

Evaluation of Apache Knox integration with CERN SSO

Author: Thomas Mauran

Supervisors: Emil Kleszcz, Borja Aparicio Cotarelo

30/07/2024

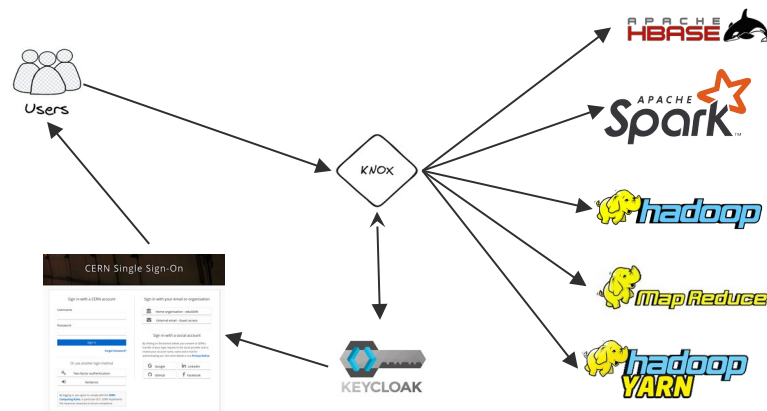


Project Scope

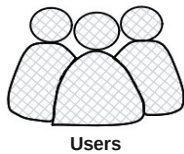
- **Objective**
 - SSO integration with Hadoop clusters for webUIs and APIs
 - Single entry point to all the cluster services
 - Use CERN infrastructure
- **Expected outcome**
 - Test, install and configure
 - Integration with SSO at CERN
 - Keycloak and OIDC
 - application-portal.web.cern.ch
 - Set up a home page with quick links
 - Explore ACL-based authorization
 - Configure auditing feature
 - Test HA support
- **More details:**
 - Latest stable release 2.0.0
 - knox.apache.org



