

SACC

2012中国系统架构师大会

SYSTEM ARCHITECT CONFERENCE CHINA 2012

架构设计 · 自动化运维 · 云计算

如何实现秒级对大数据的实时分析？

人云科技团队

- 人云科技团队组建于2010年底，团队的成员主要来自于IBM中国研究院，渣打银行，上海贝尔阿尔卡特，爱立信和EMC等著名企业。
- 人云科技团队的使命是，推出具有鲜明特色的产品，并不断做精做强，锻炼出自身的核心竞争力和技术底蕴，并且将中国的核心技术产品推广到国外！

个人介绍



吴朱华

YunTable



SACC

2012中国系统架构师大会

SYSTEM ARCHITECT CONFERENCE CHINA 2012

架构设计 · 自动化运维 · 云计算

云计算与大数据



SACC

2012中国系统架构师大会

SYSTEM ARCHITECT CONFERENCE CHINA 2012

架构设计 · 自动化运维 · 云计算

大数据的挑战和机遇

挑战

- Volume , PB级别为常态
- Variety , 结构化、半结构化、非结构化
- Velocity , 快速处理



机遇

- More Findings , 网络舆情
- Deep Insights , 用户画像
- Priceless Results , 智能交通

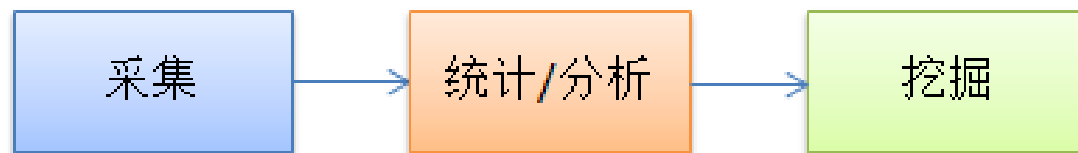
SACC

2012中国系统架构师大会

SYSTEM ARCHITECT CONFERENCE CHINA 2012

架构设计 · 自动化运维 · 云计算

大数据的处理流程



- 简介：利用多种轻型数据库来接收发自客户端的数据，并且用户可以通过这些数据库来进行简单的查询和处理工作；
- 特点和挑战：**并发数高**；
- 使用的产品：MySQL，Oracle，HBase，Redis和MongoDB等，并且这些产品的特点各不相同；

统计/分析

- 简介：将海量的来自前端的数据快速导入到一个集中的大型分布式数据库或者分布式存储集群，利用分布式技术来对存储于其内的集中的海量数据进行普通的查询和分类汇总等，以此满足大多数常见的分析需求；
- 特点和挑战：导入数据量大，查询涉及的数据量大，查询请求多；
- 使用的产品有：InfoBright，Hadoop（Pig和Hive），YunTable，SAP Hana和Oracle Exadata，除了Hadoop以做离线分析为主之外，其他产品可以做实时分析；

- 简介：基于前面的查询数据进行数据挖掘，来满足高级别的数据分析需求；
- 特点和挑战：算法**复杂**，并且计算涉及的数据量和计算量都大；
- 使用的工具：R，Hadoop Mahout。

