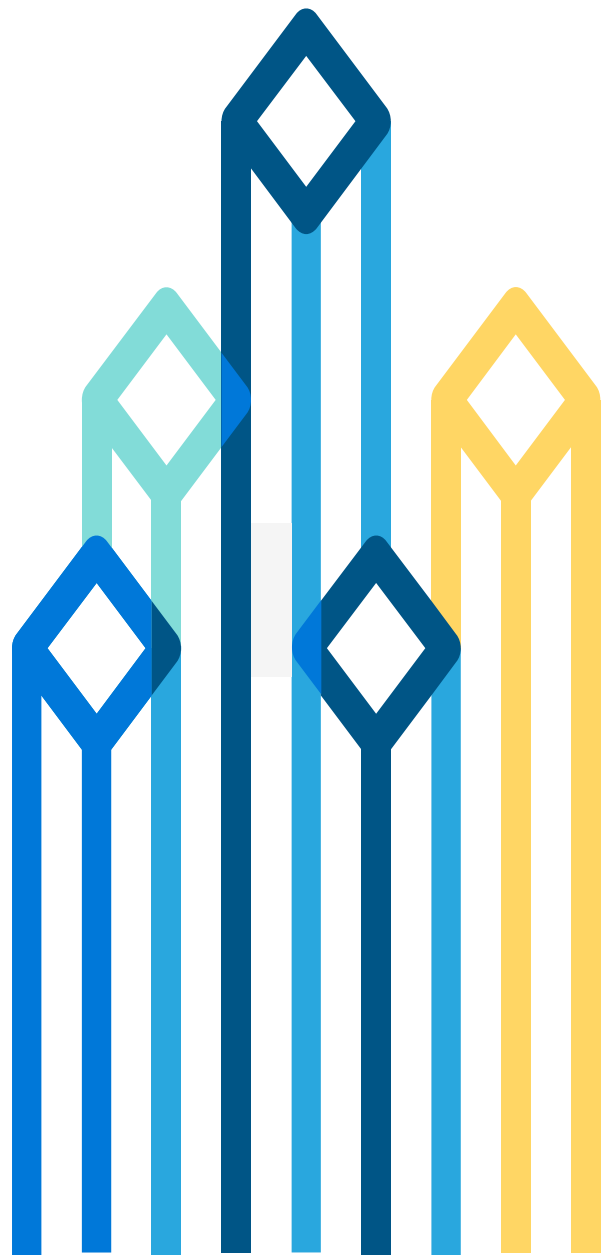


cloudera®

Cloudera 大数据行业应用

程志国
资深架构师



BIG DATA (大数据)

Volume (数量)

Variety (类型)

Velocity (速度)

= Value (价值)

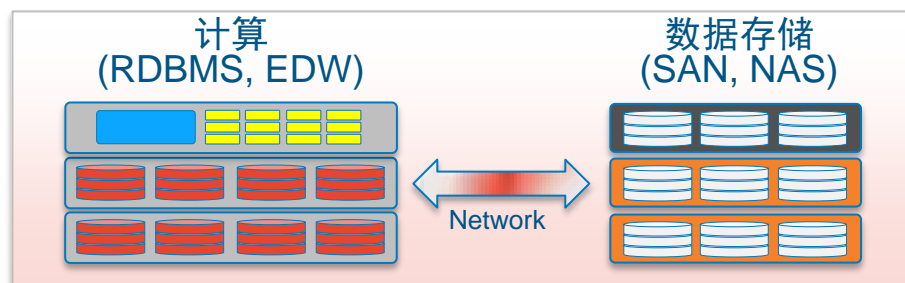
BIG DATA

= ?