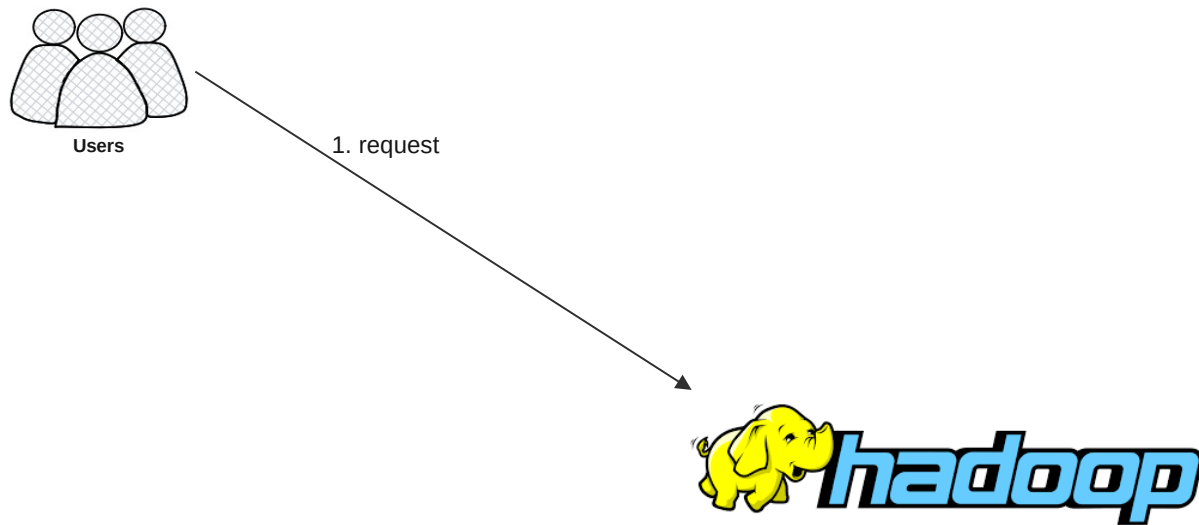
Apache Knox - Trusted Proxy



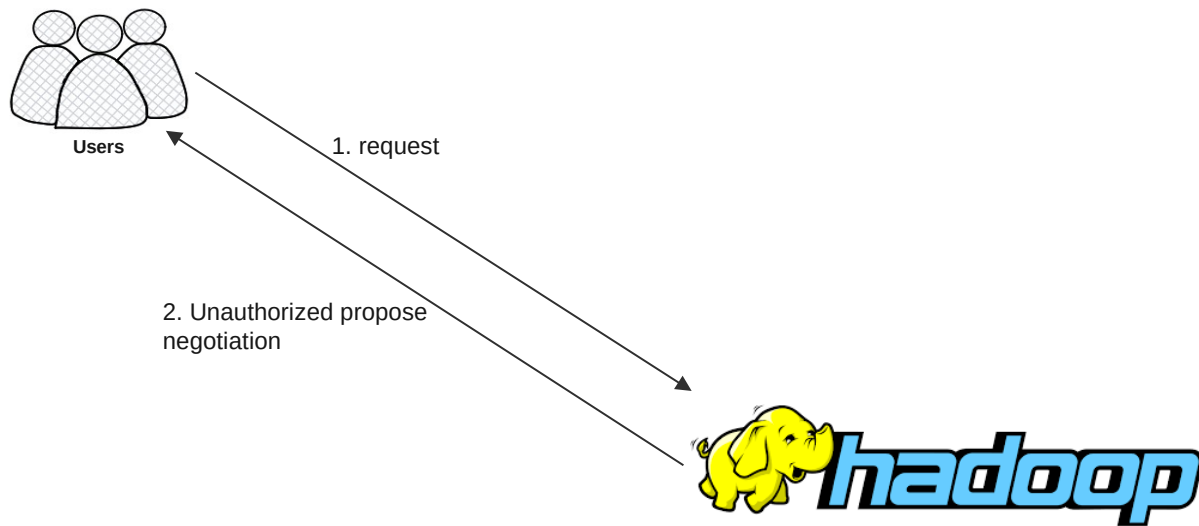
SPNEGO current setup



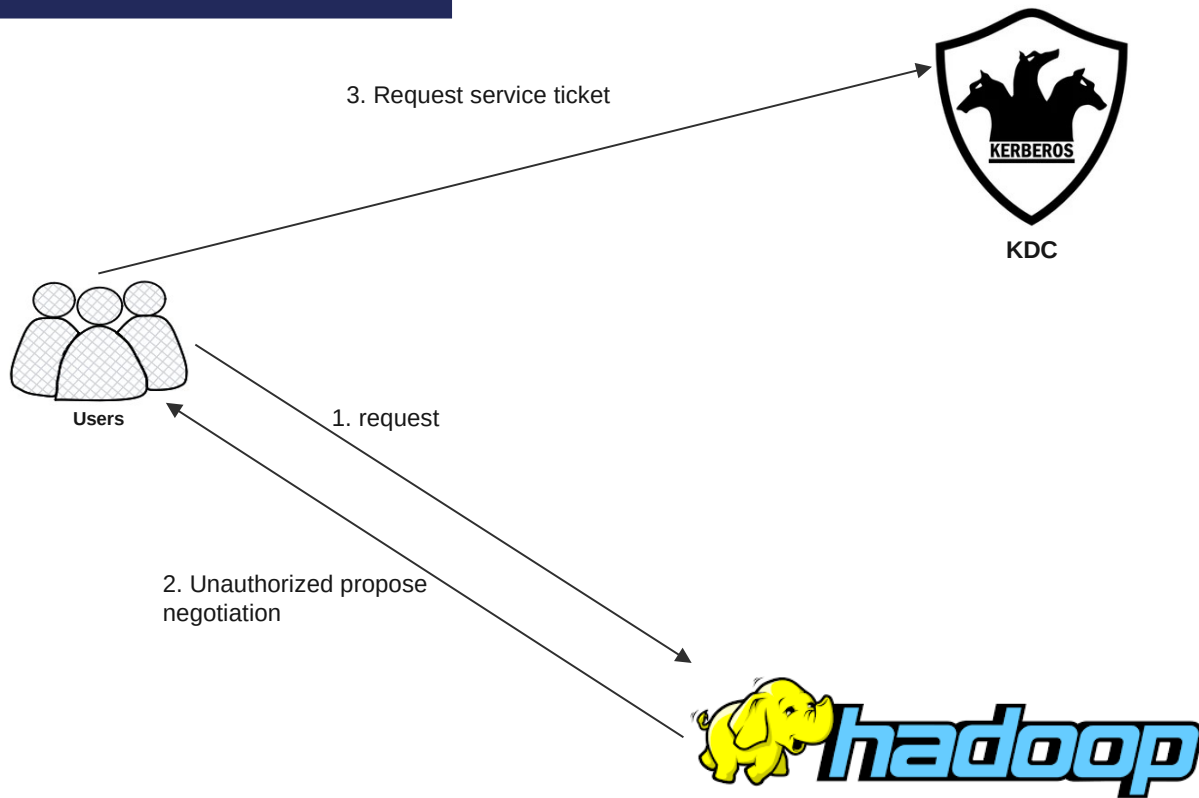
SPNEGO current setup



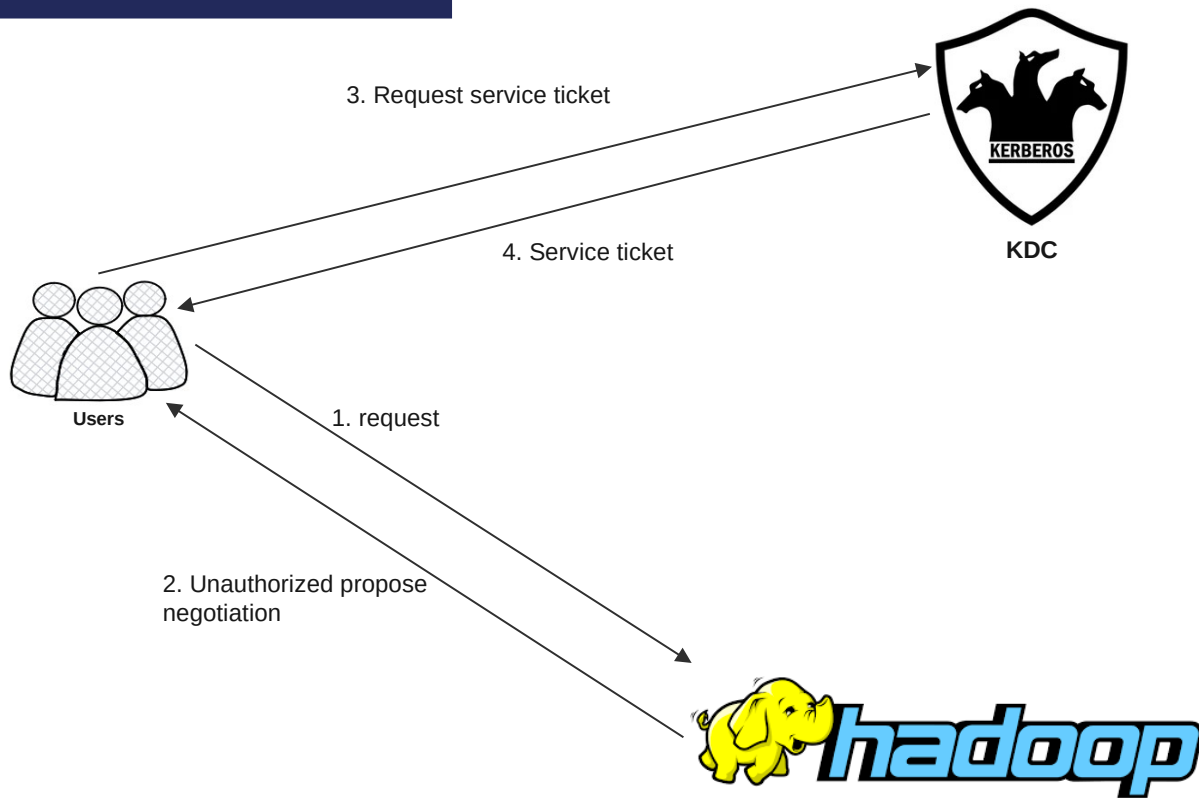
SPNEGO current setup



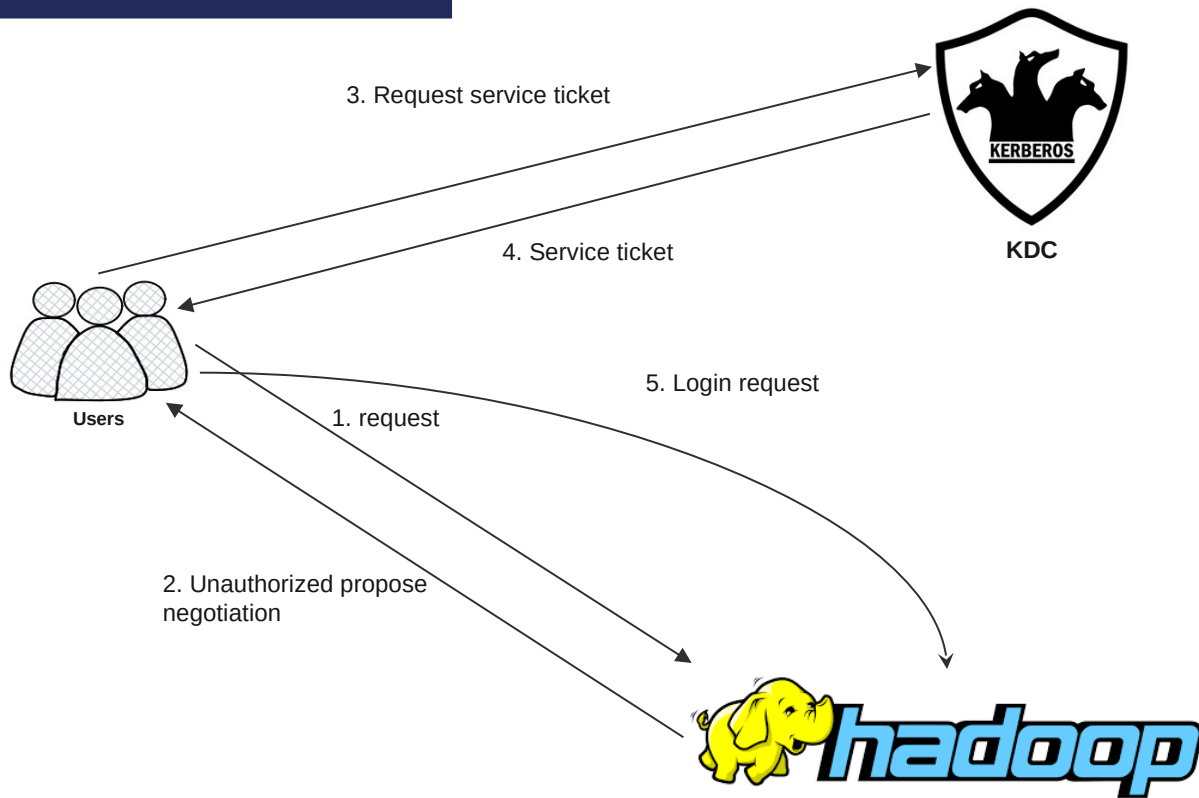
SPNEGO current setup



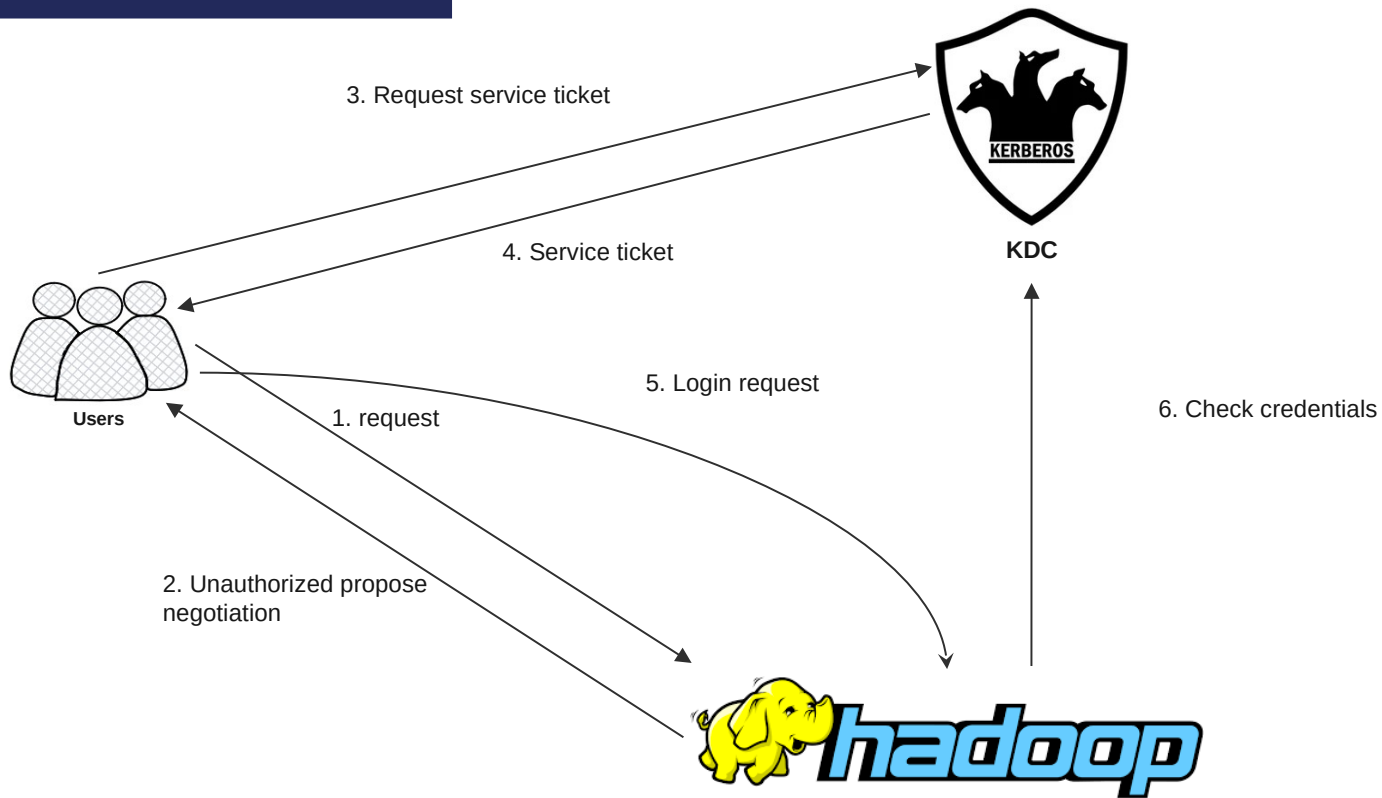
SPNEGO current setup



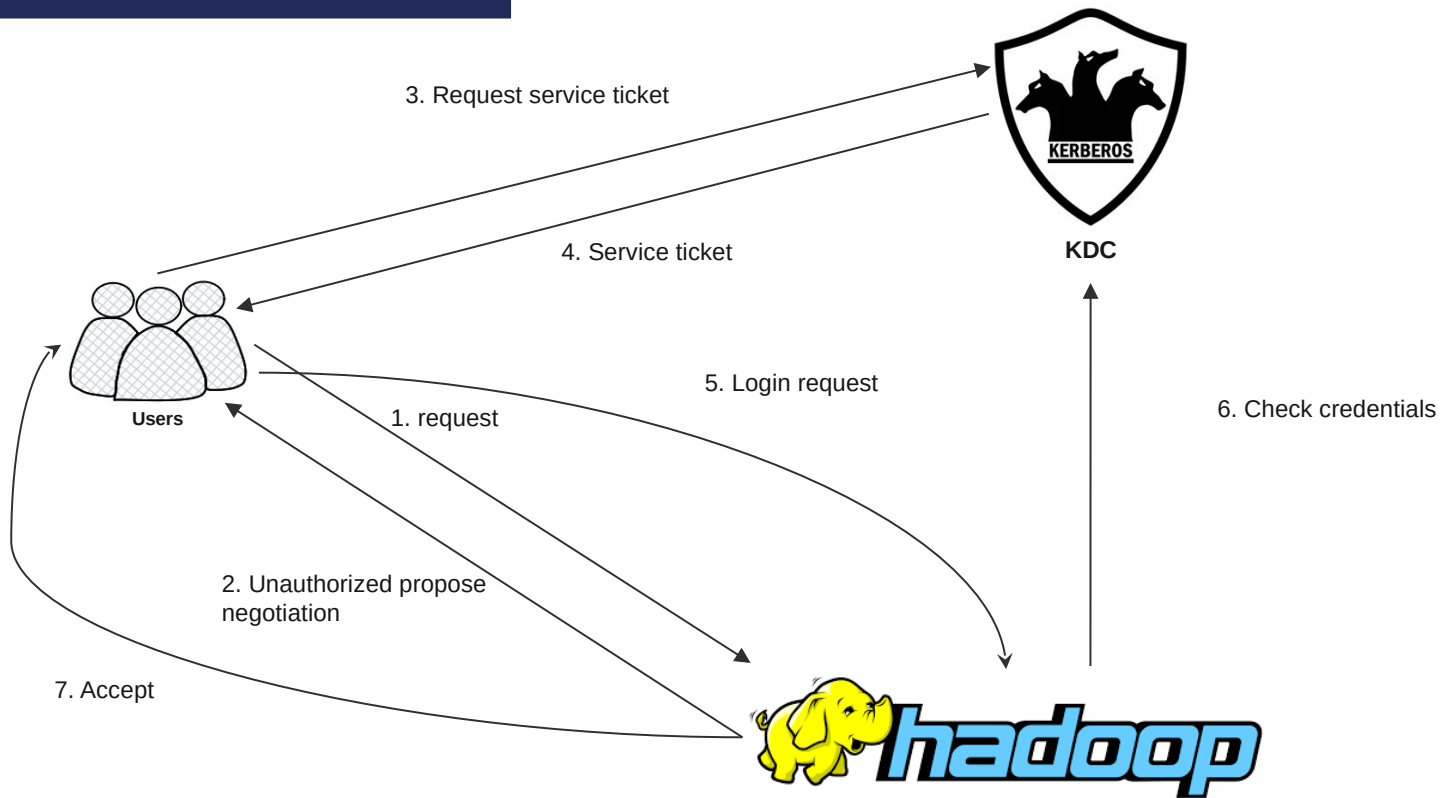
