cloudera

Executive Briefing

GDPR: Getting your data ready for heavy, new EU privacy regulations

Steve Ross, Director, Product Management
Mark Donsky, Director, Product Management

Disclaimer

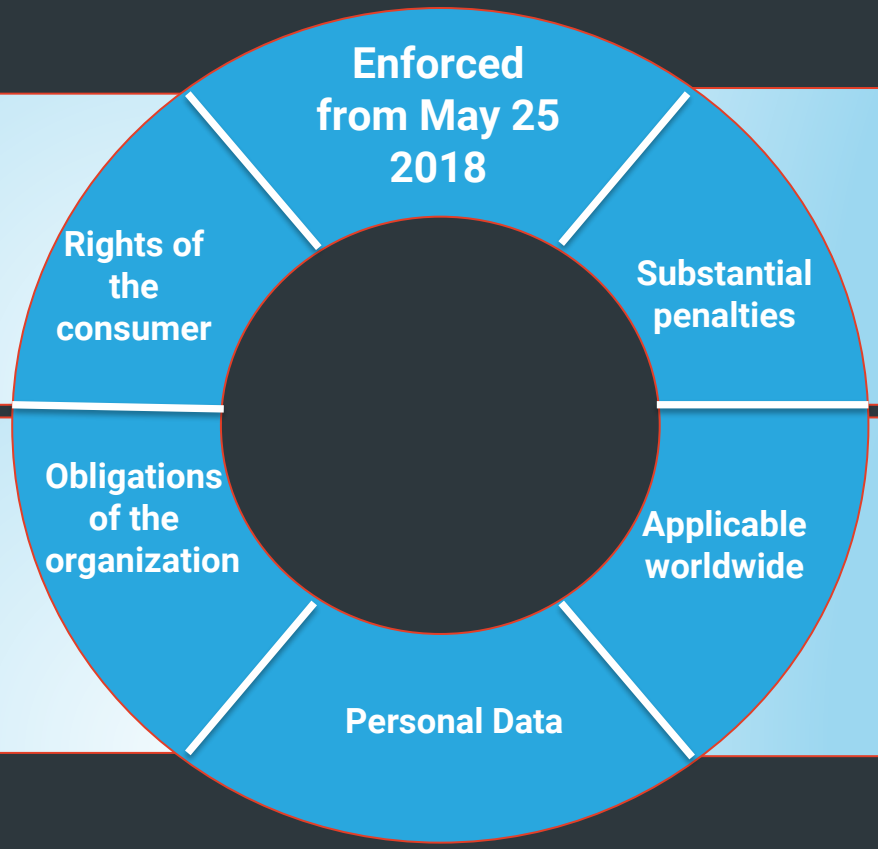
GDPR is a complex and detailed regulation.

There’s no single method or solution that will make all organizations compliant.

This presentation is intended to help organizations understand how Big Data platforms such as Cloudera software and services can be used to help comply with certain aspects of EU General Data Protection Regulation (GDPR) requirements. Applicability of any of these capabilities depends on each organization’s own requirements specific to their business. Every organization should determine its own needs with regard to GDPR and then evaluate solutions for suitability to those needs. The information contained in this presentation is not intended to be and should not be construed to be legal advice. Organizations must not rely on the information herein and they should obtain legal advice from their own legal counsel or other professional legal services provider.

General Data Protection Regulation (GDPR)

- Right to be forgotten/erasure
- Right to access information
- Right to data portability
- Right for processing to be restricted



