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

Embedding Security and Trust Primitives within Map Reduce

Samir Saklikar
RSA, The Security Division of EMC

RSA CONFERENCE
CHINA 2012
RSA信息安全大会2012

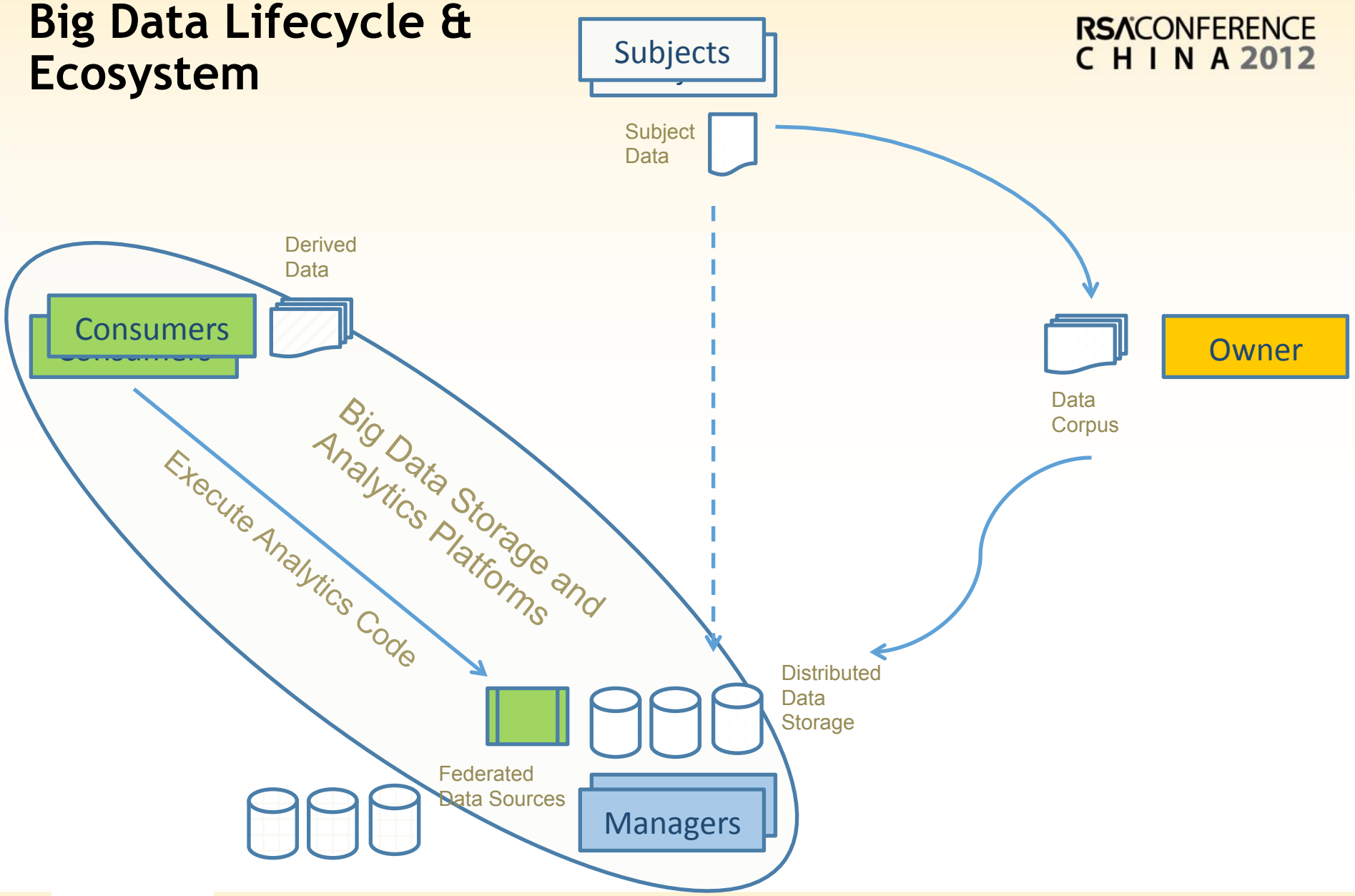
Agenda

- Overview of Big Data ecosystem
  - Different stakeholders & perspectives on trust requirements
- Different security requirements within Big Data
- Requirements & Proposal for Map-Reduce Introspection Framework
- Leverage Introspection framework for security primitives
  - Access Control
- Conclusion

Big Data Players

- Data Subjects (M Subjects)
  - About whom the Data is applicable?
- Data Owner (1 Owner)
  - May be different from the Data Subjects.
    - For e.g. the enterprise owning data about their users/systems
  - Varying levels of ownership from *absolute* to *custodian*
  - Varying levels of freedom to work with the data
- Data Managers (N Managers)
  - Store, facilitate access and enable processing over Data
  - May overlap with Data Owners
- Data Consumers (P Consumers)
  - Are interested in value out of Data
  - Often, Subjects are indirect consumers

Big Data Lifecycle & Ecosystem



Data Subjects – Assets and Concerns

- Assets
  - Profile Data (user preferences)
  - Behavioral Data (usage/consumption patterns)
  - Characterizing information (endpoint identifiers)

