RSA CONFERENCE
CHINA 2012
RSA信息安全大会2012

THE GREAT CIPHER

MIGHTIER THAN THE SWORD

伟大的密码胜于利剑

在映射化简 (Map Reduce) 中嵌入安全性和信任基本形式

Samir Saklikar
EMC 的安全产品分公司 RSA

RSA CONFERENCE CHINA 2012
RSA信息安全大会2012

专题会议 ID：TC-2003
专题会议分类：高级

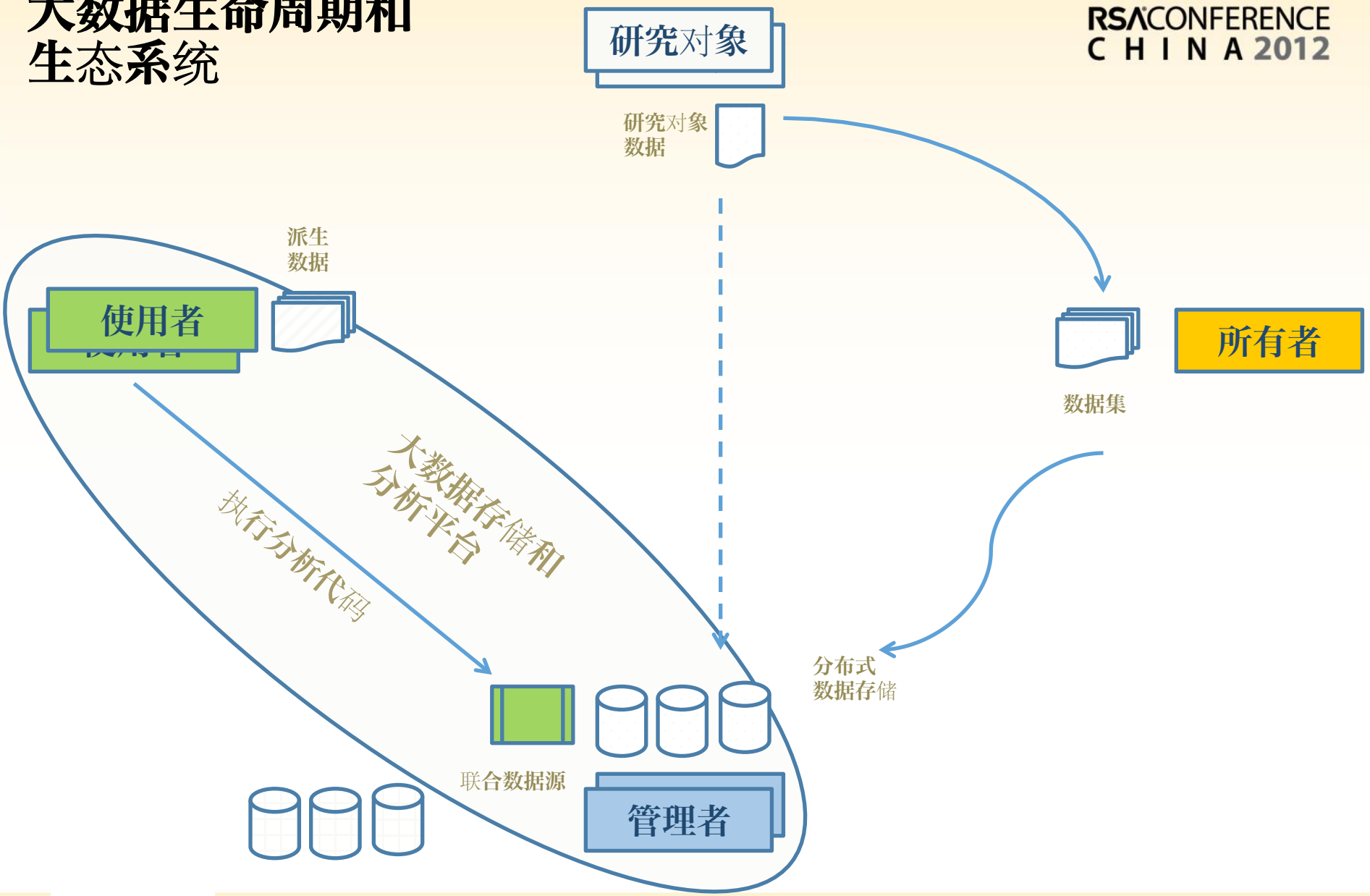
议程

- 大数据生态系统概述
  - 不同利益相关者以及关于信任要求的观点
- 大数据中的不同安全要求
- 映射化简自检框架的要求与建议
- 将自检框架用于安全基本形式
  - 访问控制
- 结论

大数据参与者

- 数据研究对象（M 个研究对象）
  - 数据适用于哪些人？
- 数据所有者（1 个所有者）
  - 可能不同于数据研究对象。
    - 例如，拥有关于其用户/系统的数据的企业
  - 从绝对所有者到保管人的不同级别所有权
  - 不同级别的数据使用自主权
- 数据管理者（N 个管理者）
  - 存储，为访问提供便利和允许对数据进行处理
  - 可能与数据所有者重叠
- 数据使用者（P 个使用者）
  - 对数据价值感兴趣
  - 通常，研究对象是间接使用者

