Strata
DATA CONFERENCE

ReactoData

HADOOP UNDER ATTACK

SECURING DATA IN A BANKING DOMAIN

WHOAMI

Federico Leven

- @ ReactoData
- Big Data + Open Source from 2012
- Big Data Meetup coordinator (http://www.iaar.site), speaker ...
- federico@reactodata.net
- Web : http://www.reactodata.net
- Twitter: @reactodata
- Linkedin : https://www.linkedin.com/in/federicoleven/

ReactoData

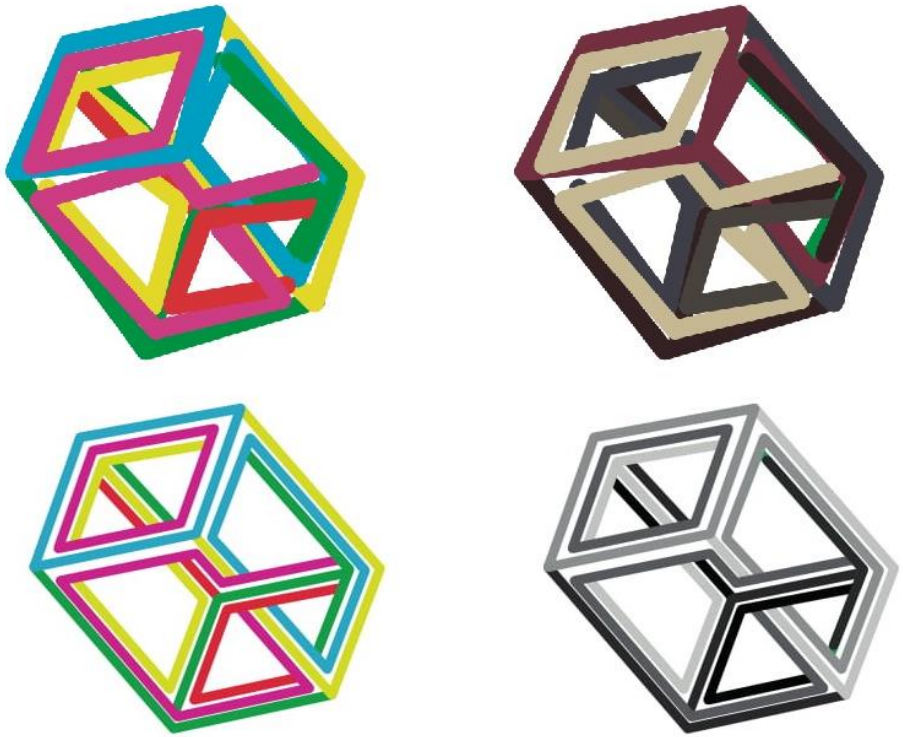
We are a startup based in Buenos and Poland, providing Big Data + Cloud solutions based on Open Source and proprietary software and Hadoop consultancy.

- Big Data and Hadoop applications development
- Machine Learning
- Cloud
- UX/UI and Mobile Apps for Big Data platforms
- Hadoop Consultancy

Agenda

- The Challenge : Best Practices + Regulations
- How to do it in Hadoop
- End-to-End Secured Architecture
- What can go wrong ?
- References
- Conclusion & Questions

The Challenge
Best Practices
and regulations



The Challenge : Data Security

The set of preventive, detective and corrective measures to protect the integrity, confidenciality and availability of the data.

CAAIN

- CONFIDENCIALITY
- AVAILABILITY
- AUTHENTICITY
- INTEGRITY

} • NON-REPUDIATION

Cain and Abel

- ❑ *ACCOUNTABILITY / AUDITING*
- ❑ *TRACEABILITY*

C(A)AIN

- **C**ONFIDENCIALITY : Data is not made available or disclosed to unauthorized parties.
- **A**VAILABILITY : Data is available when is needed.
- **A**UTHENTICITY : Data source identity is verifiable.
- **I**NTEGRITY : Data is accurate and complete over its entire lifecycle.
- **N**ON-REPUDIATION : Parties of a data transaction cannot deny having received/sent the data .

