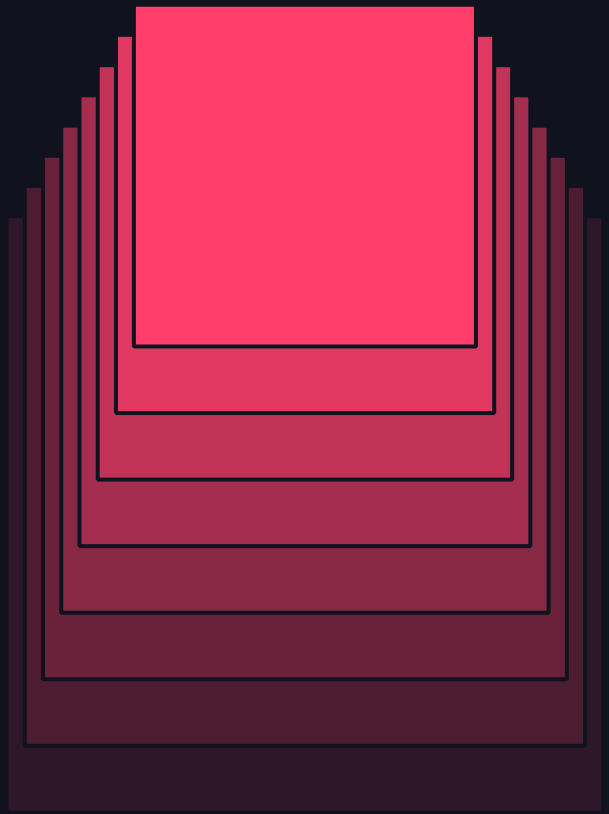
DATA+AI SUMMIT BY databricks

UC Heart of democratization for CNH



Antonio Nuzzo
Mauro Cavallotti
06-10-2024

REPLY

CNH

GREAT IRON

GREAT TECHNOLOGY

A strong **family of brands** for greater **customer focus**

FULL – LINER **GLOBAL BRANDS**

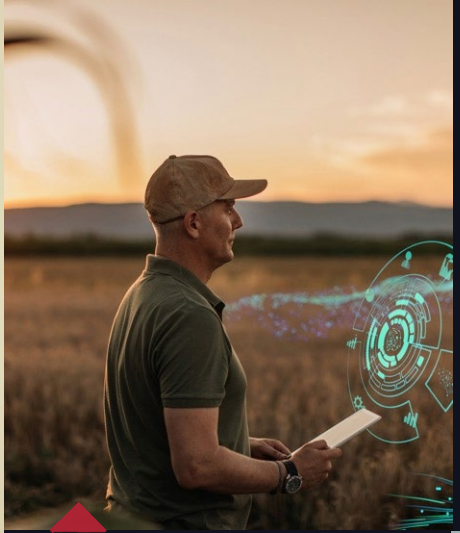
CASE IH · NEW HOLLAND · NEW HOLLAND CONSTRUCTION · CASE CONSTRUCTION

STEYR · RAVEN · flexi coil · Miller · KONGSKILDE · EUROCOMACH

FOCUSED **REGIONAL BRANDS**

Three segments with distinct features

CNH is listed on the New York Stock Exchange (NYSE: CNHI)



Construction

- A global player in construction equipment
- Extensive synergies with Agriculture segment
- Co-investment in digital / electrification

Agriculture

- World' s second largest manufacturer of agricultural machinery
- Leader in machine automation and precision technologies
- Pioneer in alternative propulsion

Financial Services

- Global financier supporting the brands, customers, importers and dealers
- Enhanced customer experience to offer a competitive edge to our brands

REPLY

CNH

Key figures

1
Company

23.6B
Consolidated Revenues

43
Plants

10
Brands

40,000+
Full-time Employees

40
R&D Centers

Note: figures are on a US GAAP $ basis and updated at the end of 2022 as per Company's 10-K Report.

REPLY

A focused **agriculture** and **construction business**

Agriculture
76%

Construction
15%

Financial Services
9%

$23.6B

CNH
Revenues
2022, Full-Year

Note: Consolidated Revenues, CNH Full-Year 2022.

