hosted by Alibaba Group 阿里巴巴集团 APACHE HBASE

Kerberos-based Big Data Security Solution and Practice in Alibaba Cloud HBase

Jiajia Li @ Intel
Chao Guo @ Alibaba

August 17, 2018

Content

Hadoop Authentication Service

01

Jiajia Li Intel

Content

1.1 Background

Background

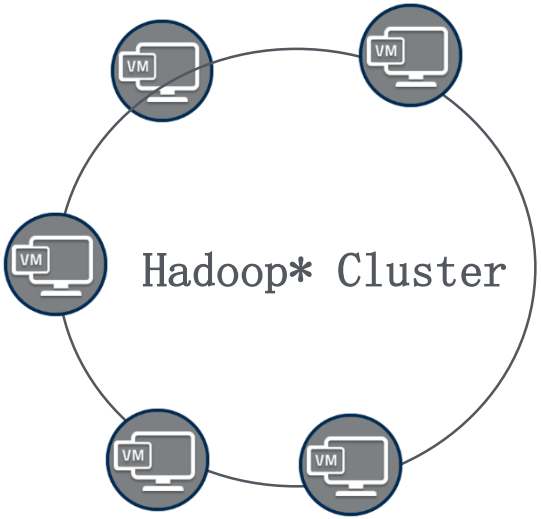
Motivations

- Jan 2017, The ransomware attacks on poorly secured MongoDB (Over 27000 MongoDB Databases Held For Ransom Within A Week), cyber crooks have started targeting unprotected Hadoop Clusters as well (Hadoop, CounchDB Next Targets in Wave of Database Attacks), until now, large amounts of Hadoop cluster data is still exposed in the public network (Insecure Hadoop Clusters Expose Over 5,000 Terabytes of Data).

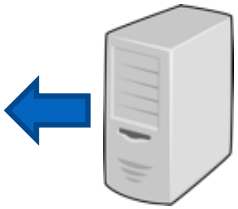
Background

Multiple Ways to Attack Insecure Hadoop Cluster

Service



I’ m Datanode



DANGER

User

login with hadoop

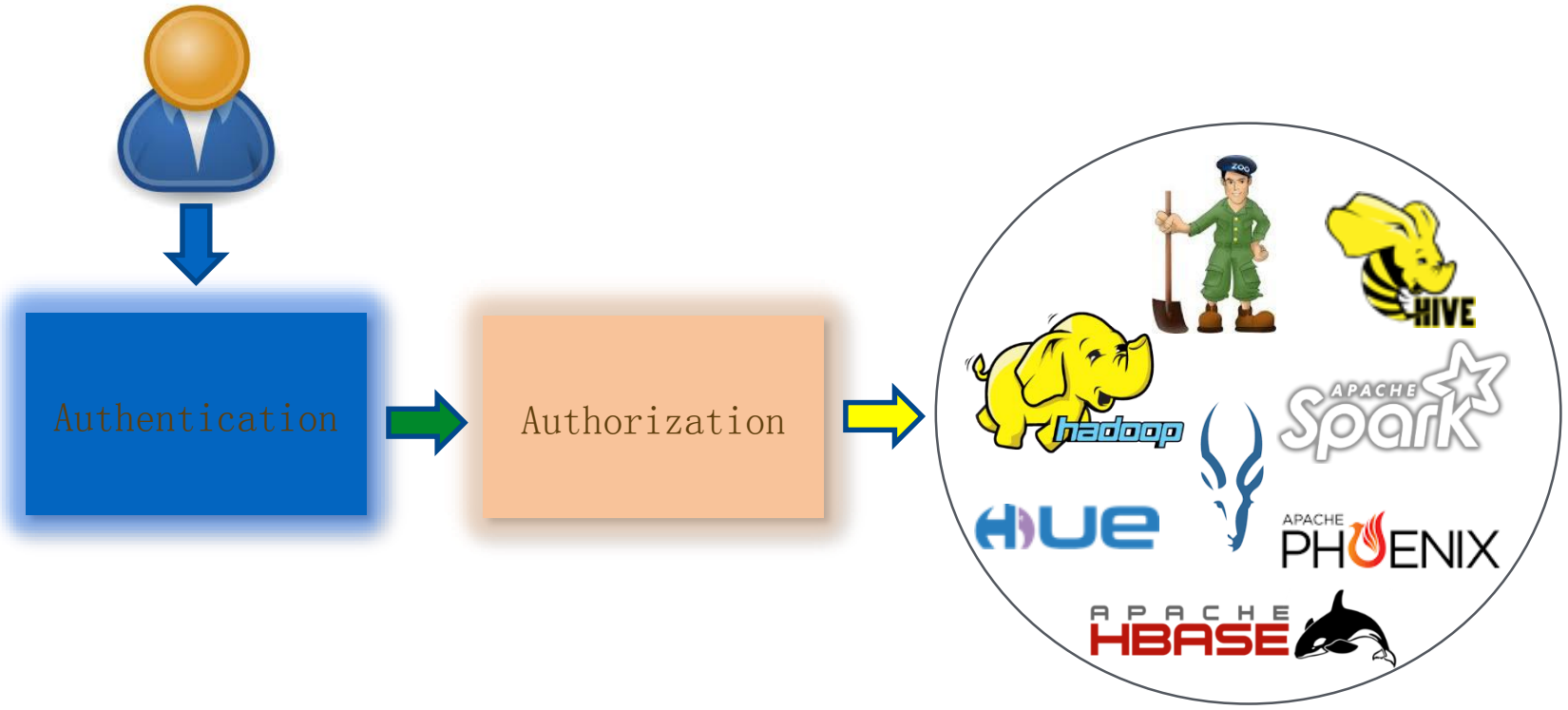
hadoop fs -rmr /

HDFS with ACL
superuser: hadoop

DANGER

Background

How to Secure a Hadoop Cluster



- Authentication
  - Kerberos is the right approach adopted for Hadoop security
  - MIT Kerberos, Azure AD, HAS
- Authorization
  - Apache Sentry(Cloudera), Apache Ranger(Hortonworks)

