英特尔中国研究院

大数据分析和数据安全

@周鑫

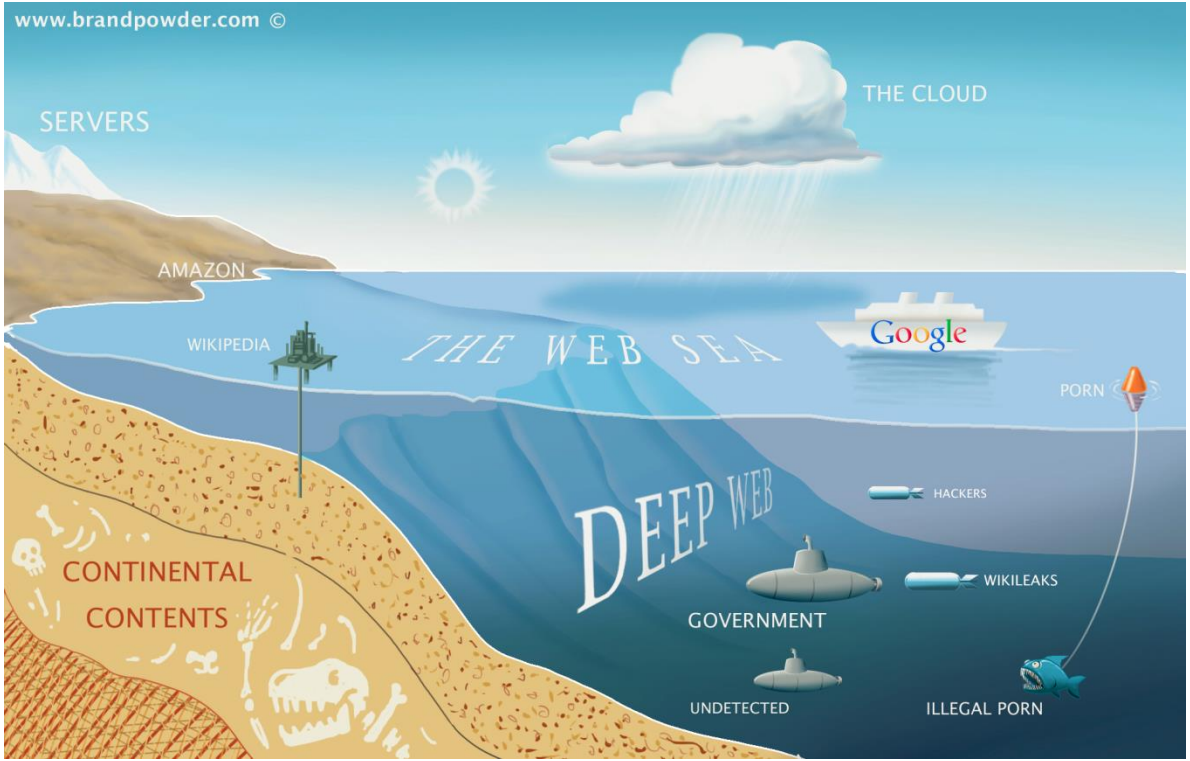
英特尔中国研究院

内容

- 大数据交易的必要性、数据安全的重要性
- 工作分享
  - 安全计算云——安全Spark平台
  - 支持隐私安全的数据分析平台

工作重点

- 关注实体行业
- 数据分析带动应用
- 数据必须走上开放、共享和交易之路



数据分析行业应用

乘法效应 + 外部效应

- 大企业
  - Google
  - Intel
  - GE
- 实体行业开始探索
  - 餐饮
  - 服装服饰
  - 零售

大数据需要数据交易

- 局部和全局
  - 大小定义的差距
    - 100MB/年 vs. xPB/年
  - 局部和全局
    - 地理范围限制 vs. 虚拟空间无限

- 集中和分散
  - 应用场景
    - 实体耦合 vs. 虚拟空间聚合
  - 拥有权和使用权
    - 不明晰

大数据需要隐私保护

- 数据利用和保护
  - “隐私”和价值 —— 敏感数据
- 全局统计特征 和 个体特征
  - 全局统计模型、分类模型
  - 数据脱敏（匿名化，去标识化），同态加密

