本文是作者在ACMUG 2016 MySQL年会上的演讲内容，版权归作者所有。

中国MySQL用户组（China MySQL User Group）简称ACMUG。ACMUG是覆盖中国MySQL技术爱好者的一个技术社区，是Oracle User Group Community和MairaDB Foundation共同认可的MySQL技术社区。

我们关注MySQL，MariaDB，以及其他一切周边的开源数据库和开源工具，我们交流使用经验，推广开源技术，为开源贡献力量。

我们是开放社区，欢迎任何关注MySQL及其相关技术的人加入，我愿意跟其他任何技术组织和团体保持沟通和展开合作。

我们期望在我们的活动中大家都能以开心的、轻松的姿态交流技术，分享技术，形成一个良性循环，从而每个人都可以有一份收获。

ACMUG的口号：开源，开放，开心

关注ACMUG公众号，参与社区活动，交流开源技术，分享学习心得，一起共同进步。

Massive data availability architecture practice

2016.12 FOR ACMUG

阿里巴巴 Alibaba.com.cn

阿里技术保障 Ali Infrastructure Service

About Me

NAME： 刘晓军、谪仙、数据块

ALIPAY： from 2011

MAIL： shujukuai@163.com

EXPERIANCE：

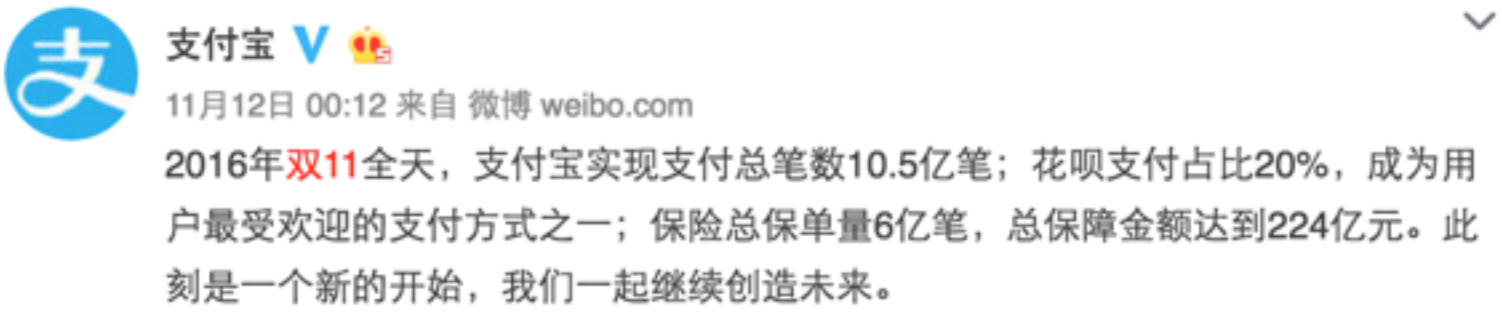
经历千万级支付到十亿级支付的蜕变

参与蚂蚁DB容量规划、实施、高可用架构设计

蚂蚁数据库稳定性建设负责人

双11、双12、春节红包等多次重大活动蚂蚁数据库负责人

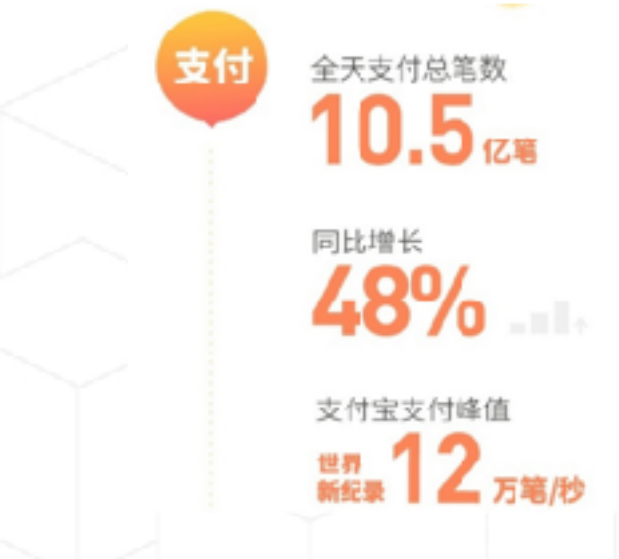
Business Challenge



支付宝

11月12日 00:12 来自 微博 weibo.com

2016年双11全天，支付宝实现支付总笔数10.5亿笔；花呗支付占比20%，成为用户最受欢迎的支付方式之一；保险总保单量6亿笔，总保障金额达到224亿元。此刻是一个新的开始，我们一起继续创造未来。