1.2 Introduction to HAS

Introduction to HAS

Challenges for the Existing Solution

Hard to integrate existing identity management systems of enterprises to Kerber

Over the past few years, multiple cloud providers have introduced Hadoop-as-a-Service and more organizations consider the cloud as a component of their Hadoop deployments

Java lacks a comprehensive Kerberos library. The Kerberos support in Java/JRE i

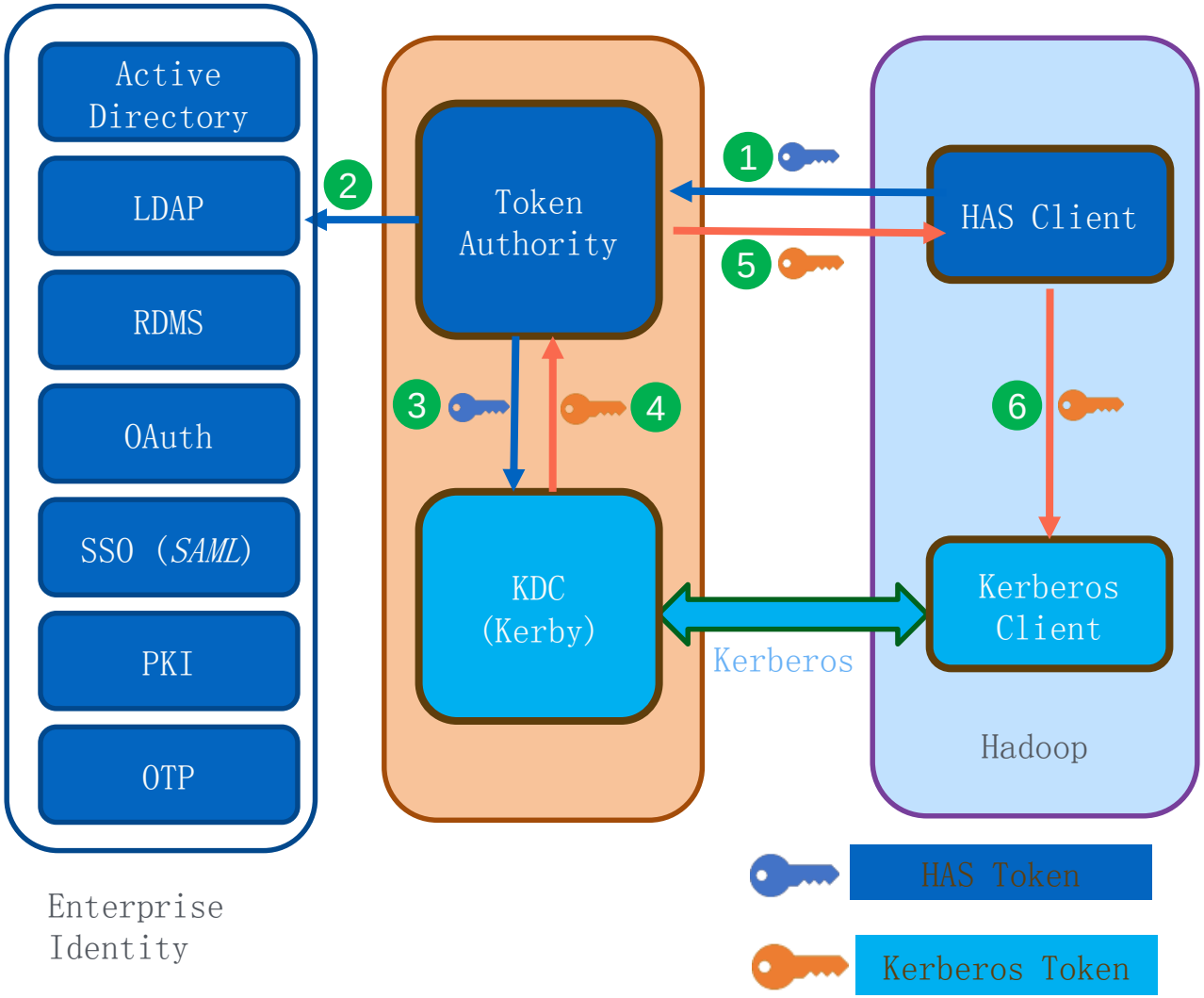
- Limited, lacking full encryption and checksum types
- Hidden from GSSAPI/SASL layers
- Evolving slow

Very difficult in Kerberos cluster deployment

Kerberos is essentially a protocol, or secure channel, doesn't have to be that complex to most or normal users, hiding the details

Introduction to HAS

HAS System Architecture



HAS is a secure and extensible authentication framework for addressing the problem of integration of enterprise identity with Kerberos centric Hadoop ecosystem.