Business Needs

02 - data consistency and accuracy:

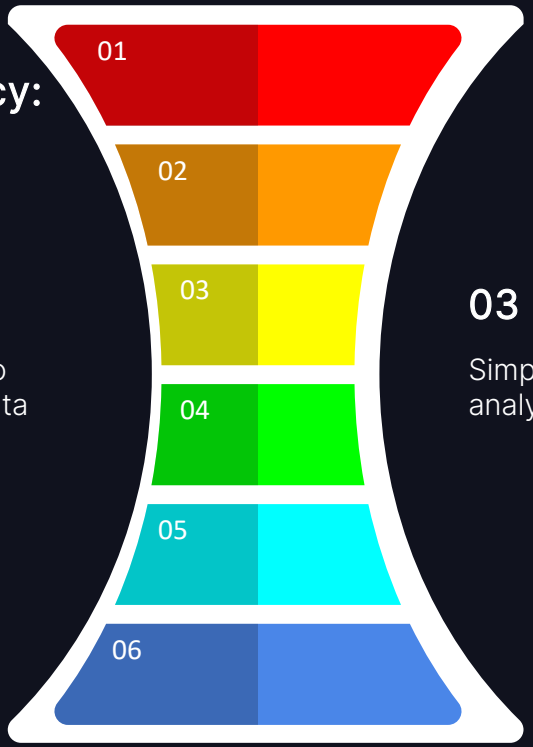
Ensuring the accuracy and consistency of data across over six regions

04 - empowering users:

Enabling self-service data access to empower employees to leverage data insights independently.

06 - Scalability

Implementing a scalable solution that can grow with the enterprise's expanding data needs and user base.



01 - Centralized Data governance

Need for a unified platform to enforce data governance policies consistently across all regions and departments.

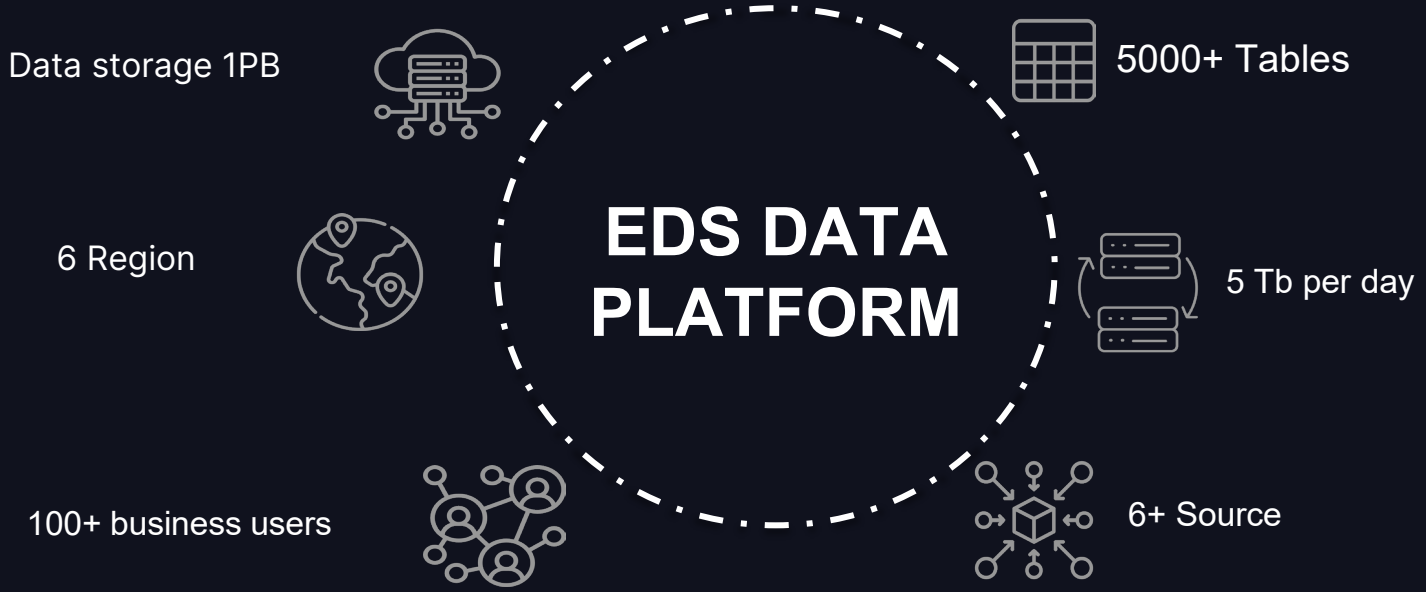
03 - Enhanced Data Accessibility:

Simplifying data access to support efficient data analysis and decision-making.