The Challenge : Threats in financial and banking domain

Emerging Technologies Challenges

- Botnet
- IoT unsecured devices
- DDoS (Distributed Denial of Service Attack)

Insider Challenges

- Unintentional actions
- Malicious users

Regulation Challenges

- Periodically new and/or stricter regulations
- US Data Protection rules
- EUR : GDPR

Target

- Sensitive data
- Access credentials

The Challenge : Best Practices in banking

Organization

- Security Officer
- End User Guidelines
- Access Policies
- Governance
- …

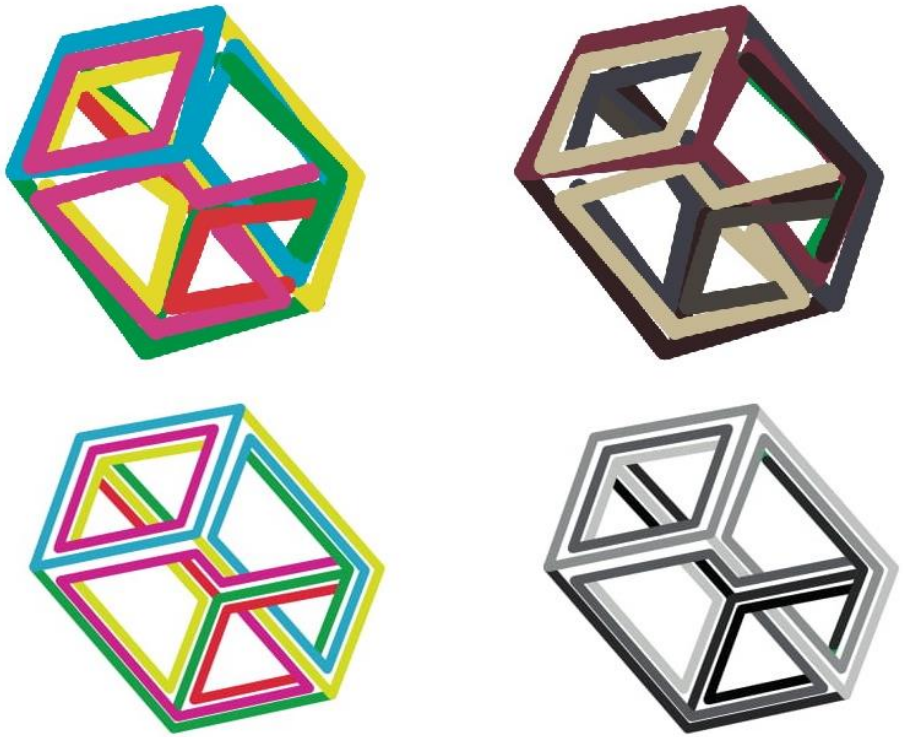
Human

- Employees Awareness
- Training
- …

Technological

- Networking
- Software Updates
- Data Protection
- Auditing
- …

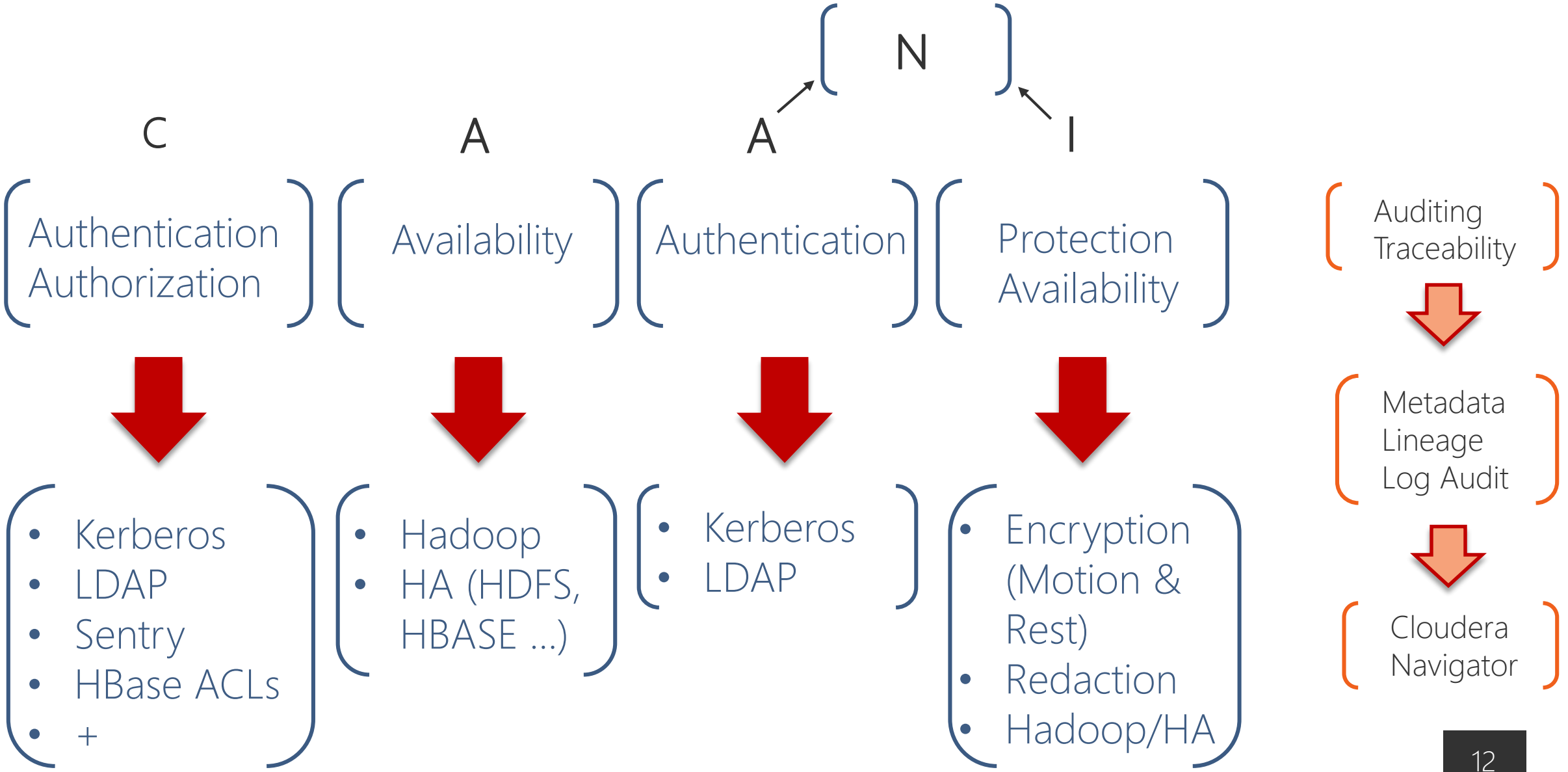
How to do it in Hadoop



From concepts to technology

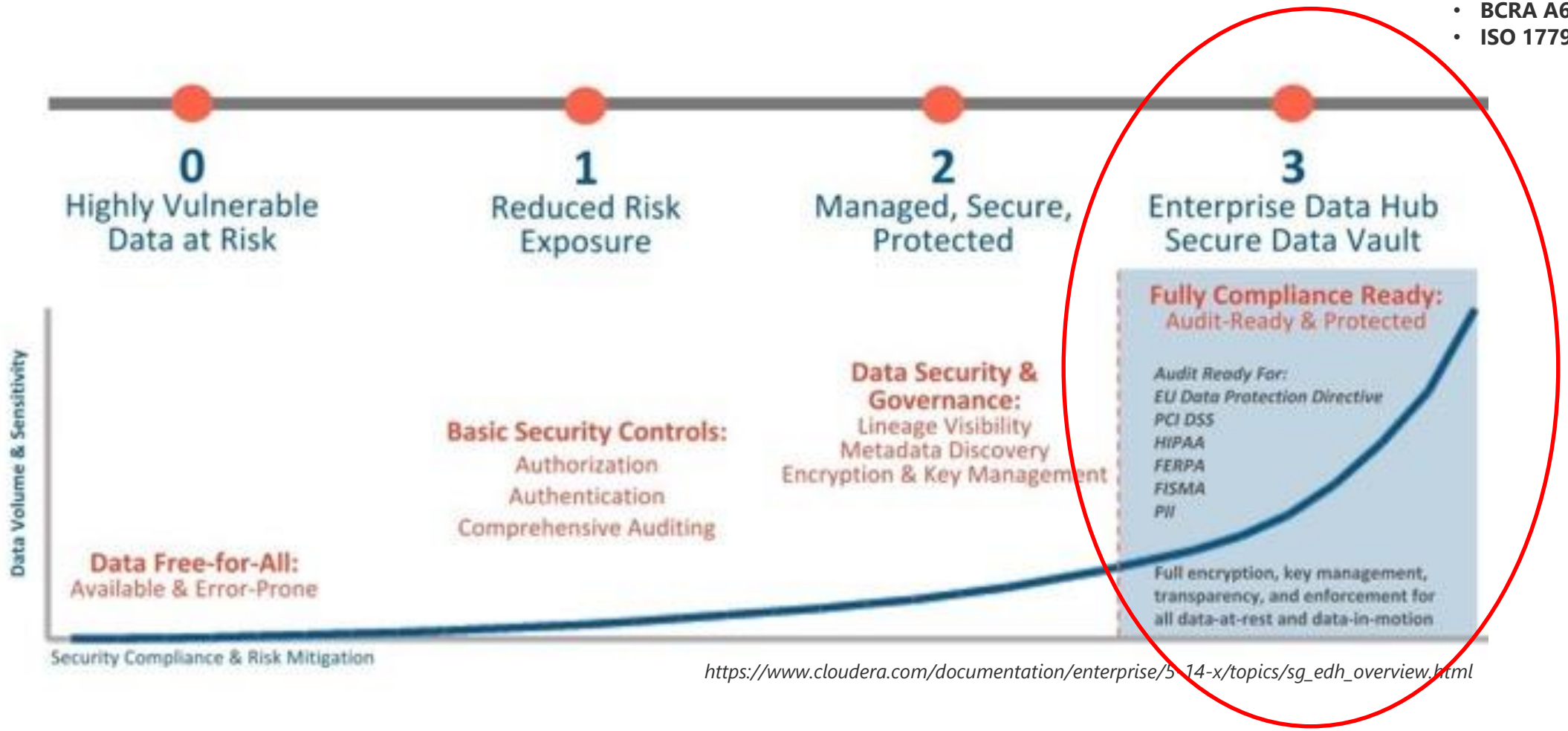
- AUTHENTICATION : Identify the user.
- AUTHORIZATION : Grant user access to the data.
- PROTECTION : Protect data from being used except by authorized users.
- AVAILABILITY : Make data accessible when needed.