Introduction to HAS

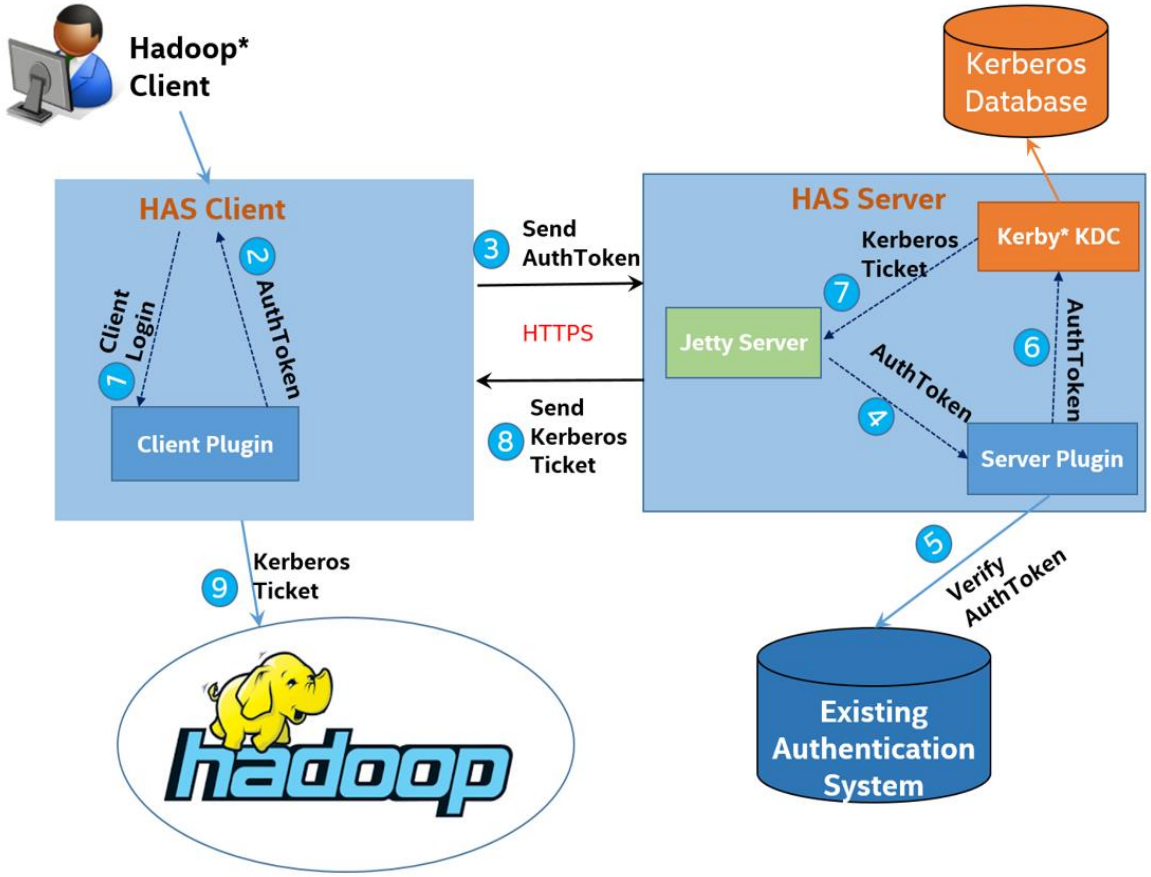
Key points of HAS implementation

- Hadoop services continuously use the original Kerberos authentication mechanism.
- Hadoop users can also continue to log in using a familiar authentication mode.

- In the new authentication mechanism, the combination of your plugin and the existing authentication system can be customized and implemented.
- Based on the new authentication mechanism, security administrators don’t need to synchronize user account information to the Kerberos database.

Introduction to HAS

HAS protocol flow



Introduction to HAS

HAS plugins

HAS client plugin HasClientPlugin:

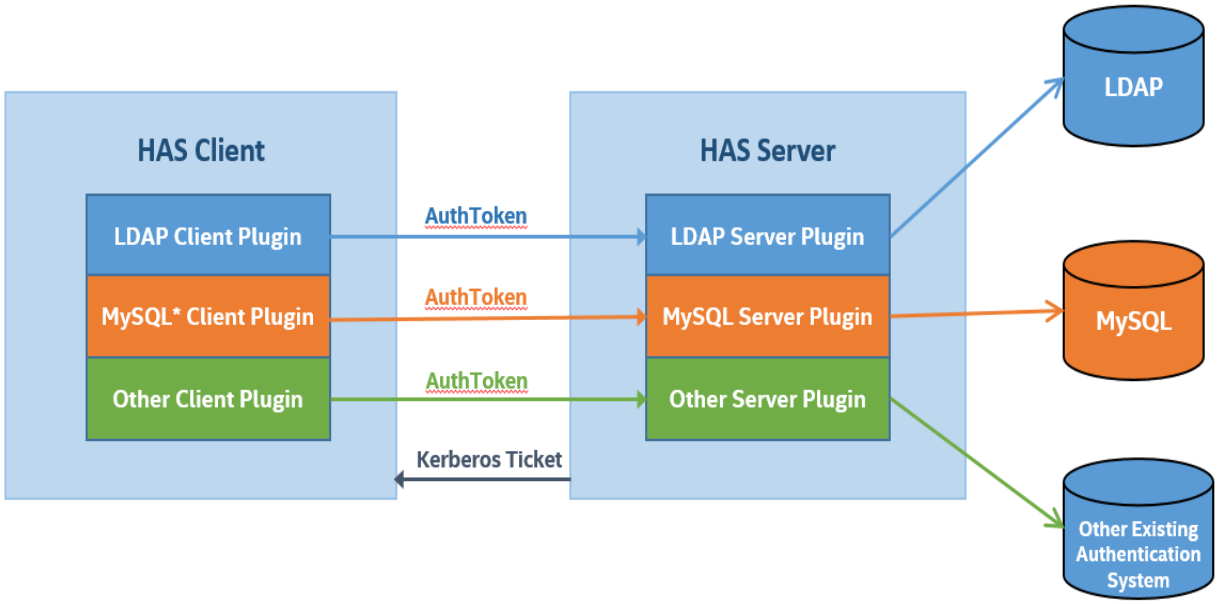
```
// Get the login module type ID, used to distinguish this module from others.
// Should correspond to the server side module.
String getLoginType()

// Perform all the client side login logics, the results wrapped in an AuthToken,
// will be validated by HAS server.
AuthToken login(Conf loginConf) throws HasLoginException
```

HAS server plugin HasServerPlugin:

```
// Get the login module type ID, used to distinguish this module from others.
// Should correspond to the client side module.
String getLoginType()

// Perform all the server side authentication logics, the results wrapped in an AuthToken,
// will be used to exchange a Kerberos ticket.
AuthToken authenticate(AuthToken userToken) throws HasAuthenException
```



1.3 Outlook and Summary

Outlook and Summary

- The new authentication mechanism (Kerberos-based Token Authentication) provided in HAS supports most components in the Hadoop ecosystem and makes little or no change to the components.

- HAS open source in the branch of Apache Kerby project (https://github.com/apache/directory-kerby/tree/has-project).

- HAS will be merged to trunk in the master JIRA (https://issues.apache.org/jira/browse/DIRKRB-671).

- According to the community plan, the HAS feature will be released in Kerby 2.0.0, and Kerby 2.0.0 will be released in the near future.

