

基于Hawq的MaxCompute SQL引擎

- Seahawks

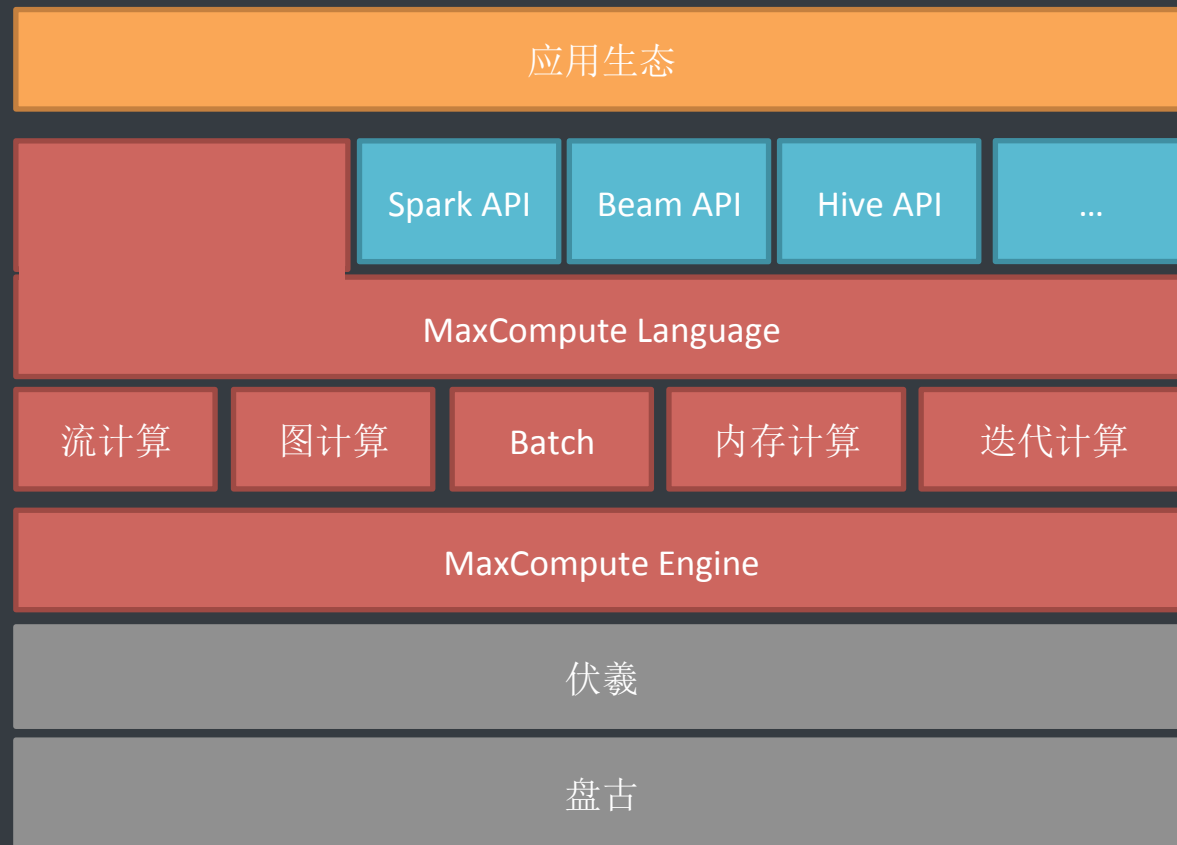
夏晨 2017.03.23

大纲

- **阿里云MaxCompute计算平台**
 - MaxCompute简介
 - MaxCompute技术特色及现状
- **Seahawks**
 - 为什么做？为何选型HAWQ？
 - Seahawks功能
 - Seahawks架构
 - Seahawks云上部署
 - Seahawks未来展望
- **Demo**
- **Q & A**

MaxCompute简介

- 大数据计算服务(MaxCompute)是一种快速、完全托管的PB/EB级数据仓库解决方案。具备万台服务器扩展能力和跨地域容灾能力,是阿里巴巴内部核心大数据平台,支撑每日百万级作业规模。
- MaxCompute向用户提供了完善的数据导入方案以及多种经典的分布式计算模型,能够更快速的解决用户海量数据计算问题,有效降低企业成本,并保障数据安全。