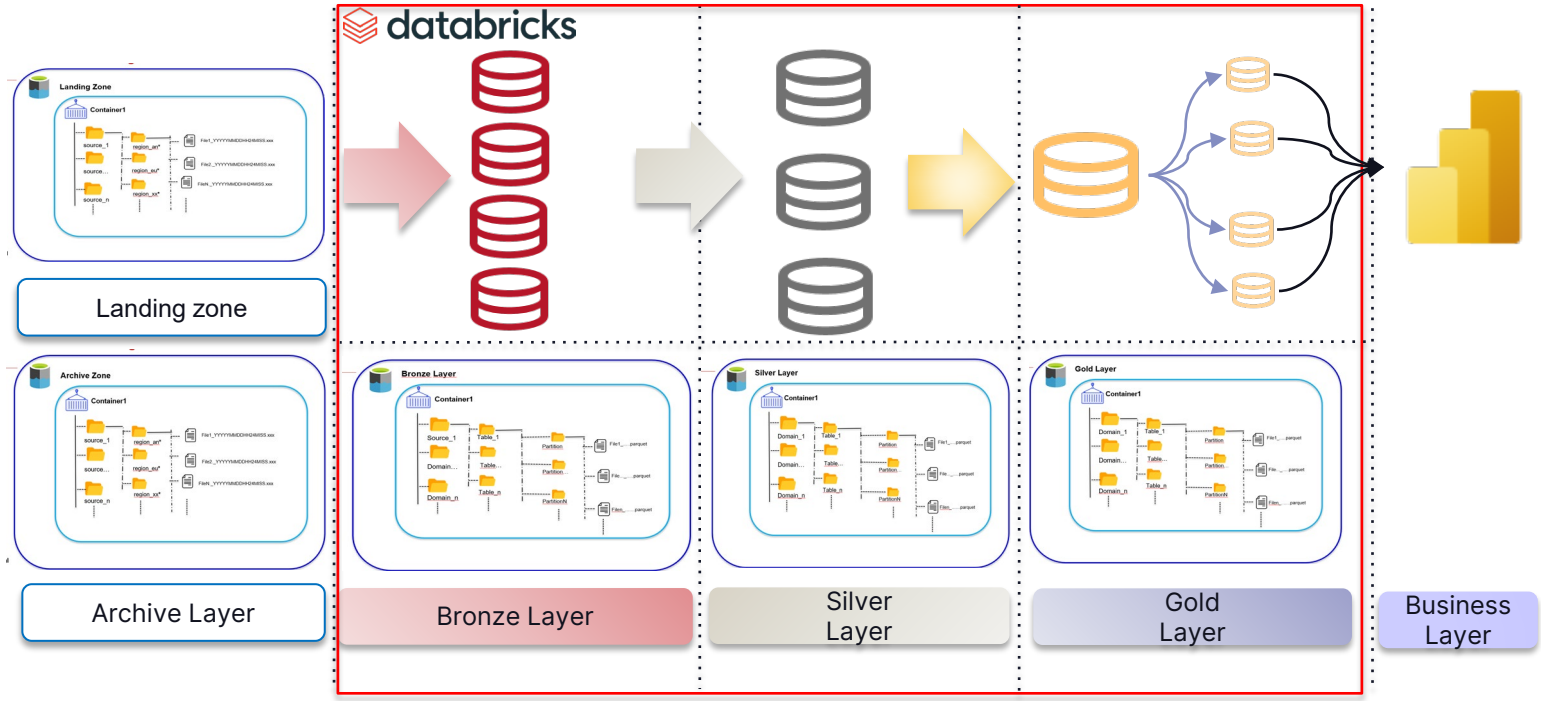
05 - Data Insights Utilization:

Maximizing the utilization of data insights for strategic decision-making and competitive advantage.

