Tencent | DataFun.

# 大数据协同中的隐私与可靠性保护

腾讯计费平台部　　张韬

# 目录 CONTENT

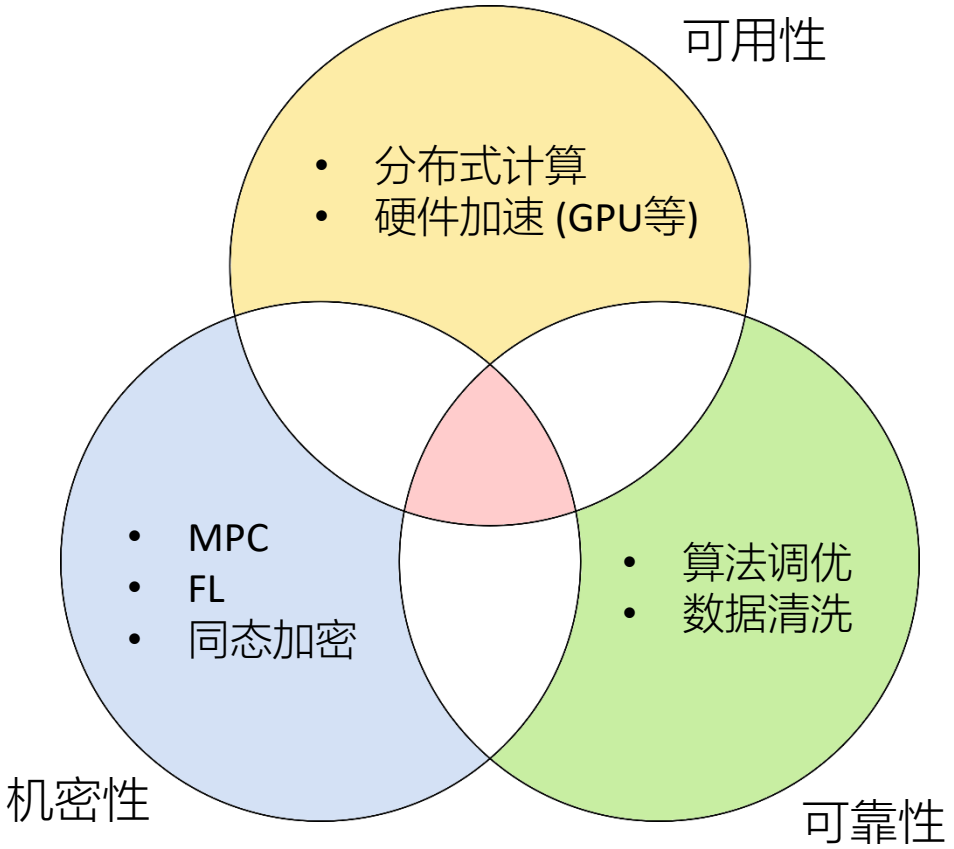
Tencent | DataFun.

# 01

# 数据协同中的安全问题

# 数据安全与实用性的两难

- 数据不能明文跨域互通
  - 数据成为资产、生产资料
  - 个人信息保护法、GDPR
- 单一平台无法获得全面的数据
  - 特征缺失
  - 样本偏差
- 密码学算法性能开销较大
- 计算逻辑正确性难以保障
  - 分布式计算、外包计算