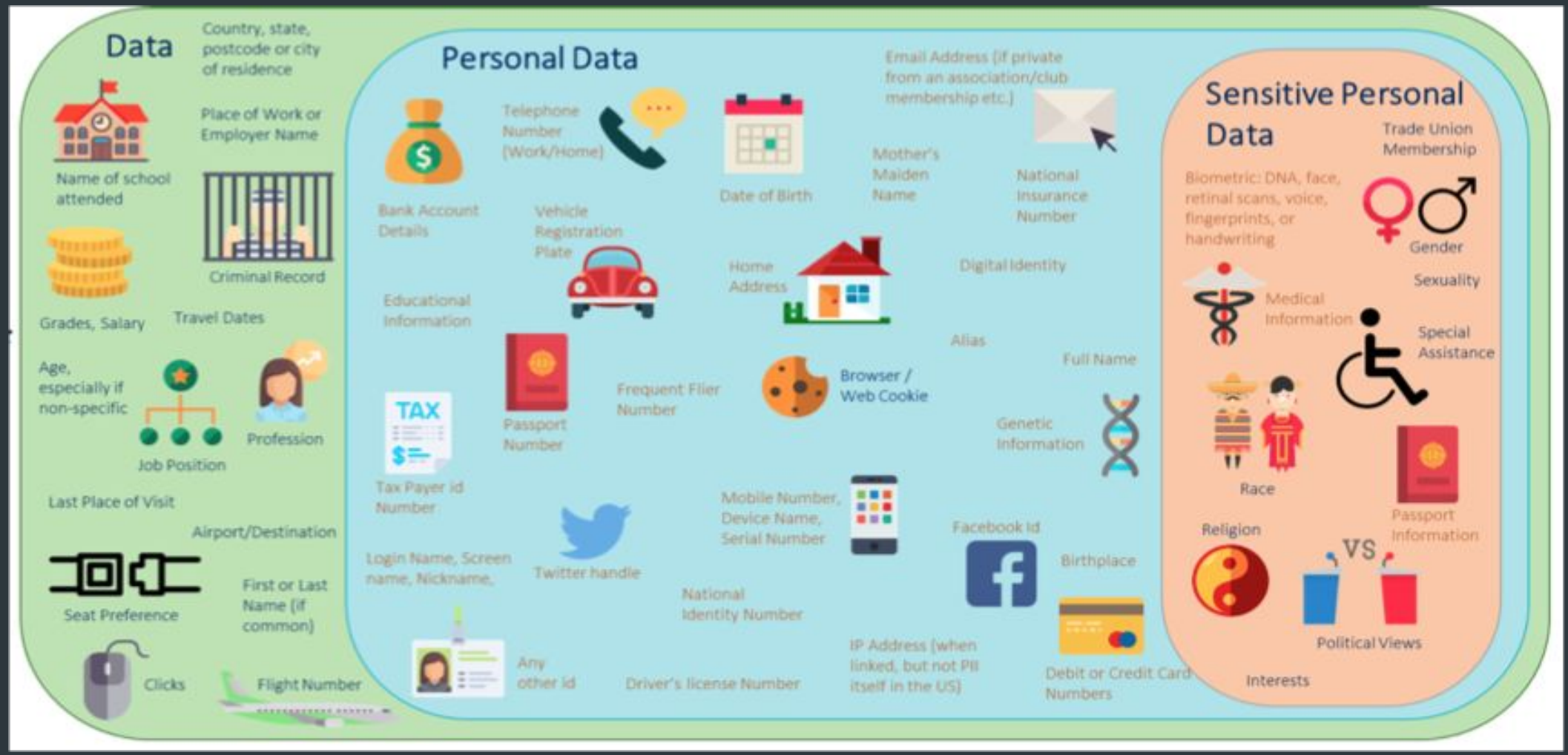
- Heavy fines for violations
- Up to 20M Euros or 4% of the annual global turnover for the preceding financial year

- Across people, process and technology
- Impacts how **personal data** is collected and used

- Any organization with any users in the EU needs to be compliant.
- Includes companies based outside the EU, processing personal data from EU residents in connection with the offering any goods or services or monitoring user behavior.
- Includes data processor and data controller

