

# Cloudera简介

Mike Olson | 主席，联合创始人，首席战略官



# Cloudera公司简介

---

创始	2008年, 由前 ORACLE YAHOO! facebook Google 员工共同创始
員工人數	500人以上
世界级技術支持	24x7的全球工作人员 积极主动与预测技術支持方案
关键任务	数以千计的企业用户 300多位付费客户
最广泛的生态系统	900多个商业合作伙伴
Cloudera University	培训40,000人以上
开源领袖	Cloudera的员工是业界领先的开发者和提供商 我们与英特尔的合作将能成功地开拓市场

# Who Are We?

## 经营者

**Tom Reilly**  
首席执行官  
HP, ArcSight, IBM

**Jim Frankola**  
首席财务官  
Yodlee, Ariba, IBM

**Peter Cooper-Ellis**  
工程副总裁  
VMware, SpringSource

**Charles Zedlewski**  
产品副总裁  
SAP, BEA

## 创新者

**Mike Olson**  
主席，首席战略官，联合创始人  
Oracle, SleepyCat, Illustra, Informix

**Dr. Amr Awadallah**  
首席技术官，联合创始人  
Yahoo!

**Doug Cutting**  
首席架构师，Hadoop的共同创造者  
ASF Director, Yahoo!

**Jeff Hammerbacher**  
首席数据科学家，联合创始人  
Facebook

## 执行者

**Kirk Dunn**  
首席运营官  
PowerFile, Inktomi, NET, IBM

**Alan Saldich**  
营销副总裁  
Riverbed, Inktomi

**Paul Bedhun**  
技术服务副总裁  
TIBCO, Inktomi, Sybase

**Sarah Sproehnle**  
教育服务副总裁  
MySQL, Sun

**Tim Stevens**  
业务及企业发展副总裁  
QuinStreet, Borland

# 强大的创新记录

2008

CLOUDERA创立  
MIKE OLSON  
AMR AWADALLAH  
JEFF HAMMERBACHER &  
CHRISTIAN BISCIGLIA



2009

CLOUDERA发行CDH  
第一个商业化  
APACHE HADOOP分发版



2011

CLOUDERA的达到  
100运行客户



2012

CLOUDERA  
ENTERPRISE 4  
成为企业中的  
HADOOP标准



2013

CLOUDERA IMPALA  
CLOUDERA NAVIGATOR  
CLOUDERA SEARCH



2014

企业级数据中心



2009

HADOOP创始人DOUG CUTTING  
加入CLOUDERA

2010

CLOUDERA MANAGER  
第一个Hadoop管理应  
用界面

2011

CLOUDERA UNIVERSITY  
扩展到140个国家

2012

CLOUDERA CONNECT  
达到300多个  
商业合作伙伴

2013

TOM REILLY成为首席执行官  
CLOUDERA CONNECT包含800  
多个商业合作伙伴

# 驱使Hadoop前进

---

## 团队



Doug Cutting:

Hadoop的创始人，Apache软件基金会董事会成员

---

19 成立项目

---

75<sub>多位</sub> Apache项目提交者

---

上万个 受管理节点，跨越各个产业

---

发行5本 与Hadoop相关技术书籍



# 广泛存在于企业界

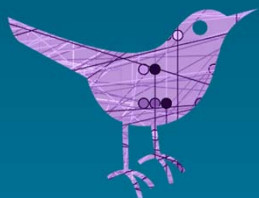
## 成绩和记录



每日有200亿在线事件是由Cloudera导入系统



Cloudera驱动美国70%的智能手机



每天有2.5亿推文由Cloudera进行过滤成为可操作的业务见解



四所顶级金融机构已经使用基于Cloudera的标准化数据管理平台



世界领先的技术公司已使用Cloudera为全球标准化的单一大数据平台



电信，国防，媒体，银行和零售商各产业中有五分之三团体机构使用Cloudera

# Hundreds of Deployments



# 成为信息驱动的商务

企业级数据中心简介





# 扩展数据需要新的方法

## 1980年代

将数据带至计算平台



以流程为中心  
企业应用：

- 已结构化数据为主
- 只包含内部数据
- 只包含“重要”的数据

相对数据量和复杂性



## 今日

将计算平台带至数据



以信息为中心的企业  
使用的所有数据：  
所有类型的多层结构，  
包含内部和外部数据

# 传统数据分析：将数据带至计算平台

4

## 复杂的架构

- 许多特殊用途的系统
- 频繁的数据移动
- 缺乏数据的整体认识

3

## 高数据分析的成本

- 现有系统负荷量过大
- 缺乏灵活性
- “商业智能积压”

