



2016 杭州·云栖大会  
THE COMPUTING CONFERENCE

云栖社区  
yq.aliyun.com

# PostgreSQL Greemplum 数据上云



主办单位:



战略合作伙伴:



曾文旌 (义从)  
数据库专家



扫码观看大会视频

- 一：数据上云碰到的问题
- 二：解决方案
- 三：方案优势



# 一：数据上云碰到的问题

1. 大数据量问题
2. 网络质量不佳
3. 阿里云各数据产品间数据互通



## 二：解决方案

	解决的问题
oss_ext	GP 通过 oss 数据上云
oss_fdw	PG/PPAS 和 oss 打通
pgsql2pgsql	PG 相关数据产品 数据上云 数据迁移
mysql2pgsql	Mysql 到 PG/PPAS/GP



## 二：解决方案

### 1. oss

- a 云上廉价的对象存储服务
- b 和云上所有数据产品打通
- c 按照存储量和请求次数收费

### 2. oss\_ext

支持 Greenplum 的告诉并行数据装载和写入

### 3. oss\_fdw

支持 PG/PPAS , 兼容 oss\_ext 格式的数据读写协议



## 二：解决方案

### 1. postgresql2postgresql

pg/gp/ppas 迁移到 pg/gp/ppas

### 2. mysql2postgresql

mysql 迁移到 pg/gp/ppas

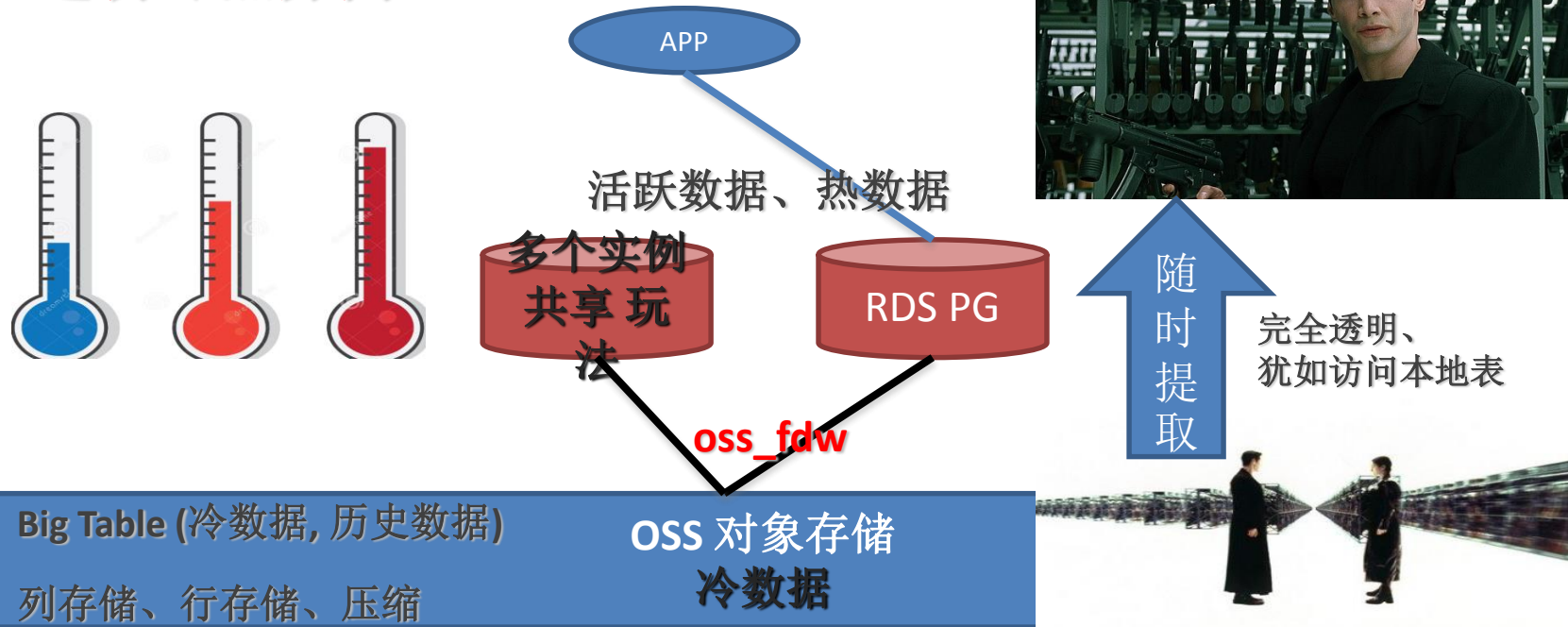


## 二：解决方案



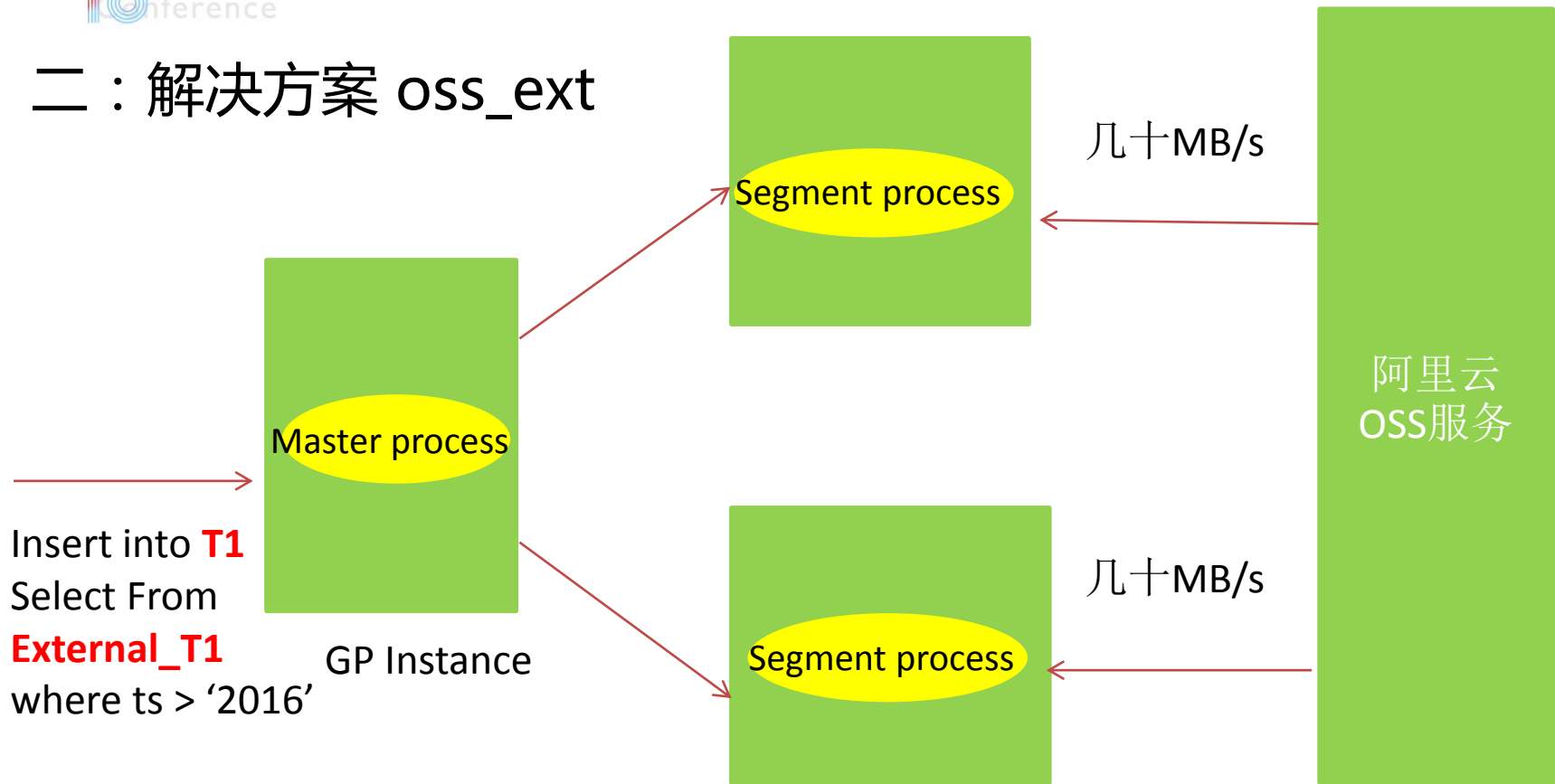
## 二：解决方案 oss\_fdw

- 透明冷热分离 解决大实例问题





## 二：解决方案 oss\_ext



## 二：解决方案 丰富的特性

### 1. 分布式并行数据加载

读写数据的性能随集群 segment 节点数线性增长