支付

全天支付总笔数

10.5亿笔

同比增长

48%

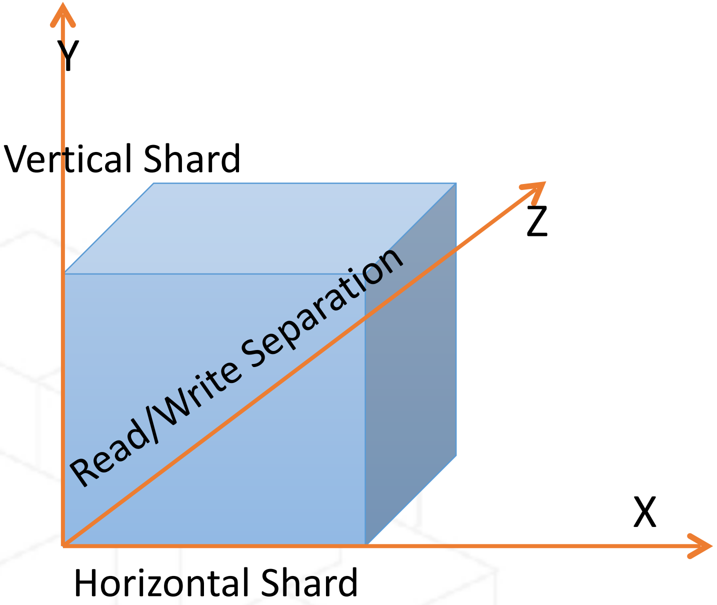
支付宝支付峰值

世界新纪录 12万笔/秒

CAPACITY

STABILITY

Capacity → Pop Scale Methods



- Read/Write Separation
  - Reading far more than writing
  - Latency tolerance
  - High availability read requirement
- Vertical Shard
  - Business dimensions (TRADE、PAY、TRANS)
- Horizontal Shard
  - User dimensions

Capacity → Bottleneck

DATABASE CONNECTIONS

OceanBase： 2M per conn)--VERSION 0.5

MySQL： 10W conns limit

ORACLE： 8-10M per conn

JVM MEMORY

OCEANBASE: Conns、 Fetchsize、 Sqltext…

MySQL： Conns、 Fetchsize、 Sqltext…

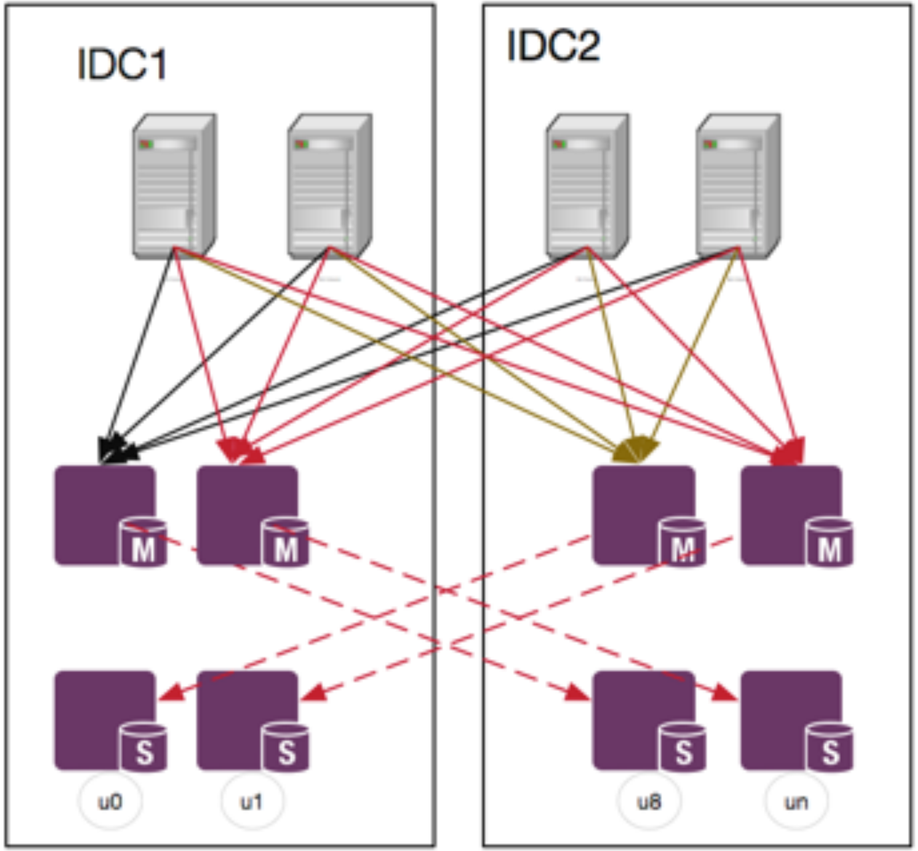
ORACLE： Conn-pools、 Conns、 Pscache、 Ps-memory