From concepts to technology



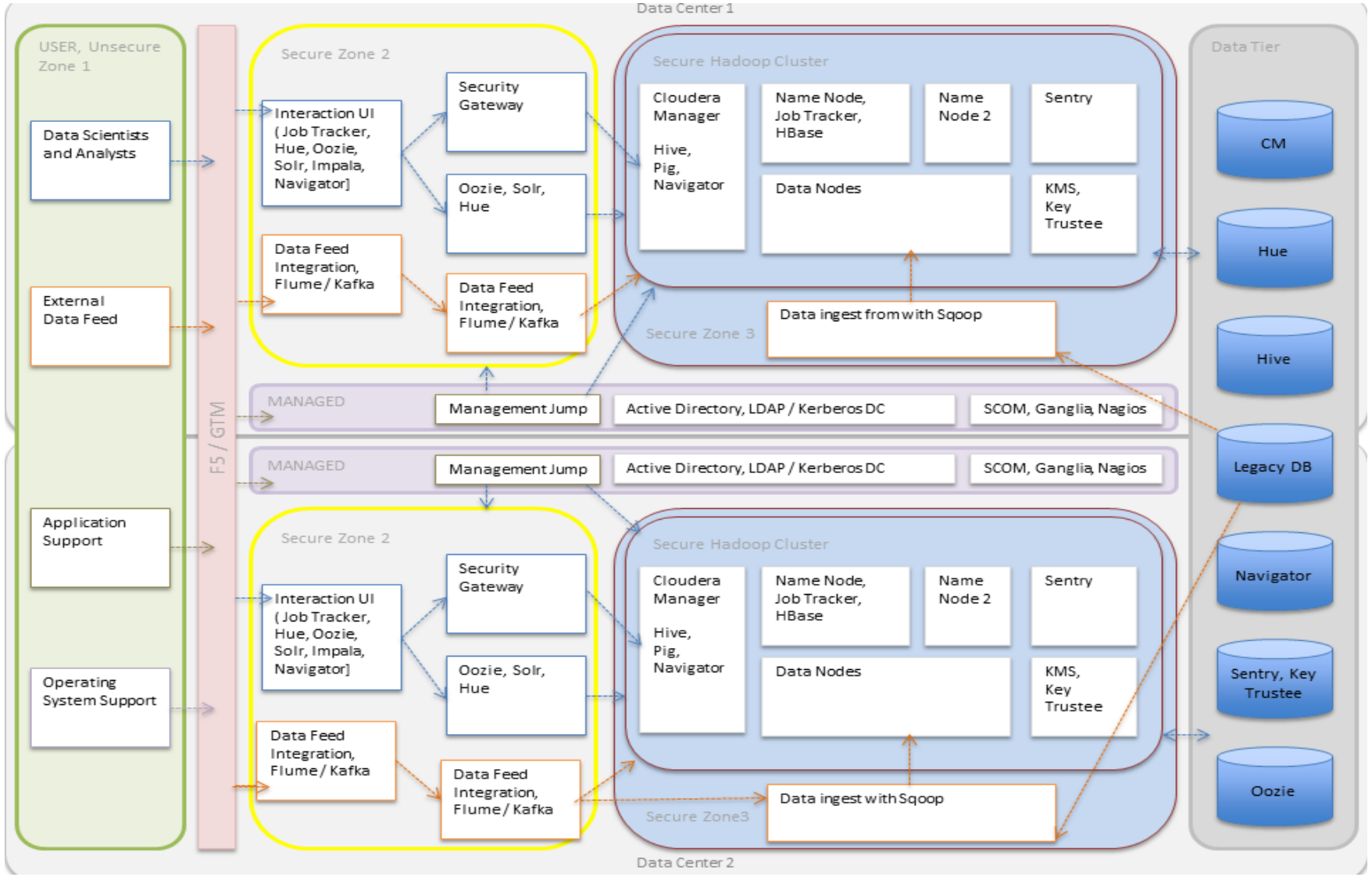
From concepts to technology

- **BCRA A6375**
- **BCRA A6495**