大数据生命周期和生态系统



数据研究对象 – 资产和顾虑

- 资产
  - 配置文件数据（用户偏好）
  - 行为数据（使用情况/使用模式）
  - 标明信息特征（终结点标识符）

- 顾虑
  - PII 泄露，导致身份盗用/隐私暴露
  - 敏感信息泄露，导致恶意使用
  - 错误的分析，导致错误的服务个性化
  - 缺乏对数据便携性和生命周期管理的控制
  - 数据货币化方面的投资非常少（如果有）

数据所有者 – 资产和顾虑

- 资产
  - 支持业务功能的大型数据集
    - 与用户/员工相关
    - 与知识产权相关
    - 业务功能
    - 信息技术
- 顾虑
  - 数据泄露和/或滥用，导致法律责任
  - 数据泄露/损坏，导致业务损失

数据管理者 – 资产和顾虑

- 资产
  - 数据管理基础架构
  - 数据分析基础架构
- 顾虑
  - 数据泄露和/或滥用，导致法律责任
  - 数据损坏，导致业务损失

数据使用者 – 资产和顾虑

- 资产
  - 数据分析功能
  - 固有数据语义
- 需求和顾虑
  - 分析功能泄露，导致 IP 丢失
  - PII 在分析过程中泄露，导致法律责任
  - 数据损坏，导致错误结果
  - 需要对各种丰富数据源的无缝访问