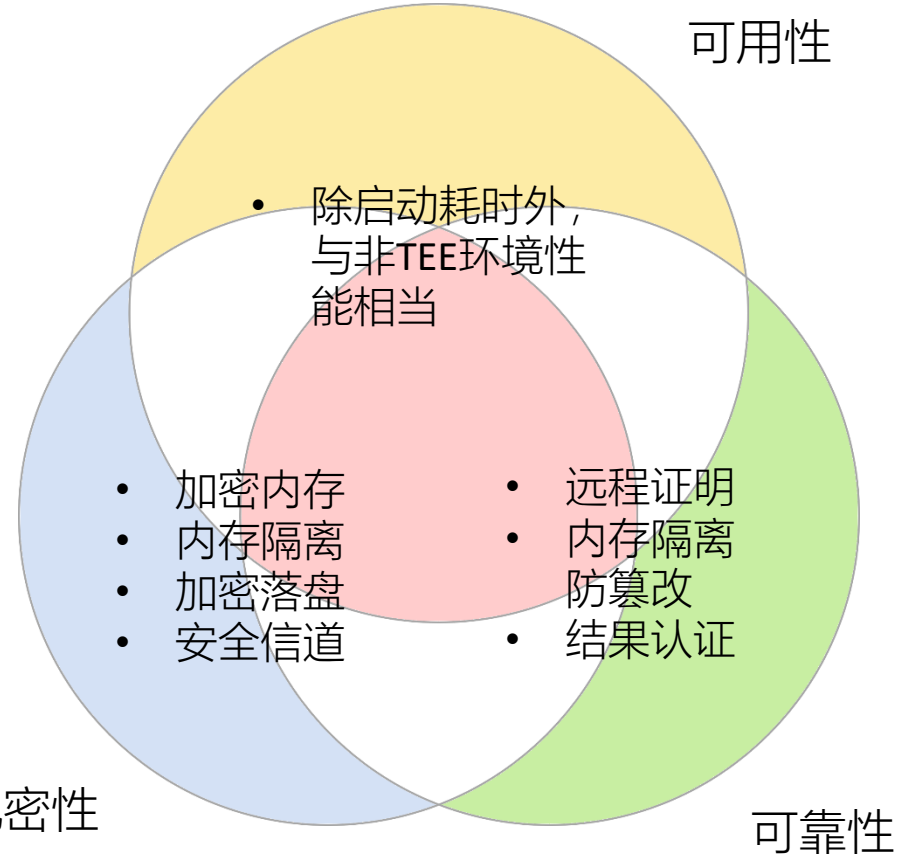
Tencent | DataFun.

# 02

# 可信执行环境 (TEE) 简介

# 可信执行环境 (TEE) 的安全特性

- 数据隐私
- 数据、计算逻辑、结果完整性
- 性能



# TEE 硬件

- Intel SGX
  - 配套基础设施完善：远程证明（IAS、DCAP）、开发者接入
  - 产品普及
- ARM TrustZone
  - 过强的信任假设：需要信任固件预置的 TrustedOS
  - 缺乏远程证明等基础设施（华为鲲鹏服务器上搭载了华为自研远程证明机制）
- RISC-V Keystone/Sanctum、蓬莱
  - 暂未有成熟服务器推出
  - 设计类似SGX
- AMD SEV
  - 专注于虚拟机隔离场景
  - 远程证明存在安全漏洞，且不具备逻辑完整性的保护 [BWS,CCS'19]
  - 寄存器、IO存在数据泄露的漏洞 [HB2017,SIGPLAN Notices 2017][LZL+,Security'19]
  - 资源调度仍然依赖Hypervisor

# TEE 的接入方式对比

- TEE SDK
  - 应用级别接入，对业务有侵入式改造
  - 攻击面最小
- libOS
  - 基础设施级别接入，业务少量感知
- 虚拟化：WebAssembly、Docker、VM
  - 基础设施级别接入，业务可无感知

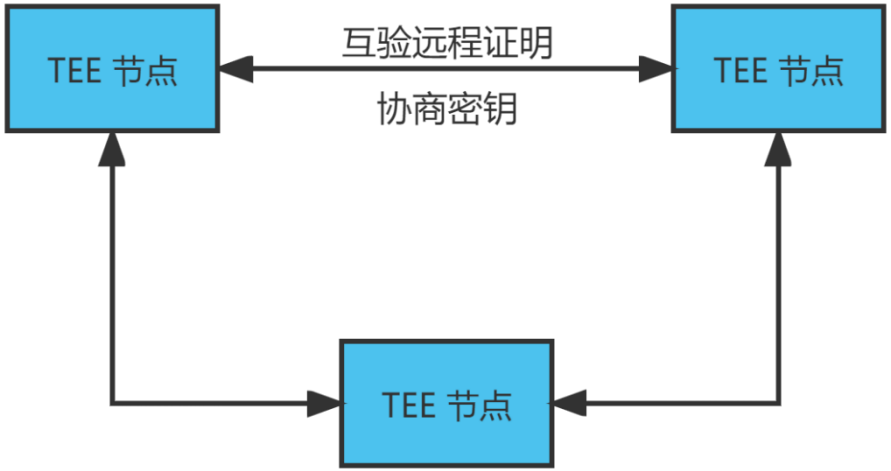
Tencent | DataFun.