cloudera

Examples of personal data



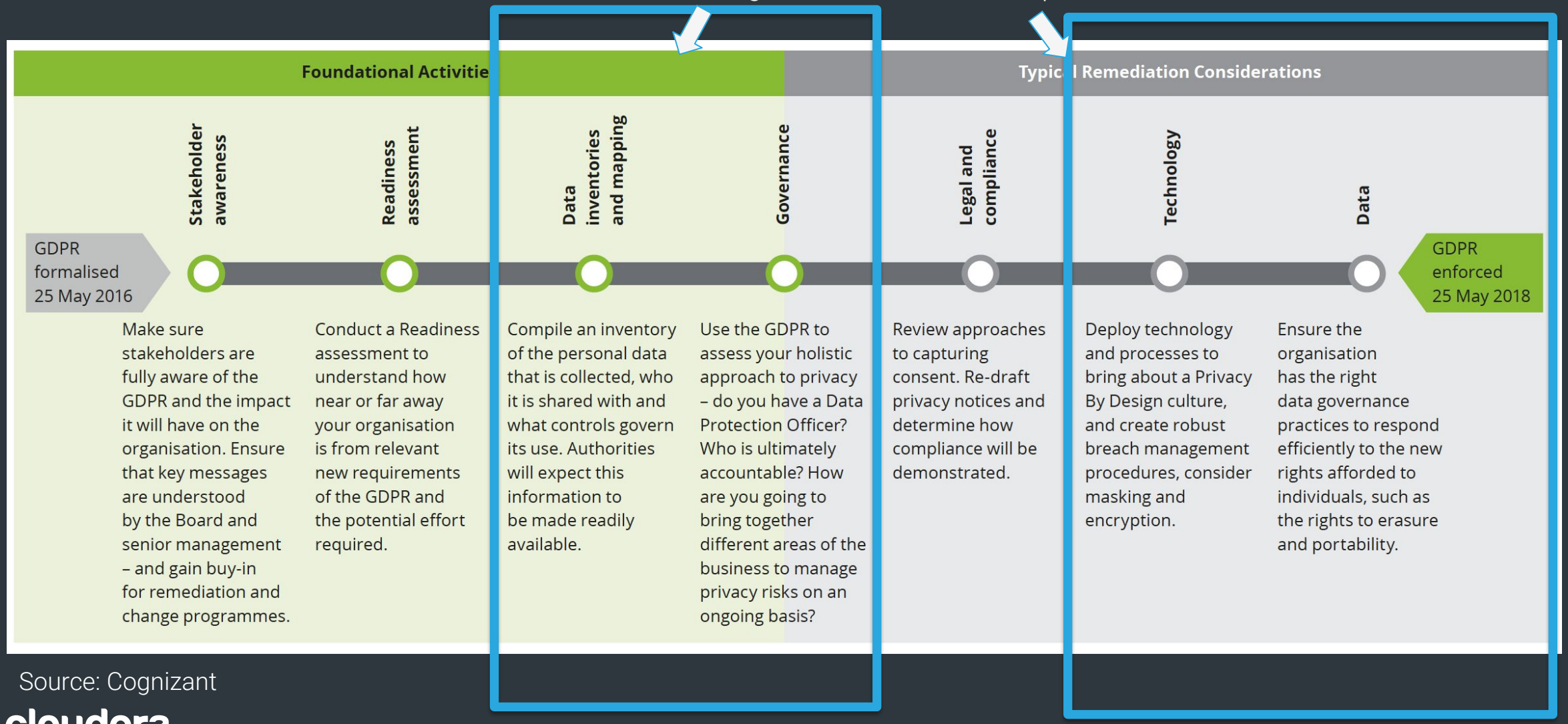
cloudera

Seven Key Principles of GDPR

cloudera

The path to GDPR compliance includes…

Big data solutions can help here



Source: Cognizant

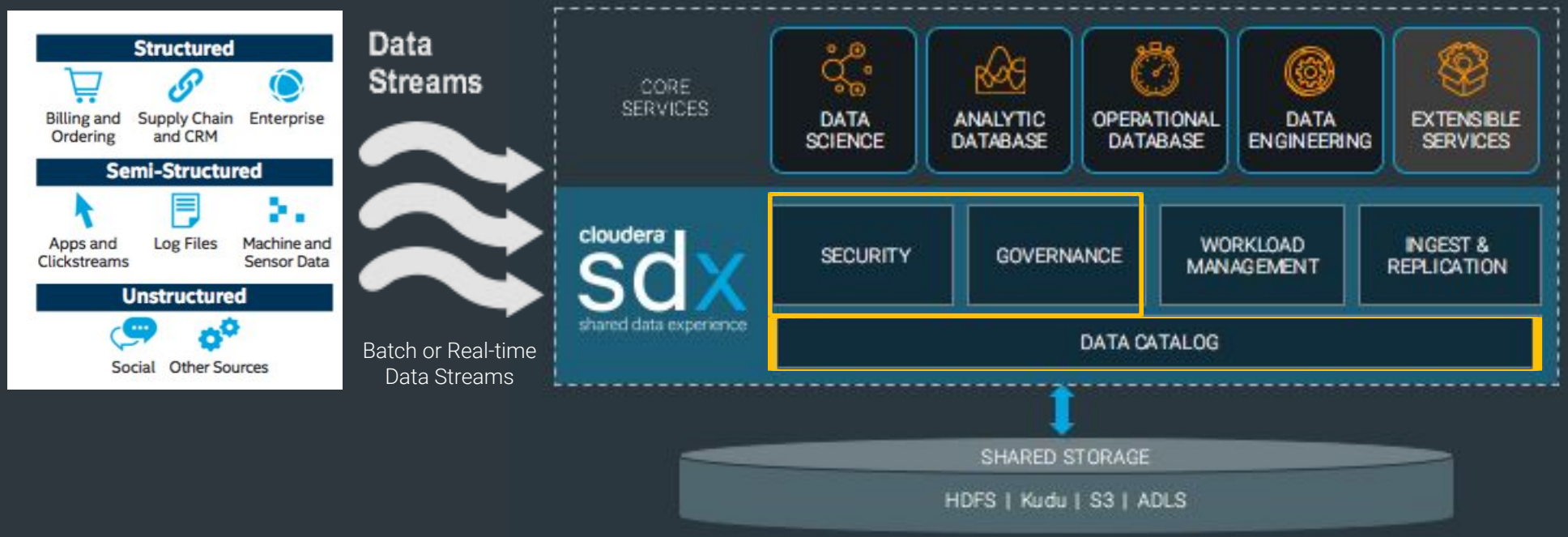
cloudera

Where are you along this path?

cloudera

How big data solutions can help accelerate GDPR compliance

A single place to secure and govern for GDPR compliance



For data that is stored on a single platform, there is a single place to secure and govern for GDPR compliance, across all analytic workloads

The Seven Principles of GDPR

1. Integrity and Confidentiality
2. Accountability
3. Lawfulness, fairness and transparency
4. Purpose limitation
5. Data Minimization
6. Accuracy
7. Storage Limitation

cloudera

How big data solutions can help

GDPR Principle	Typical Challenges	Big data solutions
Integrity and Confidentiality	Applying industry standard IT security controls to prevent unauthorised access.	Comprehensive encryption and key management. Strong authentication, fine-grained authorization.
Accountability	Demonstrating compliance. Detecting and analyzing breaches within 72 hours.	Comprehensive audit trail. Breach detection (cybersecurity solutions).
Lawfulness, fairness and transparency	Implementing a way to keep track of personal data.	Track classification and lineage of personal data elements.