- **ISO 17799/27001**

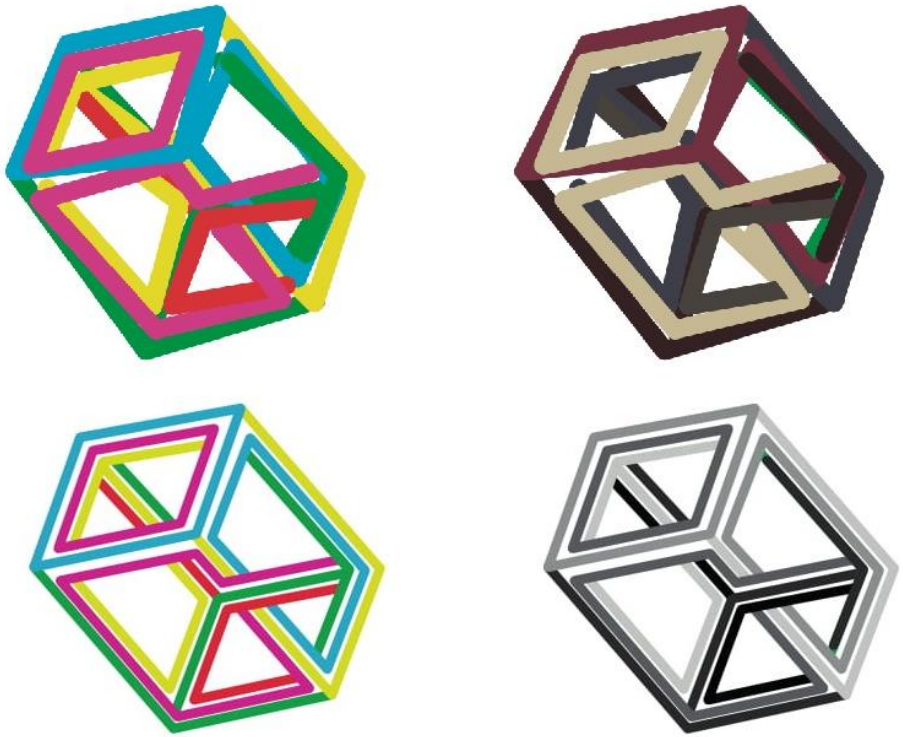


https://www.cloudera.com/documentation/enterprise/5-14-x/topics/sg_edh_overview.html