- Concerns
  - Leakage of PII, resulting in Identity Theft/Loss of Privacy
  - Leakage of sensitive information, resulting in malicious use
  - Incorrect profiling leading to wrong service personalization
  - Lack of Control over data portability and lifecycle management
  - Very less (if any) stake in the data's monetization

Data Owners – Assets and Concerns

- Assets
  - Large Corpus of Data –supporting business functions
    - Users/Employee Related
    - Intellectual Property related
    - Business functions
    - Information Technology
- Concerns
  - Leakage and/or misuse of data, resulting in legal liabilities
  - Leakage/Corruption of data, resulting in loss of business

Data Managers – Assets and Concerns

- Assets
  - Data management infrastructure
  - Data analytics infrastructure
- Concerns
  - Leakage and/or misuse of data, resulting in legal liabilities
  - Corruption of data, resulting in loss of business

Data Consumers – Assets and Concerns

- Assets
  - Data Analytics capability
  - Inherent Data Semantics

- Needs and Concerns
  - Leakage of Analytics capability, resulting in loss of IP
  - Leakage of PII via analytics, resulting in legal liabilities
  - Corruption of Data, resulting in incorrect results
  - Need seamless access to rich and varied sources of Data

Changes in Big Data Security perspective

RSA CONFERENCE CHINA 2012

Privacy and Control Tug-of-War

Data Owners
Data Managers
Data Consumers

Data Subjects

Data Consumers

Audit Agencies, *CIRT*

Data Subjects

Enterprise Users

Security Service Providers

Data Managers

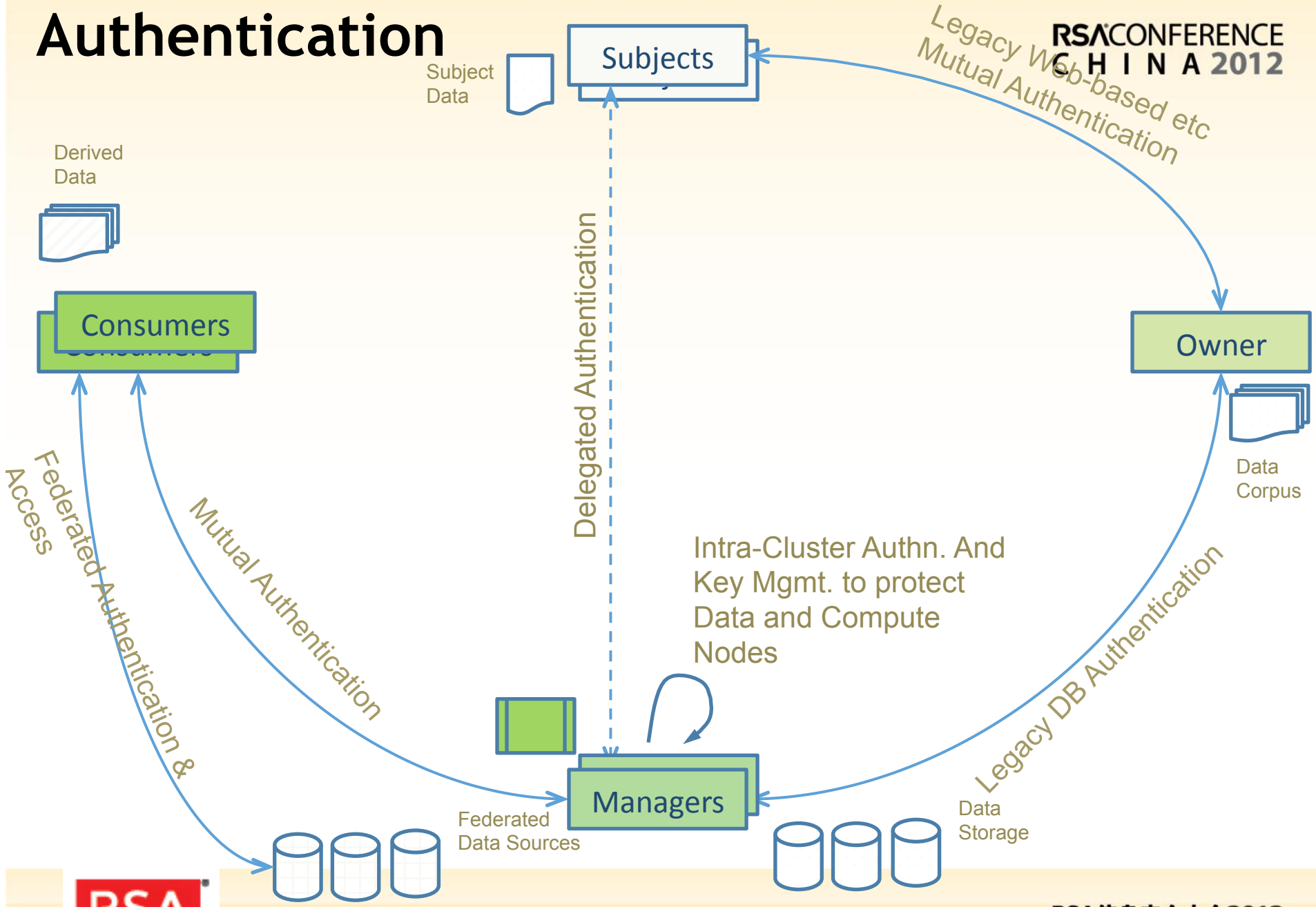
Enterprise

Data Owners

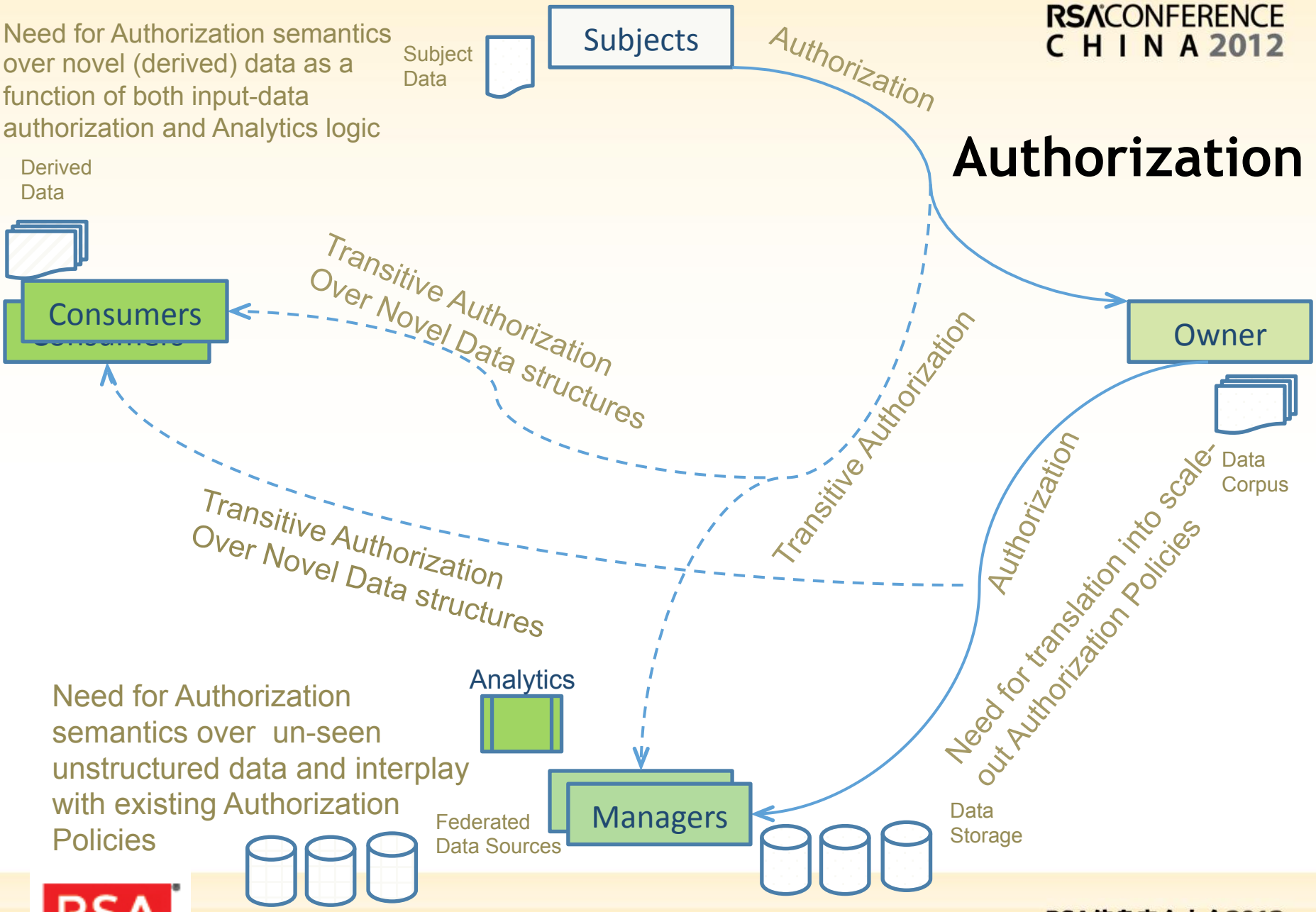
Security Requirements in Big Data

- Security Attributes
  - Authentication and Authorization
  - Asset Protection
    - Encryption
    - Content Monitoring (Data Leakage Prevention)
    - Event Monitoring (SIEM)
    - Privacy Controls
  - Auditing and Compliance
    - Policy Conformance
  - Forensics