cloudera

How big data solutions can help

GDPR Principle	Typical Challenges	Cloudera Enterprise Capabilities
Integrity and Confidentiality	Applying industry standard IT security controls to prevent unauthorised access.	Navigator Encrypt and Key Trustee: strong encryption Apache Sentry: Fine-grained authorization
Accountability	Demonstrating compliance. Detecting and analyzing breaches within 72 hours.	Cloudera Navigator: Comprehensive, inescapable audit trail Cloudera Cybersecurity solutions
Lawfulness, fairness and transparency	Implementing a way to keep track of personal data.	Cloudera Navigator: Classify/tag and track lineage of personal data elements

cloudera

Pillars of a comprehensive compliance solution

GDPR principles: Integrity, Confidentiality, Accountability, Lawfulness, Fairness, Transparency

Perimeter

Guarding access to the cluster itself

Technical concepts:
Authentication
Network isolation

Access

Defining what users and applications can do with data

Technical concepts:
Permissions
Authorization

Visibility

Reporting on where data came from and how it's being used

Technical concepts:
Auditing
Lineage

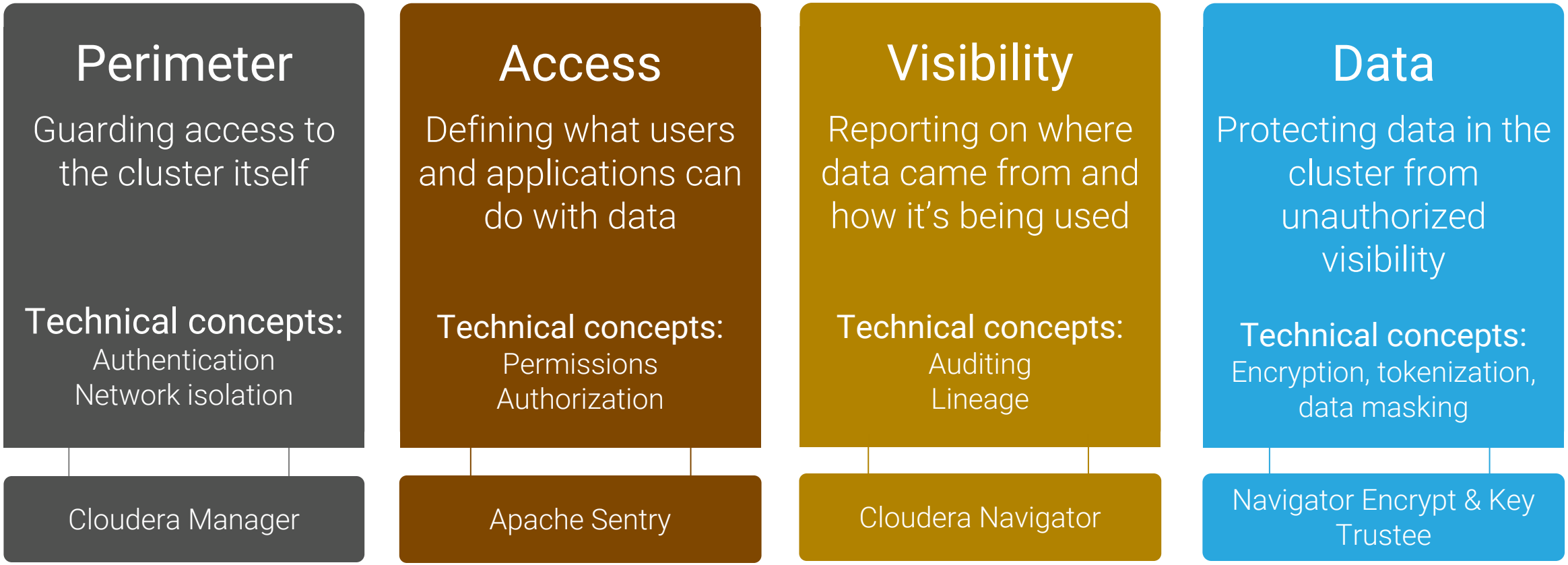
Data

Protecting data in the cluster from unauthorized visibility

Technical concepts:
Encryption, tokenization, data masking

Pillars of a comprehensive compliance solution

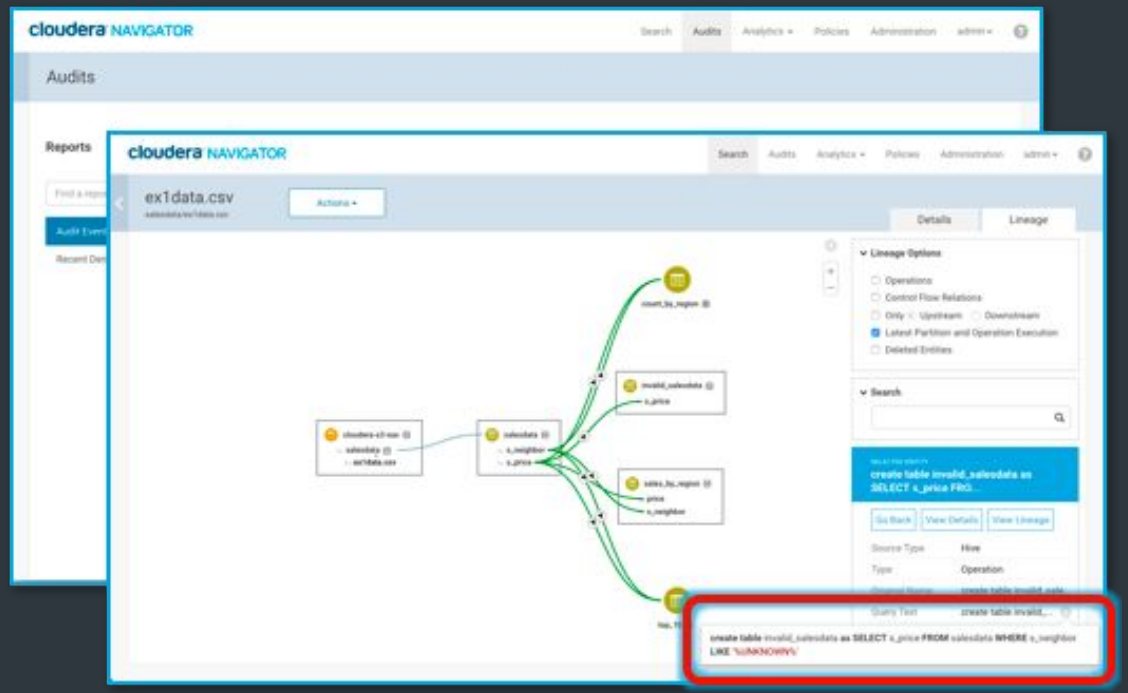
GDPR principles: Integrity, Confidentiality, Accountability, Lawfulness, Fairness, Transparency



Governance: auditing, lineage, metadata

Capabilities:

- Inescapable, detailed audit – enabling forensics
- Full tracking of personal data
- Lineage tracking and visualization



cloudera

Enterprise-grade encryption & key management

Encryption

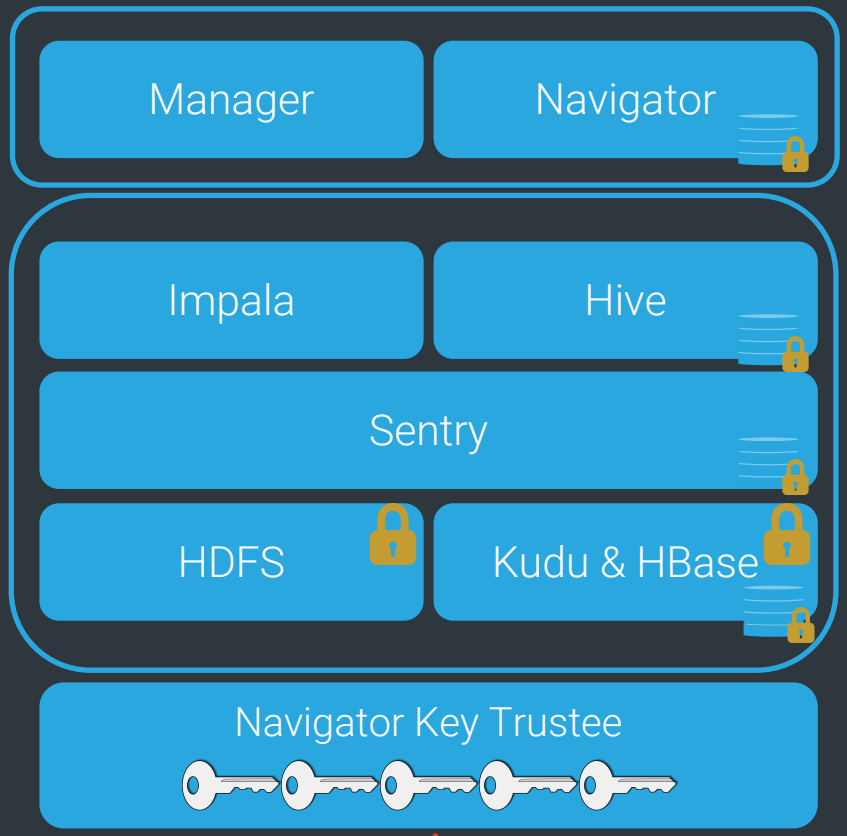
- ALL data at rest: HDFS, HBase, metadata databases, temp files, ingest paths
- ALL data on the wire