HADOOP

Hadoop: 扩展性 & 灵活性 – 存储 & 计算

传统方式



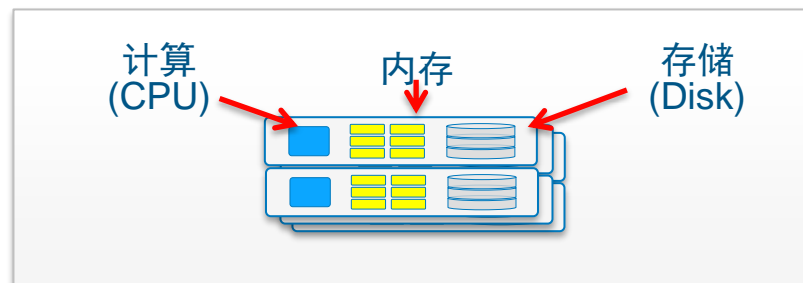
昂贵的、专用的、“可靠的”服务器
昂贵的**封闭软件**

- 扩展困难
- 网络成为不可避免的瓶颈
- 只能处理结构化 / 关系型数据
- 很难增加新的字段和数据类型

昂贵而遥不可及

\$30,000+ per TB

Hadoop方式



商业化的“不怎么可靠”的服务器
混合的**开源软件**

- 可无限平行扩展
- 网络不再是瓶颈
- 轻松摄取任何类型的数据
- 灵活的读取时检查数据类型的访问方式

经济且可以企及

\$300-\$1,000 per TB

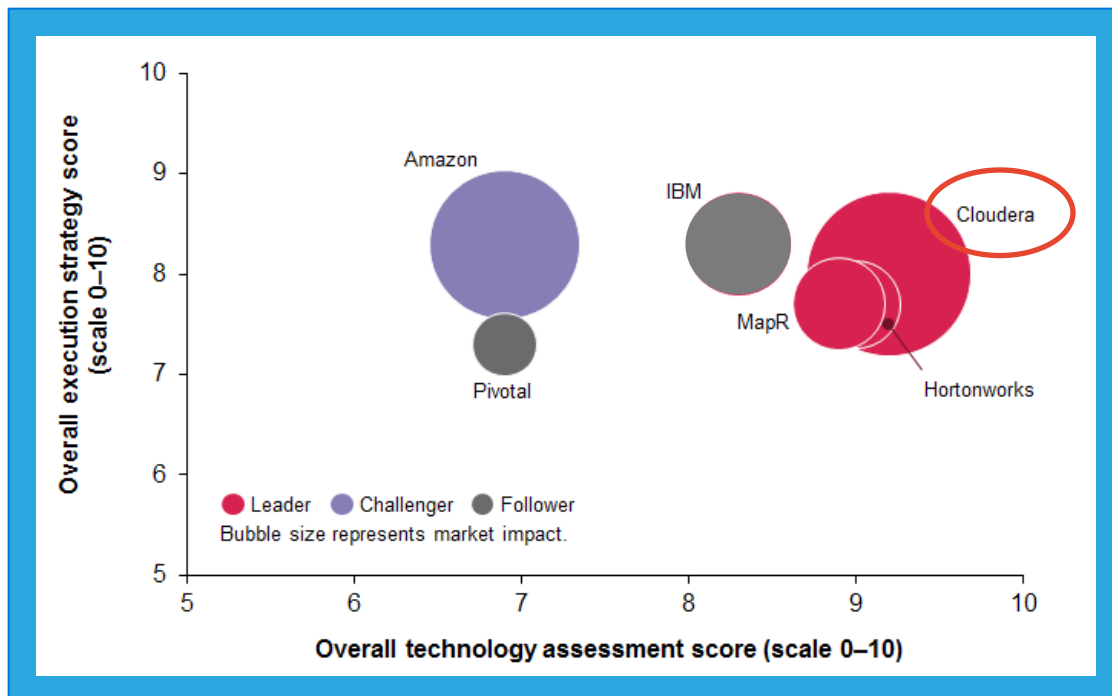
在传统数据库(RDBMS)时代,
应用创造新数据.

在Hadoop时代,
数据创造新的应用.