大数据安全观点的变化

隐私和控制拔河比赛

数据所有者
数据管理者
数据使用者

数据研究对象

数据使用者

审核机构，CIRT

数据研究对象

企业用户

安全服务提供商

数据管理者

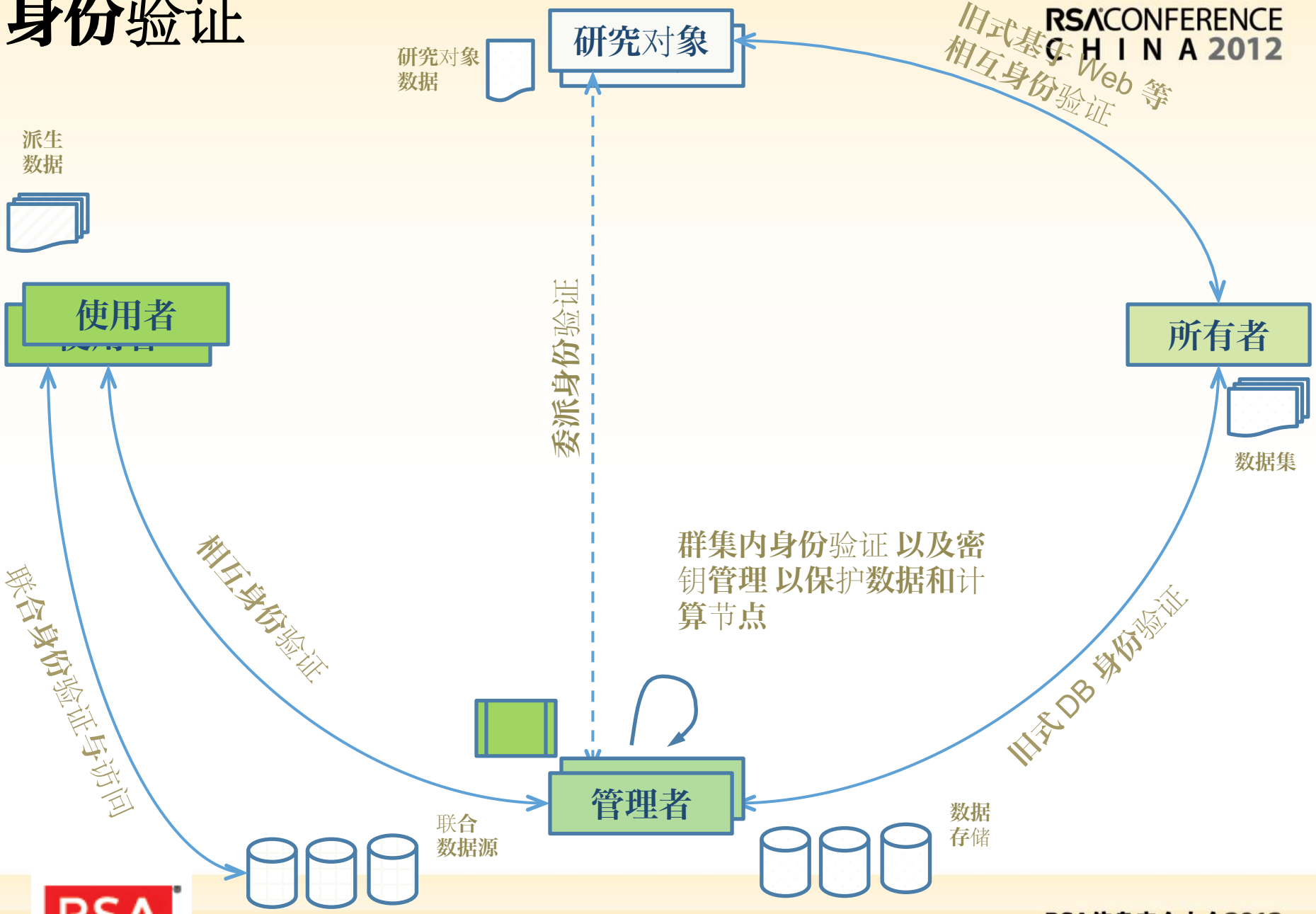
企业

数据所有者

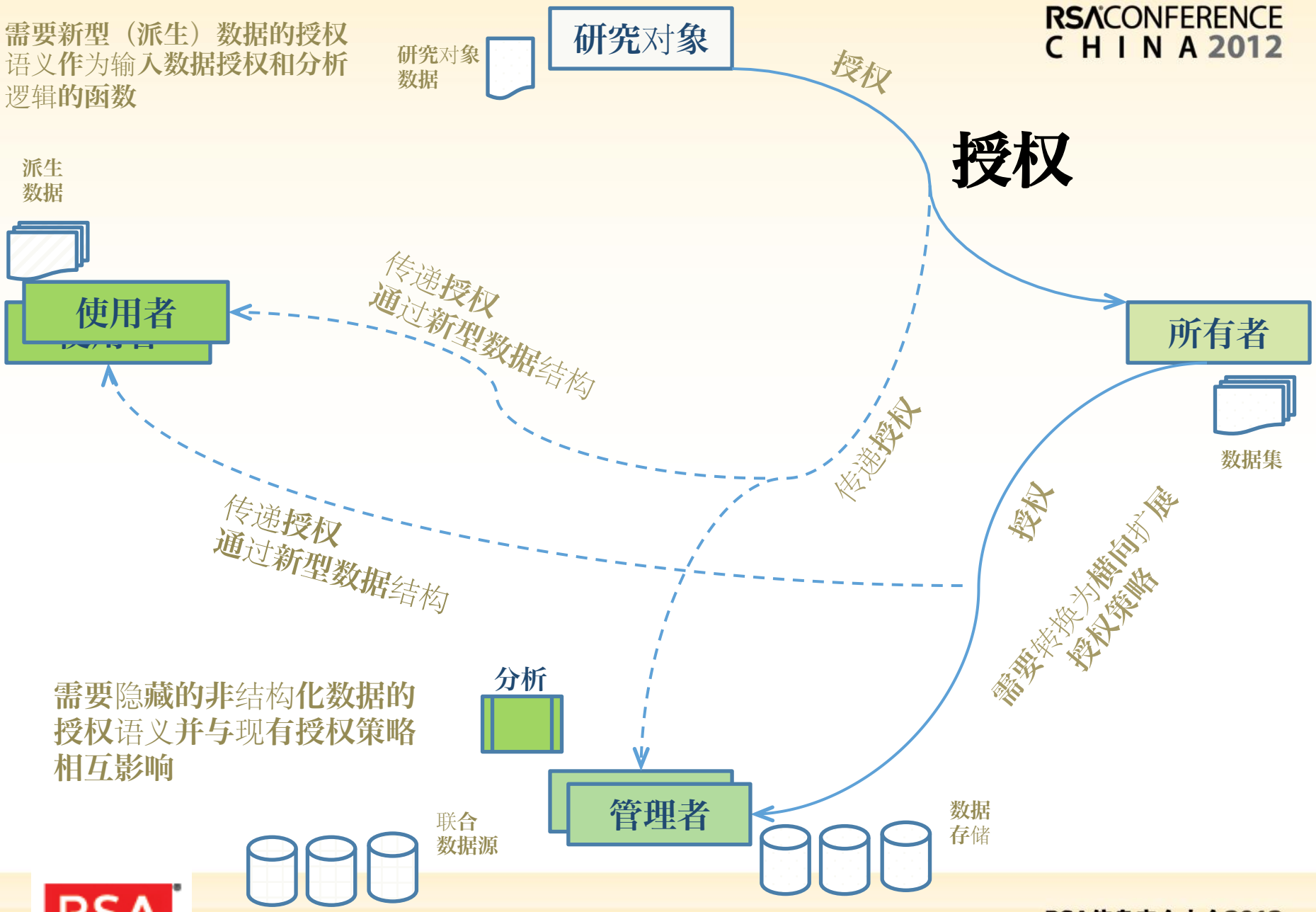
大数据中的安全要求

- 安全属性
  - 身份验证和授权
  - 资产保护
    - 加密
    - 内容监控（数据泄露防护）
    - 事件监控 (SIEM)
    - 隐私控制
  - 审核与法规遵从性
    - 策略遵从性
  - 取证

身份验证


研究对象
研究对象数据
旧式基于 Web 等相互身份验证
派生数据
使用者
委派身份验证
所有者
数据集
群集内身份验证 以及密钥管理 以保护数据和计算节点
联合身份验证与访问
相互身份验证
旧式DB 身份验证
管理者
联合数据源
数据存储