Authentication



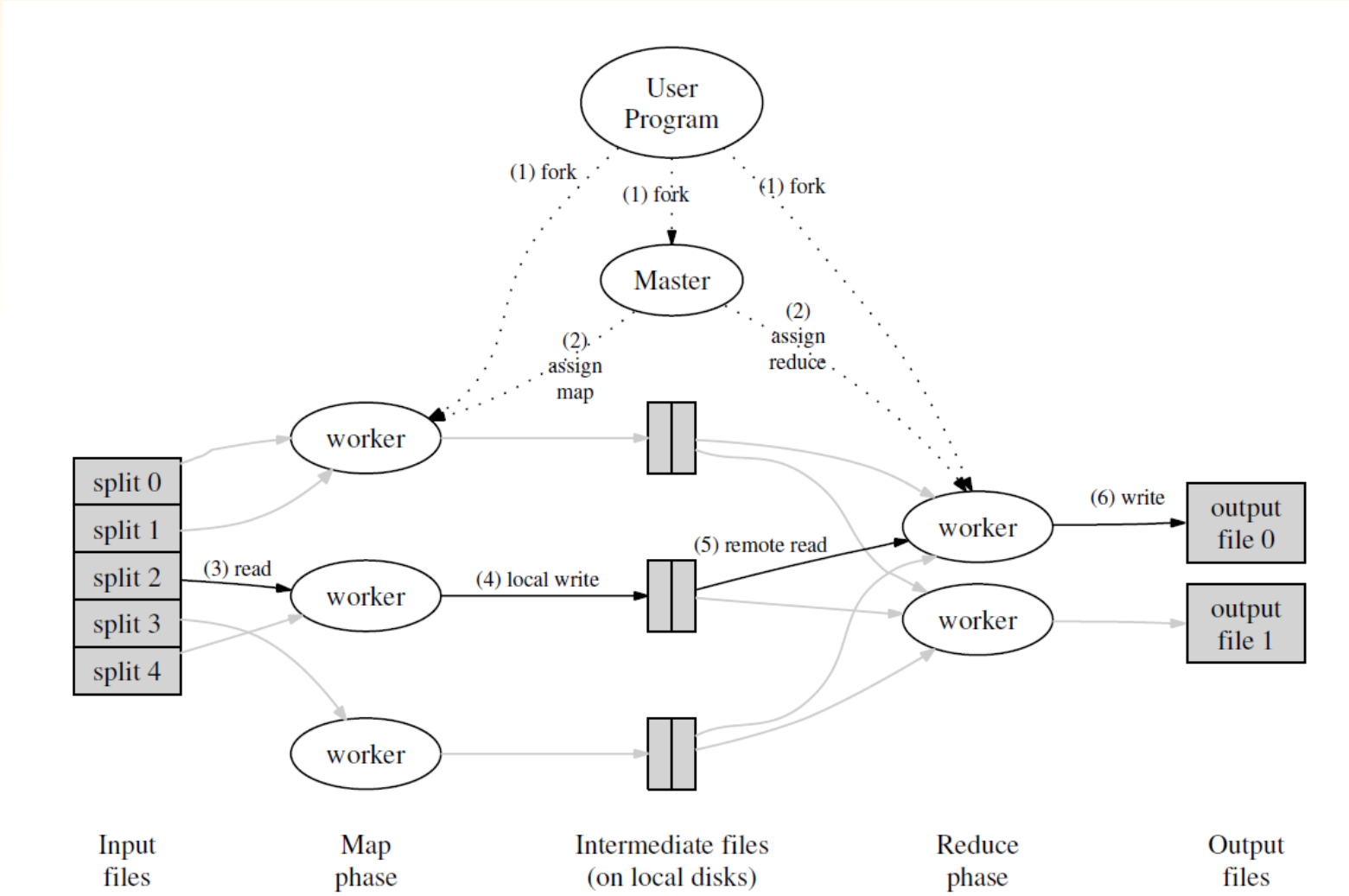
Authorization



Embedding Security in Map Reduce

- Bolted-on v/s Built-In
- Hadoop Security
  - Re-design to add in authentication, authorization
- Big Data Providers v/s Security Providers
  - OS Providers v/s Security Providers
- Need the right enablers to embed Security