2

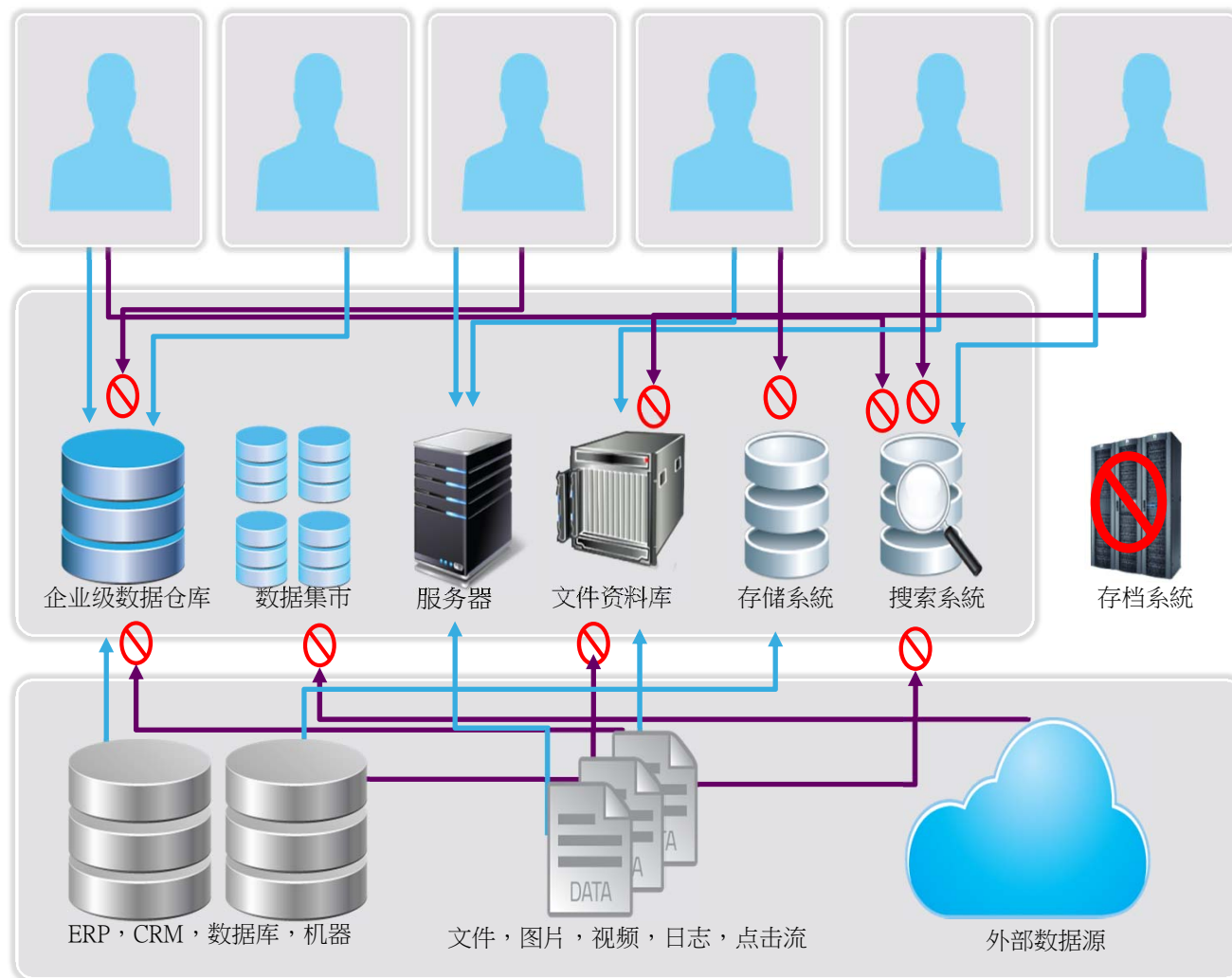
## 数据存取速度慢

- 需要前期建模
- 数据转换速度慢
- 转换程序常遗失数据

1

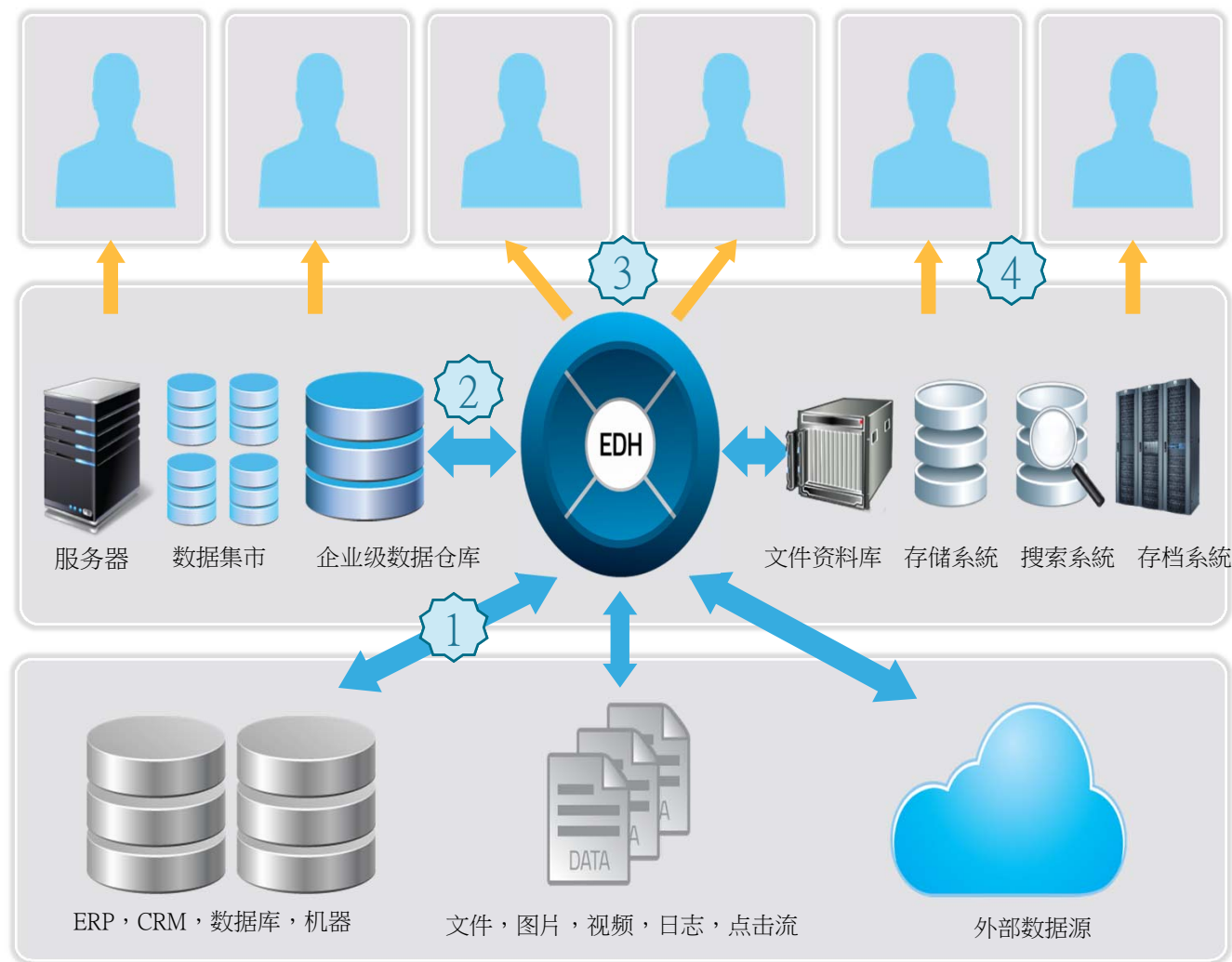
## 高数据遗失率

- 许多资料数据未能被使用
- 风险与合规
- 高存储成本



# 大数据分析：将计算平台带至数据

- 4 **多样化的分析平台**
  - 将应用程序带至数据
  - 结合多样化的工作於常见的数据（即SQL +搜索）
  - 真正的敏捷性分析
- 3 **自助服务BI探索**
  - 简单的搜索+ BI工具
  - “Schema on read” 灵活性
  - 降低BI用户请求积压
- 2 **保持Staging**
  - 所有的分析使用单一数据源
  - 保持变换后的数据的状态
  - 更快更便宜
- 1 **主动合规存档**
  - 全保真原始数据
  - 不定时间，任何来源
  - 成本最低的存储