需要新型（派生）数据的授权语义作为输入数据授权和分析逻辑的函数

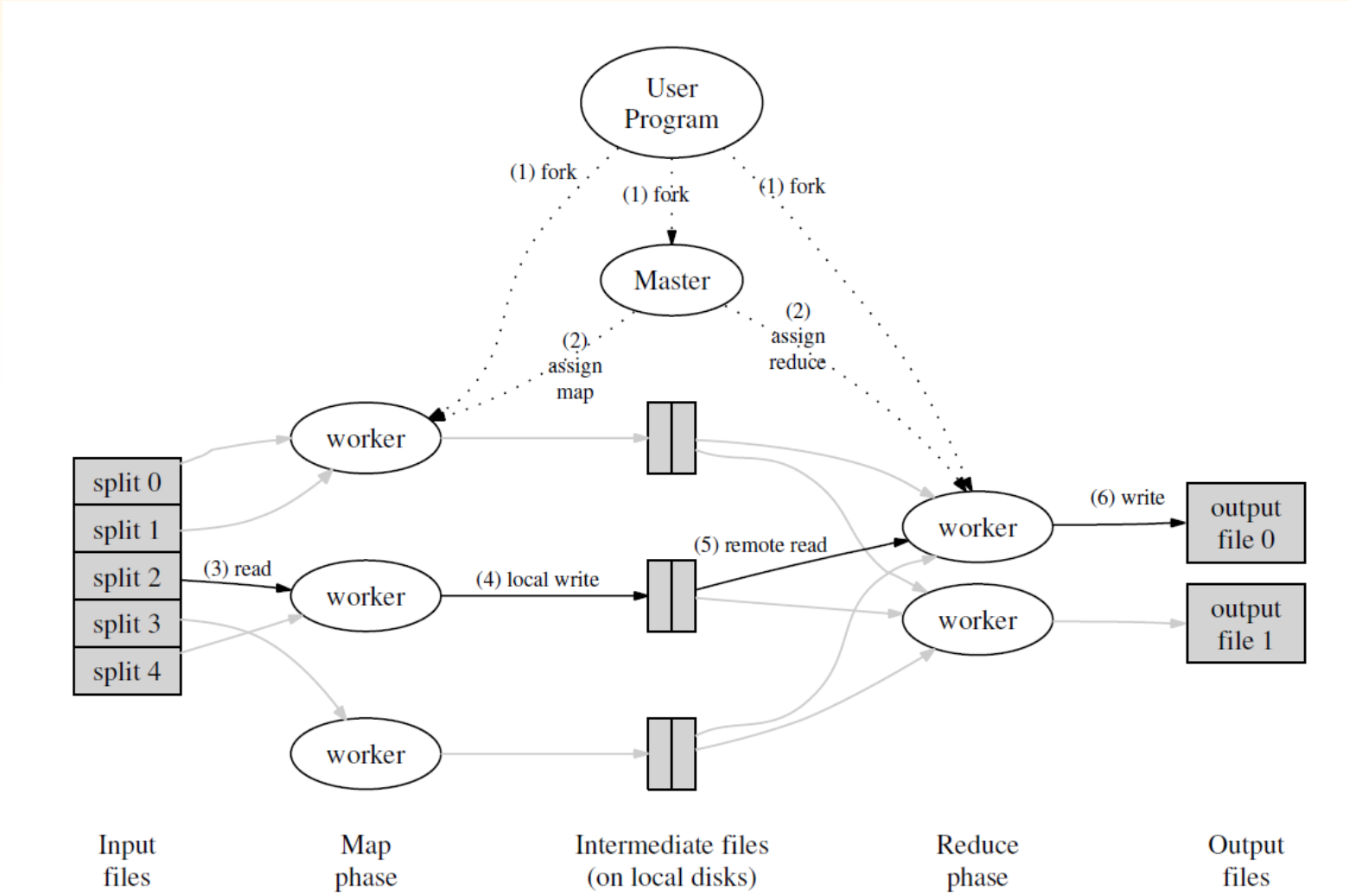


需要隐藏的非结构化数据的授权语义并与现有授权策略相互影响

在映射化简中嵌入安全性

- 附加与内置
- Hadoop 安全性
  - 重新设计以添加到身份验证和授权中
- 大数据提供商与安全提供商
  - OS 提供商与安全提供商
- 需要用于嵌入安全性的适当启用程序

