

# ArchSummit全球架构师峰会 北京站2015

打造一站式大数据分析平台  
王桐 永洪科技

# Geekbang>

## 极客邦科技

整合全球最优质学习资源, 帮助技术人 and 企业成长  
Growing Technicians, Growing Companies

**InfoQ**  
new

专注中高端技术人员的  
技术媒体



**EGO** EXTRA GEEKS' ORGANIZATION  
NETWORKS

高端技术人员  
学习型社交网络



**StuQ**  
new

实践驱动的  
IT 职业学习和服务平台



**GiT** GEEKBANG  
INTERNATIONAL  
TRAINING  
极客邦培训

一线专家驱动的  
企业培训服务



旧金山 伦敦 北京 圣保罗 东京 纽约 上海  
San Francisco London Beijing Sao Paulo Tokyo New York Shanghai

# QCon

## 全球软件开发大会

2016年4月21-23日 | 北京·国际会议中心

主办方 **Geekbang**  **InfoQ**  
极客邦科技

**7折** 优惠 (截至12月27日)  
现在报名, 节省2040元/张, 团购享受更多优惠

[www.qconbeijing.com](http://www.qconbeijing.com)



扫描获取更多大会信息

# 数据创造未来已成真理

## 没有数据化运营的企业，等于在黑暗森林中裸奔

### 不知现状

无法获知业务状况、发展规律、用户画像等，在一片漆黑中凭感觉前行。

### 不明原因

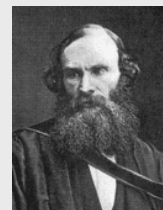
利润下滑、增长停滞、用户流失，不能探究真实原因，面对问题无法有效解决，且未来再发生时依然如此。

### 不可预测

销量、收入、用户行为，没有准确的预测，依靠经验来进行备货、促销、人员扩张。粗放经营，损耗太大。

"IF YOU CAN NOT MEASURE IT,  
YOU CAN NOT IMPROVE IT."

- LORD KELVIN



少数派 vs 全民化

数据要打通到企业的每一个个体

当前用户掌握数据技能的现状



# IT 瓶颈化

IT 部门忙得要死，业务部门急得要死





工具碎片化



# 不能很好地处理大数据

Forrester在报告中指出，有80%的用户拒绝使用企业的数据分析平台，  
因为无法形成数据化运营的闭环

现实：投入很大，期望很高  
但大数据的价值产出和预期并不匹配

如何让数据真正促进业务？

一站式大数据分析平台  
是未来企业的标配



The diagram features a central circle with the text '一站式大数据分析平台' (One-stop Big Data Analysis Platform). To the right of this circle is a dashed arc with four white circular markers. Each marker is aligned with a text label: '自服务数据准备' (Self-service data preparation), '探索式分析' (Exploratory analysis), '深度分析' (Deep analysis), and '企业级管控' (Enterprise-level control). The entire graphic is set against a solid green background.