Key Management

- Automated key replication & backup
- HSM backed key protection

Redaction

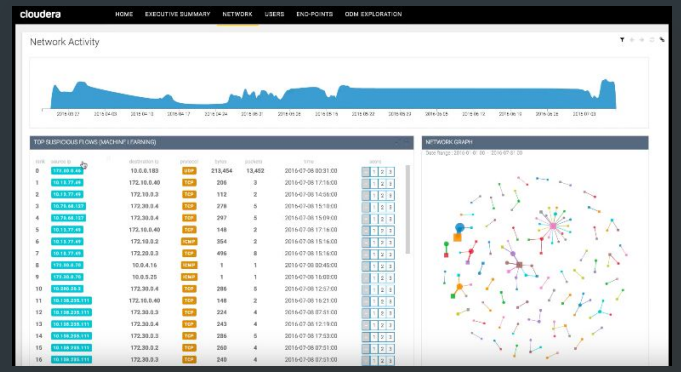
- Sensitive data in logs



cloudera

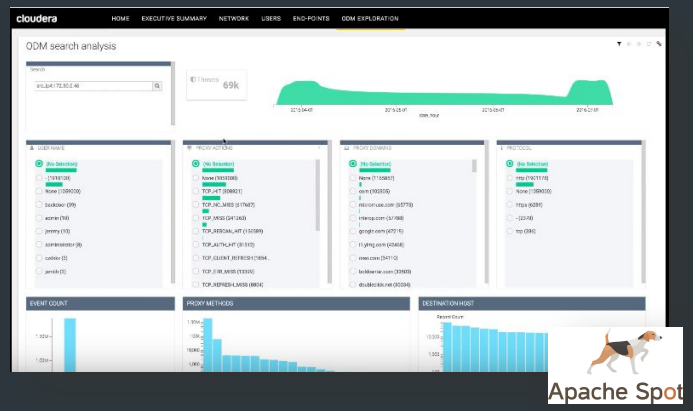
Enterprise-wide breach detection

Detect



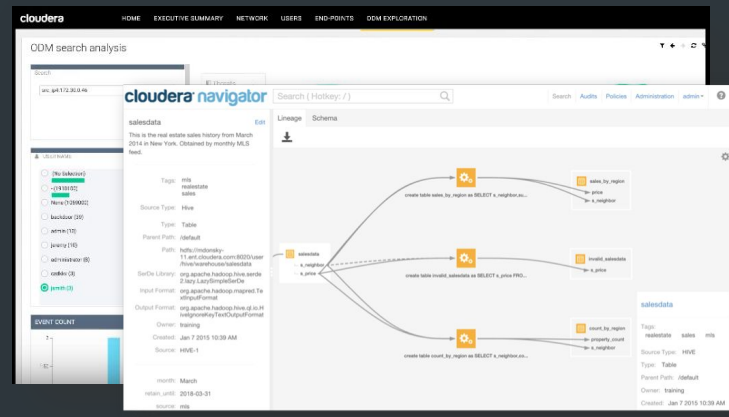
Detect advanced threat leveraging machine learning models

Investigate



Search across Apache Spot's user, endpoint, and network open data models for full context and accelerated investigation

Respond



Use Apache Spot open data models and Cloudera Navigator to see if the threat is widespread

cloudera

How big data solutions can help

GDPR Principle	Typical Challenges	Big data solutions
Purpose limitation	Track consent and data usage while allowing data scientists to mine it using tools of choice.	Data protection officer (DPO) can audit exactly how data was used. Data scientists work with flexibility while keep data governed.
Data Minimization	Removing or anonymising data where possible. Preventing unlawful data transfers outside the EU while still enabling outsourcing.	Classification tags can indicate allowed purpose. Redacted views limit what certain people can access.
Accuracy	Finding a low-overhead way to fix data.	Fast updates of individual records.
Storage Limitation	Deleting individual personal data records in Hadoop and Cloud storage, since those file systems are immutable.	Erasure of individual records. HDFS and Cloud storage options.

