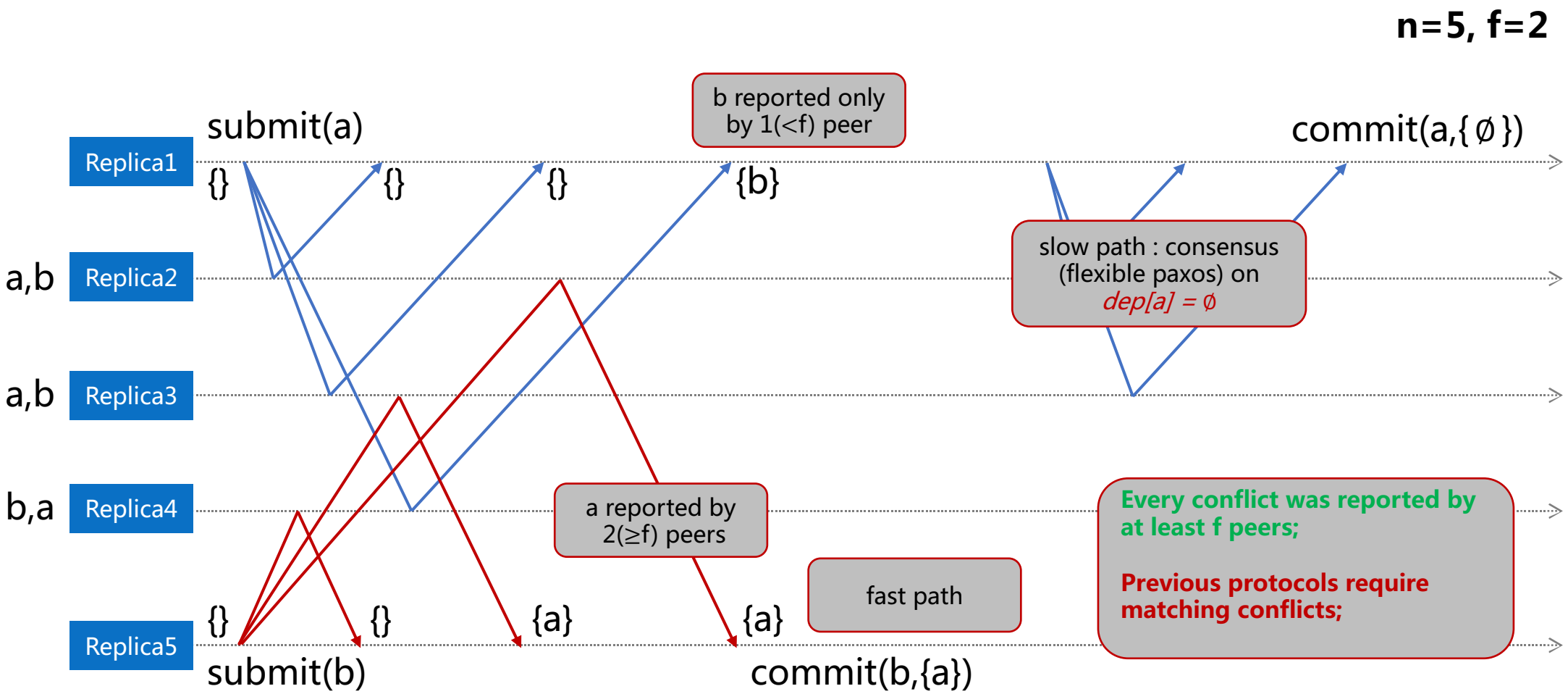
背景导读 – MSS Distributed Transaction Isolation



关键技术 – MSS Distributed Transaction – 基于 Calvin



关键技术 – Optimal Leaderless Sync Rep （ Atlas ）



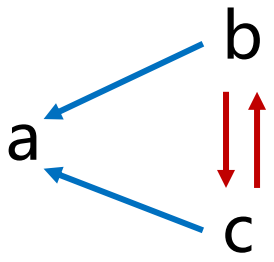
来源《State-Machine Replication for Planet-scale Systems 》

关键技术 – Optimal Leaderless Sync Rep (Conflict)

commit(a,{∅})

commit(b,{a,c})

commit(c,{a,b})



Compare logical timestamp

If b < c then execute(a), execute(b), execute(c)

If c < b then execute(a), execute(c), execute(b)

