Round Table Discussion

Security & Governance

Session#1: Merck, ING, Clearsense, Charter, HCSC
Session#2: Discover, Universal, Expedia, Honeywell, SunLife, Geisinger, Bloomberg

Customer Advisory Board
June 2017
Agenda

- Security - Focus Areas & Roadmap
- Governance - Focus Areas & Roadmap
- What’s working? Where are the gaps?
- Top 3 items to make security & governance less painful for your enterprise?
- What you would like to see improved in the next 6-12 months (wish list)?
Security – Ranger, Knox, & Platform
## Roadmap At A Glance - Security

### Security Focus in HDP

<table>
<thead>
<tr>
<th></th>
<th>HDP 2.5</th>
<th>Current</th>
<th>HDP 2.6</th>
<th>HDP 3.0 &amp; Beyond</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ranger</strong></td>
<td>• Classification (tag) based security (ABAC)</td>
<td>• Policy scalability</td>
<td>• Policy scalability</td>
<td>• Shared Services for Cloud (Multi-cluster/multi-tenant support)</td>
</tr>
<tr>
<td></td>
<td>• Dynamic Column Masking</td>
<td>• Enhanced Policy constructs (macros)</td>
<td>• Policy export/import</td>
<td>• KMIP KMS support</td>
</tr>
<tr>
<td></td>
<td>• Row Filtering</td>
<td>• Incremental LDAP sync</td>
<td>• Incremental LDAP sync</td>
<td>• Monitoring and Dynamic Block</td>
</tr>
<tr>
<td></td>
<td>• KMS HSM Integration (Safenet)</td>
<td>• Plugin policy version info</td>
<td>• Process level ACLs for Spark</td>
<td>• Process level ACLs for Spark</td>
</tr>
<tr>
<td></td>
<td>• LDAP Improvements</td>
<td>• Hive show/describe columns</td>
<td>• Support Data Plane and DLM capabilities</td>
<td>• Support Data Plane and DLM capabilities</td>
</tr>
<tr>
<td></td>
<td>• Audit Scalability</td>
<td>• authorization</td>
<td>• Policy export/import</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Reports Page Improvements</td>
<td>• SmartSense Metrics</td>
<td>• Policy export/import</td>
<td></td>
</tr>
<tr>
<td><strong>Knox</strong></td>
<td>• Knox SSO framework</td>
<td>• PAM authentication support</td>
<td>• PAM authentication support</td>
<td>• Extend Knox SSO to HDP WebUIs</td>
</tr>
<tr>
<td></td>
<td>• Knox SSO for Ambari/Ranger Uls</td>
<td>• Hadoop user-group mapping support</td>
<td>• Hadoop user-group mapping support</td>
<td>• Spark, Zeppelin, &amp; Livy support</td>
</tr>
<tr>
<td></td>
<td>• Component HA support</td>
<td>• Basic Knox UI (preview)</td>
<td>• Knox Client Improvements for Scripting and Development</td>
<td>• Knox Client Improvements for Scripting and Development</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Knox support for websockets (for</td>
<td>• Multi-factor auth. support</td>
<td>• Knox UI for managing topologies</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Zeppelin support</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Knox support for SOLR REST API</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Platform Security</strong></td>
<td>• Source Code security scans &amp; remediation</td>
<td>• Source Code security scans &amp;</td>
<td>• Source Code &amp; Web App Pen Testing &amp; Certification</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>remediation</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Governance - Atlas
# Roadmap At A Glance - Governance