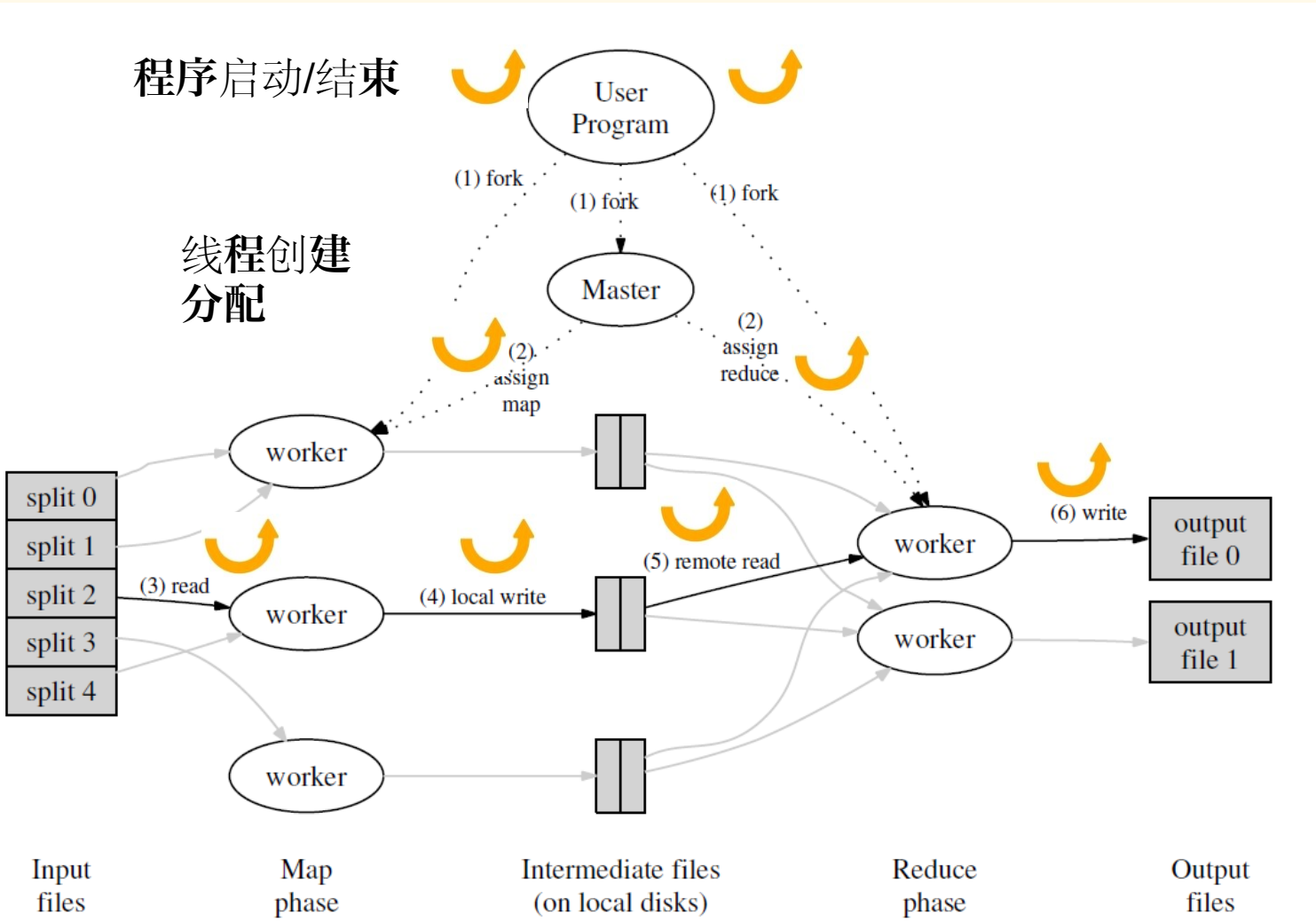
映射化简 – 快速入门


User Program
(1) fork
(1) fork
(1) fork
Master
(2) assign map
(2) assign reduce
worker
worker
worker
worker
worker
split 0
split 1
split 2
split 3
split 4
(3) read
(4) local write
(5) remote read
(6) write
output file 0
output file 1
Input files
Map phase
Intermediate files (on local disks)
Reduce phase
Output files

建议 – 映射化简自检

- 通过在以下时间点进行挂钩/回调实现可扩展自检功能
  - 线程创建和分配决策
  - 启动/完成映射化简工作
  - 从文件系统读取输出/将输入写入文件系统
  - 将中间结果从映射作业发送到分类器
  - 将核对过的中间结果发送到化简作业
- 对阻塞型/非阻塞型回调的支持
  - 非阻塞型不会导致修改数据流。更快、非侵入式、可提供较少控制
  - 阻塞型可以允许修改结果。更慢、侵入式、可提供更多控制

映射化简的自检框架


程序启动/结束
线程创建
分配
User Program
(1) fork
(1) fork
(1) fork
Master
(2) assign map
(2) assign reduce
worker
worker
worker
worker
worker
split 0
split 1
split 2
split 3
split 4
(3) read
(4) local write
(5) remote read
(6) write
output file 0
output file 1
Input files
Map phase
Intermediate files (on local disks)
Reduce phase
Output files

回调粒度

- 稀疏粒度
  - 在映射化简 API 处理的特定阶段定义
  - 影响所有数据集
  - 对基于全局策略的自检强制实施很有用
- 精细粒度
  - 根据特定的数据模式定义
  - 使数据选择能够避免影响所有数据集
  - 自然而然地融入映射化简的中间分类阶段
  - 通过对数据模式进行小段重要信息检查来利用映射化简分治方法

