Multi-Layer Authorization Framework for a Representative Hadoop Ecosystem Deployment

Maanak Gupta, Farhan Patwa, James Benson, and Ravi Sandhu

Institute for Cyber Security and Department of Computer Science
University of Texas at San Antonio

22nd ACM Symposium on Access Control Models and Technologies (SACMAT)
June 21-23, 2017, Indianapolis, IN, USA
Agenda

➢ Introduction and Motivation
➢ Multi-layer Access Control
➢ Hadoop Ecosystem Authorization Architecture
➢ Access Control Mechanisms and Policy Configuration Points
➢ Conclusion
IDC 2025:
- global “datasphere” – 163 zettabytes
- 10x than 2016

Opportunities: 21st century gold for data miners

Big Data require “Big Systems”

Security:
- Secure Storage
- Privacy Concerns (e.g. HIPPA)
- Fine granular access requirements
Hadoop Ecosystem

➢ Hadoop: resilient, cost efficient distributed storage (HDFS) and processing framework (MapReduce or YARN)

➢ Ecosystem = Hadoop core + Open-Source Projects

➢ Hadoop Data Lake

➢ Security Concerns
Multi-Layer Access Control

Data and Service Objects

Services
- HDFS NameNode,
- YARN ResourceManager
- Apache Hive

HDFS Files,
Hive Tables
Kafka Topics

Cluster Resources and Applications
- YARN Queues,
Cluster Nodes
Hadoop and Data Services Access

WebHDFS Access via Apache Knox

Hadoop Daemons Access Configuration
Data Objects Access

Hive and HDFS Access Configurations
Data Objects Access

Authorization Options and End User Impersonation
Data Objects Access

Tag Based Policy Configuration
Data Objects Access

Hive Database *
  - foodmart

Hive Table *
  - customer

Hive Column *
  - account_num

Conditions:

Select User | Access Types | Select Masking Option | Row Level Filter
------------|--------------|-----------------------|------------------
raj_ops     | select       | Mask                  | fname = 'Sheri'  

(a) Ranger Policy

Table:

<table>
<thead>
<tr>
<th>account_num</th>
<th>lname</th>
<th>fname</th>
<th>mi</th>
<th>address1</th>
</tr>
</thead>
<tbody>
<tr>
<td>11111111111</td>
<td>Nowmer</td>
<td>Sheri</td>
<td>A</td>
<td>2433 Bailey Road</td>
</tr>
<tr>
<td>11111111111</td>
<td>West</td>
<td>Sheri</td>
<td>D</td>
<td>9590 Sutton Circle</td>
</tr>
<tr>
<td>11111111111</td>
<td>Schiel</td>
<td>Sheri</td>
<td></td>
<td>8603 Holiday Hill Dr</td>
</tr>
</tbody>
</table>

(b) Hive View Table

Data Masking and Column Filtering
Geo Location Based Policies
Context Enricher and Policy Conditions

Hive Database *
- foodmart

table *
- customer

Hive Column *
- fname

Deny Conditions :
- Resources Accessed Together?
  - foodmart.customer.account_num

Select User
- raj_ops

Permissions
- resources-accessed-together : foodmart.customer.account_num

Data Combination Prohibition
Cluster Resource and Application Access

YARN Queue Access Control Configuration

- YARN acl.enable: true
- yarn.admin.acl: root
- yarn.scheduler.capacity.root.queues=default,newQueue
- yarn.scheduler.capacity.root.acl_administer_queue= root
- yarn.scheduler.capacity.root.acl_submit_applications= root
- yarn.scheduler.capacity.root.default.acl_submit_applications= rai_ops
- yarn.scheduler.capacity.root.newQueue.acl_administer_queue= maria_dev
- yarn.scheduler.capacity.root.newQueue.acl_submit_applications= maria_dev
- yarn.scheduler.capacity.queue-mappings=u:maria_dev:newQueue
- yarn.scheduler.capacity.queue-mappings-override.enable=false

(a) Capacity Scheduler configuration (YARN)

<table>
<thead>
<tr>
<th>Queue acls for user</th>
<th>Queue Operations</th>
</tr>
</thead>
<tbody>
<tr>
<td>raj_ops</td>
<td>root SUBMIT_APPLICATIONS</td>
</tr>
<tr>
<td>root</td>
<td>root ADMINISTER_QUEUE</td>
</tr>
<tr>
<td>default</td>
<td>default ADMINISTER_QUEUE</td>
</tr>
<tr>
<td>newQueue</td>
<td>newQueue ADMINISTER_QUEUE</td>
</tr>
</tbody>
</table>

(b) ACLs for different users

- root
- default
- newQueue

- child queues
- no user

Access rights:
- root
- default
- newQueue
Queue and Job Level Access Control

Cluster Nodes Configuration
Conclusion

Secure Hadoop Ecosystem

‘Defense in Depth’

Data and Service Objects

Hadoop And Services

Cluster Resource and Application