Metric Mastery: CNH Platform

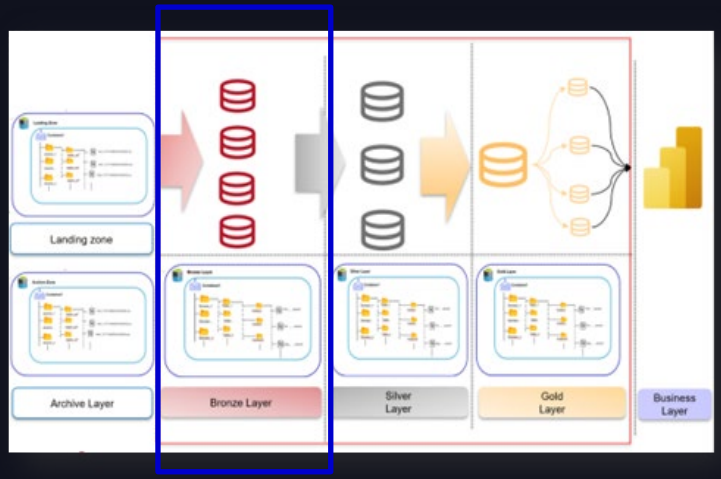


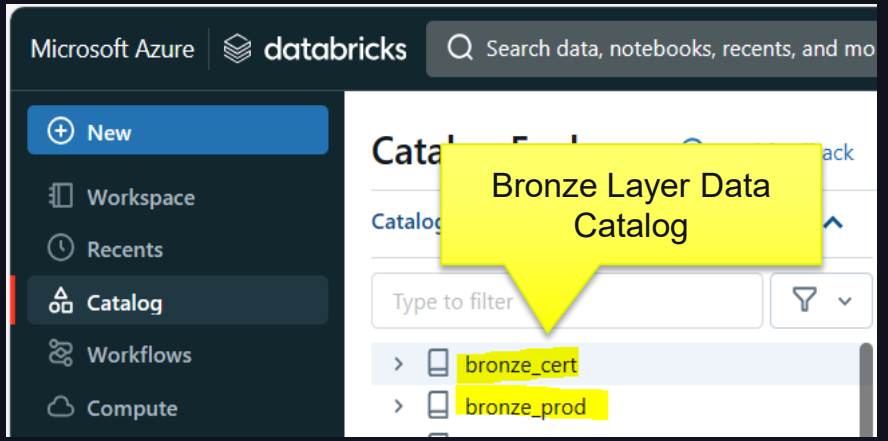
Architecture



The **EDS Lakehouse is** logically organizing data in a **Medallion Architecture** with the goal of incrementally and progressively improving the structure and quality of data as it flows through each layer of the architecture (from Bronze ⇒ Silver ⇒ Gold layer tables)