SPNEGO current setup



SPNEGO current setup



SPNEGO current setup



Why do we need it? SPNEGO limitations

- Not standard way to access webUIs at CERN
- Add extra settings in the browser config different for each web browser and OS
- Needs a valid Kerberos ticket or keytab on your local machine
- Configuration on the server side is complex



HTTP ERROR 401 Authentication required

URI: /
STATUS: 401
MESSAGE: Authentication required
SERVLET: org.apache.hadoop.http.WebServlet-49cb1baf

RFH

about a month ago · Additional comments (Customer View)

Hello.

When I access this URL I get 401 Unauthorized Access (<https://thdp6014.cern.ch:8088>). I would appreciate if I could get some help in this matter.

Thank you in advance.
Regards,

Apache Knox

Proxying Services:

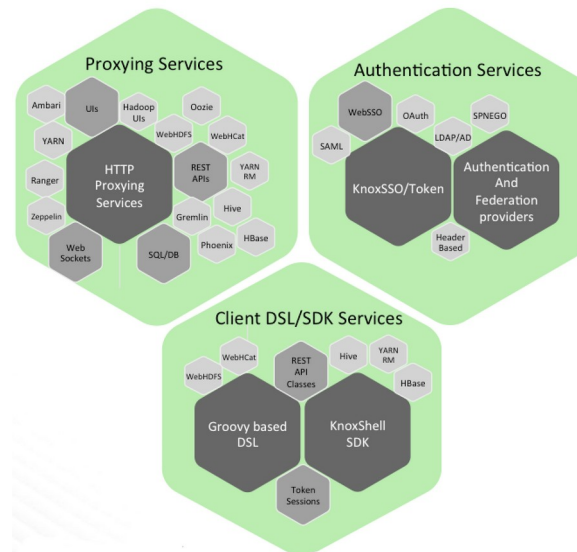
- Provides access to Apache Hadoop by proxying HTTP resources.

Authentication Services:

- Supports various authentication methods for REST API access and WebSSO for UIs:
 - LDAP/AD
 - Header-based PreAuth
 - Kerberos
 - SAML
 - OAuth

Client Services:

- Enables client development through:
 - Scripting with DSL
 - Using Knox Shell classes as an SDK
 - KnoxShell interactive environment combining Groovy shell with Knox Shell SDK classes for data interaction with your Hadoop cluster.



