

Big Data @Roche



Security requirements in enterprise Cloudera Hadoop

Paweł Dudek,
Roche Polska (ADMD)



Big Data @Roche



*What do we think about the
security?*



• *Business requirements*

• *Constraints*

• *Additional effort and time*

or

• *Data security*

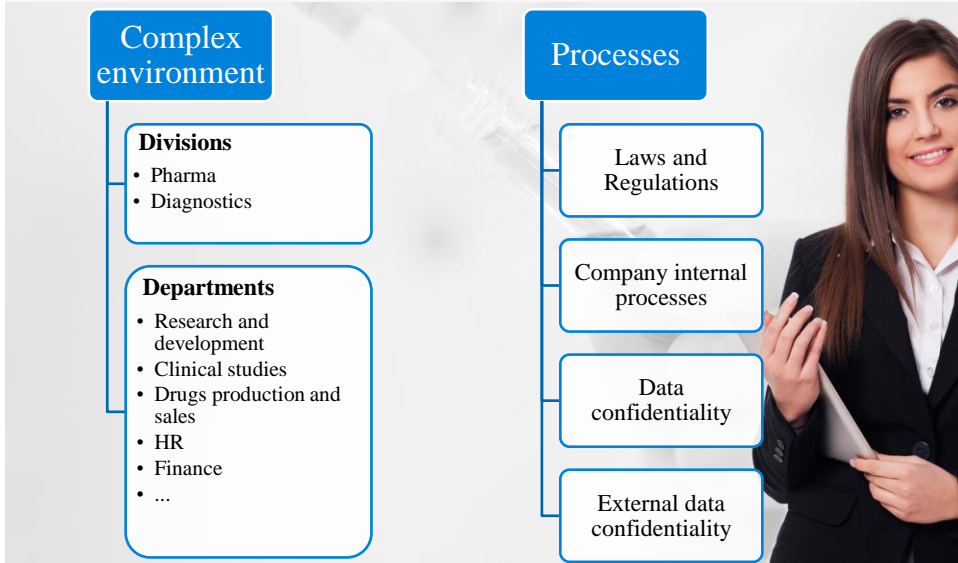
• *Guidelines and awareness*

• *Governance*




Why we need to apply security?

Requirements at Roche

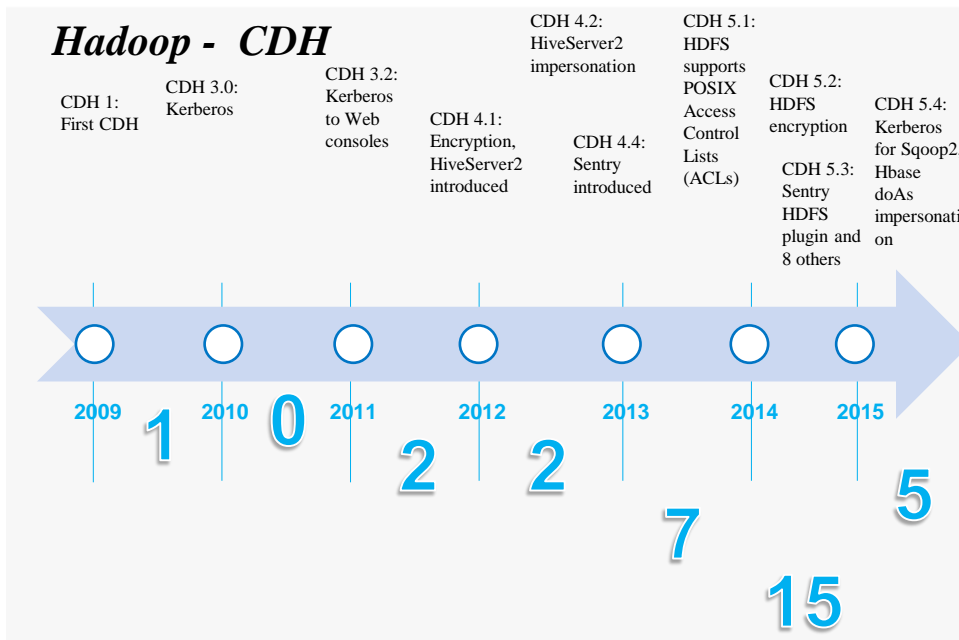


What does this mean for IT?

IT Requirements

 *Multitenancy* *Authentication,
authorization and
encryption* *Audit* *Governance and
guidelines*

What about Hadoop?



How do we address the requirements?