内容

- 大数据交易的必要性、数据安全的重要性
- 工作分享
  - 数据咖啡馆 / Data Coffeehouse
    - 安全计算云——安全Spark平台
    - 支持隐私安全的数据分析平台

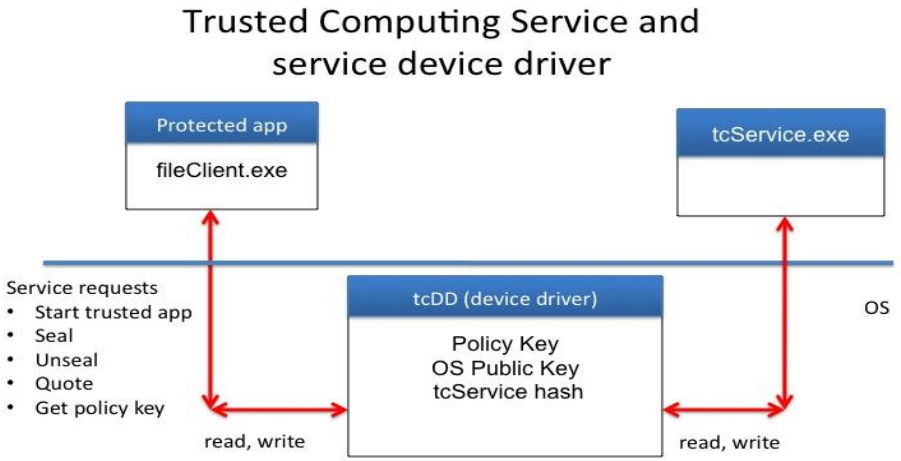
安全计算云

结合 Cloud Proxy 技术

- Intel Science and Technology Center for Secure Computing， UC, Berkeley
- Github 开源

安全的最小操作系统和可信任计算服务

- Simplified Linux
- OS support for Tao services
  - tcService
  - Kernel changes
    - encrypted swap
    - application identity tied to Tao services
- Initramfs
  - Encapsulated initial file system
  - Measured at boot
  - Contains apps
- Configuration
  - Module loading restrictions
  - Don’t mount file systems as trusted



本页材料来自CloudProxy Retreat

CloudProxy 特点

在云环境下支持安全的分布式计算

- 一致性可计算
- 数据一致性和保密性

简单可靠地保护模型

- 故障安全 (completed operations accompanied by “proof of correct operation”)
- 和现有的基础设施无关 (e.g.- CA’s)
- 能有效防护潜藏在云内部的威胁