为什么需要实时分析？

SequelMedia
盛拓传媒

IT168.com

- 数据是**有时效性**的；
- 催生新的**商业模式**；

SACC

2012中国系统架构师大会

SYSTEM ARCHITECT CONFERENCE CHINA 2012

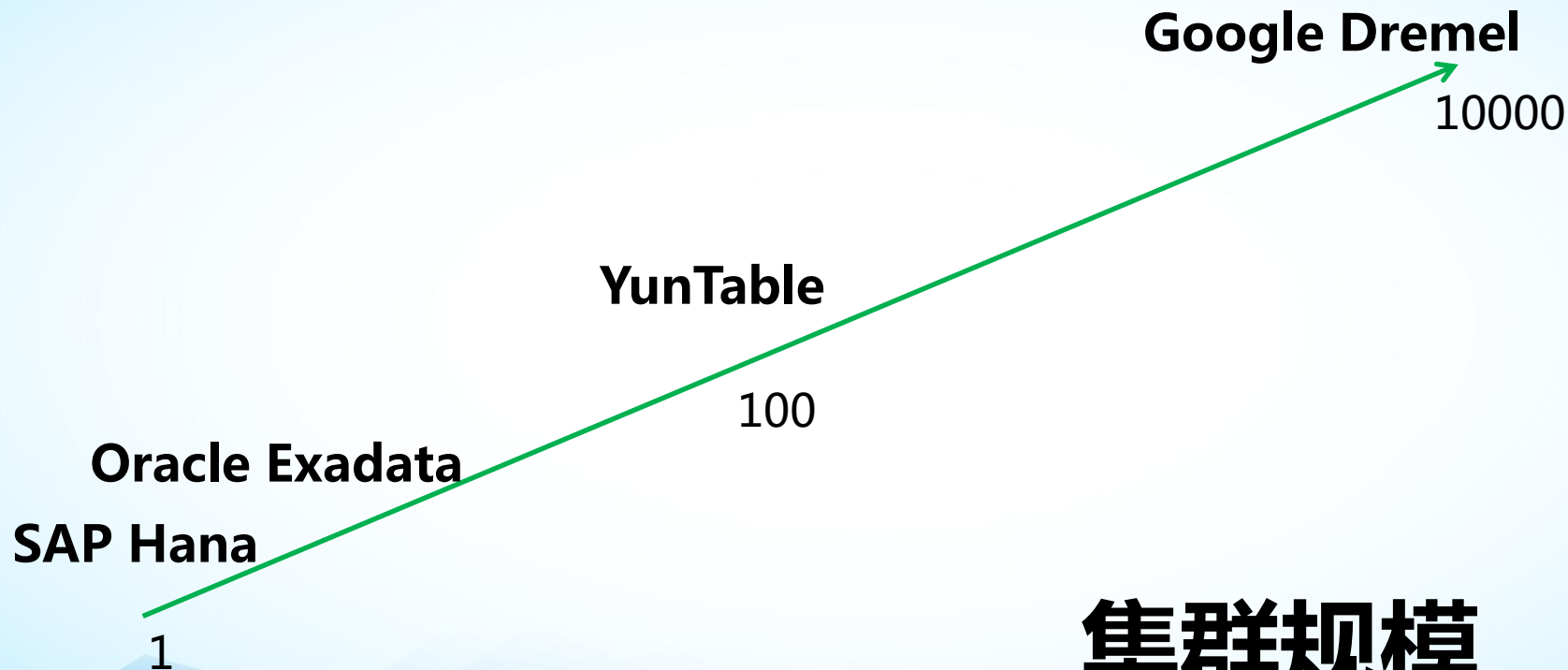
架构设计 · 自动化运维 · 云计算

- YunTable是在分布式数据库和新的NoSQL技术的基础上发展而来的**新一代大数据实时分析DB**。
- YunTable最大的特色就是**快**，它能极快地导入海量的数据，并极快地进行相关的分析，并且扩展能力很强。

功能特性

- 大数据，**秒级**内存计算；
- 采用**廉价**的x86硬件；
- 自动线性**动态扩展**至数百台集群；
- 每秒**GB**级别吞吐量，**PB**级别存储量；
- **SQL92**特性覆盖；
- 数据**安全**性；