来源《State-Machine Replication for Planet-scale Systems 》

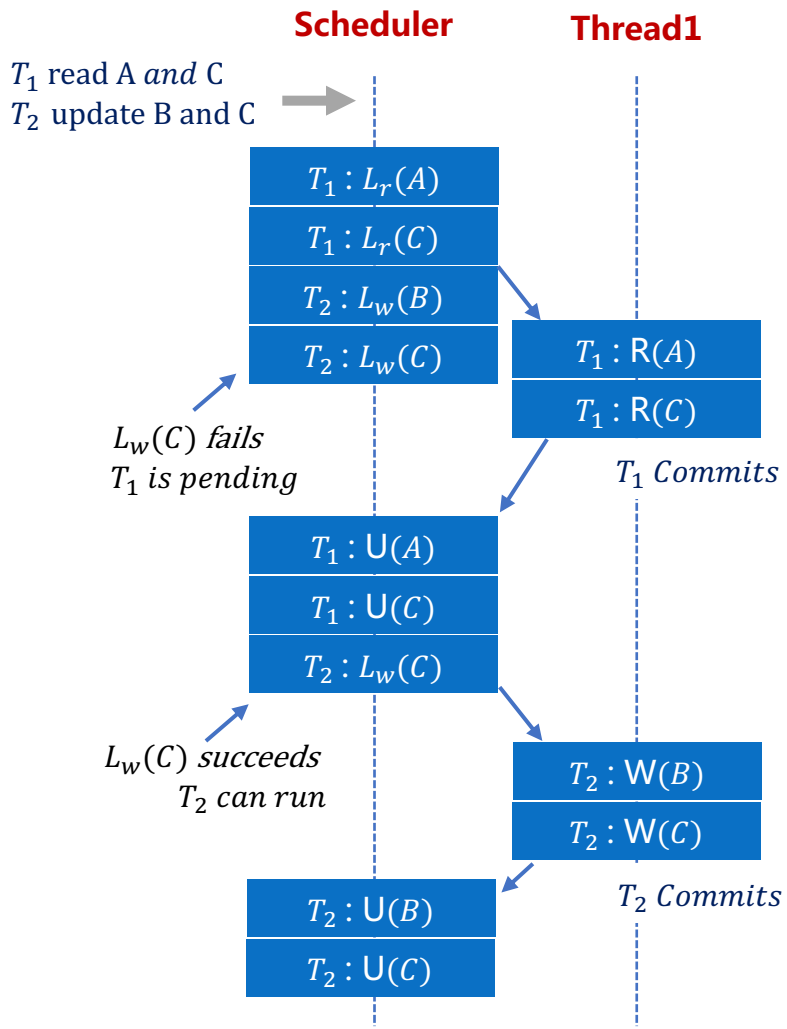
关键技术 – Distributed Transaction Execution

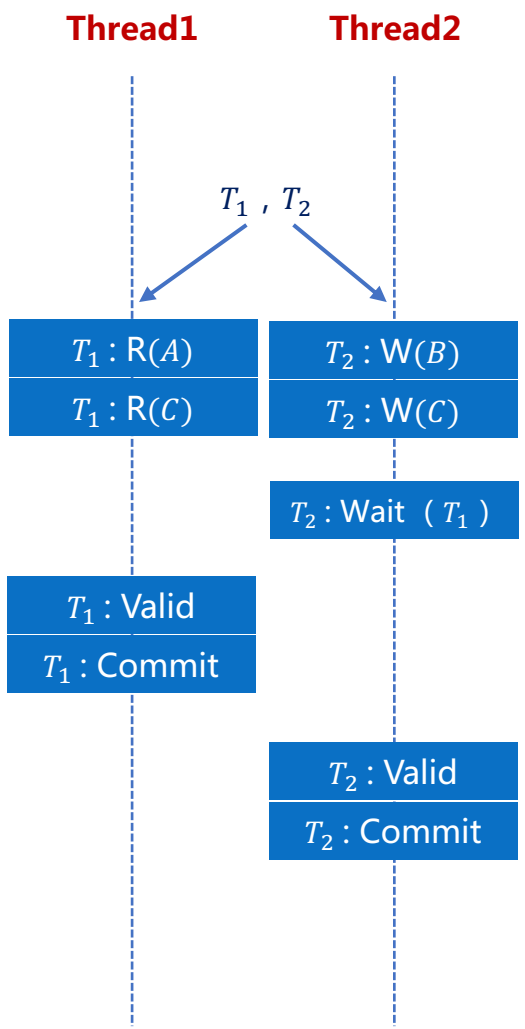
T1:
A = A + B;
C = C + B;

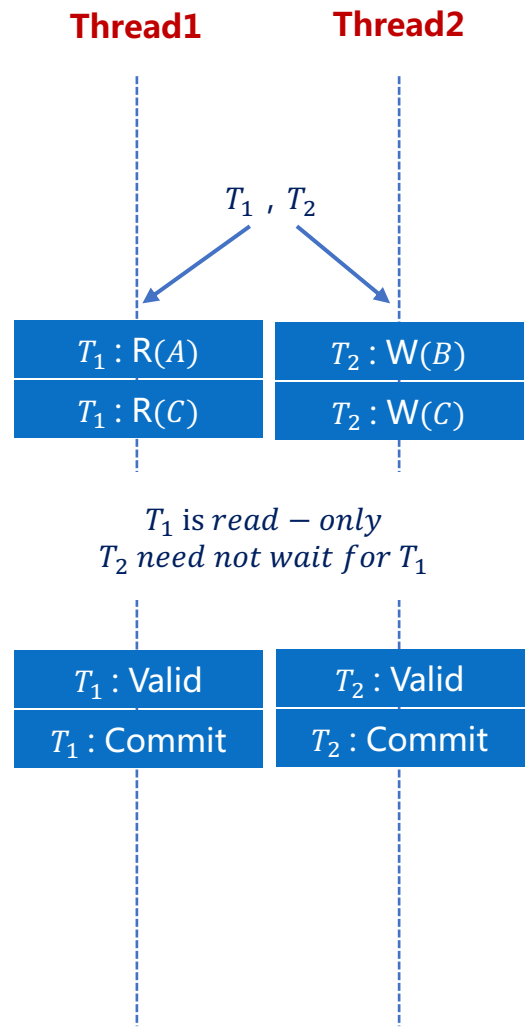
Analyze RS/WS → Identify Role → Perform Local Reads → Serve Remote Reads → Collect Remote Items → Local Execute