Map Reduce – A Quick Primer



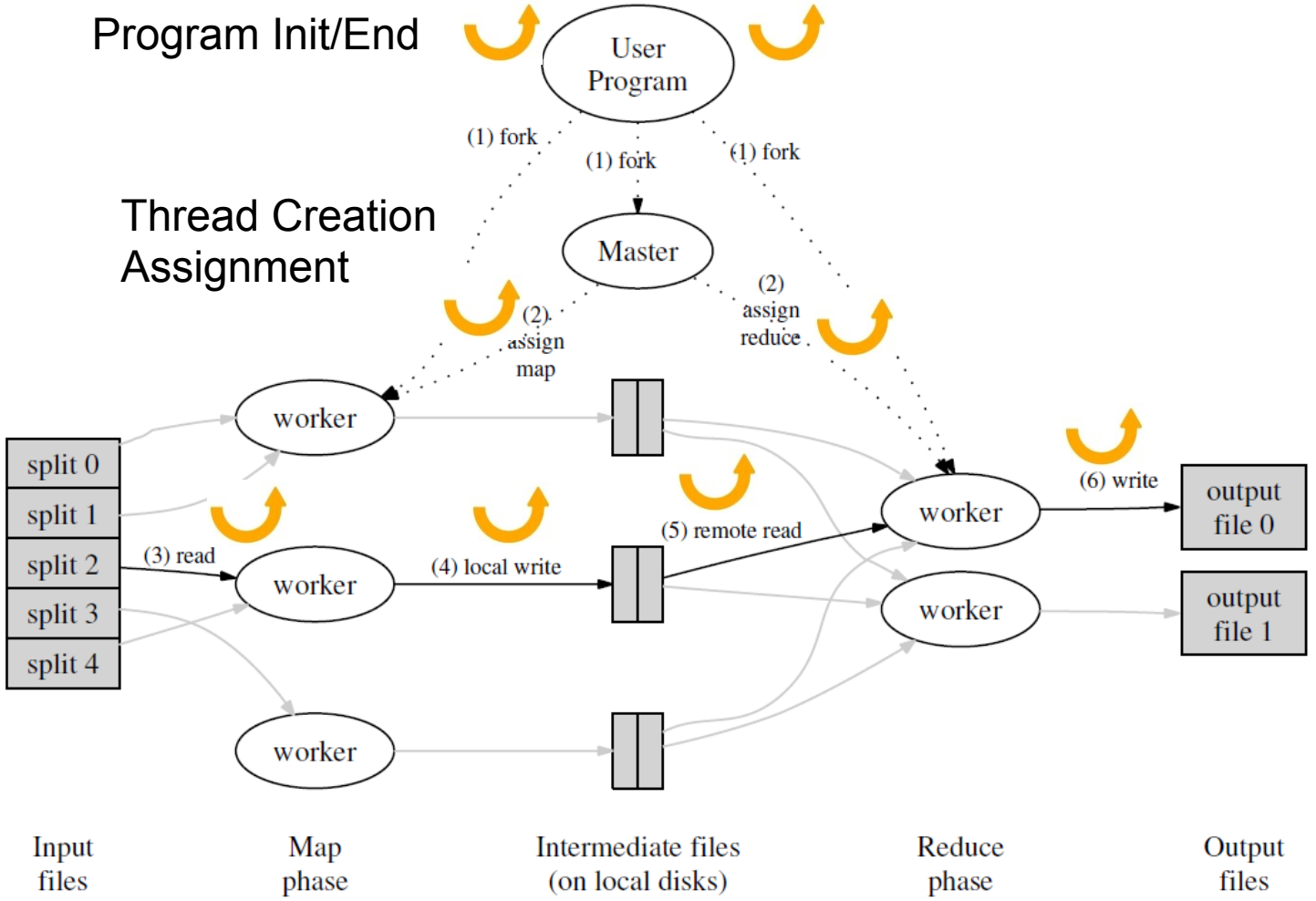
Proposal – Map Reduce Introspection

- Extensible Introspection capability via hooks/callbacks at following points
  - Thread Creation and Assignment Decision
  - Initiating/Completing a Map Reduce job
  - Reading/Writing Input/Output from/to file system
  - Sending Intermediate Results from Map job to sorter
  - Sending collated intermediate results to Reduce job
- Support for Blocking/Non-Blocking callbacks
  - Non-blocking do not result in modifying flow of data. Faster, Un-intrusive, Less Control
  - Blocking can enable results modification. Slower, Intrusive, More Control

Introspection Framework for MR

Program Init/End

Thread Creation
Assignment



Callback granularity

- Coarse Granularity
  - Defined on a particular stage in the Map Reduce API processing.
  - Affects all data sets
  - Useful for global policy-based introspection enforcement

- Fine Granularity
  - Defined on a particular data-pattern
  - Enable data selection to avoid hits on all data-sets
  - Natural fit into the intermediate sorting phase of MR
  - Leverages MR divide-and-conquer approach by small piece meal checks on the data pattern