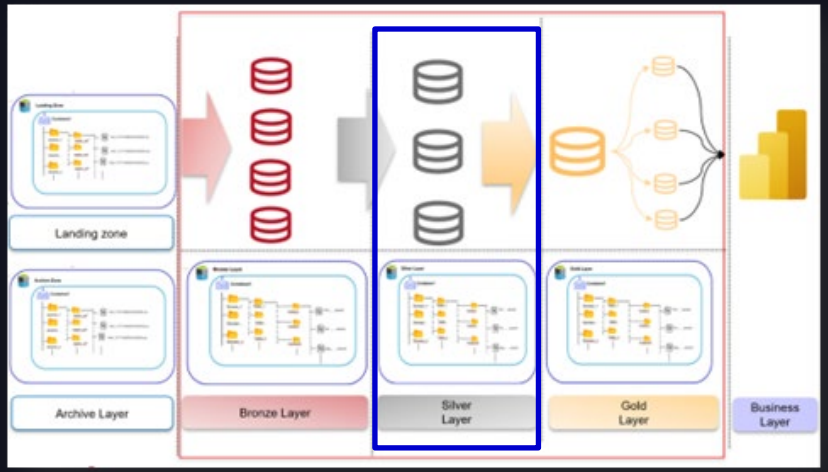
EDS - Bronze Layer

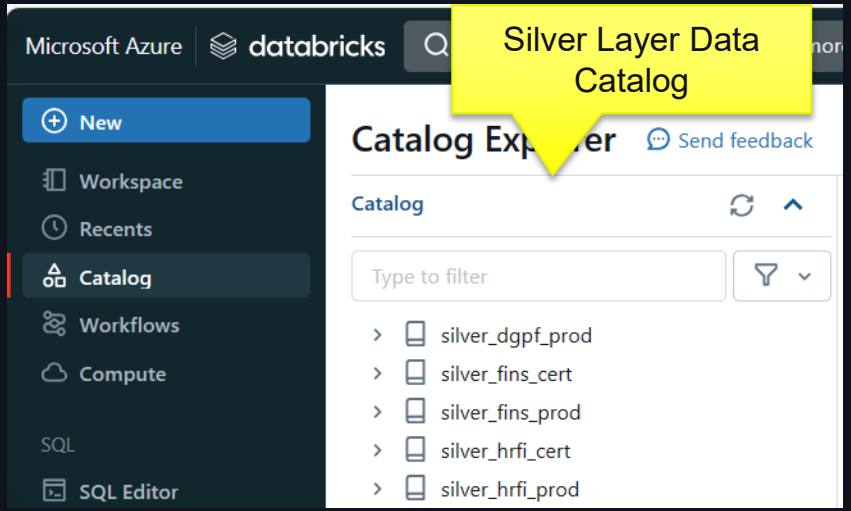




Bronze Layer is the initial stage of EDS. Bronze data maintains a 1:1 relationship with Source System: data types, column names and entities are untransformed and unchanged from whatever sources produced it. Only accepted exceptions are the ones change during Data Quality process.

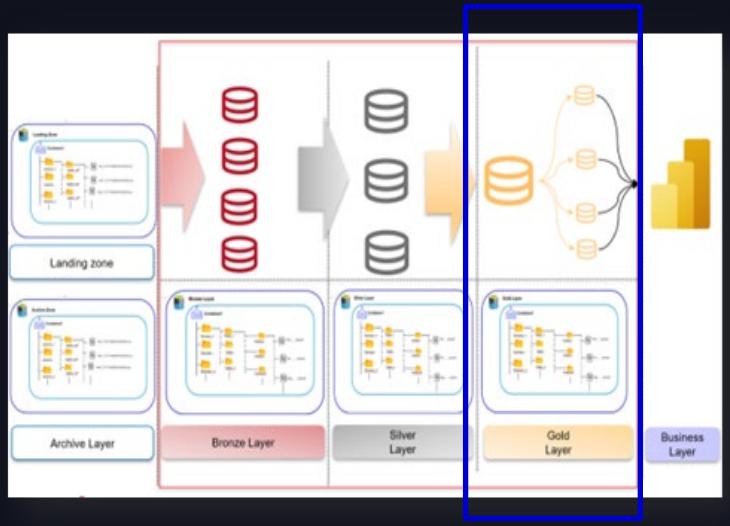
EDS - Silver Layer

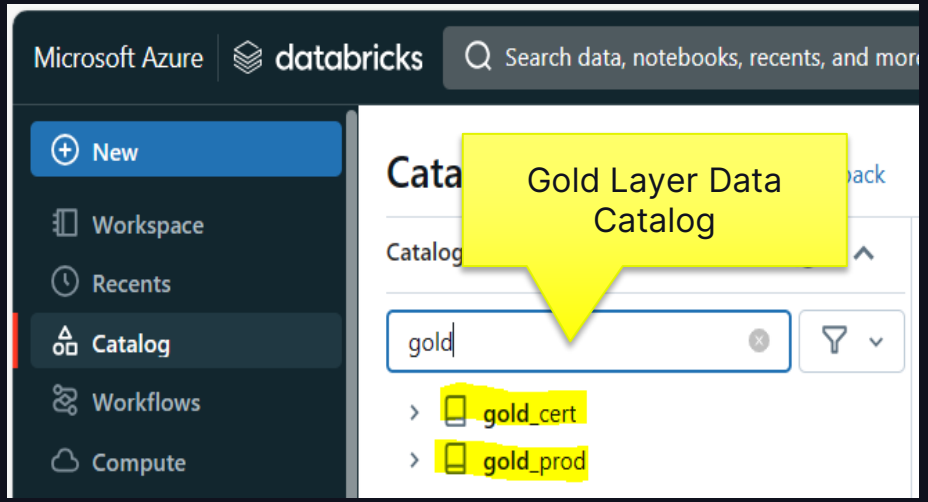




Silver Layer are data that are considered cleansed and processed to make them more accessible and valuable.

EDS - Gold Layer





Gold Layer data reach at this stage major transformation, data should be summarized or prepared in table that has a de-normalized structures to simplify data usage, Star Schema modeling

Let's play detective: what's the missing link on the last slides?