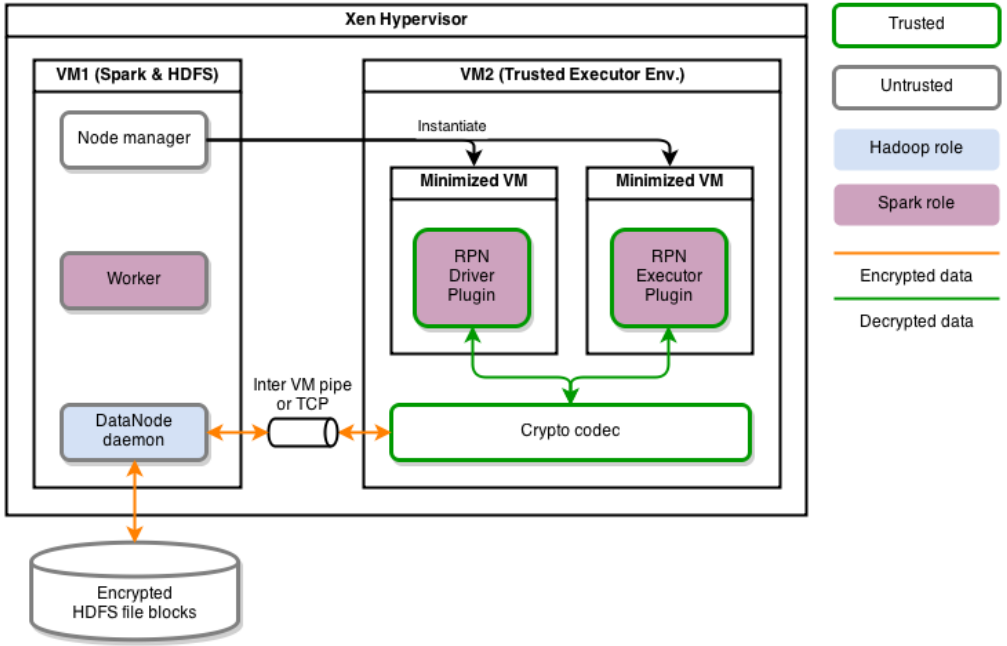
简单但是完整的应用程序框架

- 简单而安全的应用部署方案，秘钥安全
- 兼容目前各种云计算解决方案

本页材料来自CloudProxy Retreat

安全计算云——安全Spark

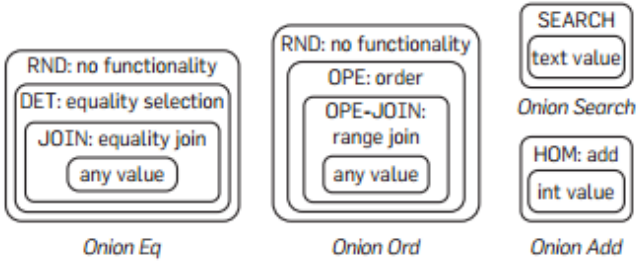
- 可以结合HDFS加密机制
- 无缝结合Spark 运行时平台
- 程序安全
- 数据安全