02 Security Practice in ApsaraDB for HBase

Chao guo Aliyun

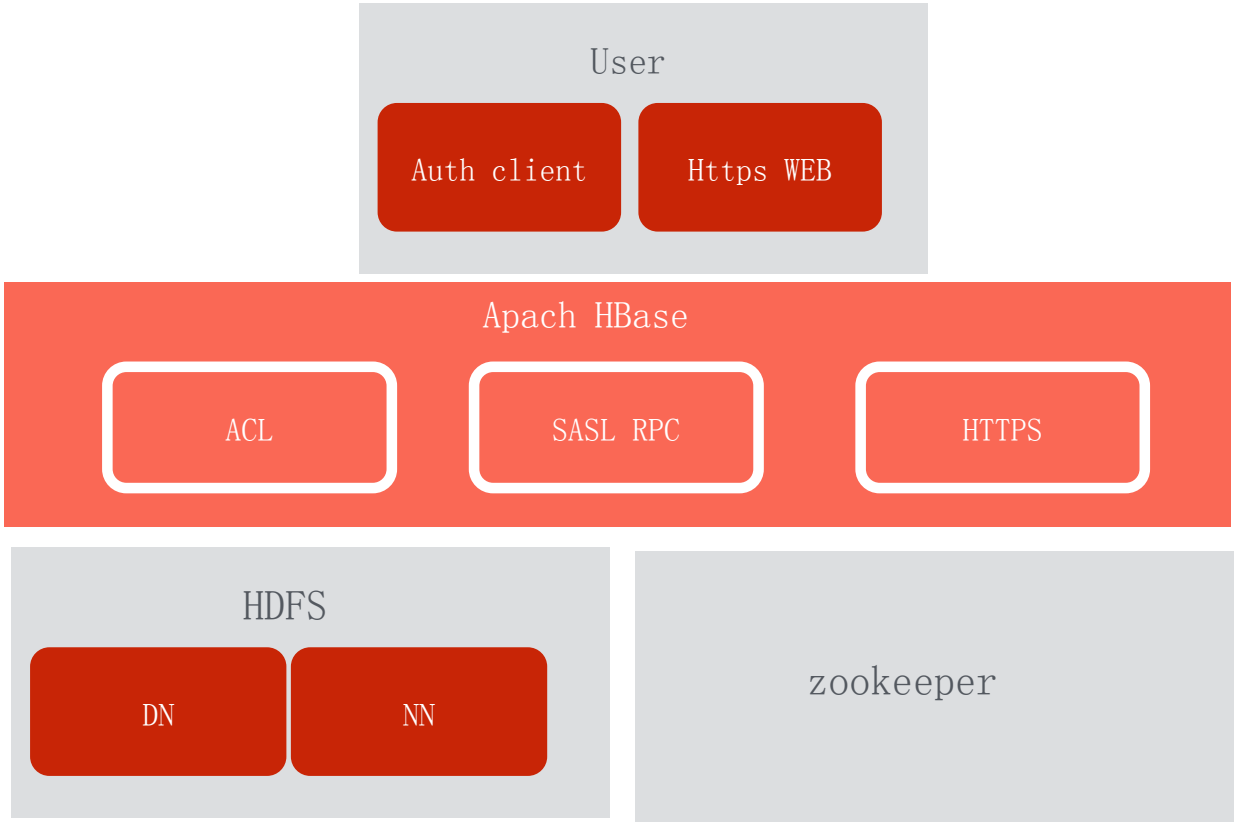
Content

2.1 Introduction to Apache HBase Security and Security of ApsaraDB for HBase

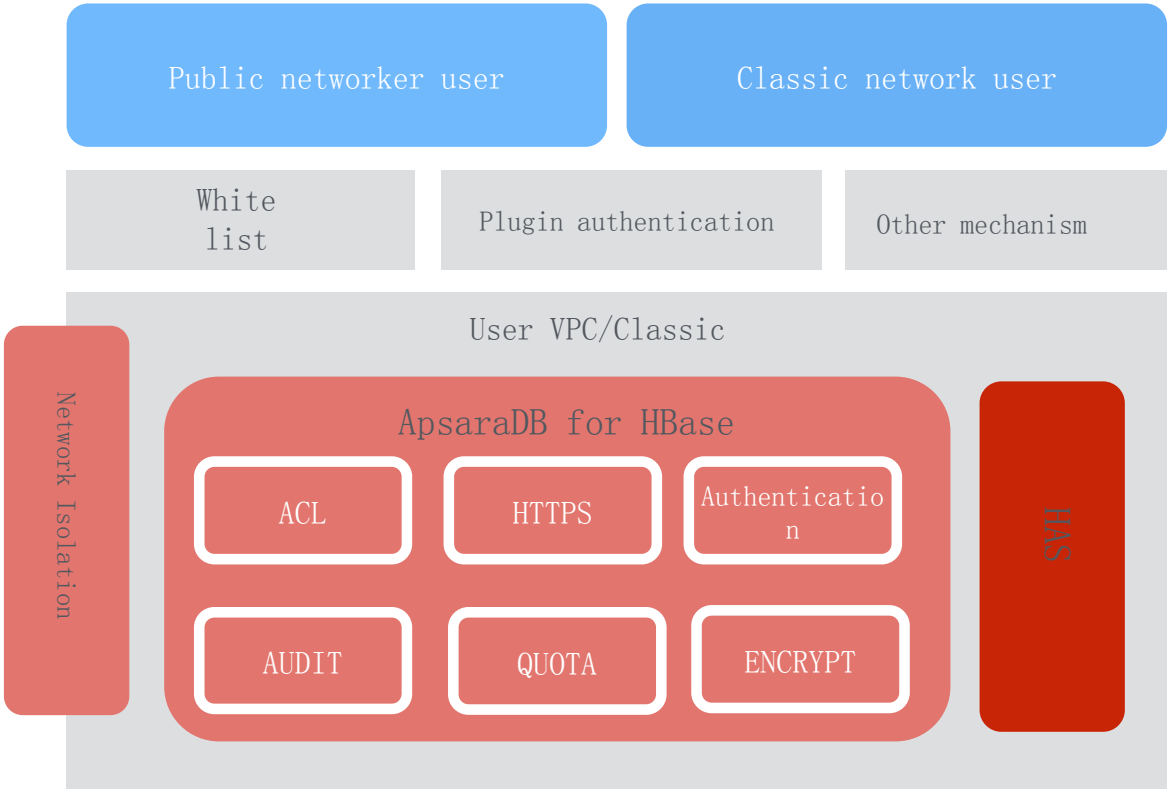
Introduce to Apache HBase Security

Apache HBase Security includes :