NET

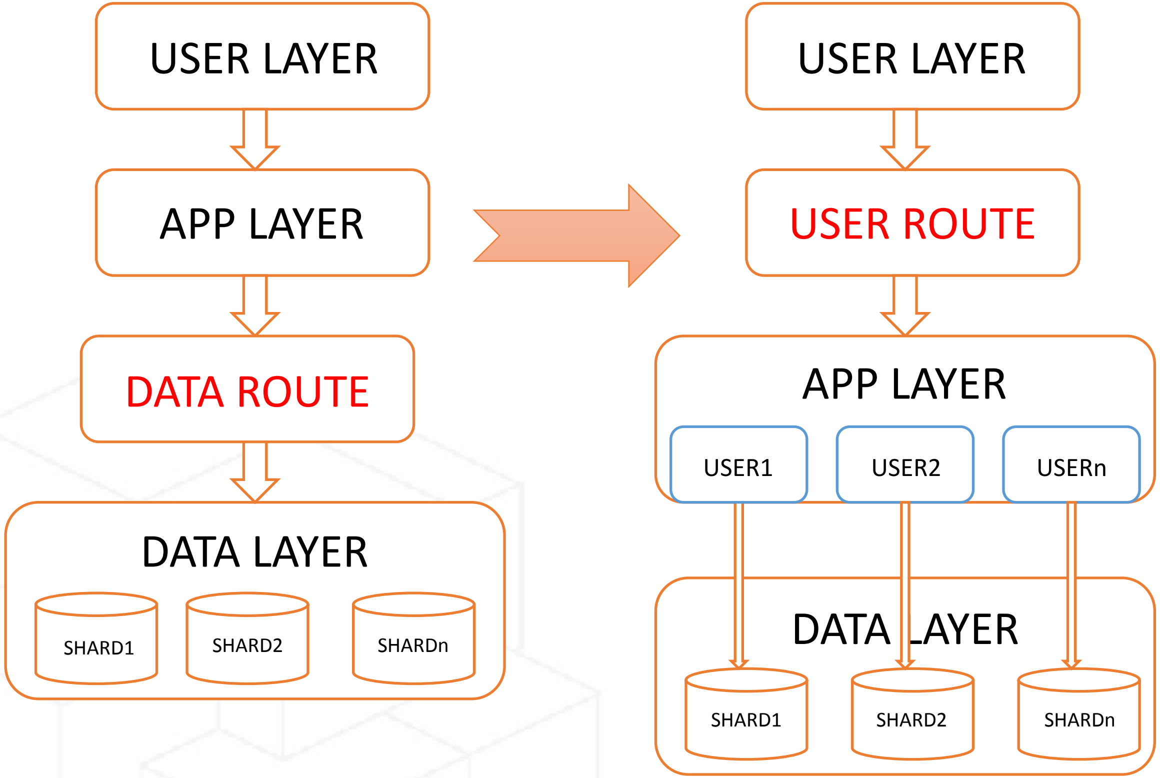
bandwidth

CITY RESOURCE

IDC、 Electric power



Capacity→ Solution



METHOD： LOCAL/REGION

All requests are able to be completed within a single REGION

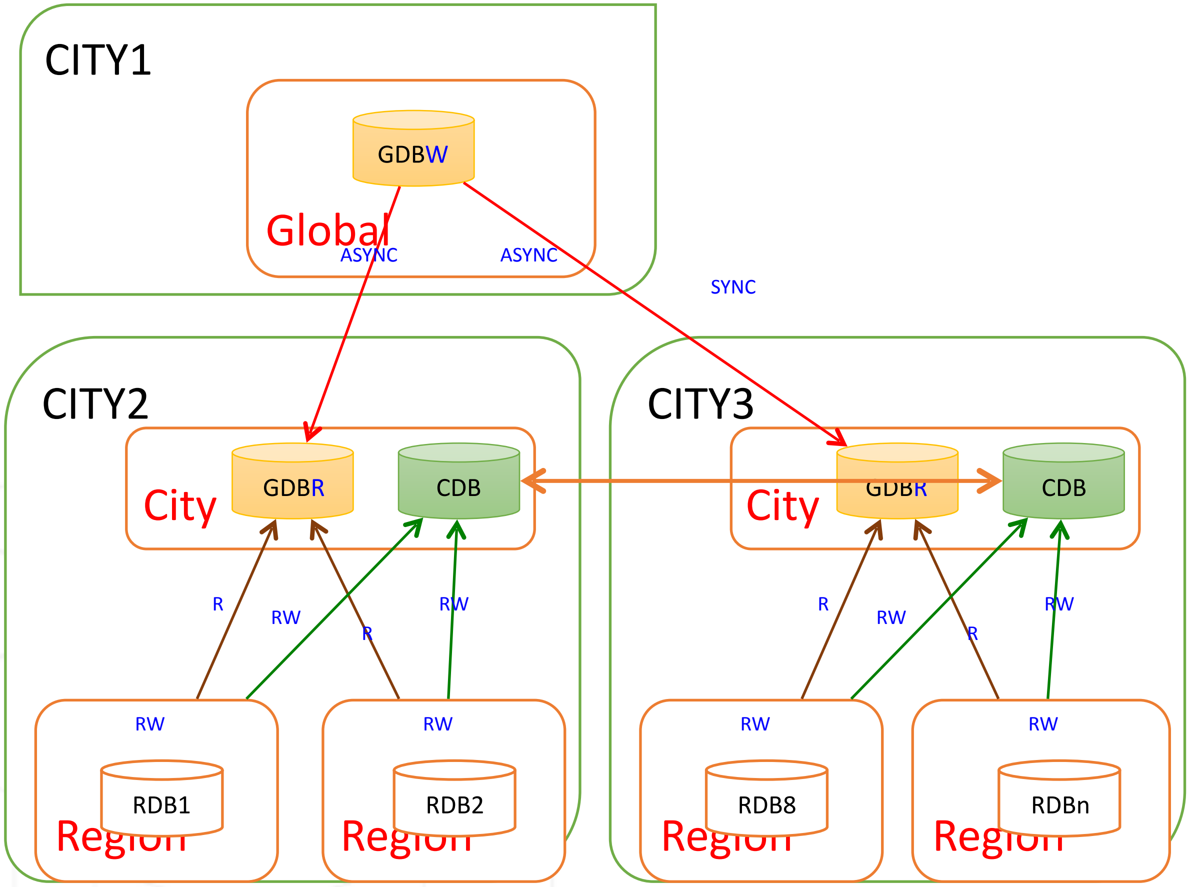
BENIFITS：

CONNECTIONS: Becomes 1/n

NET： Reduce the Internet-based Interaction between SHARDS

SCALABILITY： regions can deploy to anywhere

Capacity→DB Topology Map



How to deal with host、IDC、city failure?

Stability →Ablility and Requirements

RISK LEVEL DEFINITION		
LEVEL	**RPO**	**RTO**
A	0	MIN
B	0	10MIN
C	MIN	MIN
D	MIN	10MIN
X	OTHER	OTHER

FAILURE LEVEL DEFINITION	
RANGE	LEVEL
HOST	1
IDC	2
CITY	3

FAILURE LEVEL	MYSQL		ORACLE		OCEANBASE	
	SEMI_SYNC+HA	MP+HA	SHARE REDO+HA	MP+HA	3 COPIES	5 COPIES
HOST FAILURE	C	A	B	B	A	A
IDC FAILURE	C	A		B	A	A
CITY FAILURE	C	A		B		A

Requirements：

3A+3C

In the financial system , How to weigh availability and consistency?

