

# 分布共享数据服务理论与技术研究

魏恒峰

导师: 吕建 黄宇

南京大学软件所

May 22, 2016

# 分布共享数据服务理论与技术研究

1 研究背景

2 研究问题

3 研究方法

# 分布式应用

开放互联的网络环境下, 网络应用 (web applications) 分布部署.

**TODO:** 图: 分布式应用 (Weibo social network 举例)

# 分布式应用

应用三层架构: 表示层, 业务层, 数据层

**TODO: 图: 三层架构**

数据层中间件:

- ▶ 屏蔽底层数据具体形态
- ▶ 简化业务层开发

数据层中间件**H<sup>3</sup>L**特性:

- ▶ high availability
- ▶ high fault-tolerance
- ▶ high scalability
- ▶ low latency

# 分布式应用

应用三层架构: 表示层, 业务层, 数据层

**TODO: 图: 三层架构**

数据层中间件:

- ▶ 屏蔽底层数据具体形态
- ▶ 简化业务层开发

数据层中间件H<sup>3</sup>L特性:

- ▶ high availability
- ▶ high fault-tolerance
- ▶ high scalability
- ▶ low latency

数据形态: 共享数据 (shared data) vs. 分布数据 (distributed data)

# 数据层：共享数据

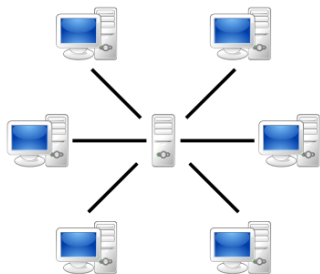
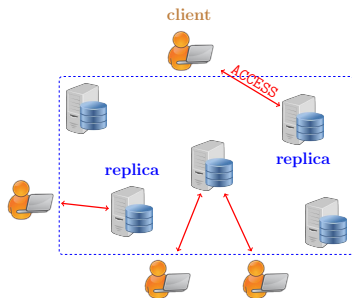


图: (集中式) 共享数据系统 **TODO: 重绘.**

- ▶ 单点故障  $\Rightarrow$  high availability
- ▶ 性能瓶颈  $\Rightarrow$  low latency
- ▶ 超荷负载  $\Rightarrow$  high scalability

# 数据层：分布数据



图：分布数据系统 **TODO:** 动画: partition+replication.

distributed data : partition + replication

# 分布数据的典型应用 (I)



图: 分布式存储系统 (开源 [左] & 商用 [右]).

应用需求 [Facebook@OSDI'10] vs. “分布数据”:

低延迟: 就近访问副本数据

高可用性, 高容错性: 备份容灾



## 分布数据的典型应用 (II)

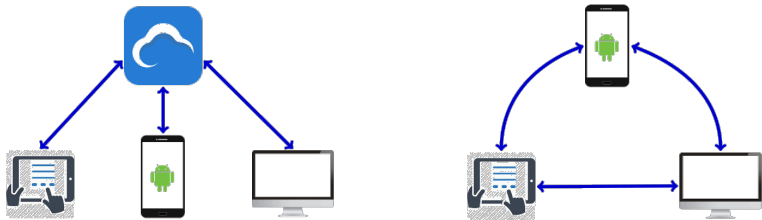


图: 个人多设备文件共享 ([基于云] C/S 结构 [左] & P2P 结构 [右]).

应用需求与特点 [Strauss@MIT Thesis'10] vs. “分布数据”:

功能需求: 文件副本

网络断连: 备份容灾; 离线可用

# 分布共享数据服务理论与技术研究

1 研究背景

2 研究问题

3 研究方法

# 面向共享数据的编程模型

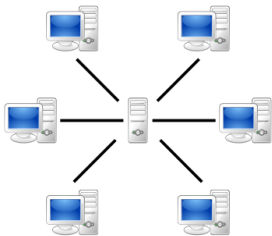


图: (集中式) 共享数据系统 **TODO: 重绘.**

共享变量编程模型 (**TODO: happy programmer**):

- ▶ 读—计算—写
- ▶ 只有一份 (one-copy) 数据
- ▶ 隐含假设: 数据随时可用
- ▶ 读操作返回最新值

# 面向分布数据的编程模型 (I)

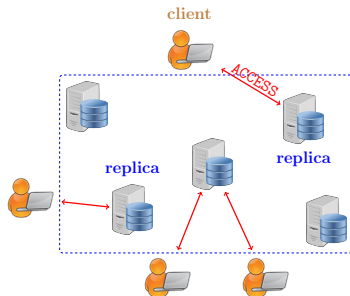


图: 分布数据系统.