## Governance Focus in HDP

<table>
<thead>
<tr>
<th>Cloud</th>
<th>HDP 2.5</th>
<th>HDP 2.6</th>
<th>HDP 3.0 &amp; Beyond</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Shared Enterprise Services for Governance (backend only)</td>
<td>• Shared Enterprise Services for Governance (Multi-cluster support)</td>
<td></td>
</tr>
</tbody>
</table>
| Integrations | • Storm/Kafka, Falcon, Sqoop hooks  
• High availability support  
• LDAP Authentication  
• Classification based security for Hive  
• Ranger Authorization | • Tag-based policy support for HDFS, Kafka, Hbase  
• Knox SSO for Atlas UI | • Support Data Plane and DLM capabilities  
• Spark job lineage and metadata extraction  
• NiFi support – Lineage and metadata  
• Hbase, HDFS hooks |
| Core & Consumability | • Business Catalog (Tech Preview)  
• New User Experience  
• High availability support | • API revamp – Swaggerized docs!  
• Simplified Basic Search UI  
• Manual entity creation (HDFS, Hbase, Kafka, custom types etc.)  
• Performance and scalability improvements  
• SmartSense Metrics | • Column level lineage  
• Tag Propagation  
• Annotation & Collaboration  
• Improved Lineage & Impact UI (Search/Filtering)  
• Export/Import of Atlas lineage and metadata  
• Query Builder & Streamlined Search  
• Business Glossary/Enterprise Catalog |
Discussion
Roundtable Questions

Security

1. Are you familiar with HDP2.6 Security features? Has your organization piloted or deployed any of them?

2. What are your experiences with the HDP security stack? What parts of our security stack do you use (Kerberos, Knox, Ranger, ..) and what do you plan to use in the near future?

3. What are challenges you see with your enterprise data security? Specific pain points would be helpful.

4. Other features / roadmap items that would help around security setup and administration?

5. Who are the users of security admin components in your organization? Which groups are responsible for managing security interfaces in Hadoop stack?

6. Do you have cloud/hybrid deployments you are considering in your near term (3-6 month) horizon? What specific security challenges do you foresee with the cloud/hybrid environments?

7. Other than GDPR are you aware of any major regulations/standards that will influence security within your big data environments in the near term?
Roundtable Questions

Governance

1. How familiar are you with Atlas and HDP governance capabilities?
2. What systems do you use currently as a catalog for metadata?
3. Do you use data classification, metadata discovery and management and data profiling currently? If so, which specific tools?
4. How important is having a business catalog or taxonomy for your governance function?
5. If familiar with Atlas, what features / roadmap items would you consider as barriers to adoption to roll out Atlas across your enterprises?
6. What tools would you like to see integrated/certified with Atlas ecosystem?
7. If you have deployed Atlas already in production or in the process of rolling it out enterprise wide, who are the users of Atlas and what skills do they currently possess? Are there any skills gaps that we should consider bridging with tool capabilities as we look ahead?
Roundtable Responses

1. Geisinger in HDP2.6, majority on 2.5

2. Adoption: Kerberos (12/12), Ranger (10/12), Knox (4/12), rest have plans to deploy Ranger or Knox in the next 3 months. 6/12 are already using HDP on AWS or Azure or have plans to deploy there

3. 4/12 customers had Atlas

4. Challenges:
   1. Multi-platform data consolidation
   2. Flexible access control
   3. Protection on Onboarding
   4. Open metadata standards and governance process
   5. Metadata federation
   6. Kerberos is a HUGE pain point still (even with Ambari managed environments)
Readout
What are 3 concrete items that can make security & governance less painful for your enterprise?

<table>
<thead>
<tr>
<th>Pain Point</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Security #1: Multi-tenancy, Multi-factor Auth</td>
<td></td>
</tr>
<tr>
<td>Security #2: “What If” mode or policy enforcement impact</td>
<td></td>
</tr>
<tr>
<td>Security #3: SSO with OAuth support</td>
<td></td>
</tr>
<tr>
<td>Governance #1: Data Discovery</td>
<td></td>
</tr>
<tr>
<td>Governance #2: Data Quality and Profiling</td>
<td></td>
</tr>
<tr>
<td>Governance #3: Metadata integrations and open metadata exchange</td>
<td></td>
</tr>
</tbody>
</table>
Thank You!