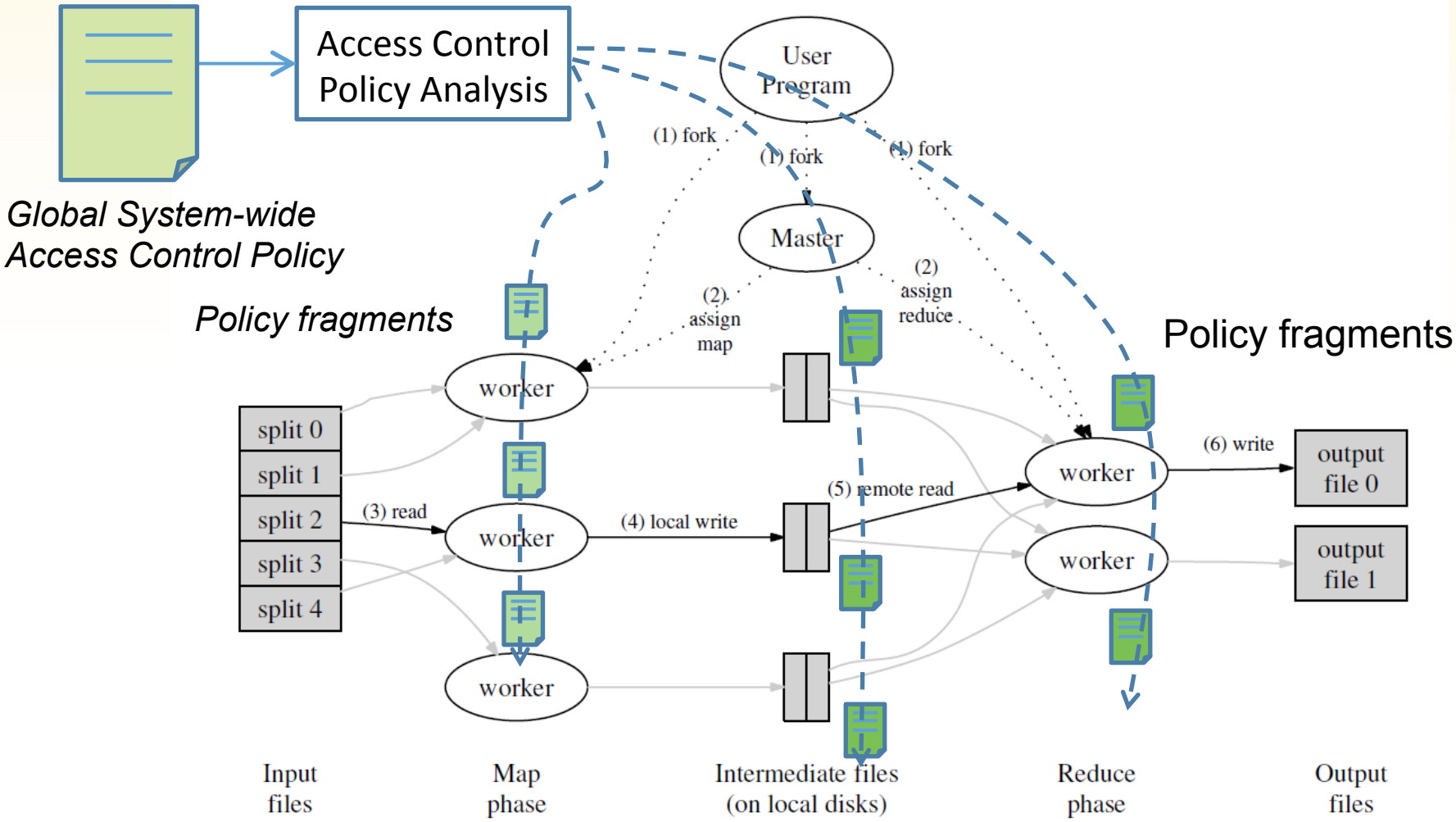
Leverage Introspection framework

- Provides the right hooks for embedding security logic
- Works at scale-out level of Map Reduce
- Enables embedding scale-out fine-grained ..
  - Dynamic access control
  - Data assurance
  - Data Privacy
  - Data Analytics Protection