cloudera

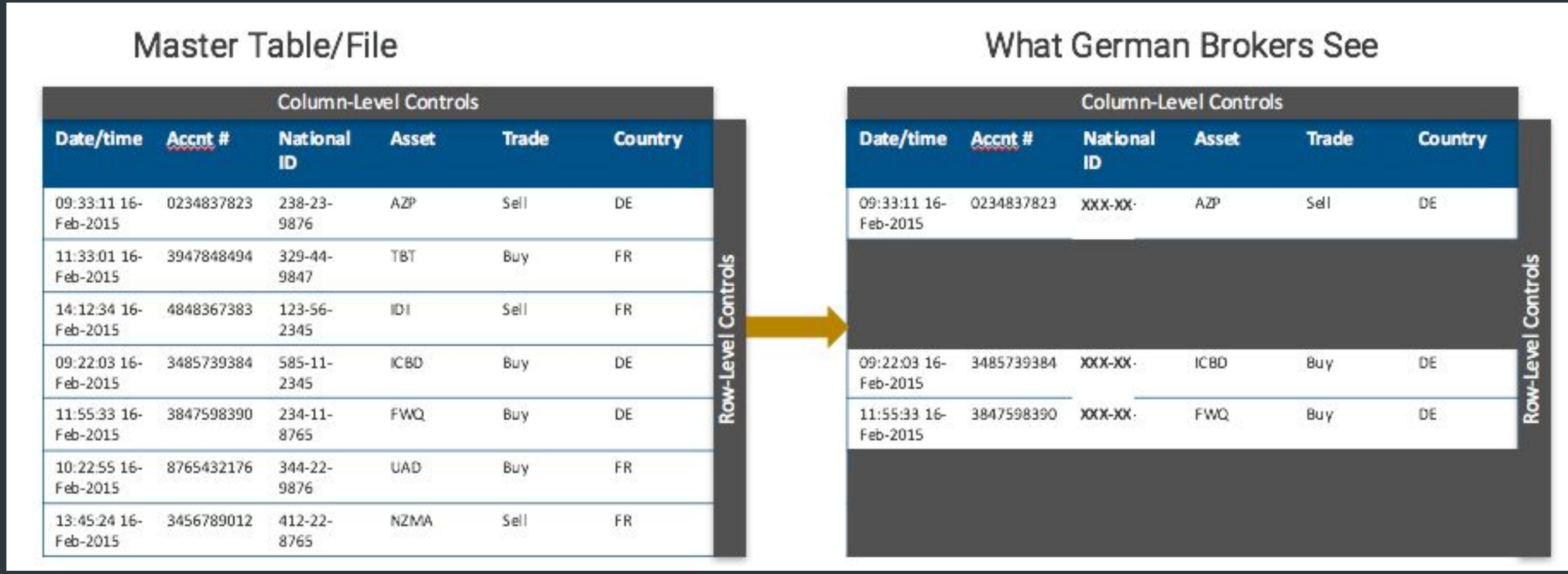
How big data solutions can help

GDPR Principle	Typical Challenges	Big data solutions
Purpose limitation	Track consent and data usage while allowing data scientists to mine it using tools of choice.	Cloudera Navigator: DPO can audit exactly how data was used Cloudera Data Science Workbench: keep data governed
Data Minimization	Removing or anonymising data where possible. Preventing unlawful data transfers outside the EU while still enabling outsourcing.	Cloudera Navigator: tags can indicate allowed purpose, time limit Apache Sentry: Redacted views
Accuracy	Finding a low-overhead way to fix data.	Apache Kudu: Fast updates of individual records
Storage Limitation	Deleting individual personal data records in Hadoop and Cloud storage, since those file systems are immutable.	Apache Kudu: Fast erasure of individual records HDFS and Cloud storage options

cloudera

Fine-grained access control and data masking

Apache Sentry for column-level permissions and views with masking and row filtering
(optional) Cloudera partners enable dynamic masking and tokenization



Date/time	Accnt #	National ID	Asset	Trade	Country
09:33:11 16-Feb-2015	0234837823	238-23-9876	AZP	Sell	DE
11:33:01 16-Feb-2015	3947848494	329-44-9847	TBT	Buy	FR
14:12:34 16-Feb-2015	4848367383	123-56-2345	IDI	Sell	FR
09:22:03 16-Feb-2015	3485739384	585-11-2345	ICBD	Buy	DE
11:55:33 16-Feb-2015	3847598390	234-11-8765	FWQ	Buy	DE
10:22:55 16-Feb-2015	8765432176	344-22-9876	UAD	Buy	FR
13:45:24 16-Feb-2015	3456789012	412-22-8765	NZMA	Sell	FR

Date/time	Accnt #	National ID	Asset	Trade	Country
09:33:11 16-Feb-2015	0234837823	XXX-XX-	AZP	Sell	DE
09:22:03 16-Feb-2015	3485739384	XXX-XX-	ICBD	Buy	DE
11:55:33 16-Feb-2015	3847598390	XXX-XX-	FWQ	Buy	DE

cloudera

The right to erasure - challenges

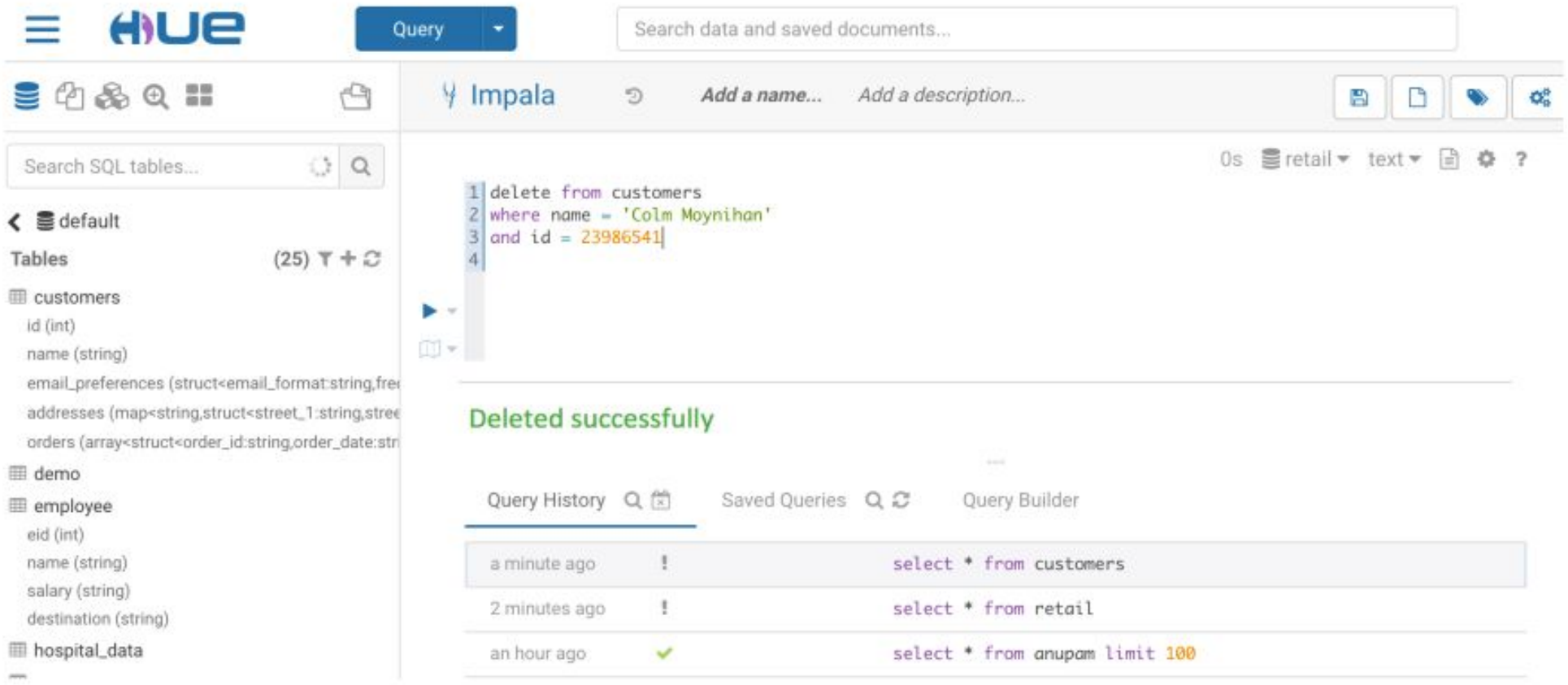
1. **Existing data architectures** may spread personal data across many objects
2. **Self-service** generates derived datasets also subject to GDPR
3. **Volume and variety** means any solution needs to scale
4. **Storage capabilities** limit erasure options
    - HDFS and cloud object stores are “immutable”