# 一站式大数据分析平台

自服务数据准备

探索式分析

深度分析

企业级管控

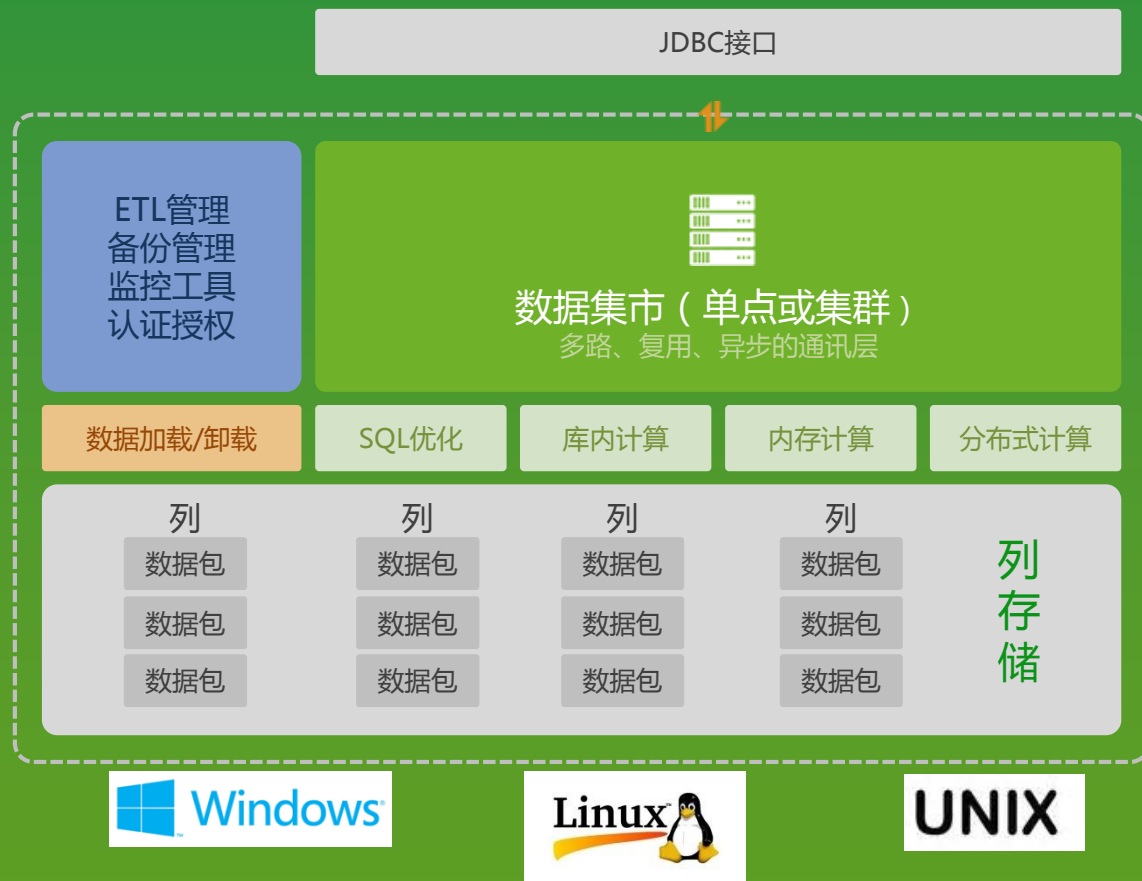
Demo：一站式大数据分析平台  
带来了什么样的体验

要做到这一点  
需要什么样的数据处理引擎



# MPP ( Massively Parallel Processing , 大规模并行处理系统 ) 数据集市

借鉴Hadoop体系架构和增量流计算理念，采用内存计算、库内计算、分布式通讯、列存储分布式文件系统等技术开发实现的一个具备分布式并行实时处理、复杂查询优化、复杂分析支持的大数据计算处理数据库系统。



## 列存储

### DSM(decomposition storage model )

- 大幅压缩数据
- 降低磁盘I/O
- 降低内存和cpu开销

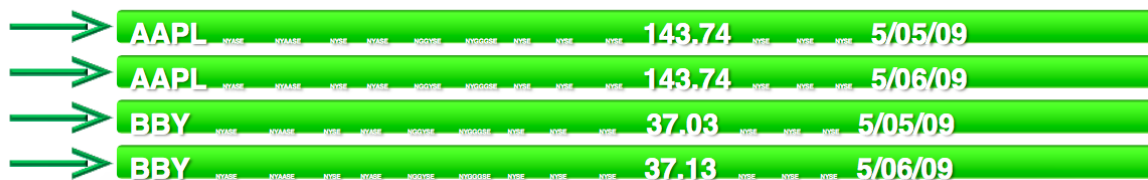
例如:计算1天中某支股票的平均价格

```
SELECT AVG(price)
FROM table1
WHERE
ID = 'AAPL'
AND date = '5/06/09'
```