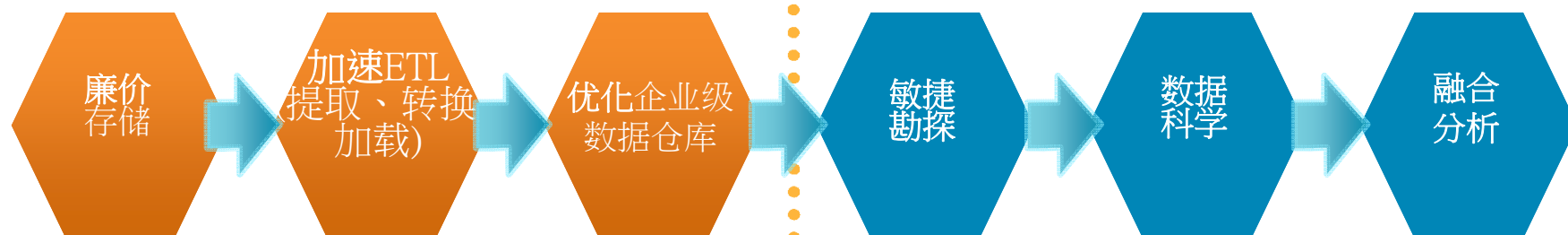


# 实现全部潜能的旅程

从战略推进到实现高投资回报率的最佳方式

运营效率  
(更快, 更大, 更便宜)

应用转型  
(新业务价值)