MaxCompute技术特点及现状

• 技术特点

- 大规模，易扩展
- 高性能，低成本
- 易用性，兼容性
- 多租户，隔离性
- 兼容生态规范

• 现状

- 规模
- 作业形态
- 用户生态



Seahawks技术选型

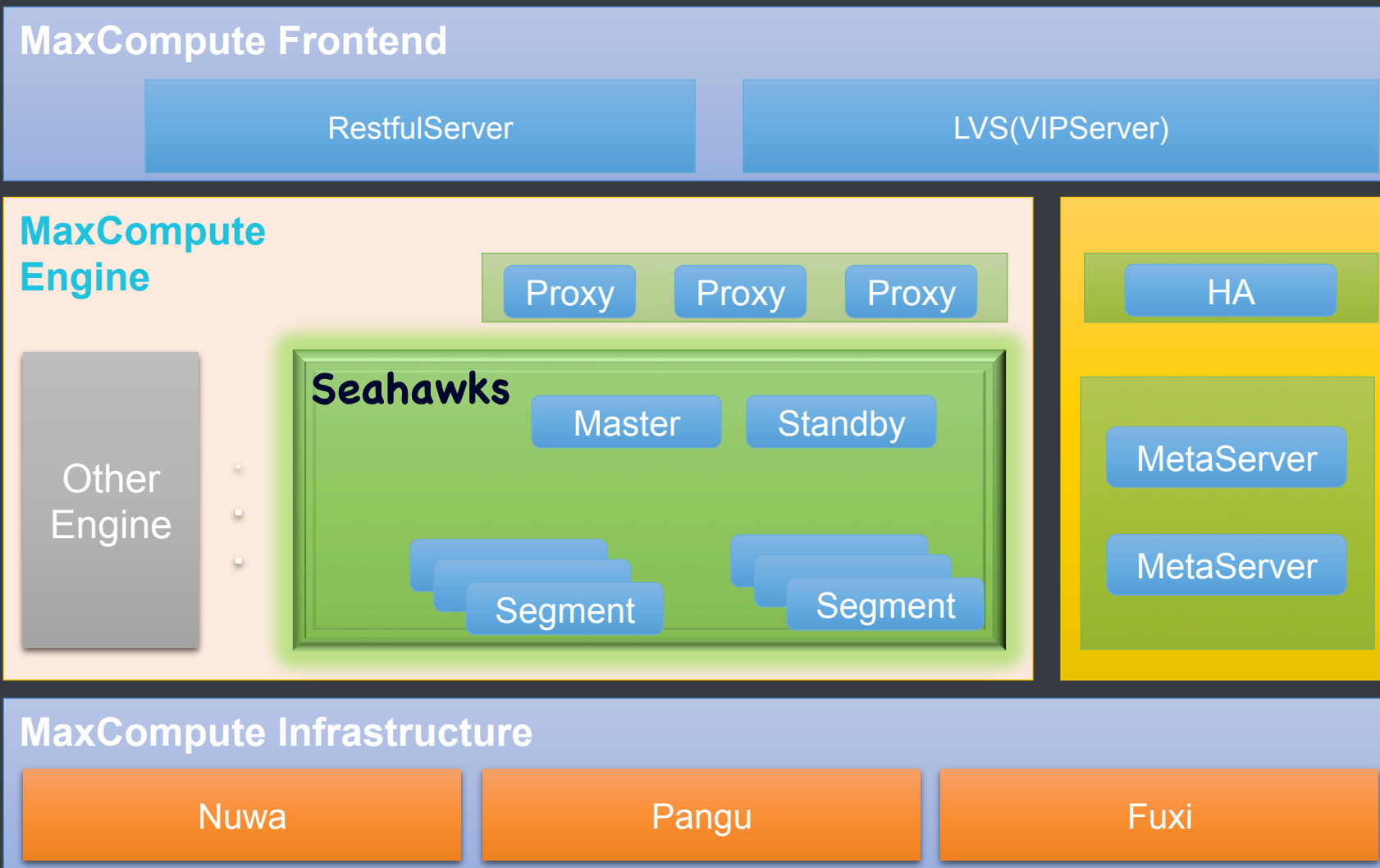
- 为什么要做Seahawks
 - MaxCompute vs Postgres
 - 生态/规模/性能
 - ...
- 为什么选型HAWQ
 - HAWQ vs GPDB
 - Pangu vs HDFS
 - Random Access File
 - ...
 - Fuxi vs YARN
 - Sort Benchmark
 - ...

Seahawks功能

- 一份数据多种计算
- 与Postgres/GPDB/HAWQ生态融合
- 与MaxCompute生态深度融合
 - 统一的资源调度
 - 统一的存储管理
 - 统一的账号认证
 - 统一的权限模型
 - 兼容的UDF生态
 - 共享的分布式元数据
 - 高可靠，高可用

