

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

July 18, 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

4 未来工作

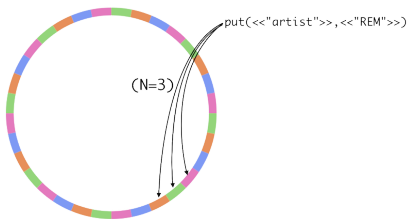
分布式应用



TODO: 动画: 分布部署

分布数据

(**TODO:** 动画: partition + replication)



distributed data : partition & replication

分布数据典型应用 (I)



图: 分布式存储系统 (开源 [左] & 商用 [右]).

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)

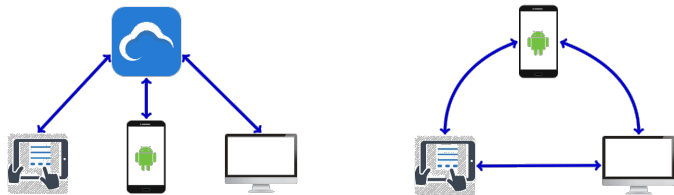
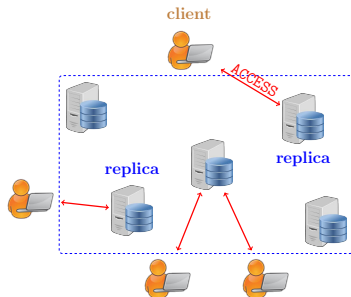


图: 个人多设备文件共享 ([基于云] C/S 结构 [左] & P2P 结构 [右]).

功能需求: 文件副本 [Strauss@MIT Thesis'10]

网络断连: 备份容灾; 离线可用

分布式应用访问分布数据



TODO: 重绘: 显示 partition & replication

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

4 未来工作

数据一致性问题

读操作语义问题:

- ▶ 共享数据环境: 返回最新写入的值
- ▶ 分布数据环境: “最新” 没有定义 (副本)
- ▶ 问题: 在分布数据环境下, 读操作允许返回什么值?

数据一致性问题

读操作语义问题:

- ▶ 共享数据环境: 返回最新写入的值
- ▶ 分布数据环境: “最新” 没有定义 (副本)
- ▶ 数据一致性问题: 在分布数据环境下, 读操作允许返回什么值?

数据一致性问题举例 (I)

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.

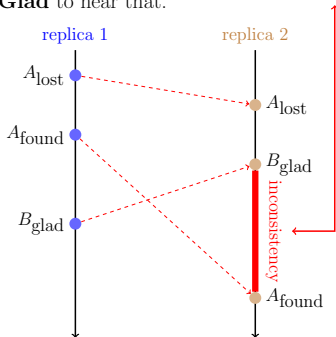


图: 社交网络中, 消息-评论乱序 [Lloyd@CACM'14].

数据一致性问题举例 (II)



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

数据一致性问题举例 (II)

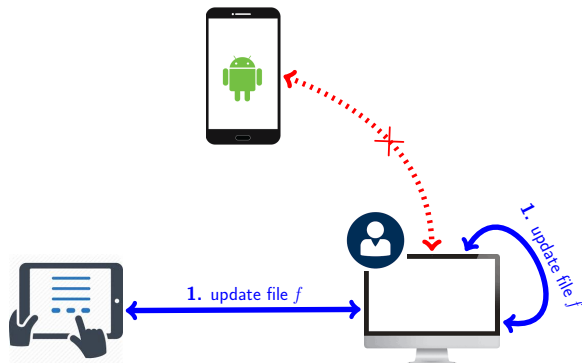


图: 多设备文件共享时, 更新丢失 ($\#N = 3, \#W = 2, \#R = 1$).

数据一致性问题举例 (II)

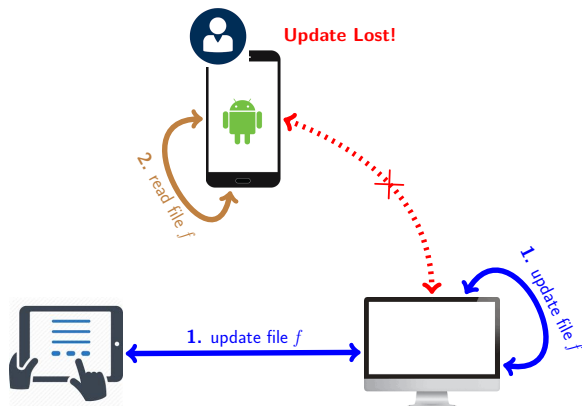
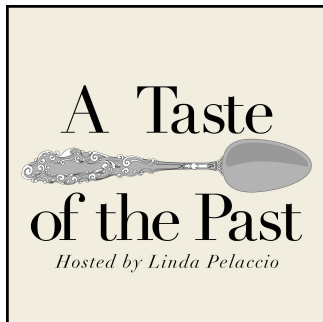