Before Unity

SHARING

- Can’t share easily table across workspaces
- Grant full access to production data

Before Unity

SHARING

- Can’t share easily table across workspaces
- Grant full access to production data

AUDITING

- No chance to verify access to the data
- no possibility to limit access resource and allocate costs

Before Unity

SHARING

- Can’t share easily table across workspaces
- Grant full access to production data

AUDITING

- No chance to verify access to the data
- no possibility to limit access resource and allocate costs

USER MANAGEMENT

- no rules based on roles for data access
- Difficulty to manage access to workspace

Before Unity

SHARING

- Can’t share easily table across workspaces
- Grant full access to production data

AUDITING

- No chance to verify access to the data
- no possibility to limit access resource and allocate costs

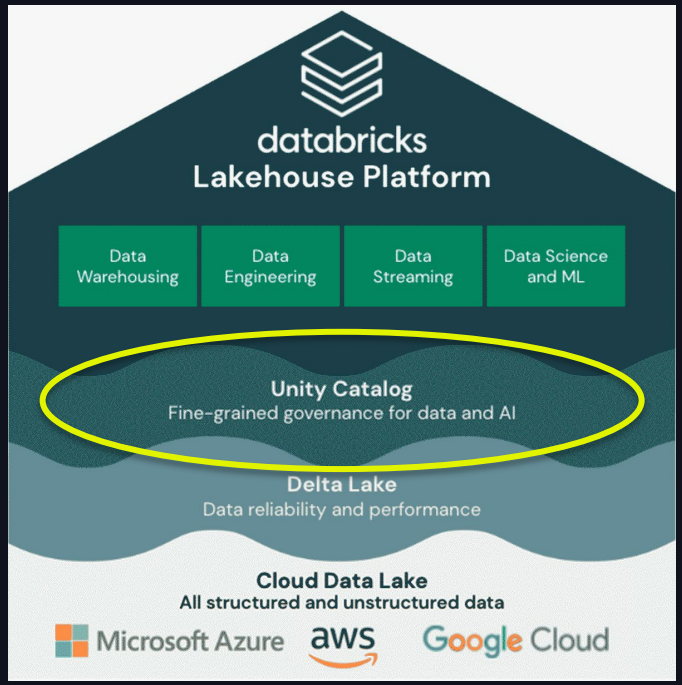
USER MANAGEMENT

- no rules based on roles for data access
- Difficulty to manage access to workspace

LINEAGE

- no possibility to handle data lineage easily
- impossibility to identify all item related to a table

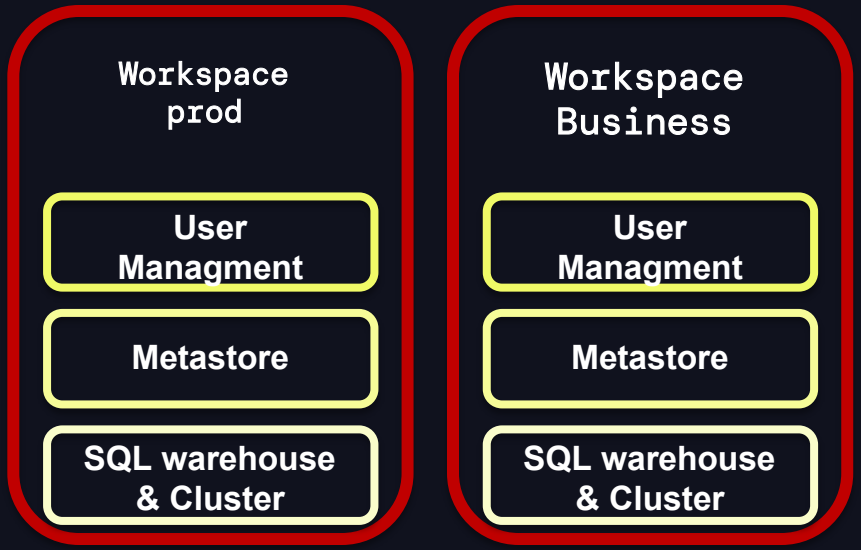
Enhancing Data Governance with Unity Catalog