ACID TO CAP/BASE

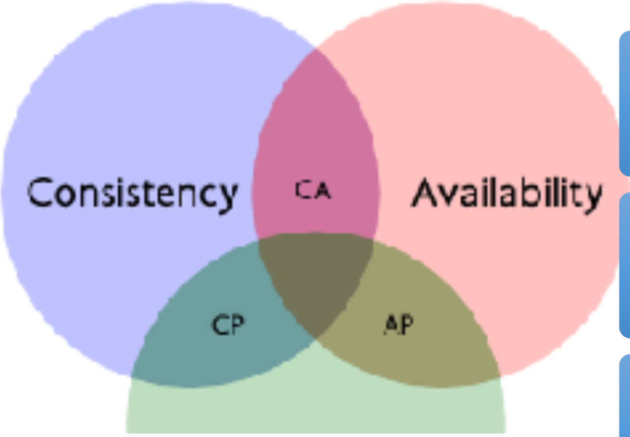
FULL ACID

AUTOMIC
- 原子性

CONSISTENT
- 一致性

ISOLATION
- 隔离性

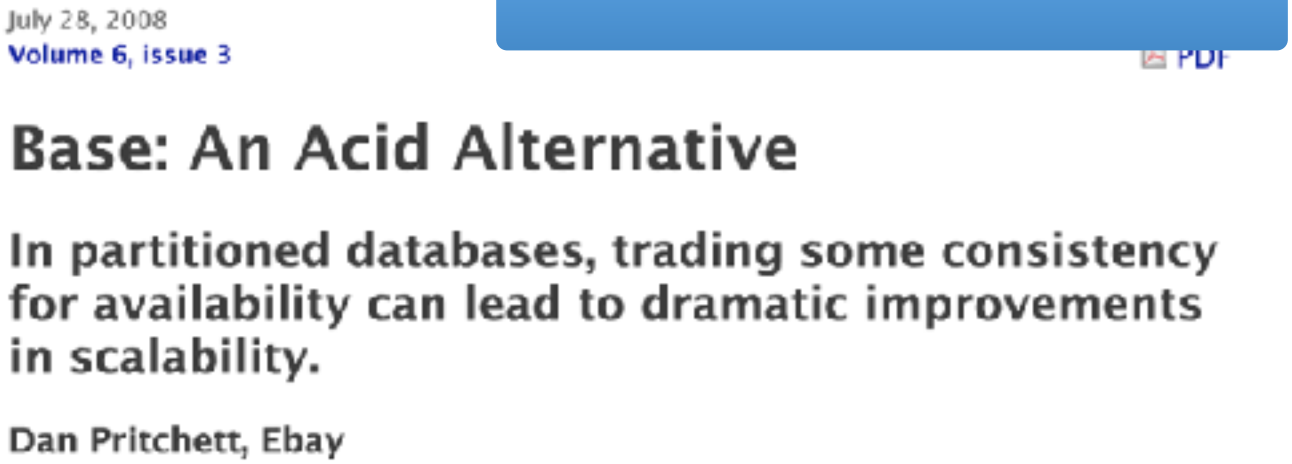
DURABLE
- 持久性



P
- Not always partition

A
- 0%-100%

C
- Many consistency levels


July 28, 2008
Volume 6, issue 3

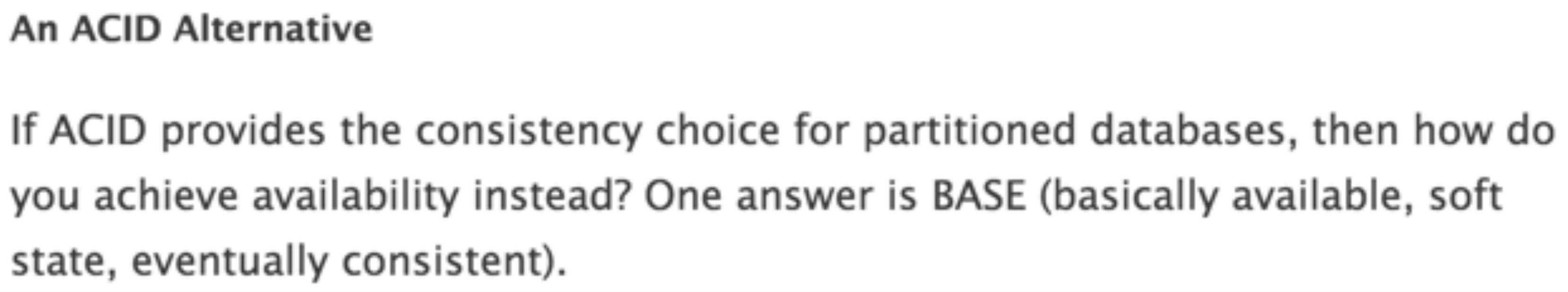
PDF

Base: An Acid Alternative