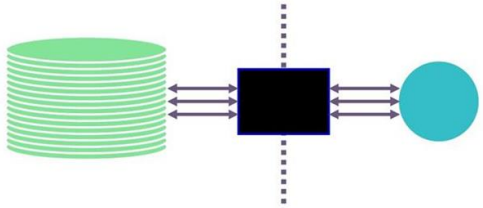
数据共享的数据保护机制

点对点数据共享下数据隐私保护：

- 具备基本的数据保护能力
- 可扩展型受到限制



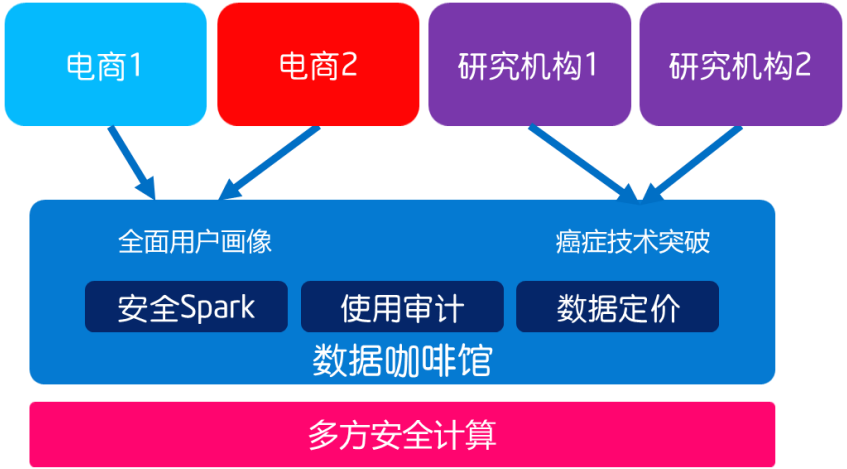
CryptDB



Differential Privacy

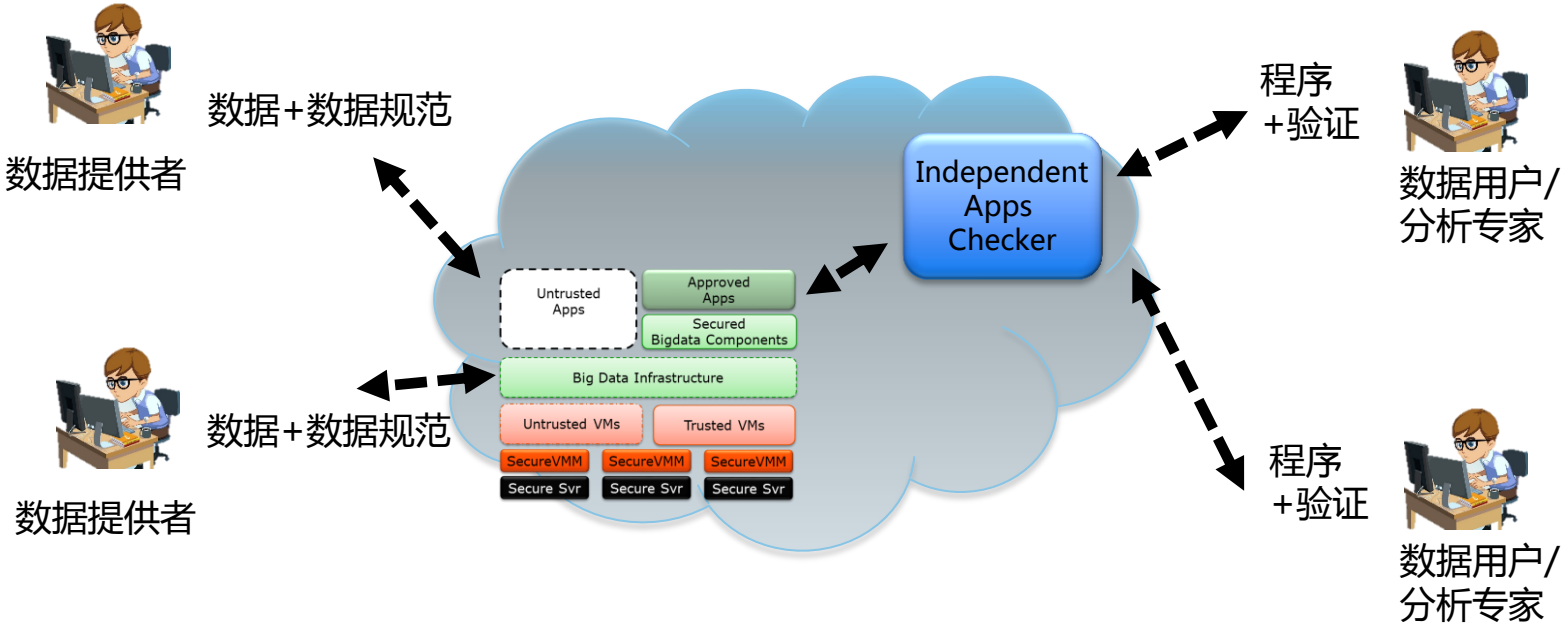
灵活的安全数据交换平台

- 多方数据融合，1+1 > 2
- 相互信任的窘境
  - 数据分析平台
  - 数据分析算法
- 持续性的数据交易
  - 数据的自动定价



数据咖啡馆

支持多用户的安全大数据分析



总结

- 随着大数据存储和分析技术的迅猛进展，数据交换和数据共享已经来临，数据保护技术亟待突破
  - 实体行业的问题更突出
- 工作方向
  - 安全的云计算平台
  - 灵活的数据交换平台

敬谢聆听

欢迎大家共同参与