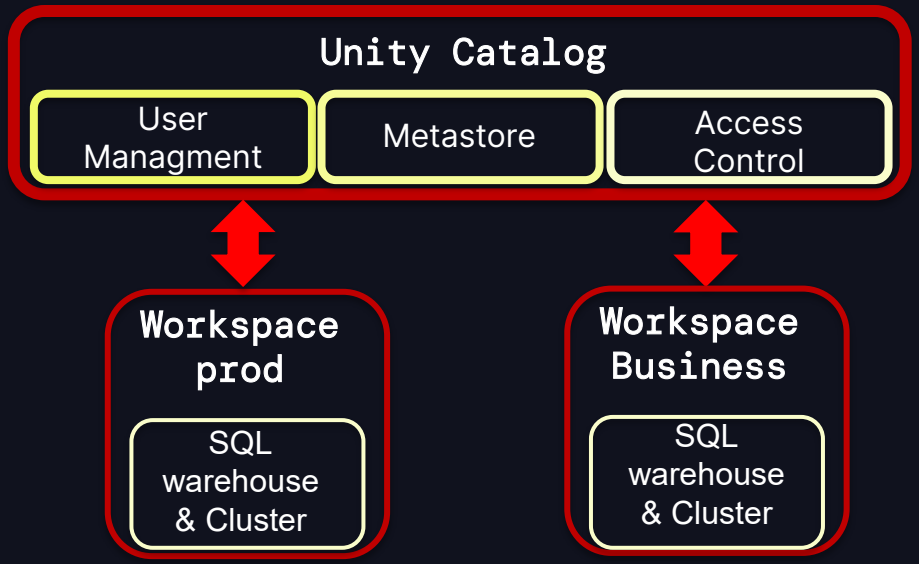
- Centralized Data Management
- Granular Access Control
- Enhanced Data Lineage
- Improved Data Discovery
- Automated Data Quality Checks

Centralized Data management

Without unity catalog



With unity catalog



Implementing Robust Data Governance Policies

Implement Data Auditing Mechanisms

Implement robust auditing mechanisms to track data access, modifications, and usage, facilitating compliance and detecting potential breaches

Data Ownership and Stewardship

Assign clear ownership for data assets to ensure accountability and responsibility for data quality and security

Conduct Regular Data Audits

Conduct regular data audits to identify and address data quality issues, security concerns, and compliance violations.

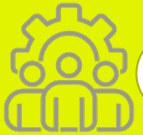
Define Data Usage Policies

Clearly outline acceptable data usage practices, including data sharing, access permissions, and data modification procedures

Implement Data Quality Standards

Establish data quality standards to ensure the accuracy, completeness, and consistency of data assets

Migration



Access control

Design and implement data security rules for access control on eds

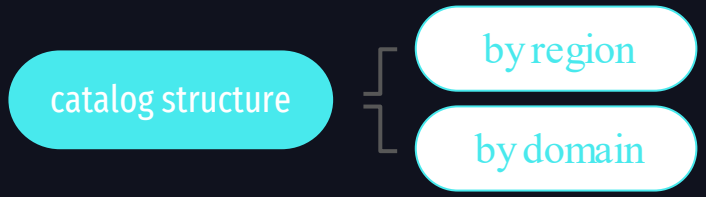
Data masking

How CNH can use data masking in databricks to enforce access control on data

Data auditing

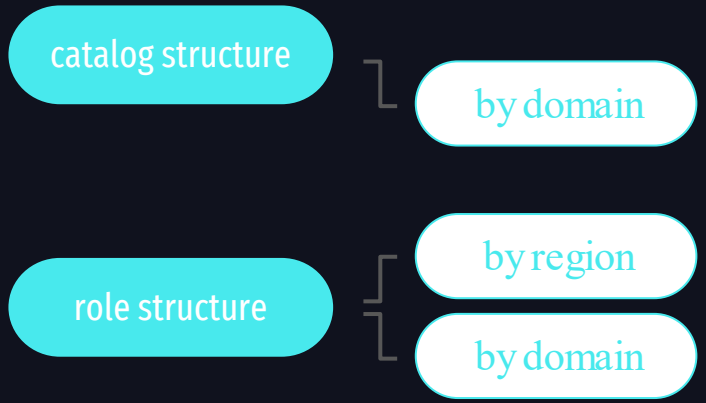
Vital component in ensuring data integrity and compliance