In partitioned databases, trading some consistency for availability can lead to dramatic improvements in scalability.

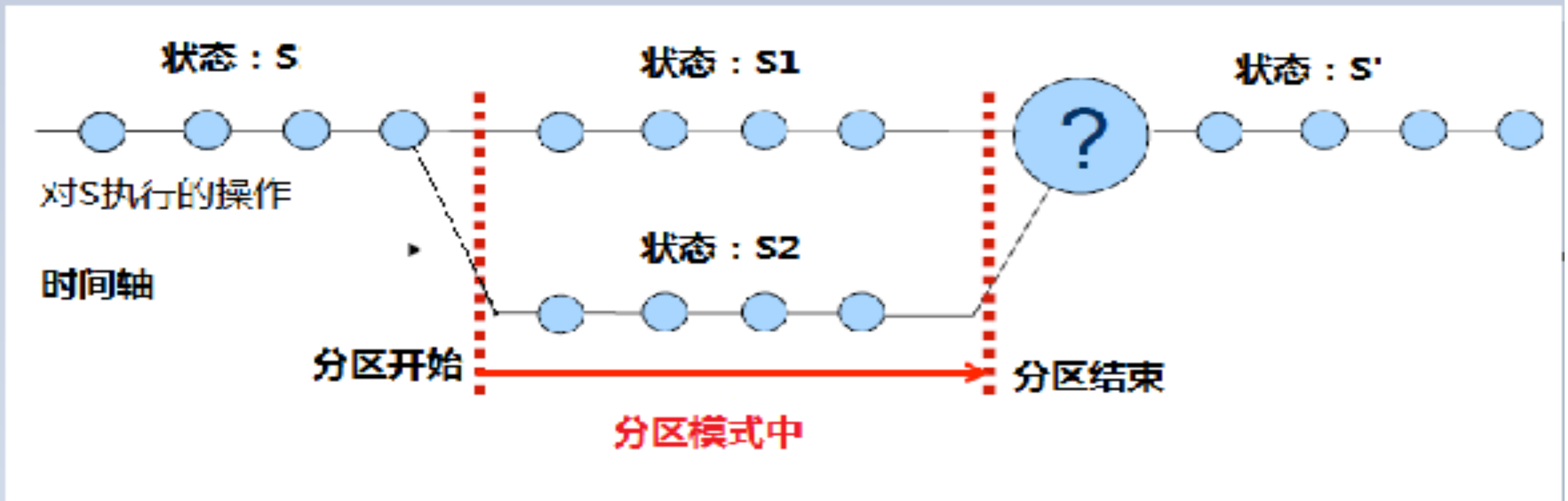
Dan Pritchett, Ebay

ACID TO CAP/BASE

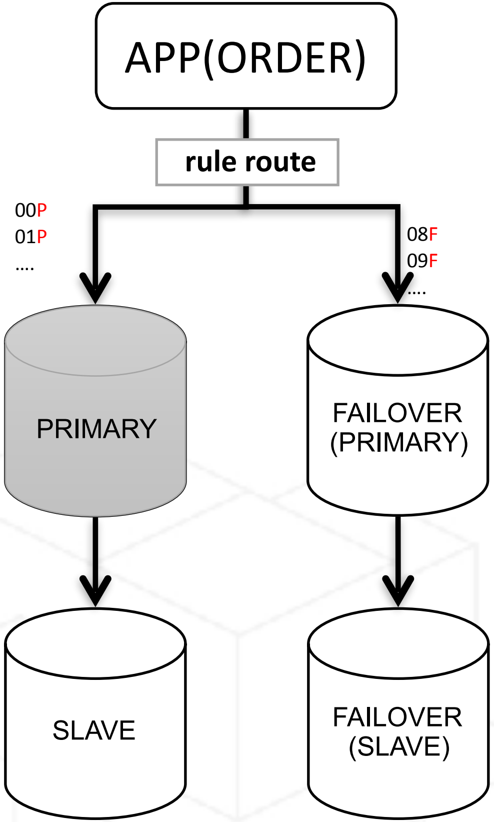


An ACID Alternative

If ACID provides the consistency choice for partitioned databases, then how do you achieve availability instead? One answer is BASE (basically available, soft state, eventually consistent).