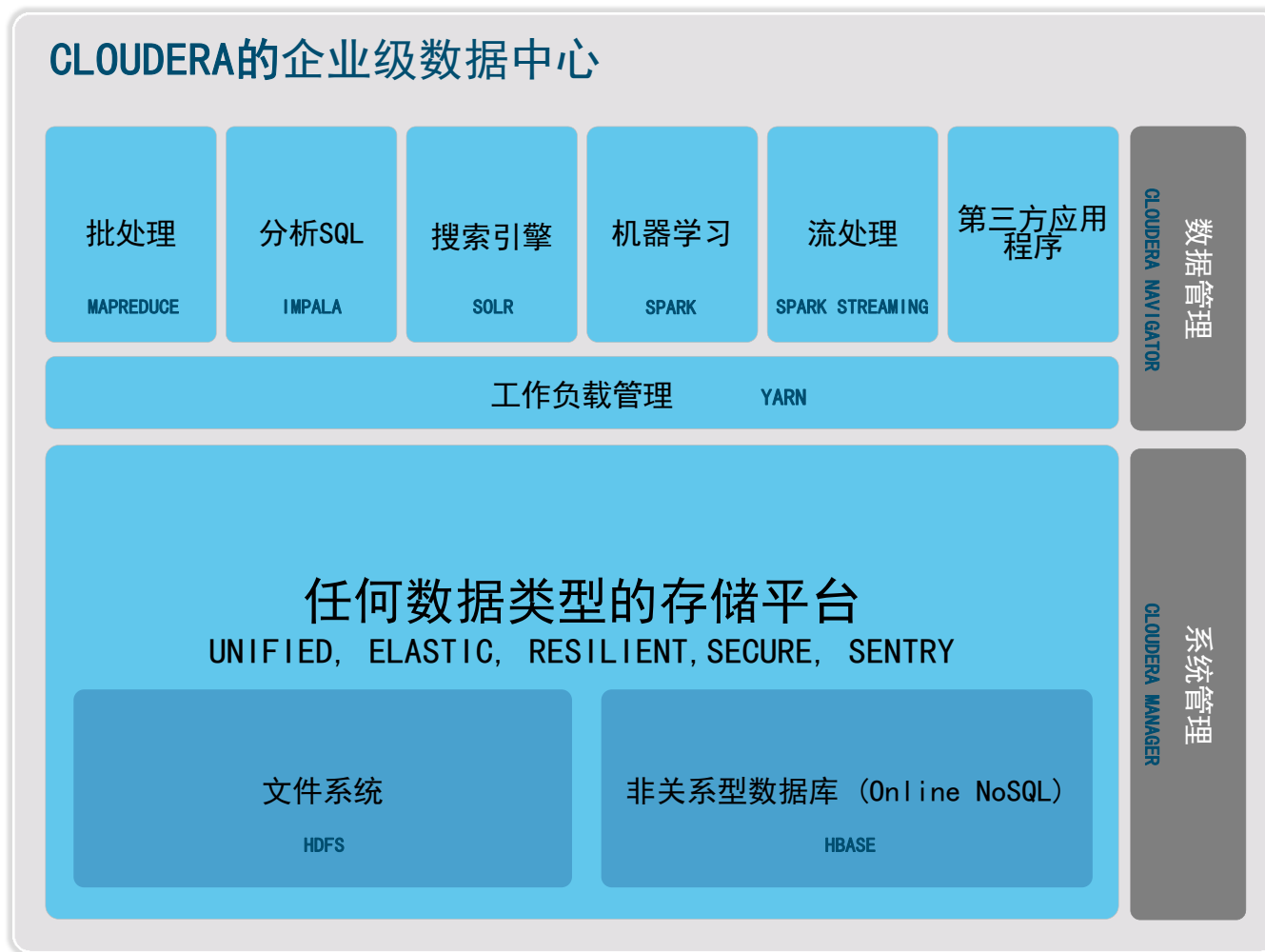
I T 部门

业务部门



# 从Hadoop至企业级数据中心

开源 可扩展性 灵活 成本效益	✓
管理	✓
开放式架构	✓
安全及管治	✓



# 适用于各个行业

## 汽车

汽车传感器报告位置以及问题



## 电信

每天有超过2.9亿智能手机与Cloudera进行联络



## 包装消费品

运动服装业使用Cloudera追踪和分析2百多万位运动员生活方式与习惯



## 金融

信用卡业务使用Cloudera避免了20亿美金欺诈费用



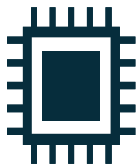
## 教育 & 研究

实验  
传感器分析



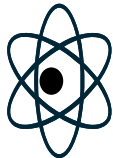
## 高技术/工业制造

制造质量  
保修分析



## 生命科学

临床试验  
基因组学



## 媒体/娱乐

Cloudera追踪超过40万个家庭的频道变化



## 在线服务/社交媒体

人与职业的匹配  
网站优化



## 医疗

病人传感器，  
监控，  
电子健康档案



## 石油与天然气

钻井勘探传感器数据分析



## 零售

使用Cloudera分析十亿客户的客流量



## 旅游 & 交通

Cloudera的系统支持80%以上的所有在线旅游交易



## 公用事业

使用Cloudera降低9000万家庭电费



## 执法与国防

Cloudera协助供应为超过1 PB的战场情报给美国陆军





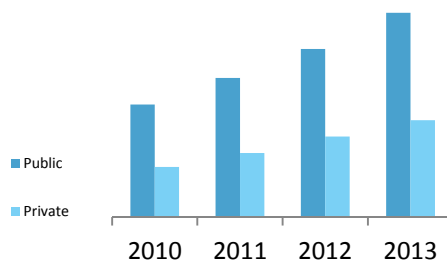
# Cloudera与英特尔联盟合作



# 全景图：数据中心的转折点

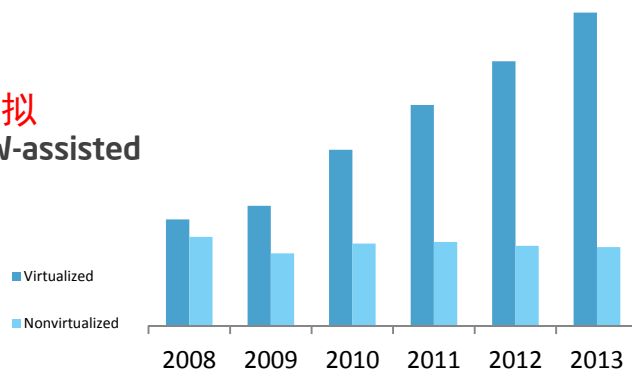
3

从集群到云端  
ASIC到IA/Fabric



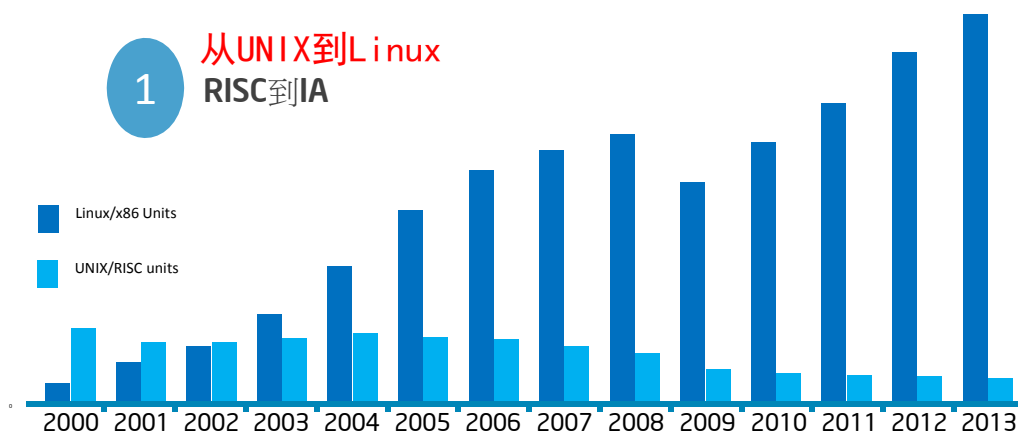
2

从实体到虚拟  
SW-only到HW-assisted



1

从UNIX到Linux  
RISC到IA



4

大数据

“2000年英特尔看到Linux来临，就对Red Hat重大投资；2005年我们看到虚拟化来临，就投资VMware；2008年我们对超级扩展计算进行重大投资。

我们认为大数据和Hadoop会是一个崭新的最高峰。”

Diane Bryant, SVP & GM  
Data Center Group, Intel



# 大事件：英特尔与Cloudera的合作联盟

---

- 英特尔7.4亿美元投资Cloudera
  - Cloudera成为英特尔最大的数据中心创业投资，代表了英特尔对大数据的承诺