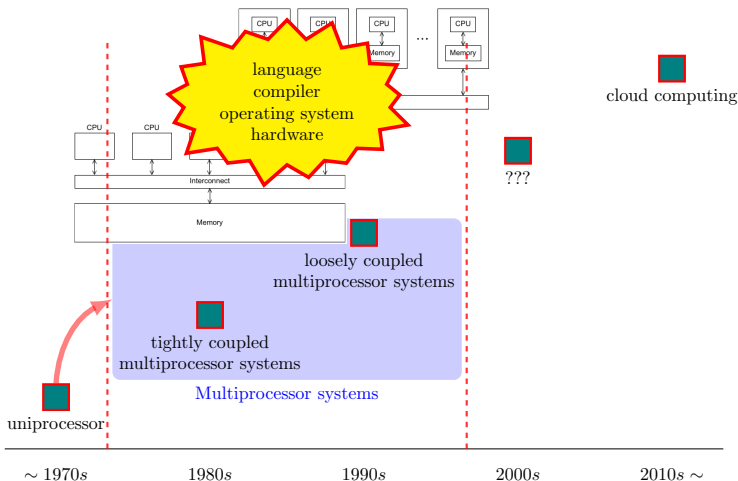
图: 多设备文件共享时, 更新丢失 ($\#N = 3, \#W = 2, \#R = 1$).

数据一致性模型

数据一致性问题研究的历史阶段



数据一致性问题研究的历史阶段



数据一致性问题研究的发展趋势

云计算凸显应用价值观



数据一致性问题研究的发展趋势 (I)

精细化: 从二元到连续谱



数据一致性问题研究的发展趋势 (I)

精细化: 从二元到连续谱



可度量: think *probabilistically* [Brewer@]



量化系统执行, 后验系统对一致性的满足程度

数据一致性问题研究的发展趋势 (II)

多样化: 从单一到融合 (mono- vs. multi-) [Terry@CACM'13]

- ▶ 融合强弱一致性: 不同操作, 不同一致性需求
- ▶ 融合一致与不一致: 容忍“有限度”的不一致

数据一致性问题研究的发展趋势 (II)

多样化: 从单一到融合 (mono- vs. multi-) [Terry@CACM'13]

- ▶ 融合强弱一致性: 不同操作, 不同一致性需求
- ▶ 融合一致与不一致: 容忍“有限度”的不一致



数据一致性问题研究的发展趋势 (II)

多样化: 从单一到融合 (mono- vs. multi-) [Terry@CACM'13]

- ▶ 融合强弱一致性: 不同操作, 不同一致性需求
- ▶ 融合一致与不一致: 容忍“有限度”的不一致



可调节: think *dynamically* [Terry@SOSP'13]

依据应用需求/系统状态调节数据一致性

我们的工作

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: 分布共享数据
- 技术途径: 三维框架

4 未来工作

分布共享数据服务理论与技术研究

1 研究背景

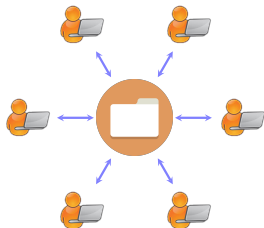
2 研究问题

3 研究方法

- 理论模型: 分布共享数据
- 技术途径: 三维框架

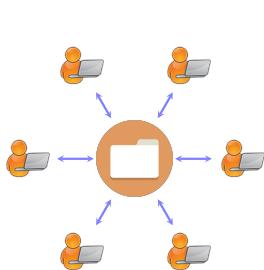
4 未来工作

分布共享数据模型

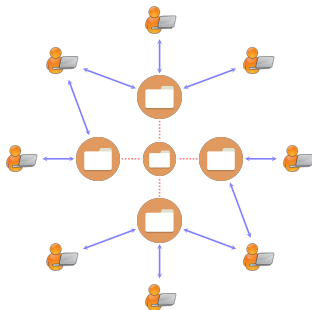


共享数据系统 (single copy)