Hadoop十年



Cloudera是公认的Hadoop平台的领导者



Source: Ovum Decision Matrix: Selecting a Hadoop Platform, 2015

强大的客户基础

完整的数据安全和数据治理

完整的合作伙伴生态环境

Cloudera

2008年成立于硅谷

第一个企业级Hadoop产品提供商 ORACLE YAHOO! facebook Google

24x7 全球支持

3/5 财富500强企业客户

全行业客户(金融、电信、零售、能源、互联网、媒体等)

2200多个生态链合作伙伴培训和认证

Hadoop生态圈最大贡献者， Intel合作加速企业数据中心的革新

Cloudera中国 2014年9月成立，上海是大中华区总部，负责产品培训、专业技术服务和产品支持，在北京和广州有本地支持

它是谁？



Cloudera和Apache开源社区

Leaders Across the Hadoop Ecosystem



Doug Cutting:
Hadoop创始人，前Apache基金会主席

22 项目由Cloudera员工创建

100 Committer 席位覆盖Hadoop生态圈

上百万 生产环境部署节点，覆盖全行业

15 Hadoop相关的原理及架构类书籍

完整的合作伙伴生态环境

Applications

Operational Tools

Data Systems

Enterprise Data Hub

Process | Discover | Model | Serve

Security and Administration

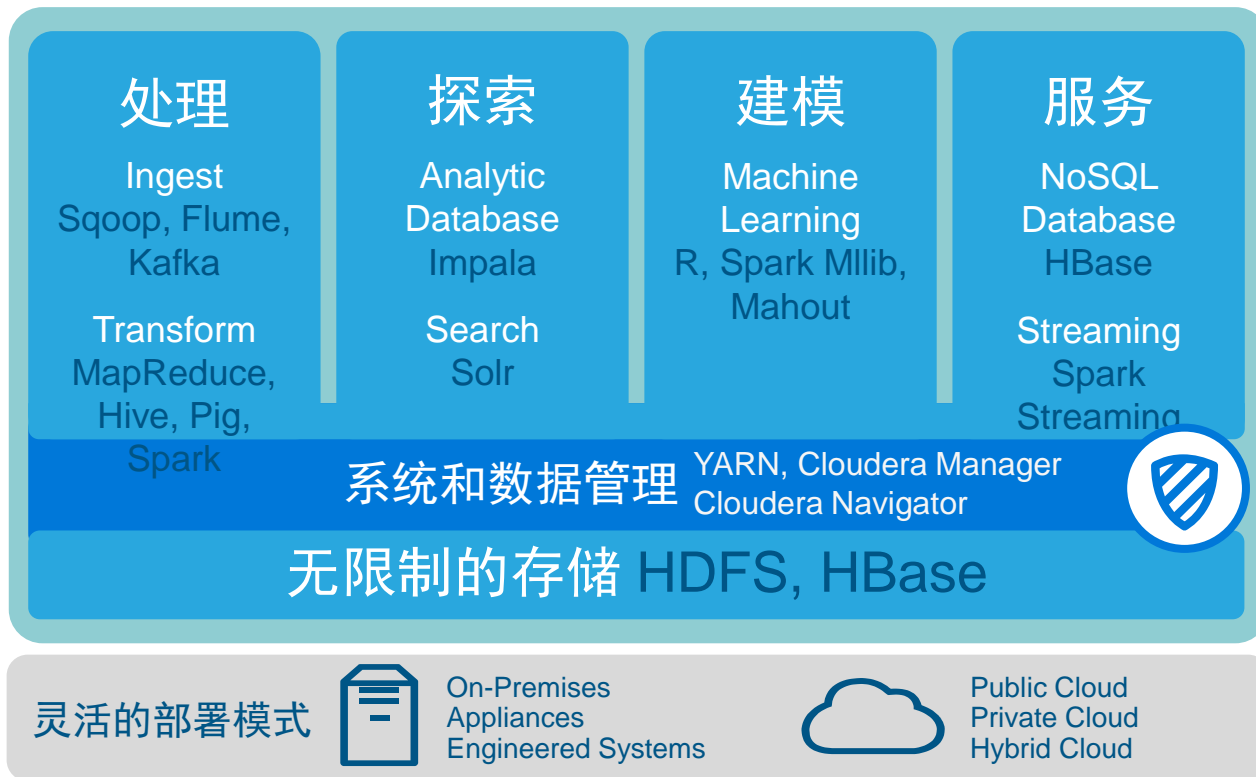
Unlimited Storage

Infrastructure

More than 2,200 partners

System Integration

Cloudera 企业数据中枢 CDH



统一的数据访问

统一的数据存储

统一的系统管理和安全

持续的创新能力

- Cloudera Manager
- Flume, Sqoop
- Search
- Impala
- Sentry
- Director, Navigator
- **Kudu**
- **Record Service**

全面的安全与管治



Cloudera全球客户精选...

Financial Services



Telecom



Healthcare & Life Sciences



Media & Technology



Retail & CP



Public Sector





金融服务

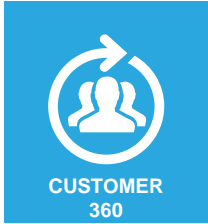
- » 数据安全性