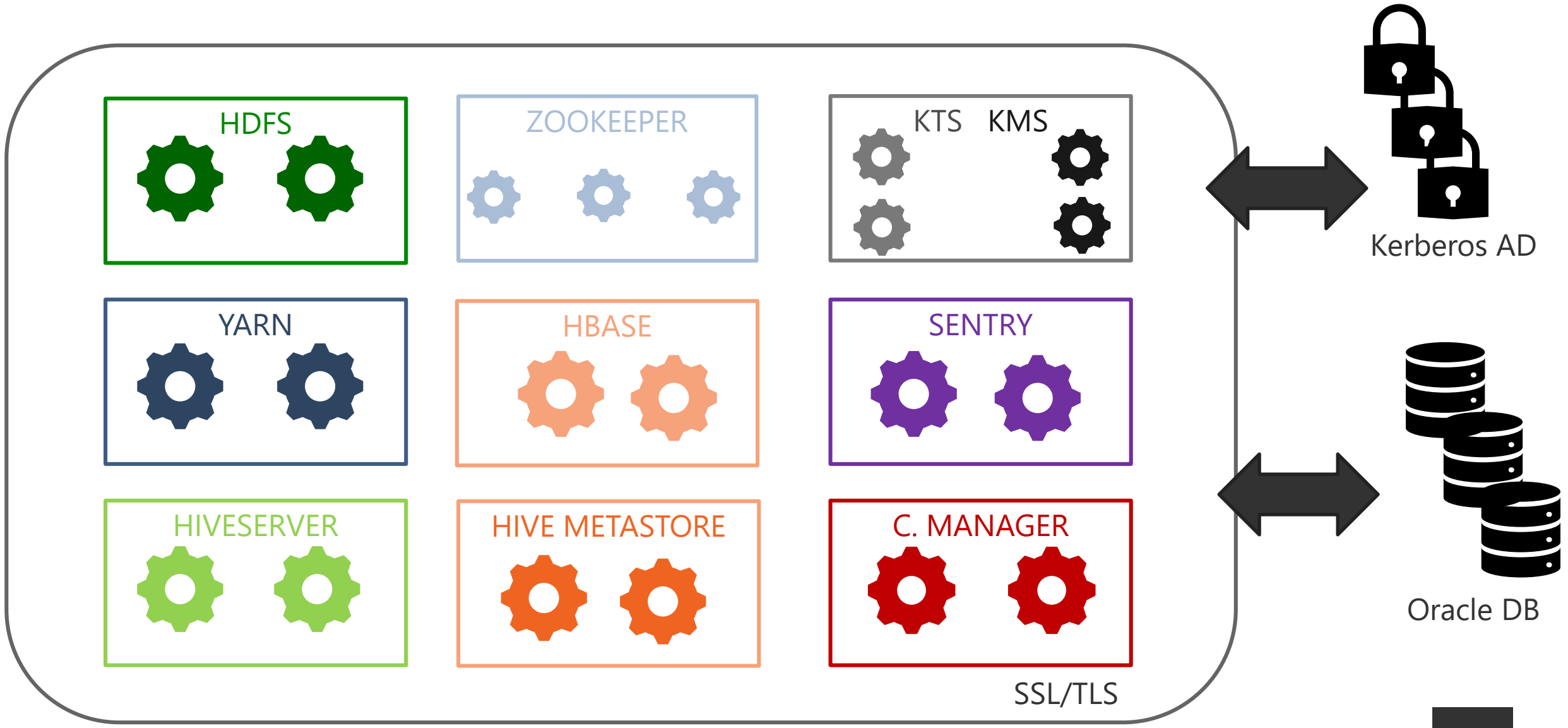
What we needed in a banking infrastructure for Hadoop