### 2. 支持自定义格式的文本文件

### 3. 支持多种 oss 文件匹配模式

### 4. oss\_fdw oss\_ext 都支持容错模式

### 5. 丰富的性能调优参数

### 6. 超时自动重试

### 7. 安全特性支持 oss id key 加密存储和隐藏显示



## 二：解决方案 正在开发的特性

1. 读写多种压缩文件，进一步降低使用成本
2. 读写模式
  - a 覆盖写模式 b 追加写模式
3. 导入导出加速，优化 CPU
  - a LLVM b 使用 SSE 指令集
4. OSS 文件前缀匹配模式



## 二：解决方案 GP已经支持的其他特性

插件化 oss\_ext、PostGIS、Oracle兼容包、DBLINK、MADlib

支持 create extension 语法

第三方合作伙伴 ETL

ORCA优化器

支持 rds\_superuser

修复BUG, copy 内存泄露

逻辑备份与恢复

扩容、缩容



## 三：方案优势

### 1. 使用 OSS

成本低 跨各数据库产品 跨可用区

冷数据转存到 oss 依然可以当做表访问

### 2. 并行导入性能高于常规数据导入导出方式

充分利用每个数据节点的硬件资源（CPU 内存 网络）

650G 20亿行文本数据导入到 32 segment 的 GP 集群约 35 分钟

### 3. 各数据产品间灵活的交换数据



## 三：方案优势

### 4. postgresql2postgresql

不落地数据迁移

pg ppas 大于9.4 支持基于逻辑复制的增量迁移

### 5. mysql2postgresql

不落地数据迁移

支持多表并发，支持基于条件的增量

### 6 已开源

[https://github.com/aliyun/rds\\_dbsync](https://github.com/aliyun/rds_dbsync)



20 The  
16 Computing  
Conference  
**THANKS**