# 03

# TEE 上的分布式计算

# 多节点协同

- 一般步骤
  - 节点间两两互验远程证明
  - 每两个节点间通信密钥不同
- 缺点
  - 交互复杂度为 $O(n^2)$
  - 需要维护其他节点信息

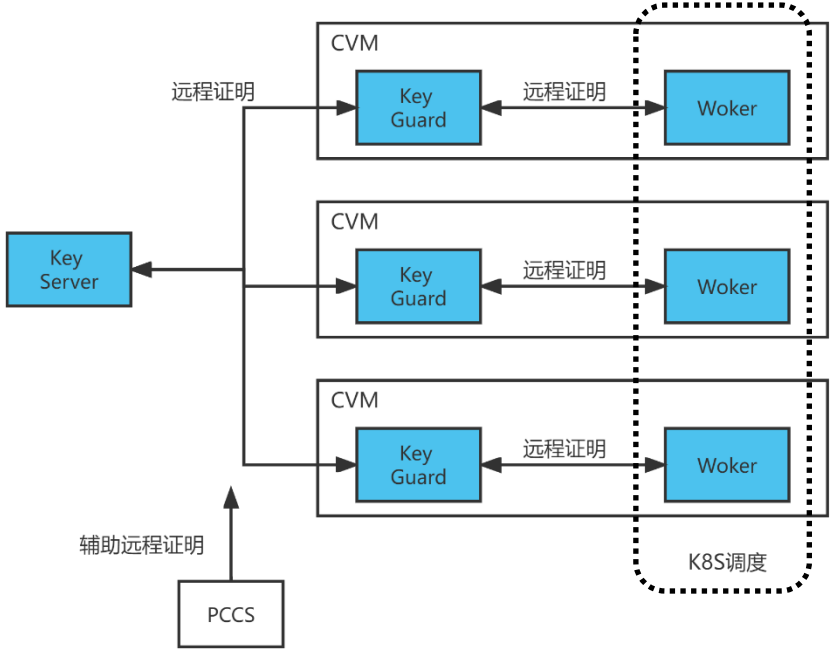


# 多节点协同

## 节点间密钥管理

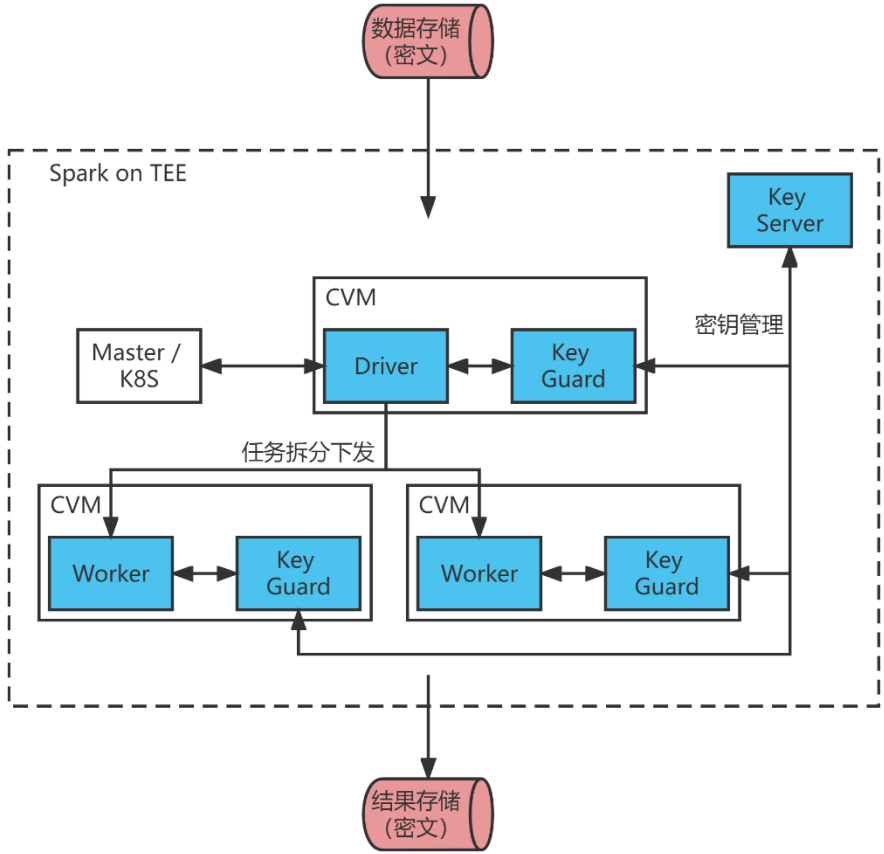
1. 数据拥有者通过远程证明验证Key Server，向Key Server注册一对加密公私钥对。
2. 数据拥有者向Worker提供签名授权使用数据。
3. Worker向Guard展示签名，并请求对应解密私钥。
4. 若Guard本地没有对应私钥，则向Key Server查询。
5. Guard发送解密私钥给Worker。
6. Worker在运行中，可使用查询到的解密私钥解密来自对应数据拥有者的入参数据。