1. Access control Label base on Access Controller coprocessor;
2. Using Secure HTTP for Web UI.
3. Kerberos authentication for RPC;



Introduction to ApsaraDB for Hbase security



ApsaraDB for HBase main function:

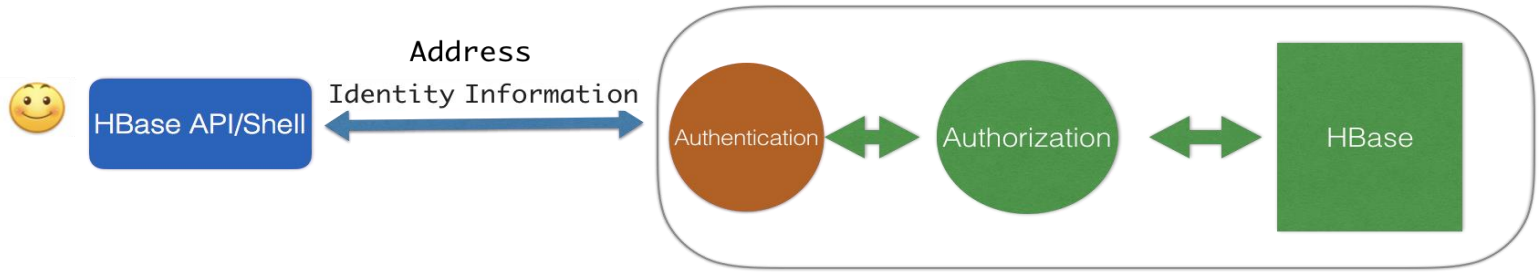
- Network Isolation and white list;
- HAS authentication;
- …

Introduction to ApsaraDB for HBase Autnentication

- Instal MIT kerberos at the locality;
- Set up local krb5.conf, the Kerberos service address, realm and so on of the HBase;
- User name and local system user name should be the same, If got no user, need create;
- Set up security configuration in hbase-site.xml,core-site.xml,hdfs-site.xml.For example:hadoop.security.authentication, dfs.namenode.kerberos.principal;
- Do kinit, run kinit throw user passwor mode or keytab, get the legal user tiket;
- Access to hbase throw hbase shell .

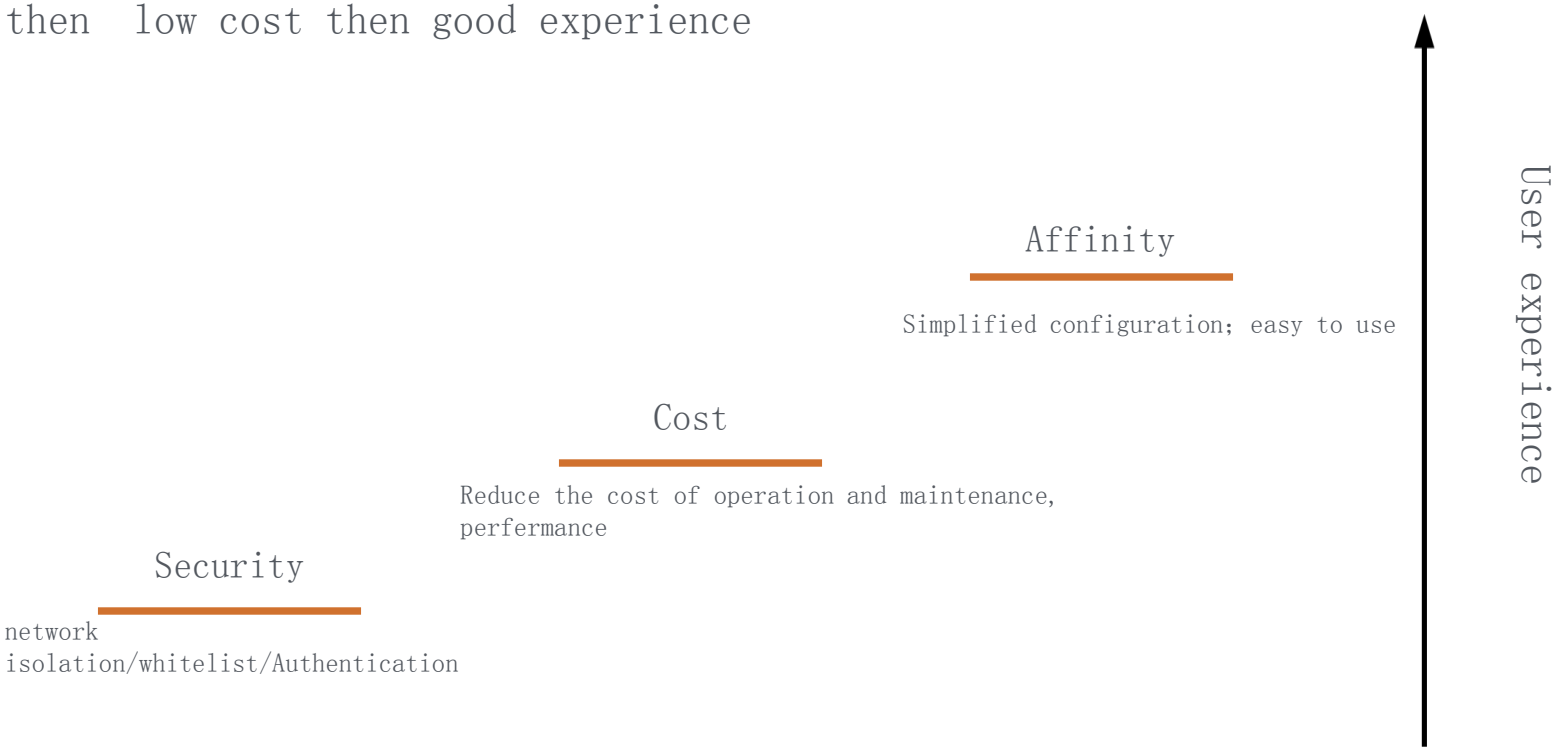
VS

- Set up hbase zookeeper address, user’s password and name;
- Access to hbase throw hbase shell .