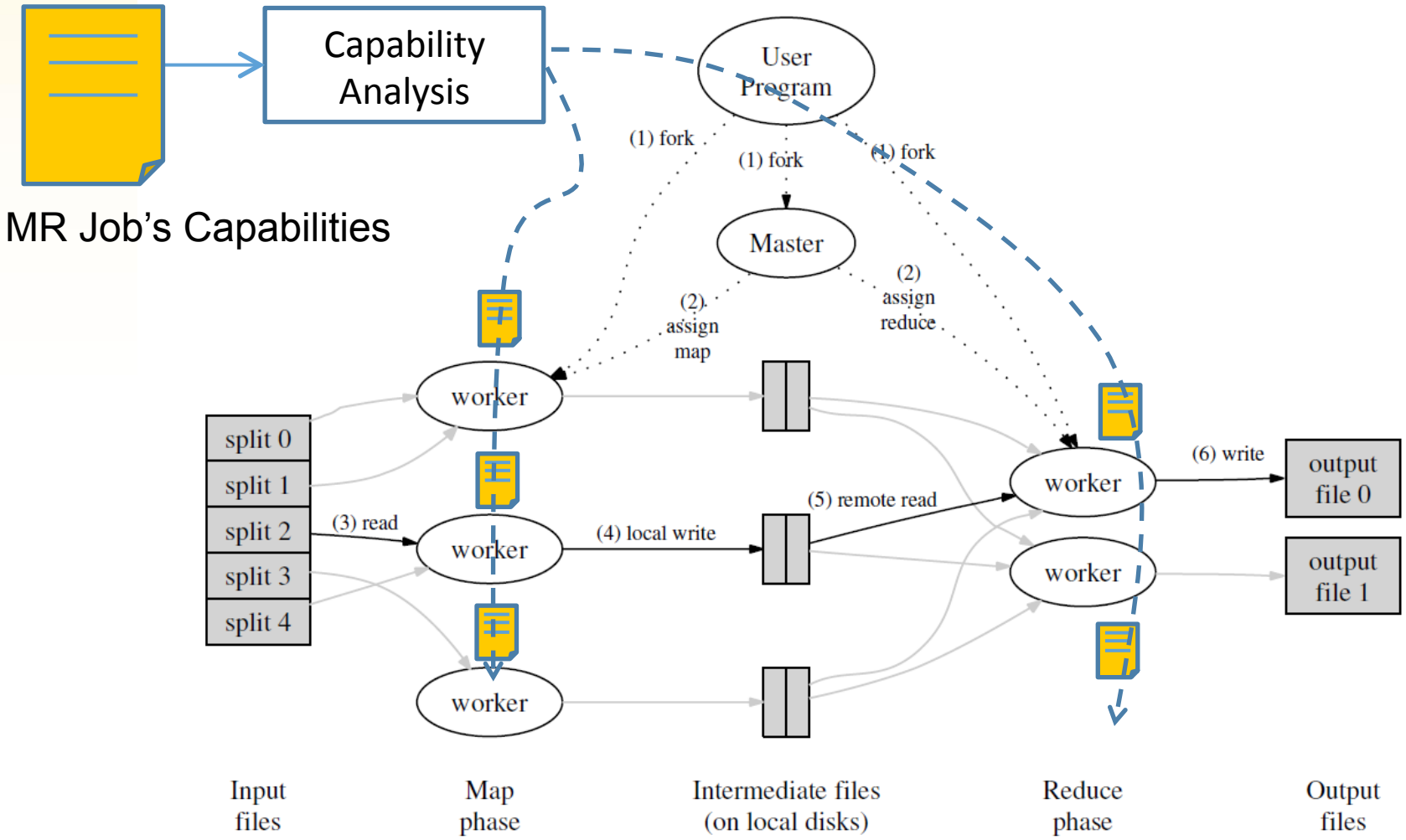
Scale-out Dynamic Access Control

- Granular access control policy delivery to Decision Points (multiple Map/Reduce jobs)
- MR Job's capability delivery to decision points for inputs to access control policy
- Support for Content-based access control attributes
- Support for Access Control Policy decision points
- Policy enforcement via results modification
- Identifying attributes for output data (inputs to next round of access control decisions)