	Local RS	Local WS	Identify Role	Perform Local Reads	Serve Remote Reads	Collect Remote Items	Local Execute
Deterministic Scheduler	{A}	{A}	Active	A := a		A := a; B := b;	P1 (A)
Deterministic Scheduler	{B}		Passive	B := b	Send B := b to P1 Send B := b to P3		P2 (B)
Deterministic Scheduler	{C}	{C}	Active	C := c		C := c; B := b;	P3 (C)

关键技术 – Deterministic Optimistic Concurrency Control







L_r : *Acquire read lock*; L_w : *Lock write lock*;
U : *unlock*; R : *read*; W : *write*

来源《Optimistic Transaction Processing in Deterministic Database》

关键技术 – Doc Data Model 示例

逻辑文档	在 MSS 中的存储

DocumentKey1={
SubKey1={
SubKey2=Value1
SubKey3=Value2
},
SubKey4=Value3
}
	DocumentKey1,T10 → {}
	DocumentKey1,SubKey1,T10 → {}
	DocumentKey1,SubKey1,SubKey2,T10 → Value1
	DocumentKey1,SubKey1,SubKey3,T10 → Value2
	DocumentKey1,SubKey4,T10 → Value3

关键技术 – Relationship Data Model 示例

msup

关系数据表 DDL & DML

```
CREATE TABLE test (
    key INT PRIMARY KEY,
    floatVal FLOAT,
    stringVal STRING );

INSERT INTO test VALUES (10, 4.5, "hello”);
```

MSS 中的元数据

test Table ID	1000
key Column ID	1
floatVal Column ID	2
stringVal Column ID	3

MSS 中的实际数据

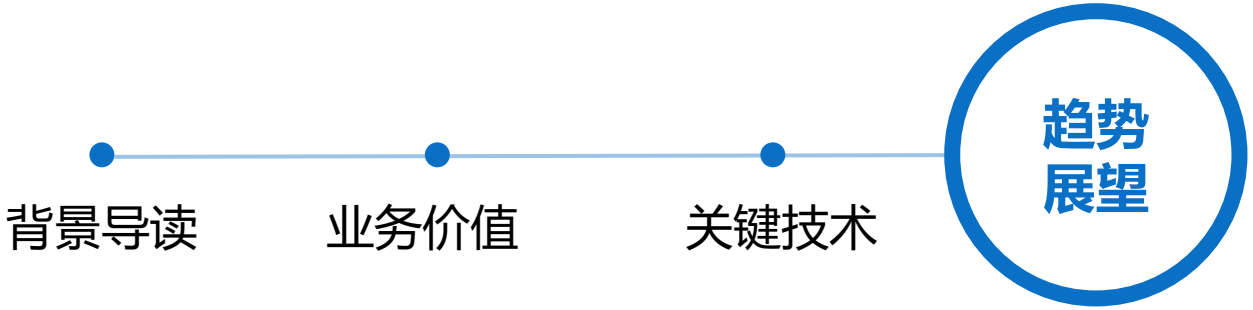
Key	Value
/1000/10/2	4.5
/1000/10/3	"hello”

查询实现

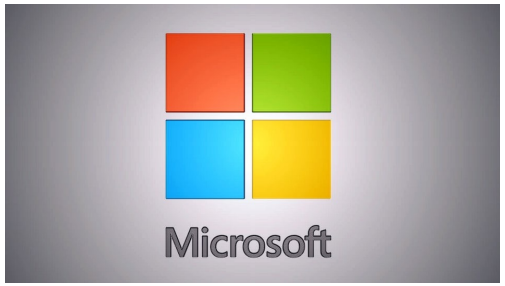
SELECT * FROM test WHERE key = 10;

Scan(/test/10/, /test/10/Ω)

目录



趋势展望 – 市场头部玩家



（图片均来自官网）

趋势展望 – 潜在场景

aPaaS：Application Platform as a Service

关注msup公众号

获取更多AI落地实践

麦思博(msup)有限公司是一家面向技术型企业的培训咨询机构，携手2000余位中外客座导师，服务于技术团队的能力提升、软件工程效能和产品创新迭代，超过3000余家企业续约学习，是科技领域占有率第1的客座导师品牌，msup以整合全球领先经验实践为己任，为中国产业快速发展提供智库。