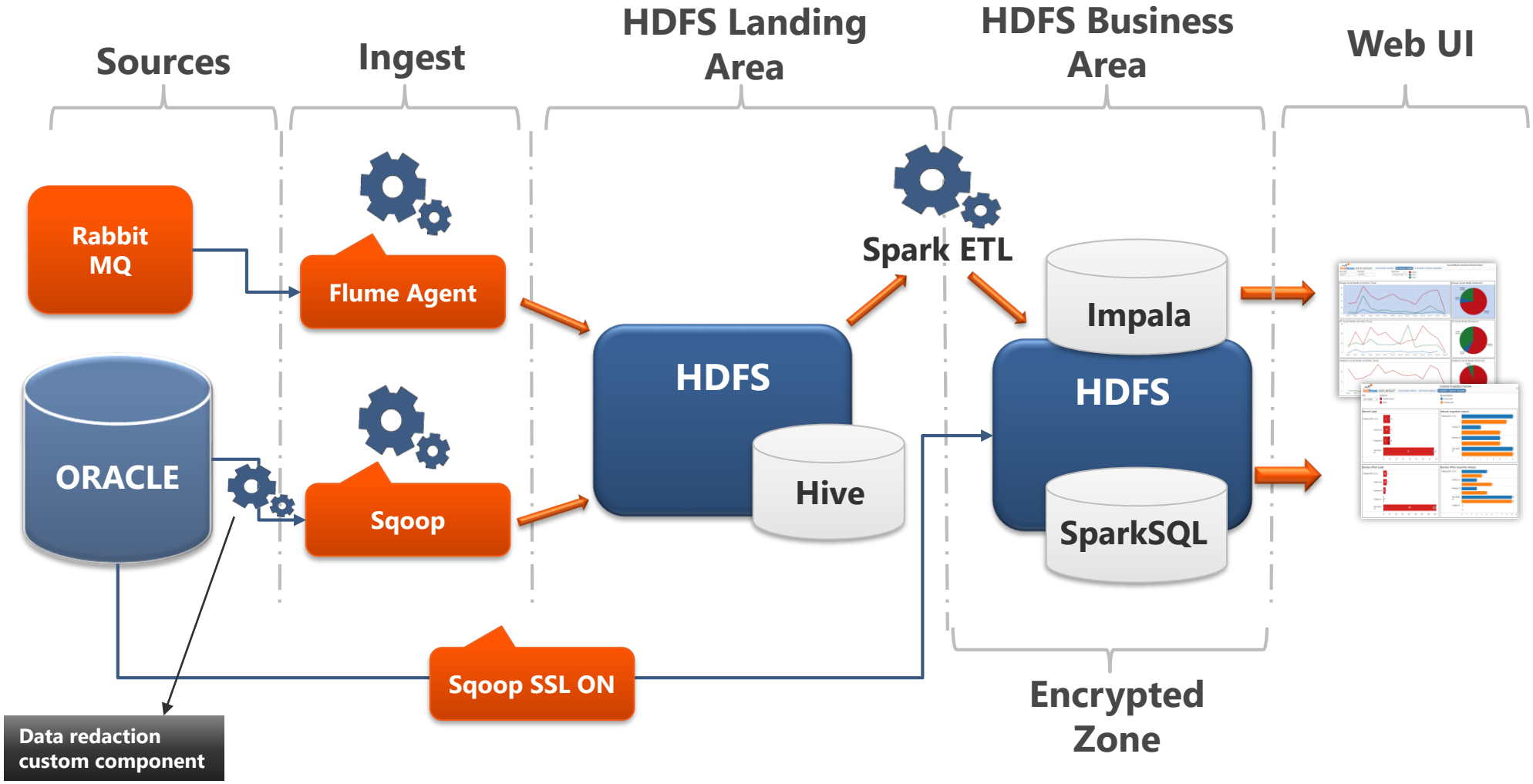
End-to-End Secured Architecture



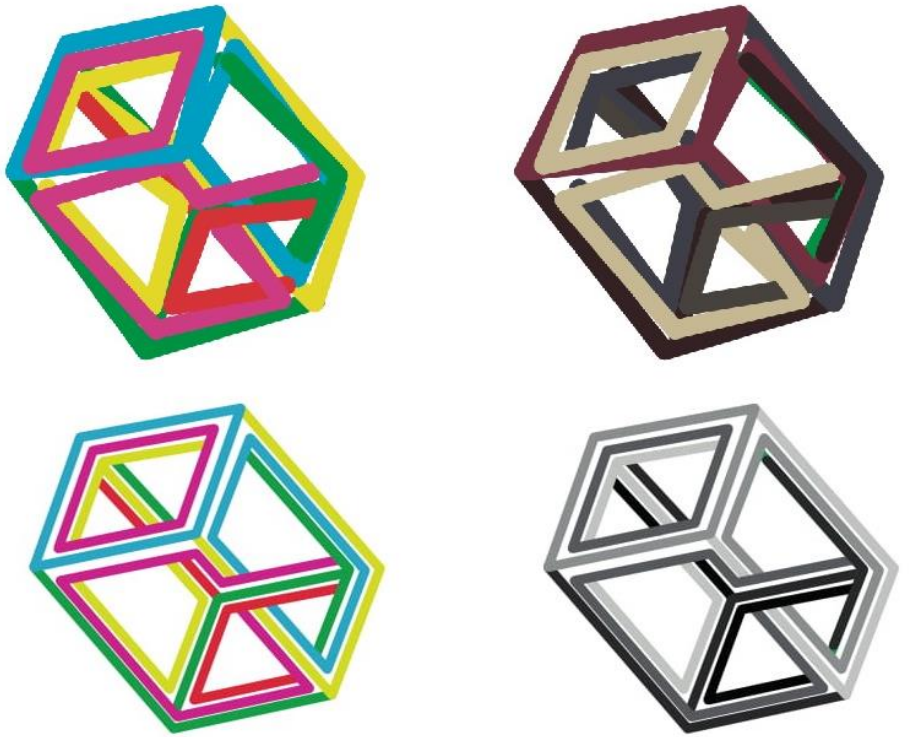
Example Production deployment (CDH 5.13)