Integration with CERN SSO



Users



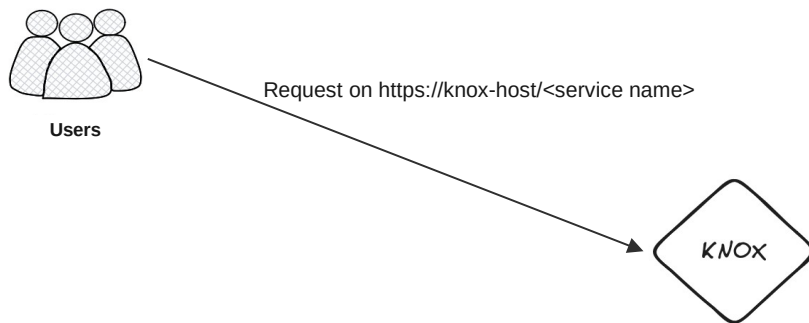
Integration with CERN SSO



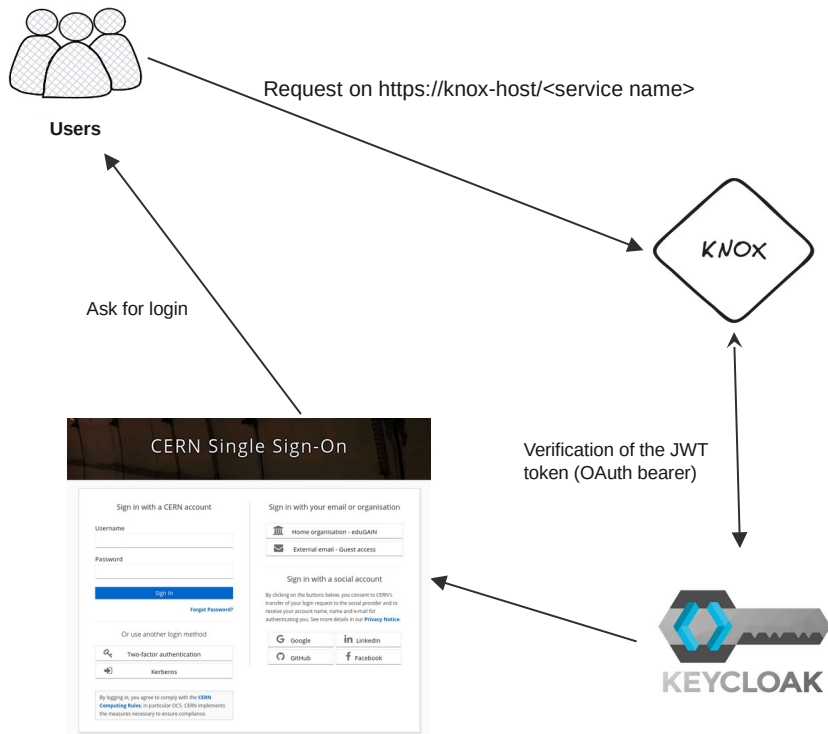
Users



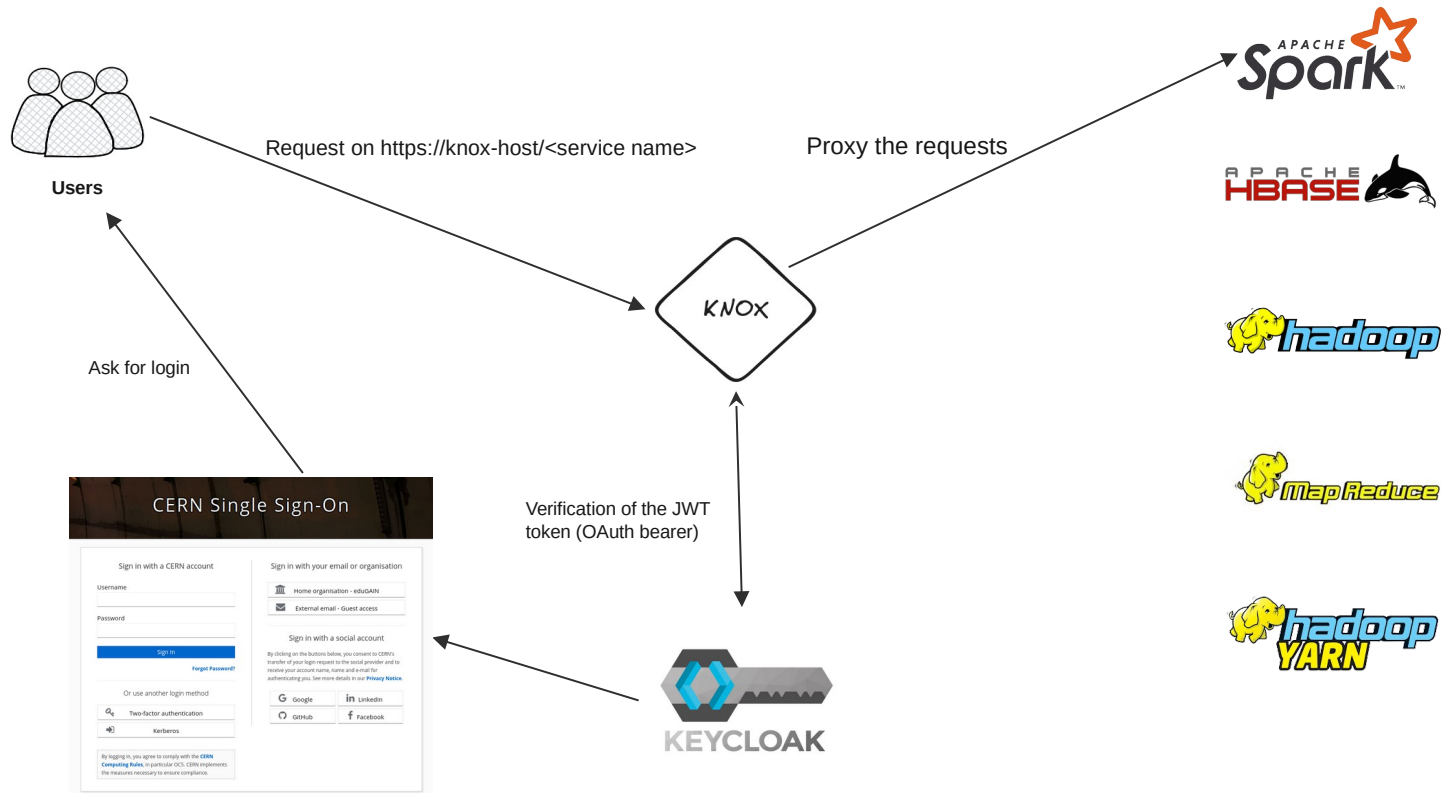
Integration with CERN SSO



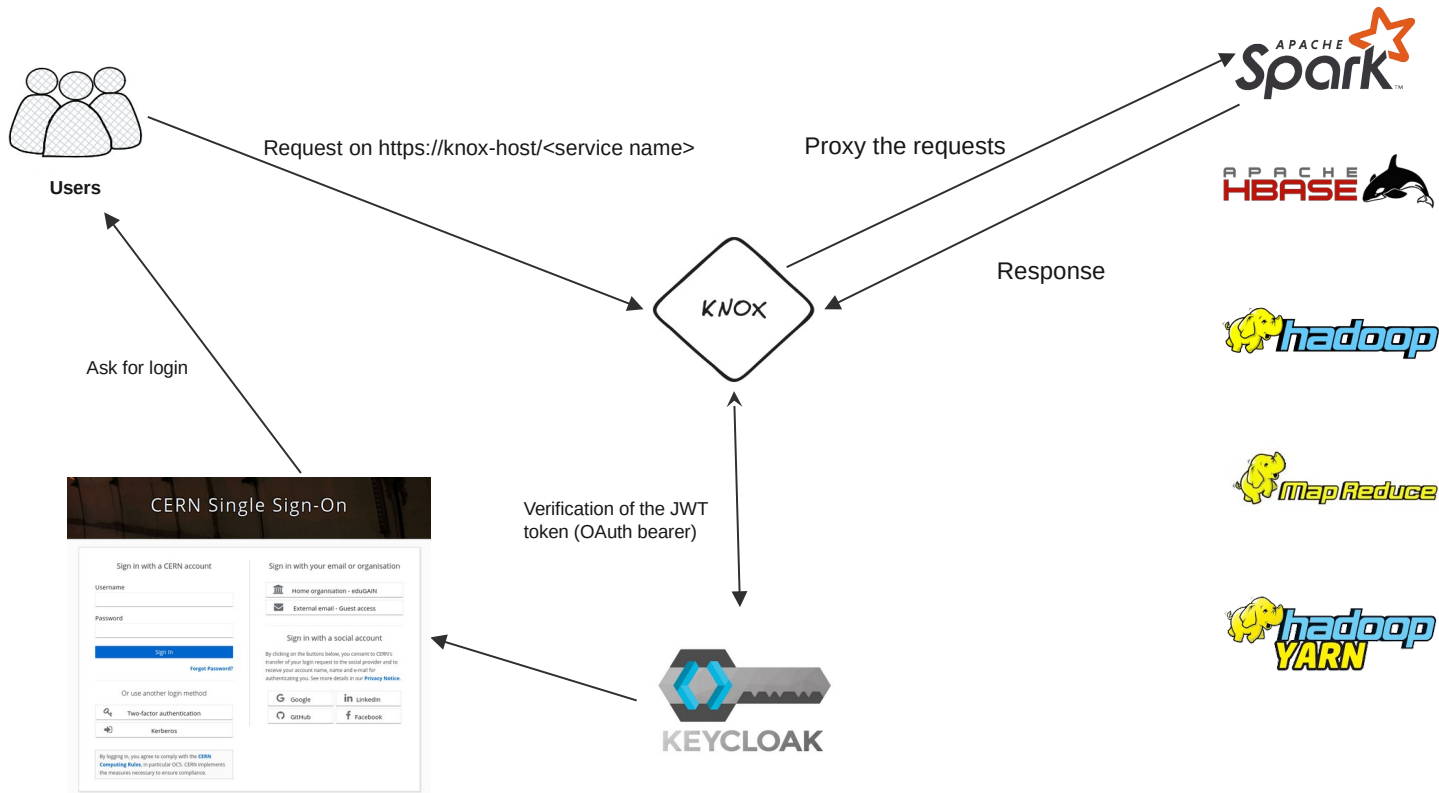