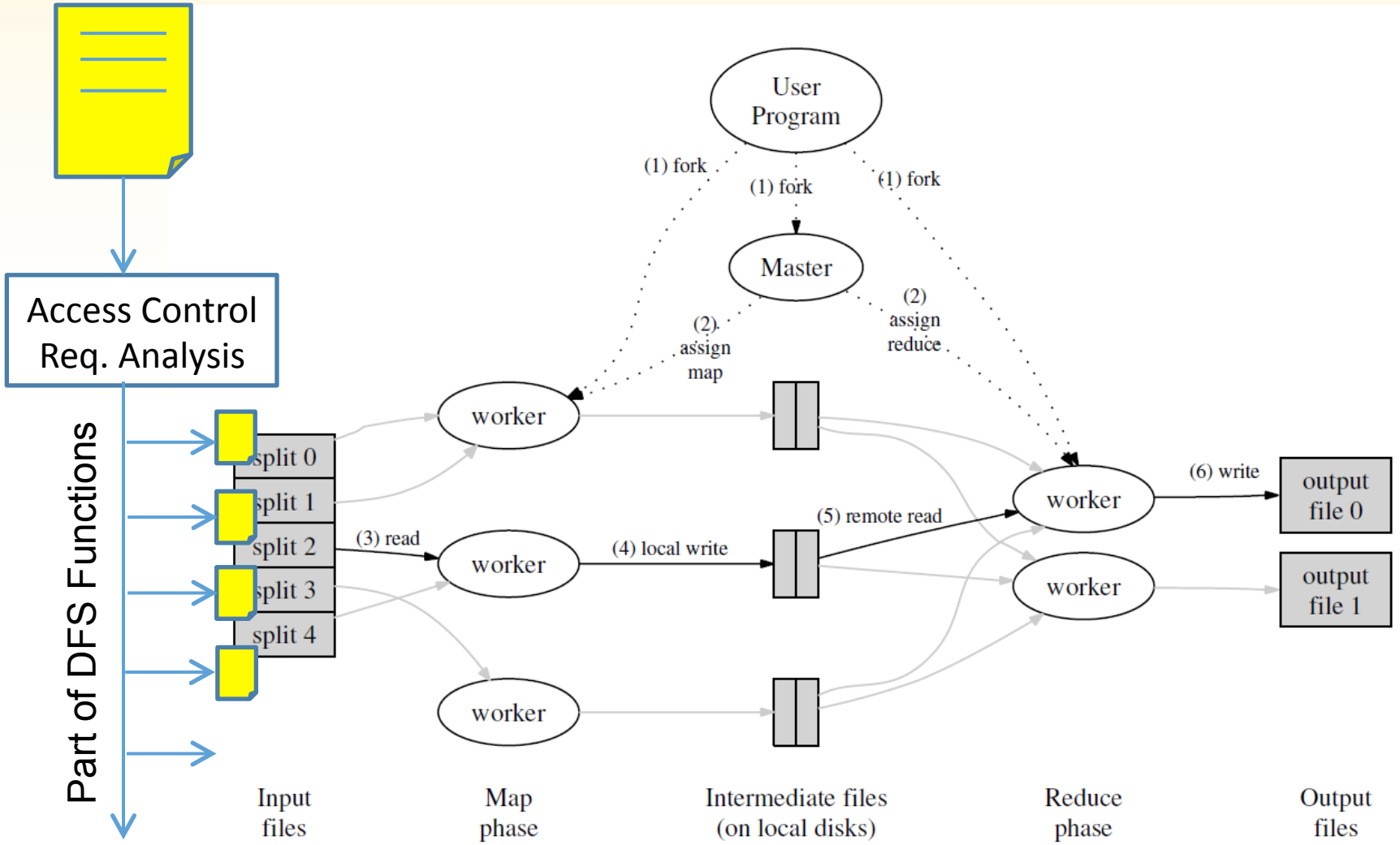
Granular Policy Delivery



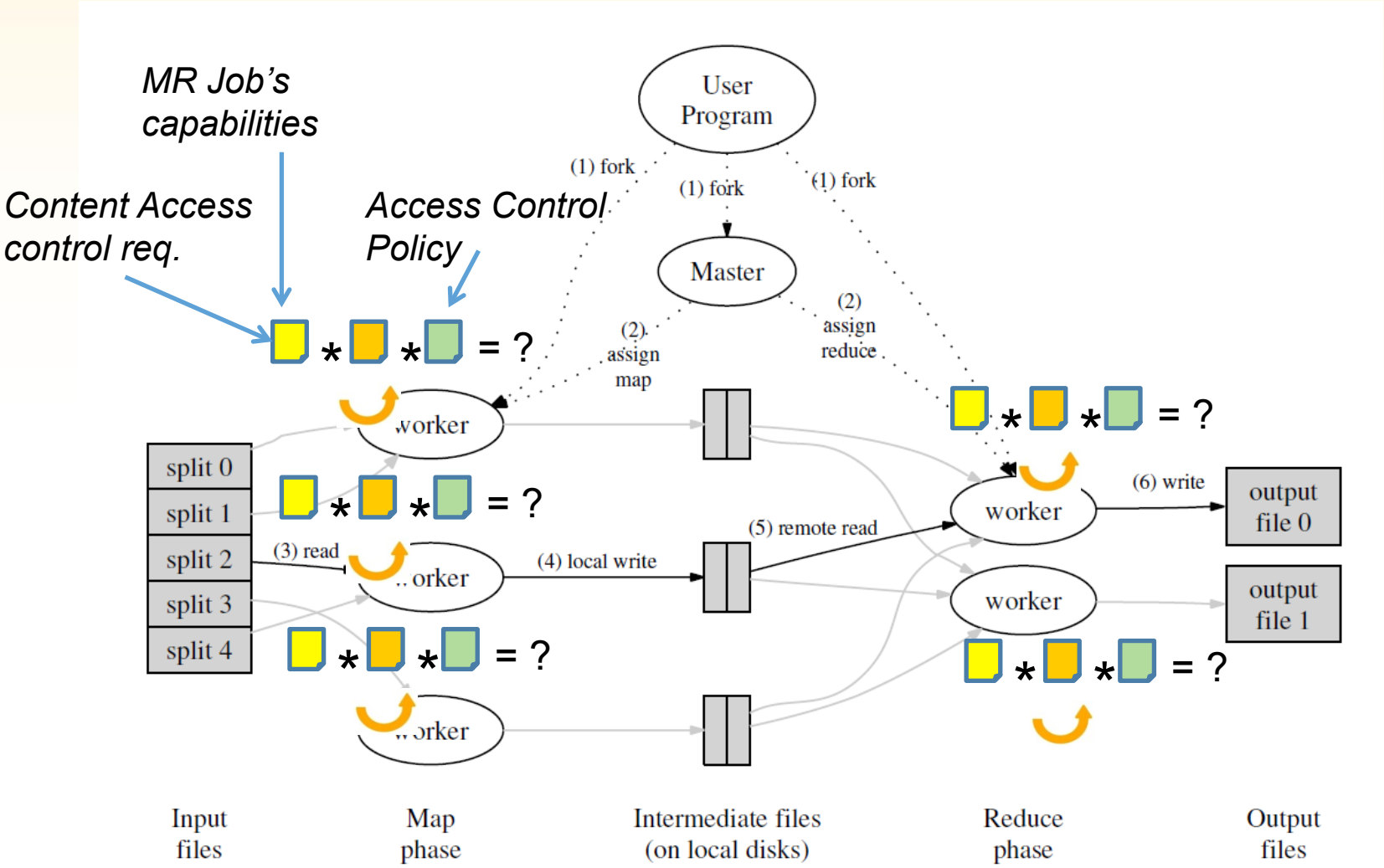
Granular Capabilities Delivery



Content based Access Control Requirements



Scale-Out Policy Decision



Conclusion

- Review of Security and Trust challenges in Map Reduce programming environment
- Need for embedding security and trust primitives within Map Reduce
- Proposal for an extensible introspection framework for Map Reduce
- Scale-out dynamic access control using introspection

Thank You

RSA CONFERENCE
CHINA 2012
RSA信息安全大会2012