Eg1:Pipelined Data



Strong consistency

Basically Available： --data partition

Affect the old orders

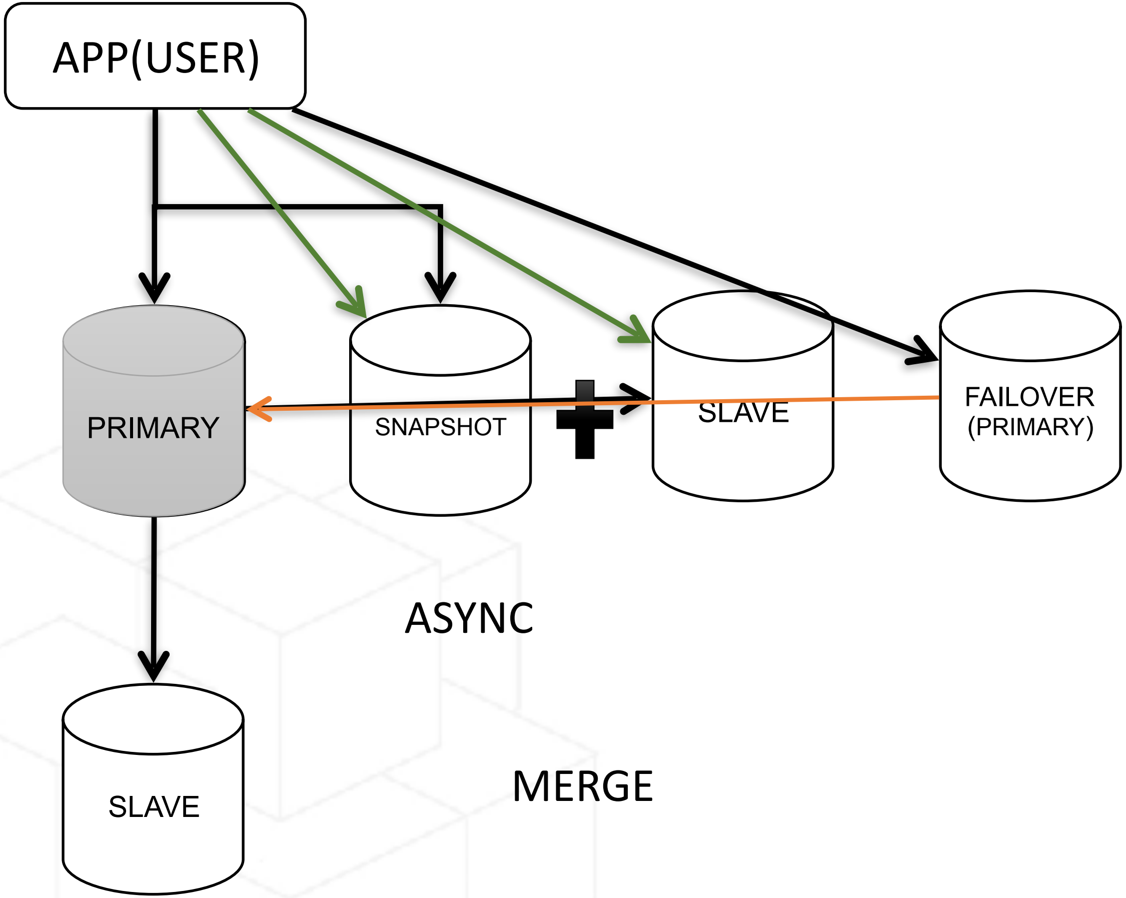
Does not affect new orders

Soft state：

depending on the design

Eventual consistency

Eg2: State Data



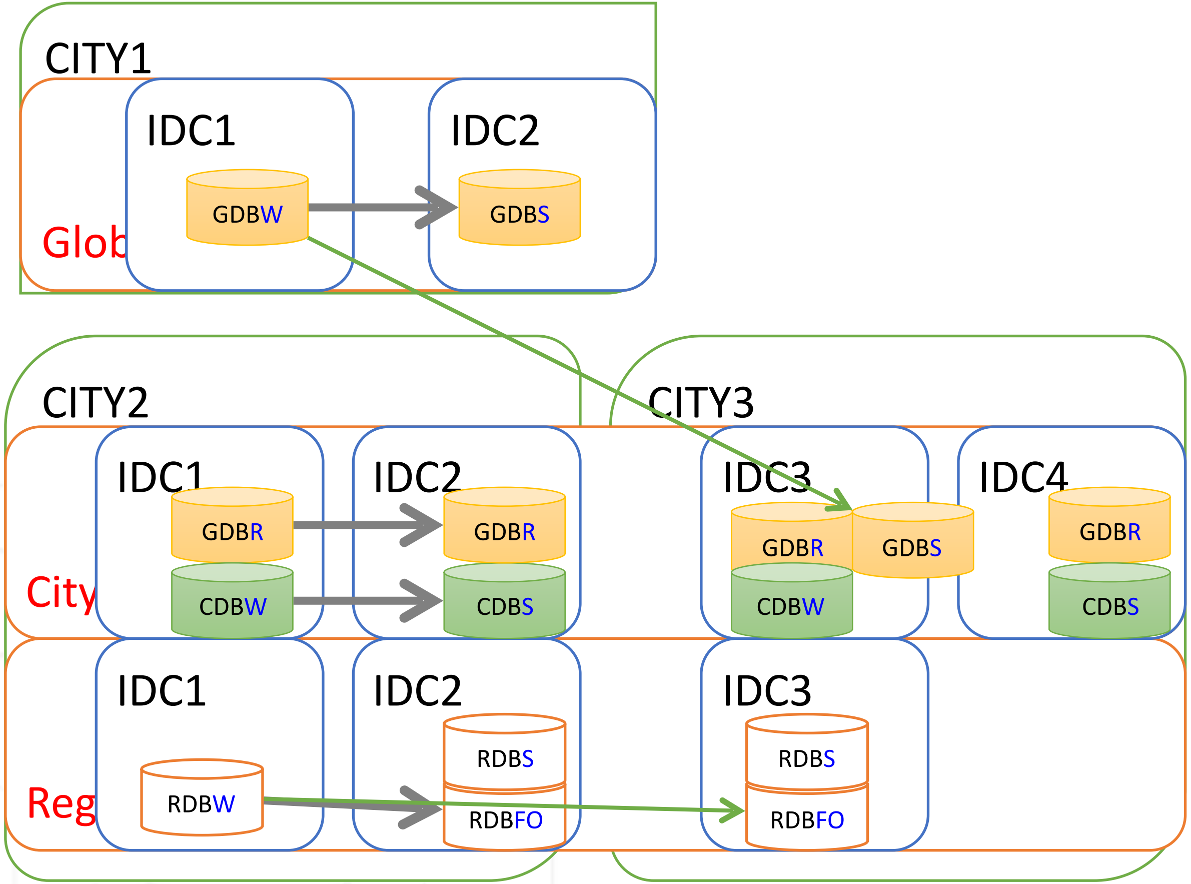
Strong consistency

Basically Available:
Affect a part of users or these function

Soft state：
Merge data

Eventual consistency

Stability → Failure



Failure Complexity(★★★★★)

Host failure: G/C/R

IDC failure: G/C/R

City failure: G/CR

Deploy Complexity(★★★★★)

M/S/S in different IDC and city

Read database in every IDC

Local failover in different IDC

Remote failover in different city

Future Challenges

- Simple Infrastructure
  - Infrastructure breakthrough
    - Max protection、Distributed database…
- Cost optimization
  - Efficient scalability through elastic computation
  - OLAP and OLTP systems with hybrid deployment
  - Stores computational separation
  - Business scenarios based optimization
- Business diversification
  - Payment
  - Social
  - International

入伙通道

weibo： 数据块V

wechat： shujukuai

mail： xiaojun.lxj@alibaba-inc.com