产品定位



集群规模

SACC

2012中国系统架构师大会

SYSTEM ARCHITECT CONFERENCE CHINA 2012

架构设计 · 自动化运维 · 云计算

三大核心技术

并行处理

行列混合存储

压缩

这三个技术是处理海量结构化数据的必备利器，
因为这样I/O处理最优化，并能对热数据提供接
近**内存计算**的性能。

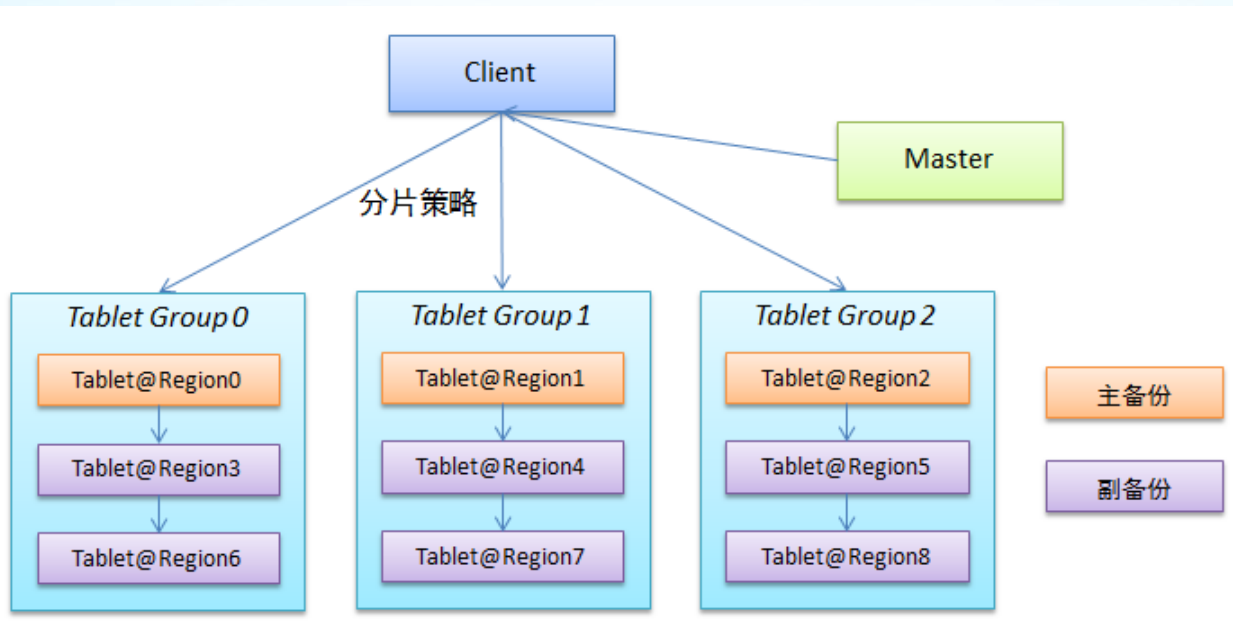
SACC

2012中国系统架构师大会

SYSTEM ARCHITECT CONFERENCE CHINA 2012

架构设计 · 自动化运维 · 云计算

分布式架构



清晰的分布式架构能自动伸缩。

性能测试

客户是国内著名的互联网广告企业，这个Case的集群规模**只用了5台X86服务器**，数据规模在**百亿**行级别，性能和之前相比有**几十倍**的提升。

	Infobright (之前)	YunTable
导入数据	2万行每秒	30万行每秒
常见操作	15分钟以上	20秒以内

客户案例

- 互联网广告行业；

海量广告监测；

- 金融行业；

证券分析系统；

SACC

2012中国系统架构师大会

SYSTEM ARCHITECT CONFERENCE CHINA 2012

架构设计 · 自动化运维 · 云计算

YunTable的路线图

1. 2012.6-2012.10 : 完善**SQL**高级特性 , **专精**大数据实时分析
2. 2012.6-2012.10 : 整合封装**Hadoop**
3. **2012.10 发布企业支持版**
4. 2012.11- : **BI**模块 并新增一些用于**数据挖掘**的功能模块 , 包括对R的支持

Why choose backend?

■ 世界级的专注企业级后端领域公司 ?

■ 一个专注产品, 专注核心技术的团队

■ 一切从开源小项目开始



SACC

2012中国系统架构师大会

SYSTEM ARCHITECT CONFERENCE CHINA 2012

架构设计 · 自动化运维 · 云计算

一起携手努力 实现技术理想



SACC

2012中国系统架构师大会

SYSTEM ARCHITECT CONFERENCE CHINA 2012

架构设计 · 自动化运维 · 云计算

Q&A

SACC

2012中国系统架构师大会

SYSTEM ARCHITECT CONFERENCE CHINA 2012

架构设计 · 自动化运维 · 云计算