共享变量编程模型与分布数据不匹配:

- ▶ 数据副本 (replication)
- ▶ 节点/通讯故障
- ▶ 读操作语义无定义

# 面向分布数据的编程模型 (II)

## 消息传递编程模型:

- ▶ 读/写 + 通信 (communication)
- ▶ 从哪里读
- ▶ 写哪里去
- ▶ 如何理解返回值
- ▶ 如何处理失败

## 消息传递编程模型的缺点: (TODO: unhappy programmer)

- ▶ 数据层分布数据细节暴露
- ▶ 业务层难于编程
- ▶ 业务层难于保证正确性

# 分布共享数据服务

分布共享数据服务: 在分布数据之上提供共享数据的假象

核心功能:

核心问题: 数据一致性问题

# 数据一致性问题

# 数据一致性问题

Alice: I've **lost** my ring.

Alice: I **found** it upstairs.

Bob: **Glad** to hear that.

Alice: I've **lost** my ring.

Bob: **Glad** to hear that.

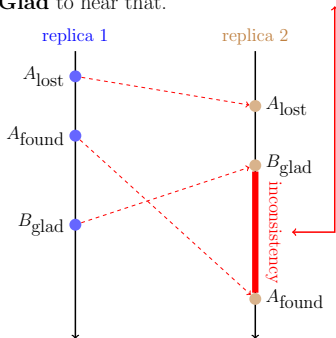


图: 社交网络中, 消息-评论乱序 [Lloyd@CACM'14].



# 数据一致性问题



图: 多设备文件共享时, 更新丢失 ( $\#N = 3, \#W = 2, \#R = 1$ ).

# 数据一致性问题

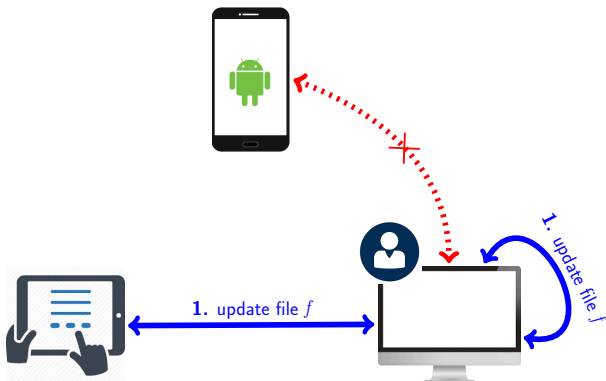


图: 多设备文件共享时, 更新丢失 ( $\#N = 3, \#W = 2, \#R = 1$ ).

# 数据一致性问题

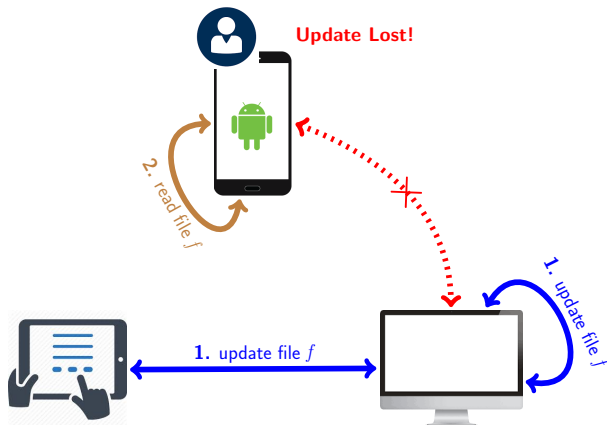


图: 多设备文件共享时, 更新丢失 ( $\#N = 3, \#W = 2, \#R = 1$ ).

# DSM Service

**TODO:** 图: DSM Service

针对数据一致性问题:

- ▶ 提供共享变量抽象
- ▶ 提供并发语义
- ▶ 屏蔽节点/通讯故障

# 分布共享数据服务理论与技术研究

1 研究背景

2 研究问题

3 研究方法

- 理论模型: 分布共享内存
- 技术途径: 三维框架

# 分布共享数据服务理论与技术研究

1 研究背景

2 研究问题

3 研究方法

- 理论模型: 分布共享内存
- 技术途径: 三维框架

# “分布共享内存” 服务

TODO: 图: 分布共享内存服务

# 分布共享内存: 概念辨析

基本定位: 传统概念应用于新型平台  
分布共享内存模型 + 分布数据系统



# 分布共享内存: 松耦合多处理器平台

多处理器并行算法:

- ▶ 单处理器
- ▶ 多处理器

# 分布共享内存: 概念辨析

问题空间: 传统问题, 新平台, 新挑战

# 分布共享数据服务理论与技术研究

1 研究背景

2 研究问题

3 研究方法

- 理论模型: 分布共享内存
- 技术途径: 三维框架