- » 预防金融欺诈
- » IT 成本减少



第一个 PCI 认证的 Hadoop 平台

- 可优化EDW并提高对金融欺诈的监测和防
- 每天以PCI合规的方式安全保障10 PB 数据
- 安全信息事件管理(SIEM) – 监测对敏感数据集的访问，对用户行为有完整的审计轨迹。





电信

- » 服务改进
- » 流程改进
- » IT 成本减少



重新平台化1,600个运营数据库和系统

- 业务和消费者的数据分布在几十个不同的客户数据库中
- 一个日常的 ETL 工作 (处理10亿个客户记录) 以前需要24小时
- **提高数据速率15倍**
(用过去1/3的时间能处理5倍的数据)
现在只需1.5小时即可完成
- BT现在可以访问最新的数据, 所有客户数据都已集中。





- 零售 / 电商
- » 客户 360°
 - » 改进流程
 - » 更好的客户服务
 - » 预测性的分析

MARKS
SPENCER

零售客户行为的360° 视角

- 集成多种不同数据来源 (点击行为轨迹、店内POS、在线订单，以及社交媒体)
- 理解放弃的在线购物车行为
- 优化运营投资，不同渠道的业务收入关联
- 增加了客户洞见，用于改进供应链计划
- 提高了客户解答能力，并可预测退货





技术 / 制造业

- » 物联网
- » 产品改进
- » 服务改进



用NetApp Filers (“Phone Home”)收集的机器数据使预测性 / 主动性支持成为可能。

- 因为数据处理提高64倍，现在可以满足最严苛的支持SLA。
- 处理机器产生的数据源于600,000多个每周交易
- 模式匹配查询有助于侦测漏洞
- 磁盘上历史数据有助于分析产品问题的根源





RISK

政府行业

- » 恐怖行为侦测
- » 360° 视角
- » 试验性分析

美国政府 – 不同的政府安全部门 & 情报部门

提前侦测和防止恐怖威胁和袭击

- 必须识别全球互联网上可疑的行为, 提供给700个商业和联邦机构
- 提供实时信息、警告和指导, 以防御来自网络的进攻
- PB级别的平台用于网络安全分析





DATA-DRIVEN
PRODUCTS

医疗

- » 客户360°
- » 预测性分析
- » 改进的服务



可提前诊断败血症进行成功的治疗，
挽救生命

- 已经挽救了100多人的生命
- 多个系统的数据集中
- 多租户环境中超过2PB数据，支持100s客户
- 既安全又可供探寻分析
- 减少了患者重返医院



总结

1. 这既是技术的转变也是文化的转变。
2. 先爬、再走、然后跑。
3. 漫漫长路始于足下。

分析

Select

Insert

变化

Update

Delete



cloudera®

cloudera®

程志国

资深架构师

+86 186 0124 3530