Secure data pipeline example



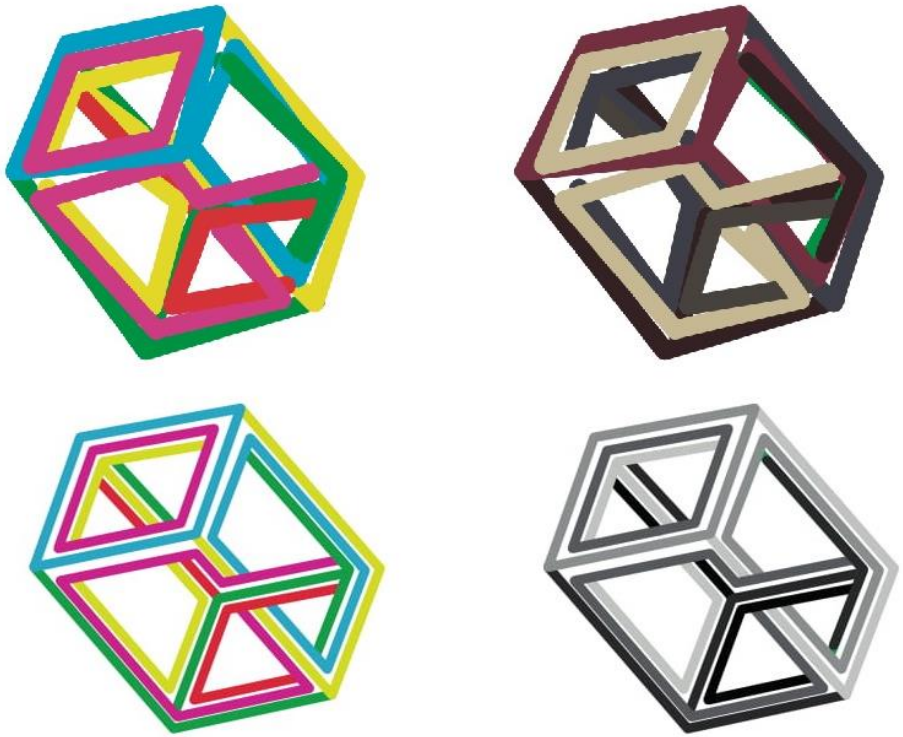
What can go wrong ?



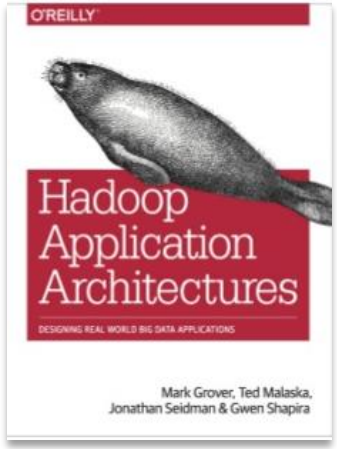
What can go wrong ? Some good news and some bad news

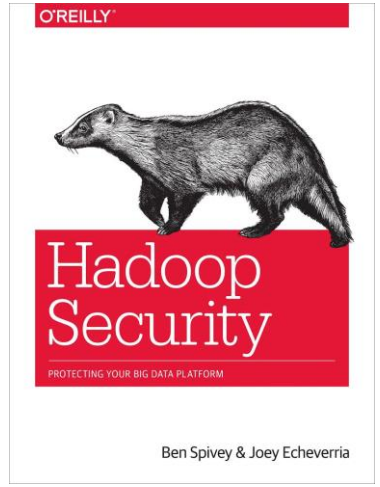
- UNSECURE APPLICATIONS WILL NOT WORK ON SECURE ENVIRONMENTS 
- Sentry HDFS synchronization does not support Hive Metastore HA (CDH 5.9) 
- Sentry HA not supported (CDH 5.9) 
- To use CM Kerberos wizard, you need a high level privileges user 
- SparkSQL does not respect Sentry permissions (Latest) 
- Enabling Sentry turns off Hive impersonation (CDH 5.9) 
- Spark Streaming cannot consume from secure Kafka (CDH 5.9) 

References

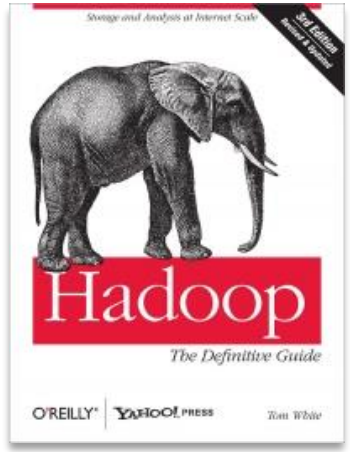


References





- ✓ http://www.bcra.gob.ar/Pdfs/Texord/t-rmsist.pdf
- ✓ http://www.bcra.gov.ar/pdfs/texord/t-seguef.pdf
- ✓ https://en.wikipedia.org/wiki/ISO/IEC_27002
- ✓ http://web.iram.org.ar/index.php?vernorma&id=2439
- ✓ https://www.cloudera.com/documentation/enterprise/latest/PDF/cloudera-security.pdf
- ✓ https://www.cloudera.com/documentation/enterprise/5-9-x/topics/security.html
- ✓ https://www.forbes.com/sites/gregorymcneal/2014/05/26/banks-challenged-by-cybersecurity-threats-state-regulators-acting/#228d745597f7



Strata
DATA CONFERENCE

ReactoData

Thank you !

Questions, suggestions or complaints ?

"No Hadoop was harmed in the making of this presentation"