User experience

User’s needs start basically from security then low cost then good experience



2.2 ApsaraDB for HBase Optimization base on HAS

ApsaraDB for HBase Optimization base on HAS

Basical Introduction

Why choice HAS for us?

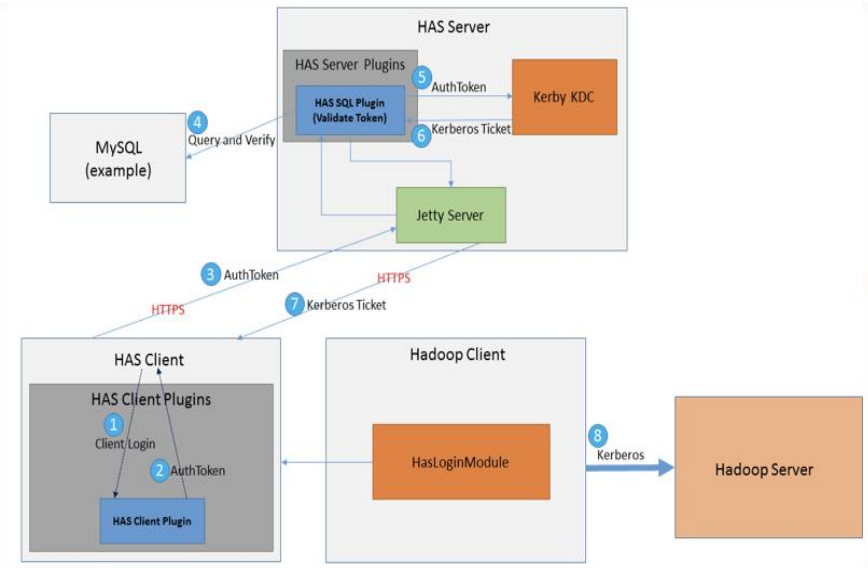
- Kerberos is the only means to enable hadoop security
- Compatible with existting security mechanism :ACL、keberos …
- Easy to deploy
- Easy to use for client
- Good scalability
- Performance is ok
- Low operating cost;
- …

ApsaraDB for HBase Optimization base on HAS

ApsaraDB for HBase’s authentication improvement

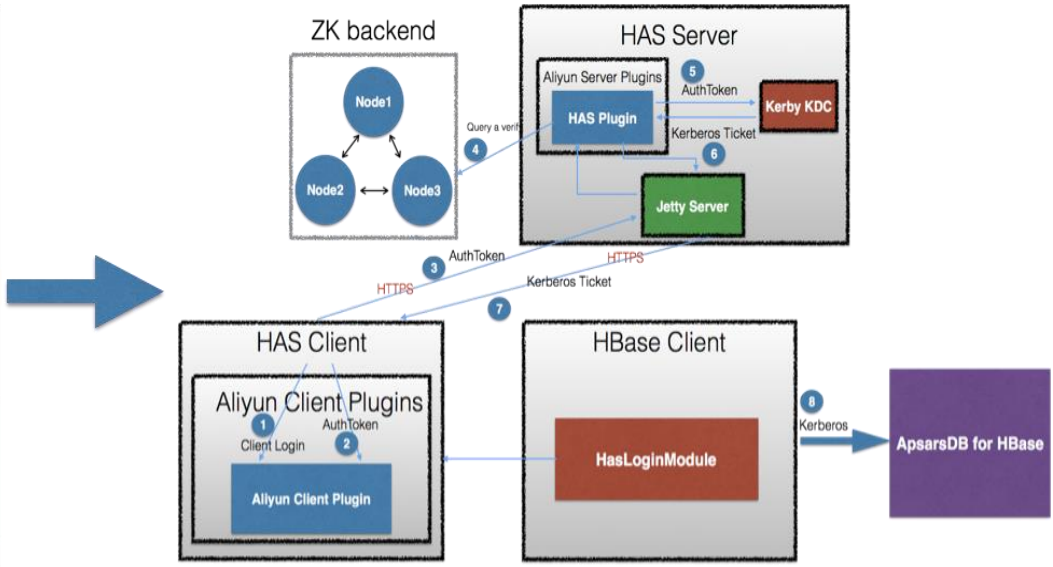
Security and practice

Whitelist for hosts and other securities



Account password management

Using plugin for password and account management

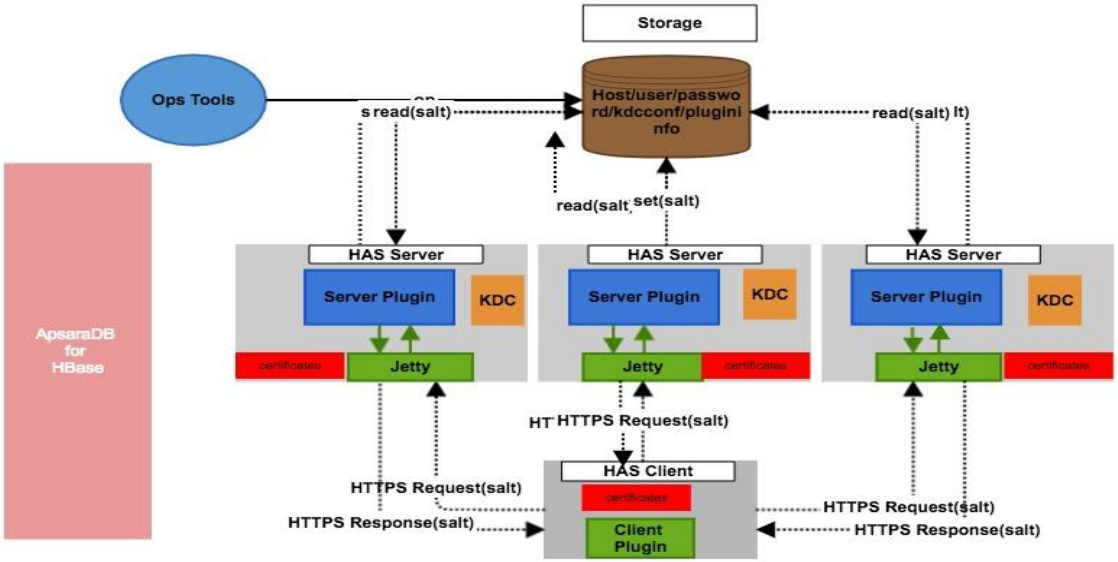


High availability backend

Impement zookeeper like backend for storage of user name/password/whitelist host/kdcconf