Erasing individual records on HDFS and cloud storage

- Concentrate personal data in a small number of “lookup tables”
- Replace personal data in most locations with anonymized or pseudonymised data
- Instead of deleting records upon request, add them to a “to be deleted” list
- Execute a periodic batch job to remove “to be deleted” records by rewriting entire files/tables

Issue: The re-write step could render the cluster unusable for a period of time

Erasing individual records on Kudu



cloudera

Laptop vs Centralized Data Science

“Laptop Data Science”

Typical Big Data Environment

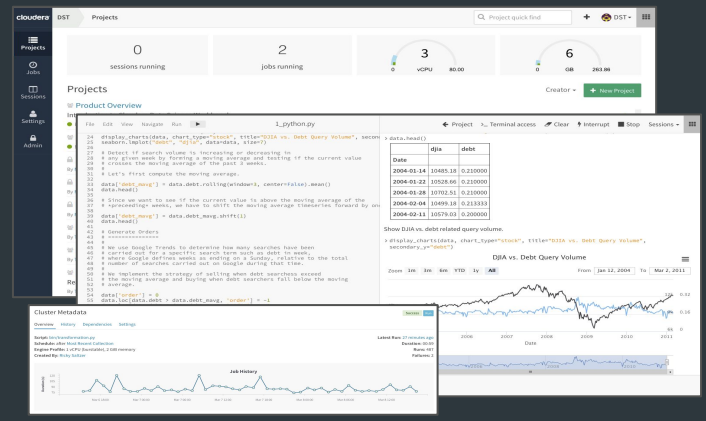
Data scientists pull data to their laptops so they can run their own tools

- Copy personal data to laptops
- Fails GDPR compliance audit
- Potential data breach

Centralized Data Science

cloudera®

Data Science Workbench



- Personal data remains governed
- Purpose limitation enforced

How big data solutions can help

GDPR Principle	Typical Challenges	Big data solutions
Integrity and Confidentiality	Applying industry standard IT security controls to prevent unauthorised access.	Comprehensive encryption and key management. Strong authentication, fine-grained authorization.
Accountability	Demonstrating compliance. Detecting and analyzing breaches within 72 hours.	Comprehensive audit trail. Breach detection (cybersecurity solutions).
Lawfulness, fairness and transparency	Implementing a way to keep track of personal data.	Track classification and lineage of personal data elements.
Purpose limitation	Track consent and data usage while allowing data scientists to mine it using tools of choice.	Data protection officer (DPO) can audit exactly how data was used. Data scientists work with flexibility while keep data governed.
Data Minimization	Removing or anonymising data where possible. Preventing unlawful data transfers outside the EU while still enabling outsourcing.	Classification tags can indicate allowed purpose. Redacted views limit what certain people can access.
Accuracy	Finding a low-overhead way to fix data.	Fast updates of individual records.
Storage Limitation	Deleting individual personal data records in Hadoop and Cloud storage, since those file systems are immutable.	Erasure of individual records. HDFS and Cloud storage options.

cloudera

What's Next

- May 25, 2018 is 78 days away
- If you haven't yet started the foundational activities, start them now!
- Come talk to us this afternoon at "Meet the Experts: GDPR" to learn more

cloudera

Questions?

cloudera

Thank you

Steve Ross: sross@cloudera.com
Mark Donsky: md@cloudera.com

cloudera