### 列式数据库- 读取3列



### 行式数据库- 读取所有列



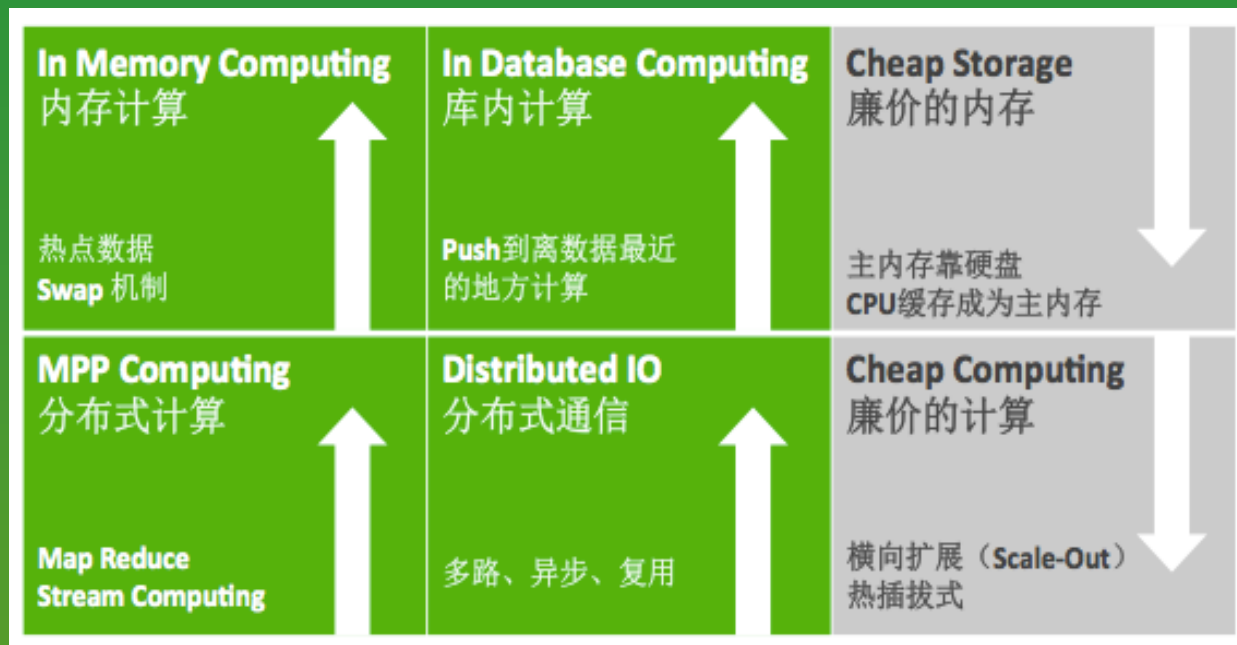
## 不仅仅是列存储，存储也为查询优化

### 库内计算

库找寻出最优化的计算方案，继而把所有开销较大的、昂贵的计算都移动到数据存储的地方直接计算。这一技术大大减少了数据移动，降低了通讯负担，保证了高性能数据分析。

### 内存计算

得益于列存储技术和并行计算技术，大大压缩数据，并同时利用多个节点的计算能力和内存容量。通过内存计算，CPU直接从内存而非磁盘上读取数据并对数据进行计算。内存计算是实现大数据分析的关键应用技术。

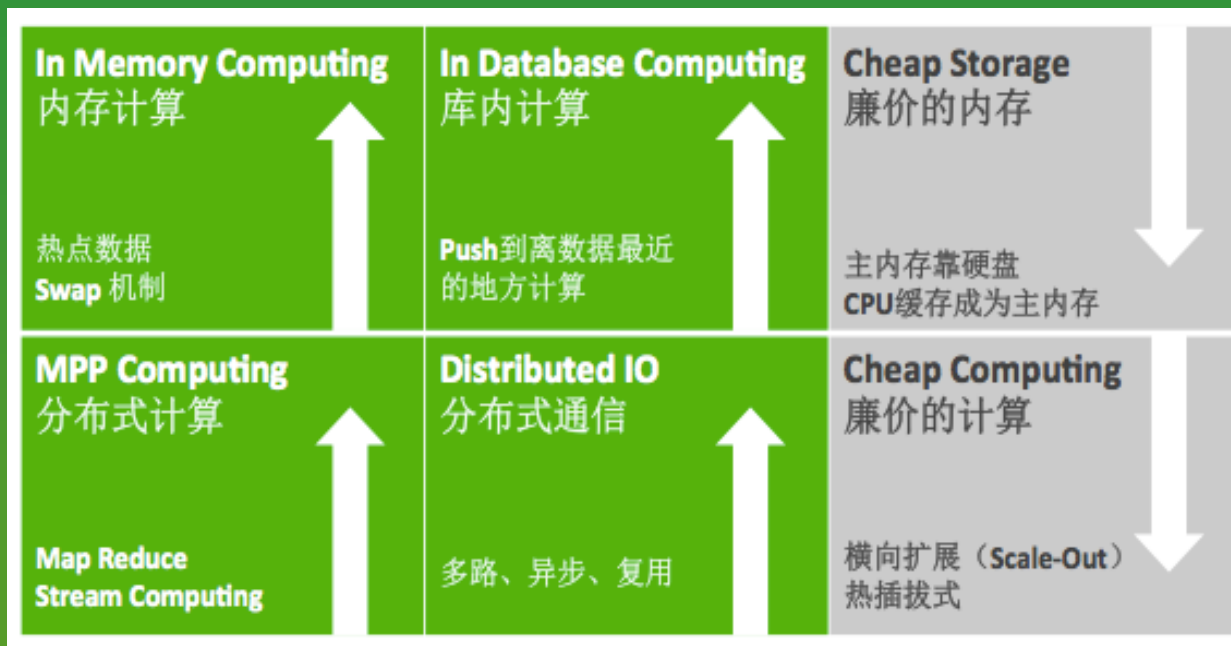


## 分布式计算

Map Reduce架构是一个通用处理并行计算的架构。利用多节点的并行计算提高计算效率。

## 分布式通讯

利用可复用的TCP/IP连接，多路，异步的信息传输等技术使得，中间计算结果在集群中稳定且高效地传输，是整个集群能否达到实时计算的关键。



一站式大数据分析平台  
打造人人都是业务专家的  
强大智慧企业

# 永洪 一站式大数据分析平台领军者

300多家合作伙伴，1000多家客户，20000多个付费用户



转发朋友圈，获取干货大数据技术白皮书



# Thanks!

大数据 小数据 一天实现可视化分析

---

永洪科技 助力智慧企业成就未来

---