- Key Server集中管理密钥和授权
- Key Guard为同一个物理节点或CVM上的Worker提供密钥查询和授权的常驻服务
- Worker为执行业务逻辑的协同计算节点
- 以上3个组件均运行在TEE可信域内



# TEE 上的 分布式计算

- 适配TEE硬件
  - Spark、 TensorFlow等计算框架适配
- 计算节点间密钥同步
  - Key Server验证节点远程证明
  - Key Server-Guard密钥管理
  - 使用者授权密钥使用
- 计算集群对接密文数据库



# 性能对比

- 区块链隐私交易场景
  - 全同态加密：性能较低
  - 零知识证明、同态加密：支持场景有限
  - 其他MPC：性能、泛用性有提升空间

| | 合约执行时间 | 链下辅助执行时间 |
|---|---|---|
| 全同态加密 | 6 ms | 无 |
| Bulletproofs<br>零知识证明 | 2.48 ms | 19.87 ms |
| Paillier<br>同态加密 | 0.7 ms | 无 |
| SGX隐私合约 | 0.34 ms | 无 |

隐私交易性能对比

- 隐私求交集 (PSI) 场景
  - 密码学算法：性能、泛用性有提升空间

| | | TEE | BlindRSA | | OT | |
|---|---|---|---|---|---|---|
| | | | 512 RSA | 1024 RSA | 512 RSA | 512 RSA+SM4 |
| 20 executor * 4G | 1亿 X 1亿 | 8min26s | 26min21s | 112min34s | 15min44s | |
| 40 executor * 4G | | 5min26s | 17min21s | 75min8s | 10min35s | 15min59s |
| 20 executor * 4G | 10亿 X 10亿 | 40min | | | 153min | |
| 40 executor * 4G | | 25min | 127min | 552min | 83min | |

PSI性能对比
128MB内存旧机型
单TEE节点

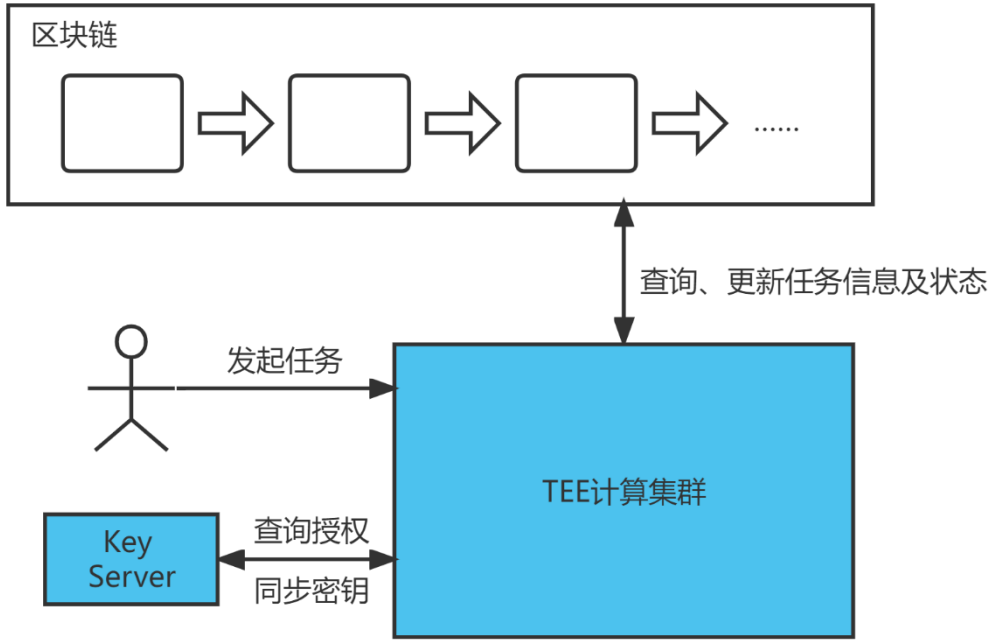
Tencent | DataFun.

# 04

# 区块链协调的数据协同

# 数据协同中的监控与协调

- 数据用途
  - 存证计算逻辑（模型）
  - 存证数据提供方
- 数据用量
  - 存证数据哈希或签名
  - TEE统计数据量
- 任务状态及结果
  - TEE更新任务状态、结果上链
  - TEE的输出均由TEE签名认证



# 腾讯云数链通

- 政务
  - 各部门、各地区数据共享
- 金融
  - 征信、风控等


Tencent | DataFun.

# 非常感谢您的观看

Tencent | DataFun.