ApsaraDB for HBase Optimization base on HAS

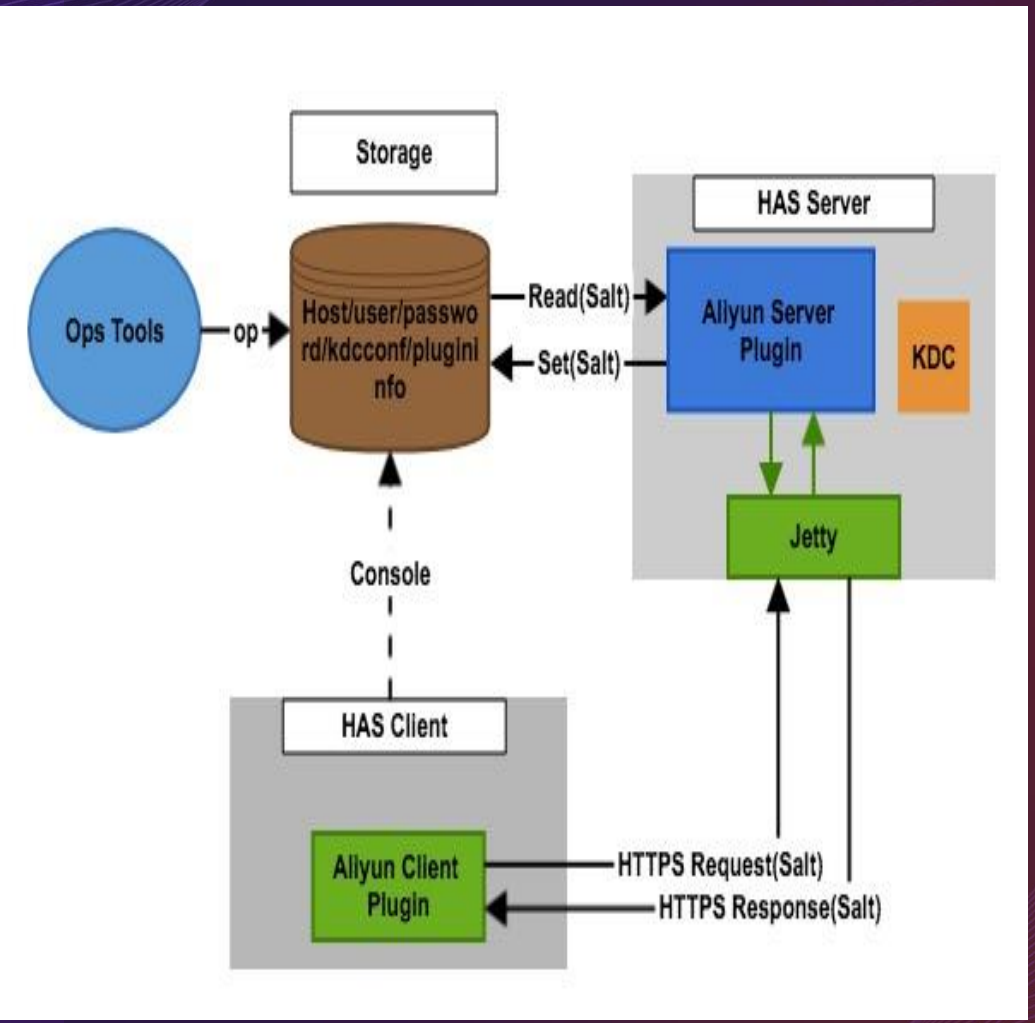
Security and practice



- HA for HAS server
- White list for host access
- Plugins use HTTPS(Initialiazation/usage)
- Configurable salting algorithm
- Backward compatible kerberos and all Hbase security
- Ops tools : one button to deploy

ApsaraDB for HBase Optimization base on HAS

Account password management



We have done some service on has like:

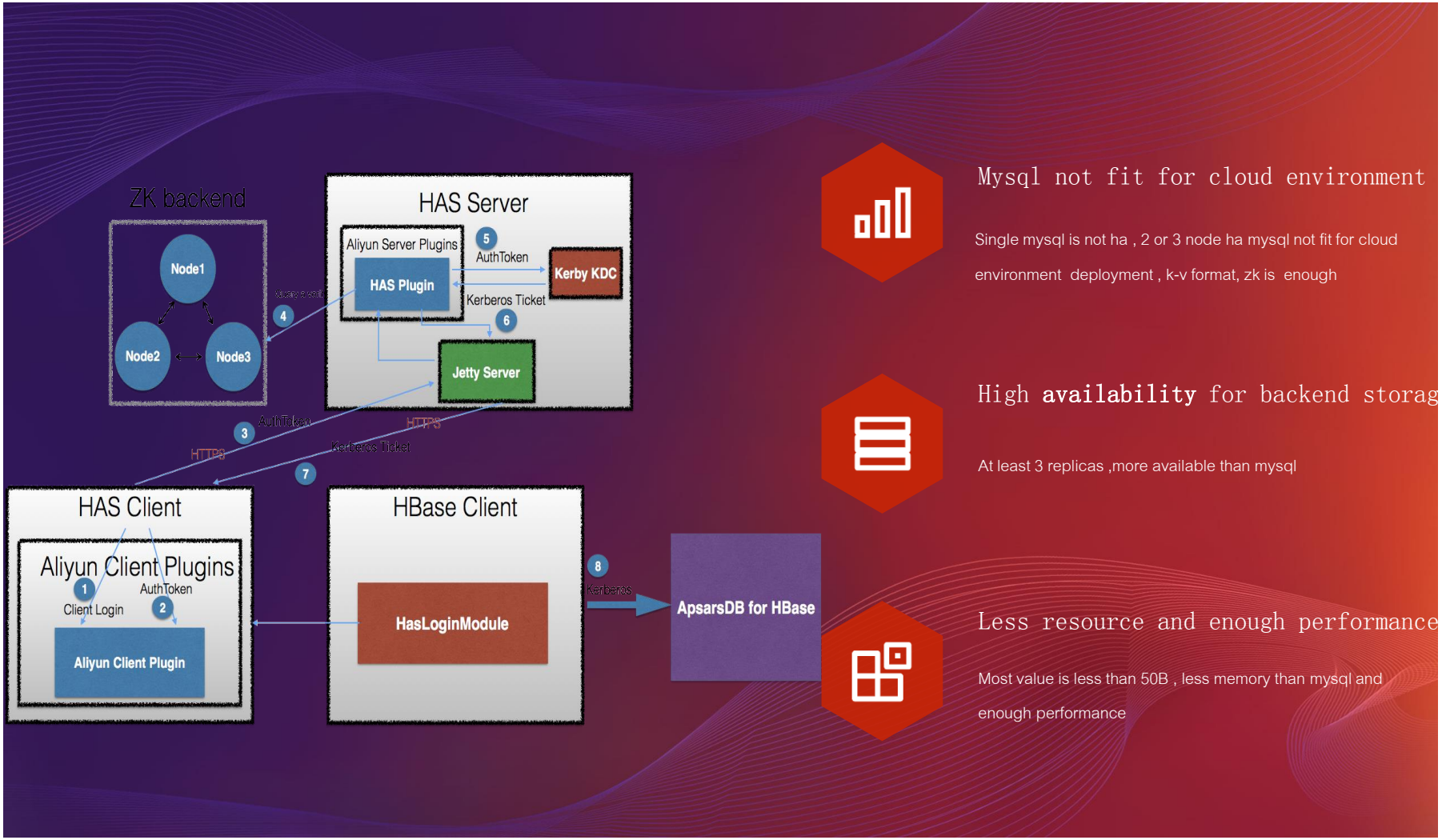
Aliyun Client/Server plugin mode

- Client can initialize with their user/password/hosts
- Client pass user/password throw https
- Server plugin verify from storage with salt
- The entire process is safe
- Configure free,easy to use;

Other plugin mode:

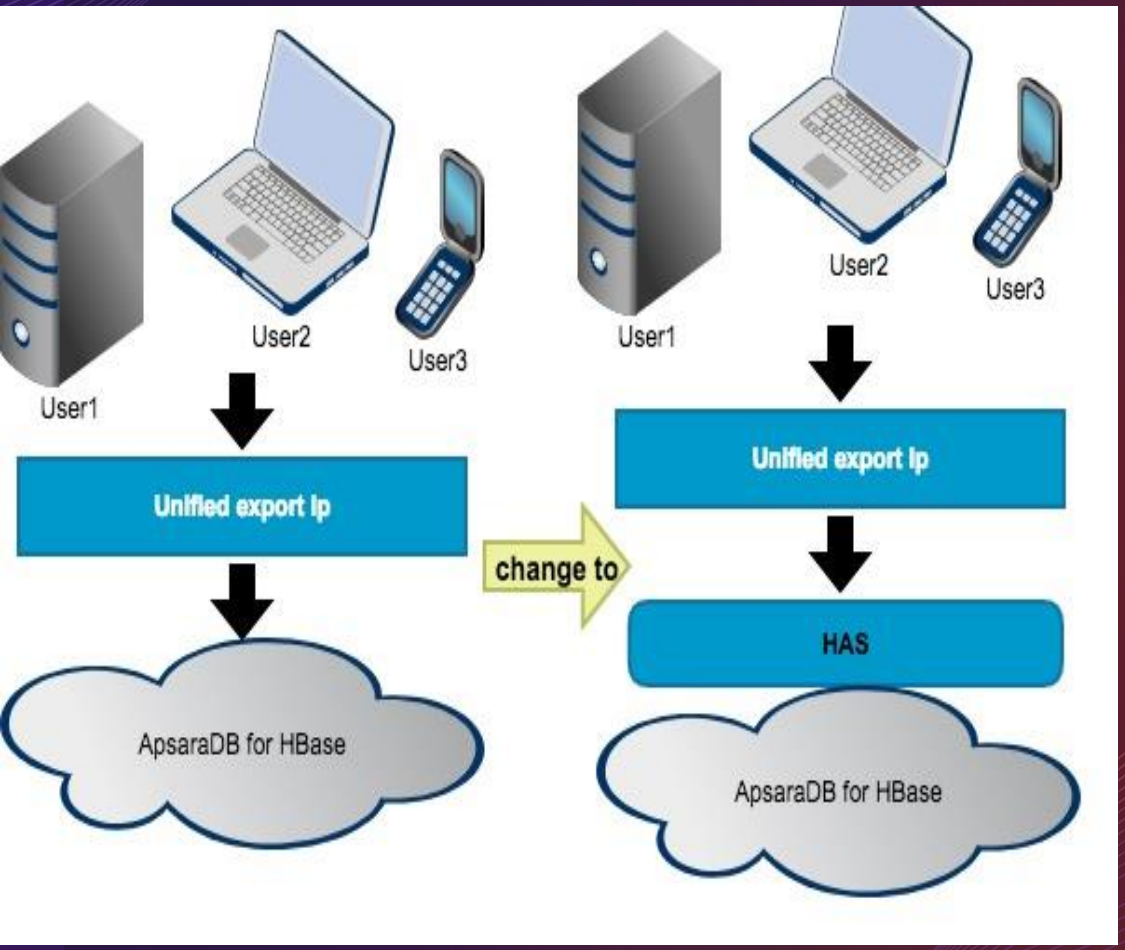
ApsaraDB for HBase Optimization base on HAS

High availability backend



ApsaraDB for HBase Optimization base on HAS

Possible user case



User consulation case:

- All users with same export Ip
- ACL is not suitable
- Users need different authentication

On Cloud almost one small cluster for on user;
But private cloud and big cluster is different;

2.3 Outlook and Summary

Outlook and Summary

- For the security of ApsaraDB for HBase, we got network ioslation 、 whitelist and other mechanism . As authentication ,we adopt HAS.

- We have do some optimization for HAS such as ha、 customized plugins.

- The secure ApsaraDB for HBase is ready now.
  https://www.aliyun.com/product/hbase?spm=5176.8142029.388261.280.702b614asnEokD

hosted by Alibaba Group 阿里巴巴集团 APACHE HBASE

Thanks

Dingding group for HBase

Personal Wechat