Integration with CERN SSO



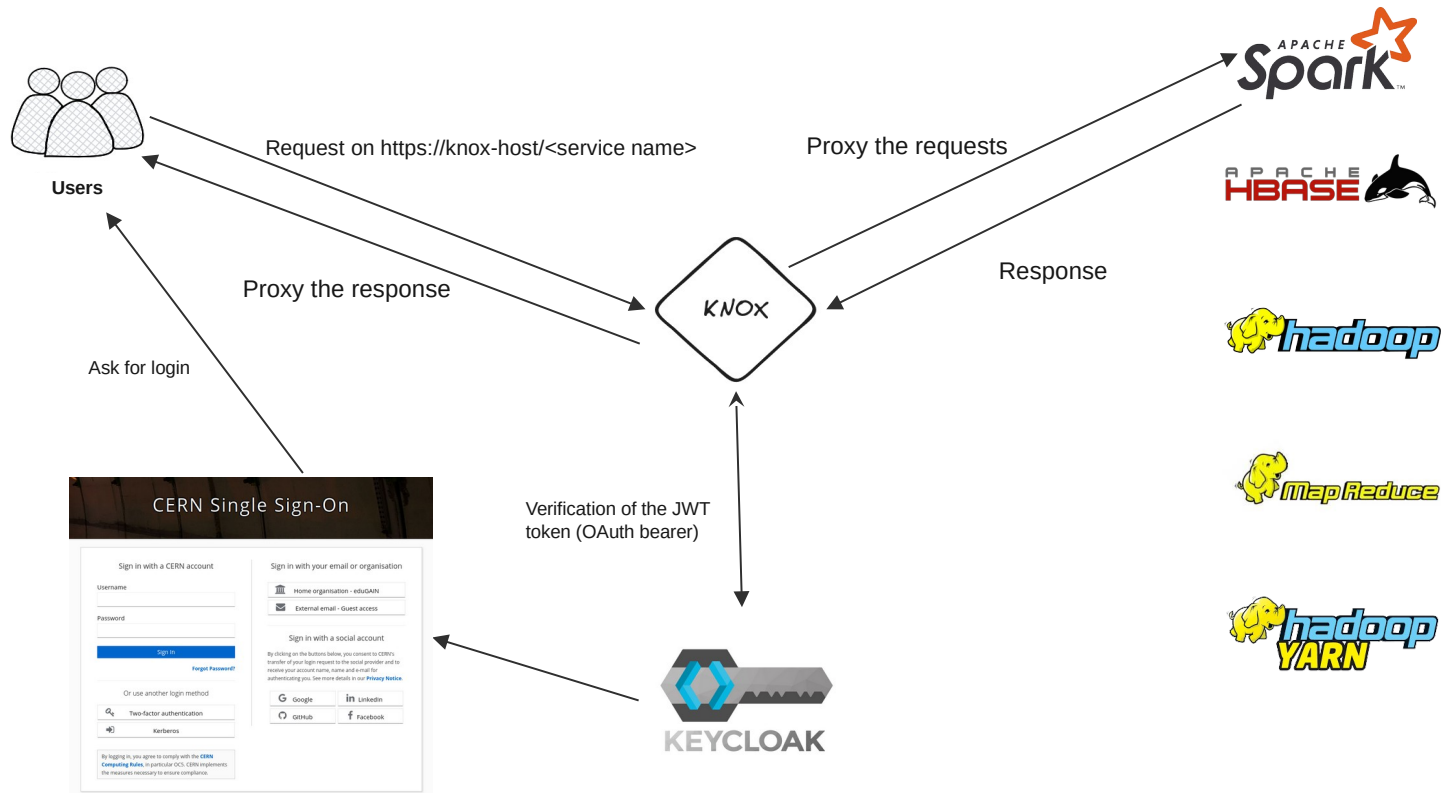
Integration with CERN SSO



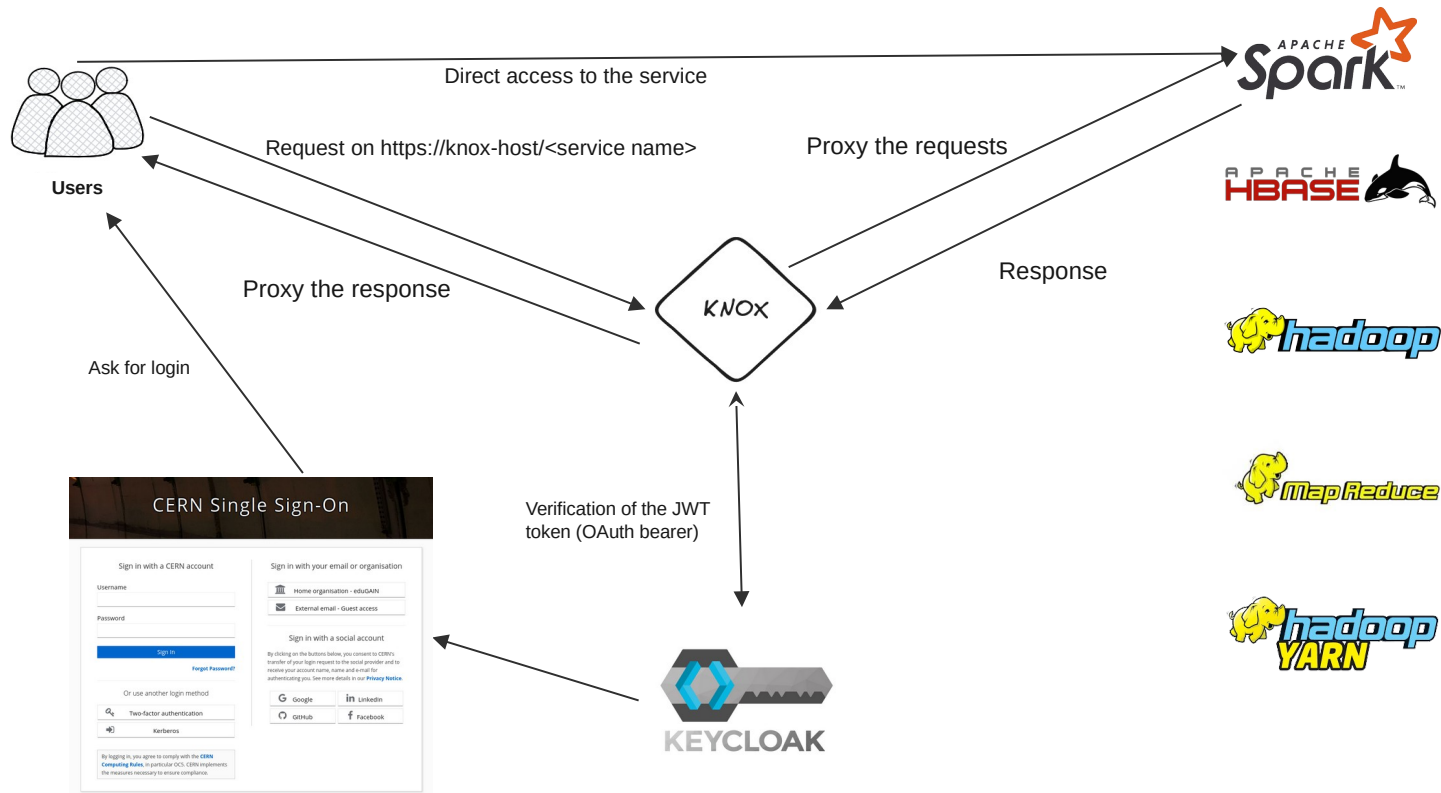
Integration with CERN SSO



Integration with CERN SSO



Integration with CERN SSO



Conclusion

- **Apache Knox allows consistent and standardize way of accessing critical services**
- **Experimentation phase was successful with CERN SSO, HA and AuthZ on all services.**
- **Automatization of the setup is in progress**
- **Next step: Deploy and test it on the staging cluster**



Thank you!

