Hadoop Infrastructure

Security

Cluster Security

Security is an important topic which impacts all elements of cluster operations and processes. The following information covers a wide range of subjects from an infrastructure perspective, so please feel free to comment or ask questions if necessary through the feedback options on the site.

Cluster Kerberos

The security of Hadoop / HAAS clusters in BT is managed by Cloudera using kerberos authentication and we use the standard AD vintella systems to manage this activity.

The link below explains the workings of kerberos in a very simple way which should help you understand the authentication process.

http://www.roguelynn.com/words/explain-like-im-5-kerberos/

Some basic points to understand

- The Realm is IUSER.IROOT.ADIDOM.COM
- The Authentication and ticket granting are the AD servers
- Everything needs a ticket e.g. users, servers, services etc.

When you log into the gateway nodes you can check the status of your ticket by issuing the klist command. This will show you the user name, realm and valid and expire dates.

If you need to run non interactive workloads you can create keytabs to use with your jobs. Some useful information around how to set up and use keytabs is located here.

Client side kerberos is covered in the next section.

Client Side kerberos

BT’s Hadoop clusters are exposed to end clients via gateway nodes. To inject data remotely (direct into HDFS) your client system must have a route to port 14000 on the gateway node for the cluster you wish to connect to. Your Linux client must also be running Vintela/VAS. This is the only way you can interact with Kerberos on BT Linux systems.   Change notes coming soon.

HDFS Permissions

HDFS permissions on HAAS service groups are monitored daily and an email is sent to the group owner and BT data security team to advise on non-compliance.

When a HAAS service is provisioned, the top level folder /user/HAASxxxxx is set to 770 (drwxrwx---) on Owner and Group, where group is a HAAS AD domain group. This means that when a user needs access to a HAAS instance to work on data sets this request is actioned via order gateway and fully approved by the service owner.

An alternative method to allow HDFS data access is for the owner to configure HDSF ACL's, this facility is explained in the HaaS cookbook.

You should not change file permissions on this folder to provide additional user access, as this can lead to potential security issues and circumvents the processes above.

HDFS Transparent Encryption

The HDFS Encryption at Rest (EAR) cluster service is now avaialable on the ROBT cluster 1A with ACF approval. Overall, EAR supports the storage of In Strictest Confidence data and that's what it should be used for.

A standard HAAS encrypted service consists of the following components and is created on an existing service. Therefore there is no need to order any new HAAS services to use this functionality.

- An encrypted zone (EAR directory) located in the high level service directory e.g. /user/HAASA1234/EAR.
- Encryption is completely transparent and all keys are managed via the cluster services, therefore no need to store any additional details or alter job code.
- Encryption zones are owned by the service owner and managed using HAAS service AD groups as per standard procedures.
- Encryption ACL's are created using the values above to enable decrypt and read activities on the zone and its contents.
- The encryption algorithm is AES-CTR mode with 256 bit Keys with a separate key for each service.

To order a HAAS encrypted service please:

- Raise service request, following the instructions in this FixIT article - selecting Hadoop/HaaS in step 4 - to raise a Bridge case. This will ensure your case is sent to the correct queue.
- The request should contain the HAAS service ID, ACF Tag & the Personal Functional Account if different from the service owner.

Search...

Security Standards and Policies

Current BT security standards and policies can be found here

Search...