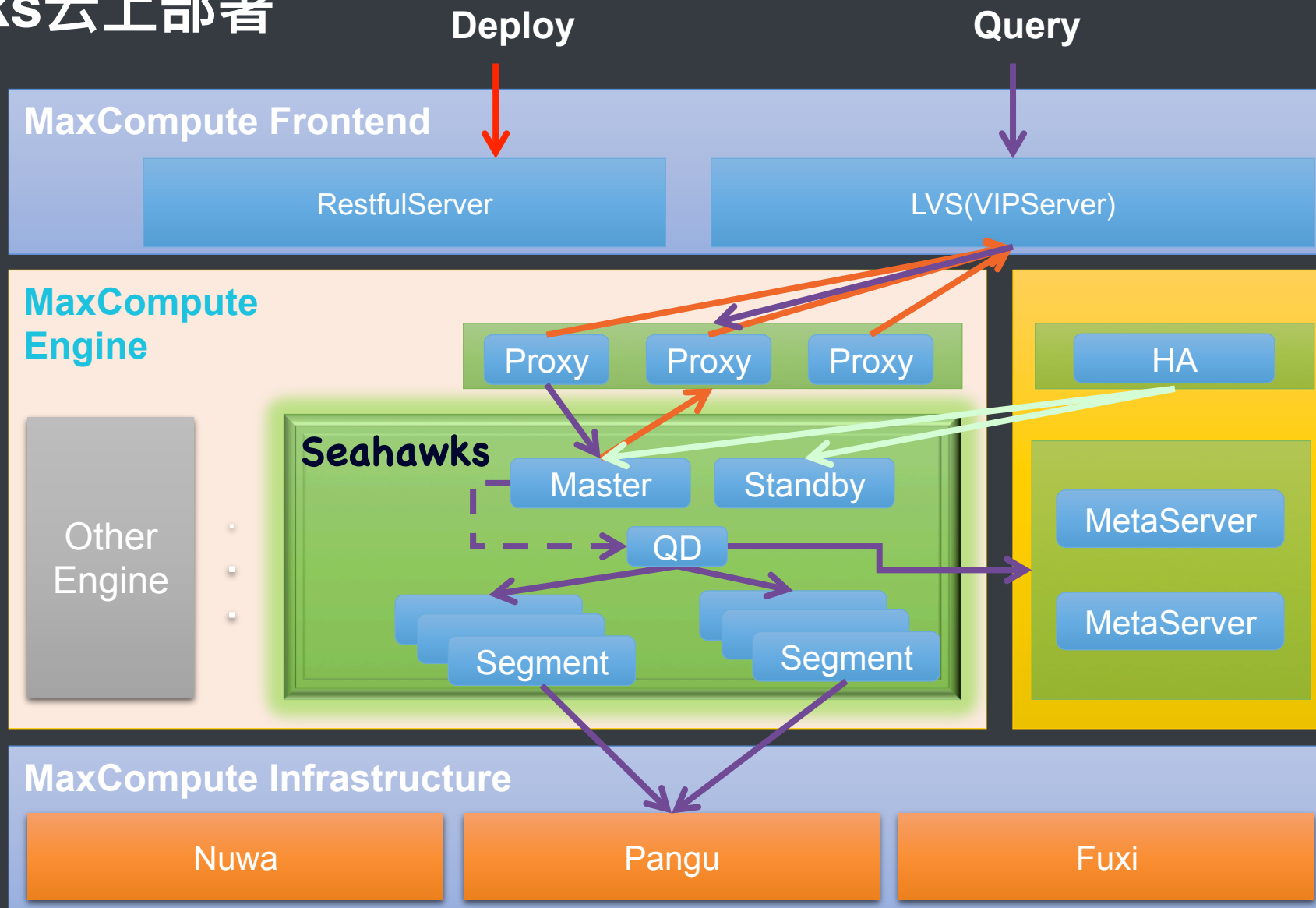
Seahawks架构

FuxiApp



Seahawks云上部署

FuxiApp



Seahawks未来展望

安全可靠

- 丰富的权限管理
- 多集群备份
- 跨机房容灾
- 定制化的高性能沙箱
- 分布式高可用元数据

可运维

- 集群管理：可视化监控
- 资源管理：CPU隔离 多租户
- 任务管理：任务队列 执行进展 日志
- 版本管理：多版本 热升级

丰富功能

- Heatable/Index
- Update/Delete(Pangu RAF)
- Query进度显示
- MaxCompute UDF(百万)
- 动态扩容缩容 RestAPI
- 更丰富的数据联通(RDS/Presto)
- 高性能 (MaxCompute Runtime)

Demo

- **快速部署Seahawks服务**
 - 单机服务
 - MaxCompute集群（100节点segments服务）
- **HA测试**
 - 自动化测试
 - 随机Kill master/segment/standby
- **查询MaxCompute海量数据**
- **MaxCompute Logview**
- **Heaptable/Index**

Q & A

欢迎加入阿里云计算平台

电邮: chen.xiac@alibaba-inc.com

微信/钉钉: 13989818302

为了无法计算的价值 |  阿里云