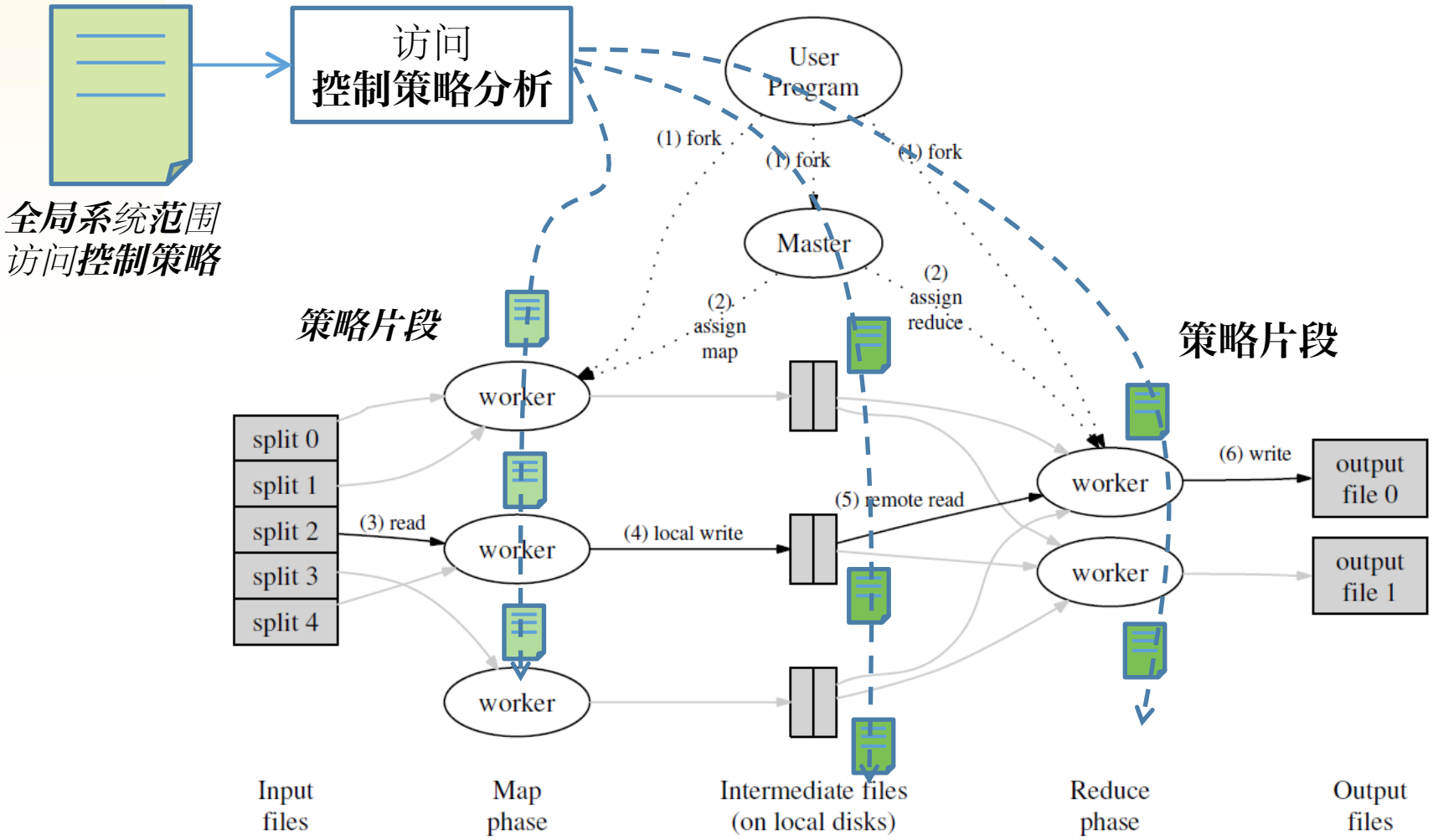
利用自检框架

- 提供用于嵌入安全逻辑的正确挂钩
- 在映射化简的横向扩展级别工作
- 允许嵌入横向扩展精细化 ...
  - 动态访问控制
  - 数据保障
  - 数据隐私
  - 数据分析保护

横向扩展动态访问控制

- 为决策点（多个映射/化简作业）提供精细访问控制策略
- 为访问控制策略输入的决策点提供映射化简作业的功能
- 支持基于内容的访问控制属性
- 支持访问控制策略决策点
- 通过结果修改的策略强制实施
- 识别输出数据（下一轮访问控制决策的输入）的属性