分布共享数据模型

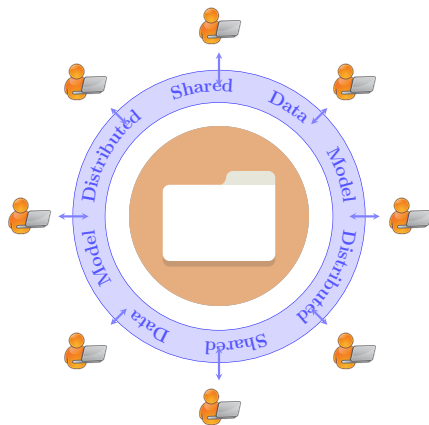


共享数据系统 (single copy)



分布数据系统 (replicas)

分布共享数据模型



分布共享数据模型: 在分布数据之上提供共享数据的假象

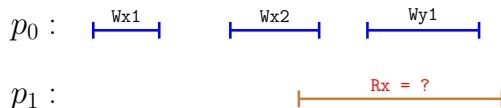
分布共享数据模型

p_0, p_1 : 客户进程 x, y : 共享变量

分布共享数据模型

p_0, p_1 : 客户进程 x, y : 共享变量

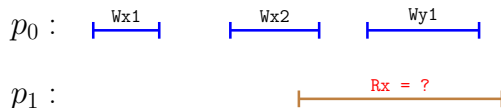
多进程并发读/写共享变量:



分布共享数据模型

p_0, p_1 : 客户进程 x, y : 共享变量

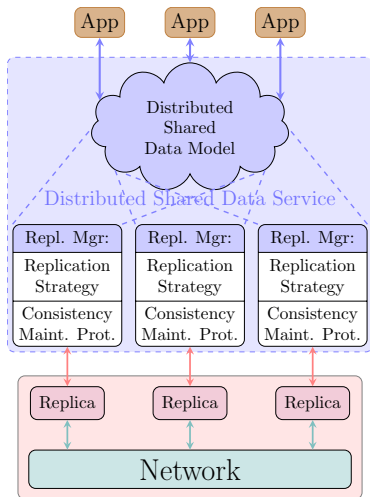
多进程并发读/写共享变量:



数据一致性问题: 读操作允许返回什么值?

不同一致性模型 $\xrightleftharpoons[\text{定义}]{\text{规定}}$ 不同合法返回值

分布共享数据服务



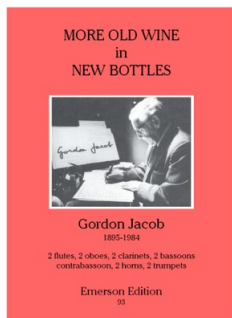
分布共享数据服务

分布共享数据服务:

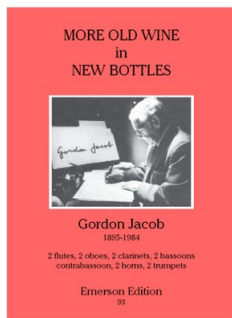
分布共享内存模型 (多处理器系统)
[传统概念]

+

分布数据系统
[新平台]



我们的工作



实现分布共享数据服务:

1. 精细化, 可度量
2. 多样化, 可调节

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: 分布共享数据
- 技术途径: 三维框架

4 未来工作

分布共享内存中的数据一致性问题

数据一致性问题的三个层面:

- 1. 虚拟共享数据有什么? ▶ 数据类型
- 2. 上层接口语义是什么? ▶ 一致性模型
- 3. 底层消息传递为什么? ▶ 一致性保障

研究框架

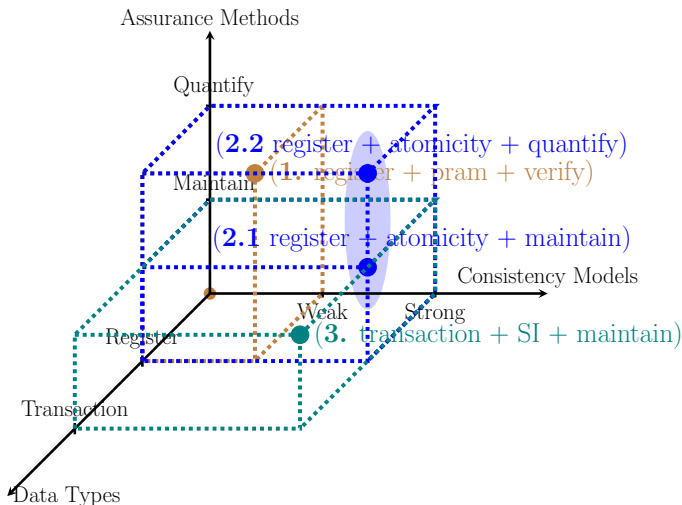


图: 数据一致性及保障技术研究框架

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
- 4 未来工作

分布数据一致性问题研究的发展趋势 (I)

分布数据一致性问题研究的发展趋势 (II)