- 英特尔与Cloudera通过开源驱动创新
  - 经由基础技术合作加速Hadoop的演变
  - 让开源开发者在Hadoop平台上进行创新
- 英特尔使得Hadoop在IA架构产品上的运行达到最优性能
- 英特尔与Cloudera协作建立广泛的大数据解决方案合作伙伴生态系统

# 大目标: 汇聚在独一开源平台



- 唯一从硬件层面增强性能和安全性的分发版
- 领先的安全功能，包括加密，访问控制和审计
- 1000位开发人员长期致力于开源软件：Linux，KVM，Xen，Java，OpenStack，Hadoop
- 50 Hadoop开发人员
- 12开放源代码提交者

- 最稳定，兼容，成熟的Hadoop发行版
- 领先的SQL功能与性能(Impala)
- 最周全彻底的管理功能
- 150 Hadoop开发人员
- 100开放源代码提交者

# 使用英特尔架构提高Apache Hadoop的性能

提升速度  
50%

计算



提升速度  
80%

存储和内存



提升速度  
50%

网络



此测试是基于1TB排序于10个节点集群总共所需时间

# 共同领导促进合作伙伴生态系统



- 半导体市场的领导者
  - 长期以来成功的历史：投资于软件平台以及与软件平台合作
  - 全球影响力
  - 在中国和印度拥有领先市场的Hadoop发行版
- 拥有最多付费和免费Hadoop用户
  - 始终如一地提供领先业界的软件功能
  - 为中国和印度地区之外大数据管理系统市场的领导者



# 总结：快速达到商业目标，提高安全性，降低复杂性



## 透过开源软件加速创新

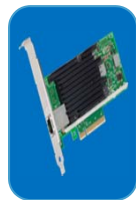
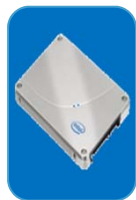
- 维持大数据开放平台
- 继续加强Apache的Hadoop和相关项目

## 使Hadoop在英特尔架构上达到最佳运行性能

- 横跨计算，存储，网络来优化性能
- 透过硬件增强确保平台的安全性

## 培育大数据合作伙伴生态系统的进化

- 建立使用模式以及行业标准的基准测试
- 开发参考架构和全行业的解决方案



# Cloudera的对外公告

---

- Cloudera将建立在中国的业务营运
- Cloudera将在中国进行直接销售，技术支持和技术服务团队
- Cloudera和英特尔将合作提供中国最强大，最安全，功能最完善，性能最高的Hadoop发行版
- Cloudera和英特尔将合作以提供最佳的技术支持, 服务中国客户



**cloudera<sup>®</sup>**

Ask Bigger Questions

谢谢!

Mike Olson

@mikeolson

mike.olson@cloudera.com