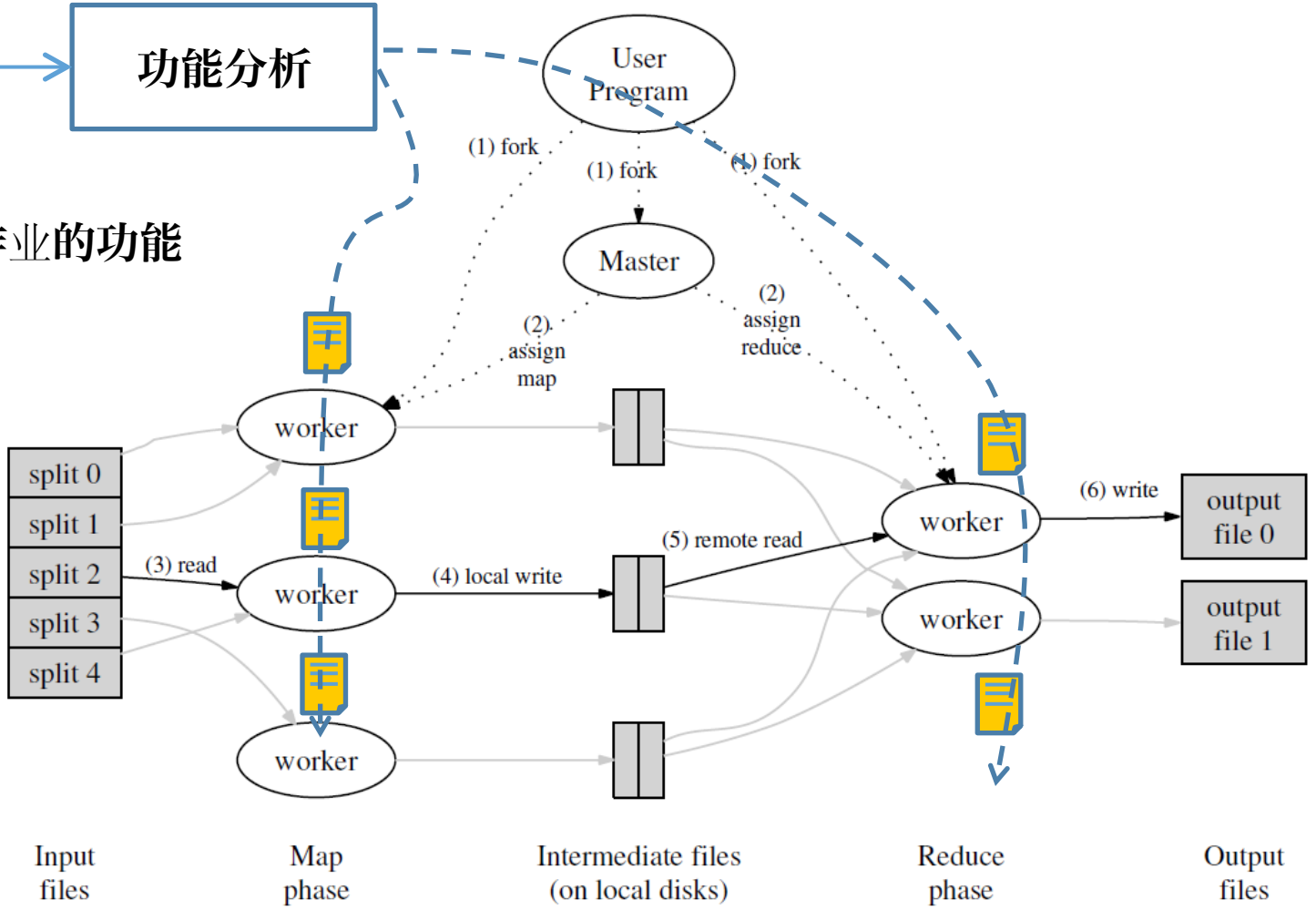
精细策略提供


访问
控制策略分析
全局系统范围
访问控制策略
策略片段
策略片段
User Program
Master
(1) fork
(1) fork
(1) fork
(2) assign map
(2) assign reduce
worker
worker
worker
worker
worker
split 0
split 1
split 2
split 3
split 4
(3) read
(4) local write
(5) remote read
(6) write
output file 0
output file 1
Input files
Map phase
Intermediate files (on local disks)
Reduce phase
Output files