Roche Cluster security analysis



Perimeter

- Guarding access to the cluster itself
- Authentication (Kerberos and Centrify) and network isolation



Access

- Define what users and applications can do with data
- Authorization and Roles/Permissions



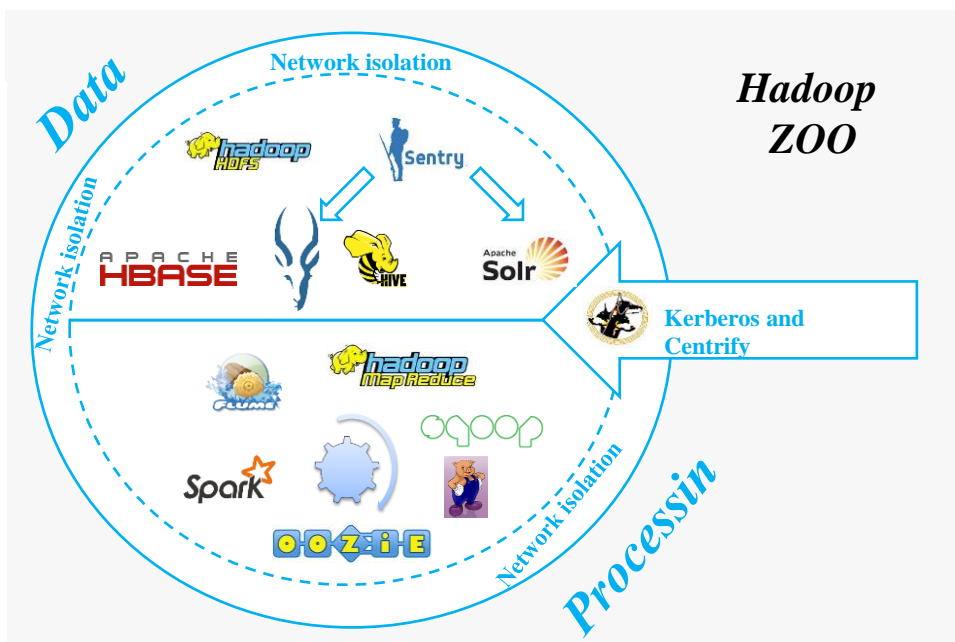
Visibility

- Reporting on where data came from and how it's being used
- Auditing, Lineage – Cloudera Navigator



Data

- Protecting data in the cluster from unauthorized visibility
- Encryption tokenizing data masking



Multitenancy approach for different components

HDFS

- Authorization with HDFS ACL

Hive and Impala

- Authorization with Sentry

HBase

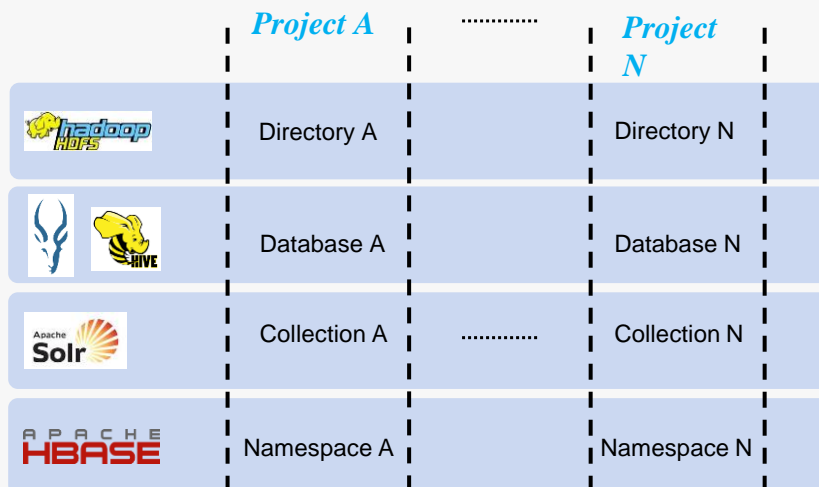
- Authorization with Hbase ACL

Solr

- Authorization with Sentry



Multitenancy in practice



Governance and guidelines



Governance model

- Cooperation between PoC/project team and platform team
- PoC/Project metadata
- Platform analysis (current usage, new CDH versions evaluation, security analysis)



Guidelines – building awareness

- Project structure
- Security concepts
- Development and usage tips
- Awareness is build thru: documents, knowledge base, presentations



What challenges are we facing?

Our journey



- *CDH 'kerberization'*
- *Moving Kerberos (local -> global)*
- *Introducing Sentry Service*
- *Introducing HDFS ACL*
- *Making Solr secure*
- *Third party tools integration (SAS, Informatica, reporting tools)*



Challenges



Evolving security in CDH

- Frequent security and structure concept changes
- New security models – evolution of Sentry



Tools maturity – enterprise ready

- Granularity of permissions needed
- New security features not fully working or requires cluster restart to apply changes
- Kerberos support for different version of tools (hive, sqoop, hbase)



Integration with third party tools

- Impersonation
- Sentry support



What we expect from the future

- *Single point of security management*
- *Authentication and authorization support for all the tools*



- *High level components to apply authorization*
- *Third party tools integration ready for CDH authentication and authorization*
- *More granular permissions*



How to approach the security?

Big Data @Roche*Our lessons learnt**Security Awareness*

Know what projects are doing and share the knowledge

*Think positively*

Security makes your data secure and gives opportunity for data governance

*Be agile*

Changing environment, be flexible and ready for future enhancements

*Be kerberized ☺**Big Data @Roche**Contact*

it.roche.pl



pawel.dudek@roche.com



*Doing now what patients need
next*