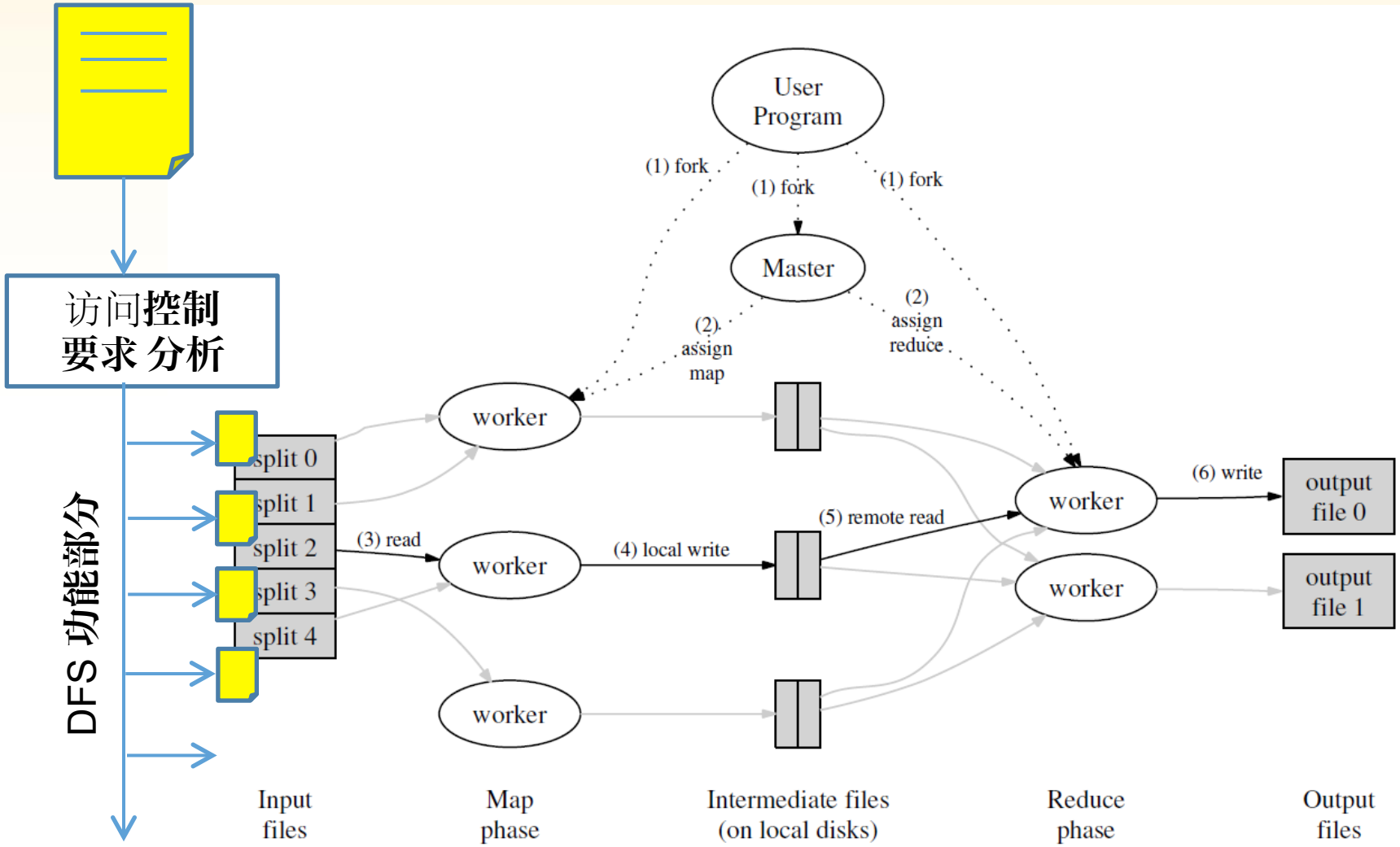
精细功能提供

功能分析

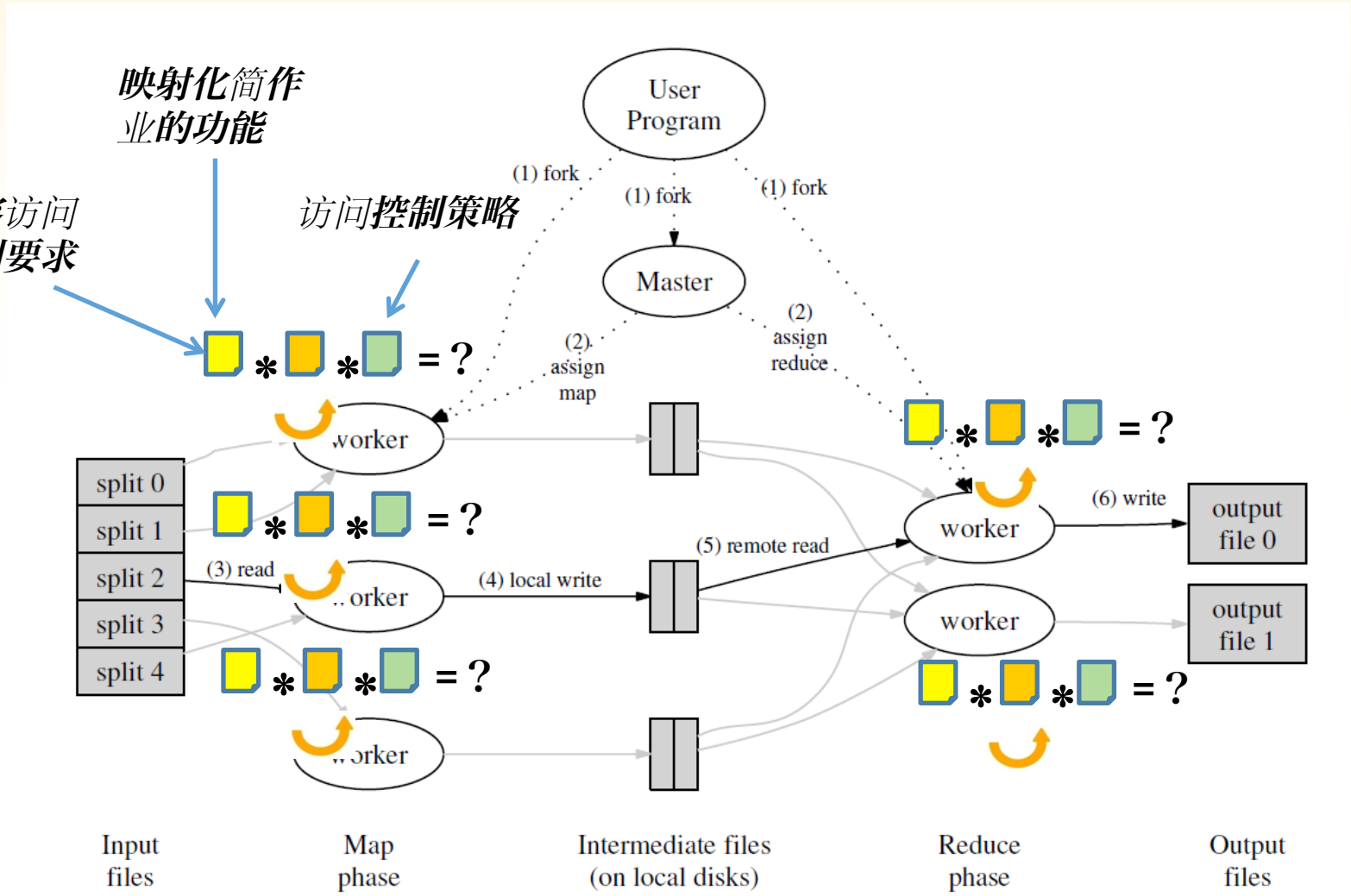
映射化简作业的功能



基于内容的访问控制要求



横向扩展策略决策


映射化简作业的功能
内容访问控制要求
访问控制策略
User Program
(1) fork
(1) fork
(1) fork
Master
(2) assign map
(2) assign reduce
= ?
worker
split 0
split 1
split 2
split 3
split 4
(3) read
(4) local write
(5) remote read
(6) write
output file 0
output file 1
Input files
Map phase
Intermediate files (on local disks)
Reduce phase
Output files

结论

- 回顾映射化简编程环境中的安全和信任挑战
- 需要在映射化简中嵌入安全性和信任基本形式
- 映射化简的可扩展自检框架建议
- 使用自检的横向扩展动态访问控制

谢谢大家！

RSA CONFERENCE
CHINA 2012
RSA信息安全大会2012