Strategy

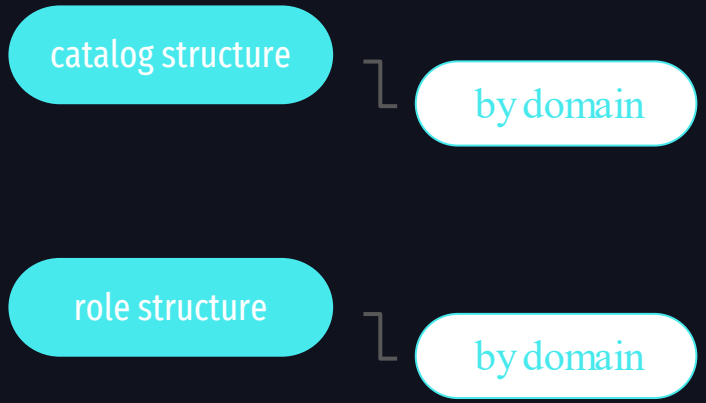


Strategy

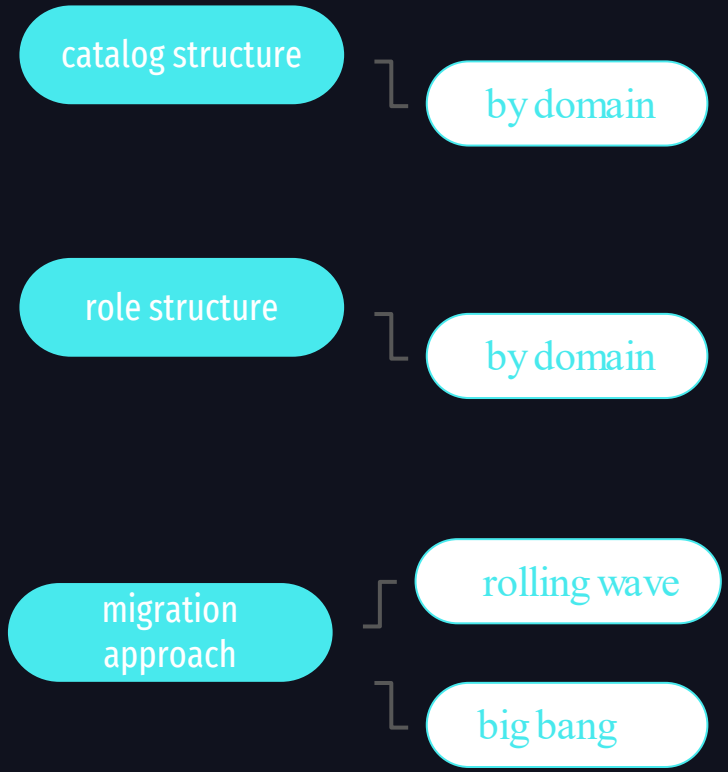
Strategy



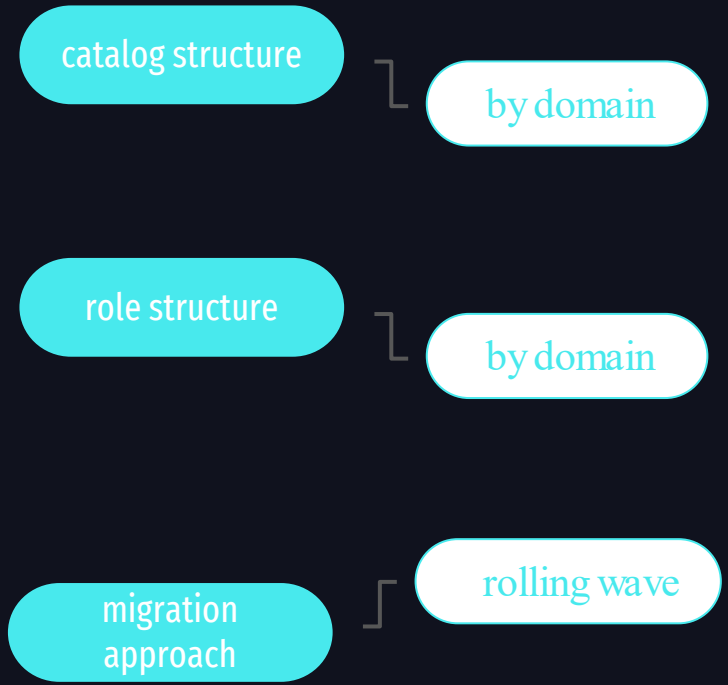
Strategy



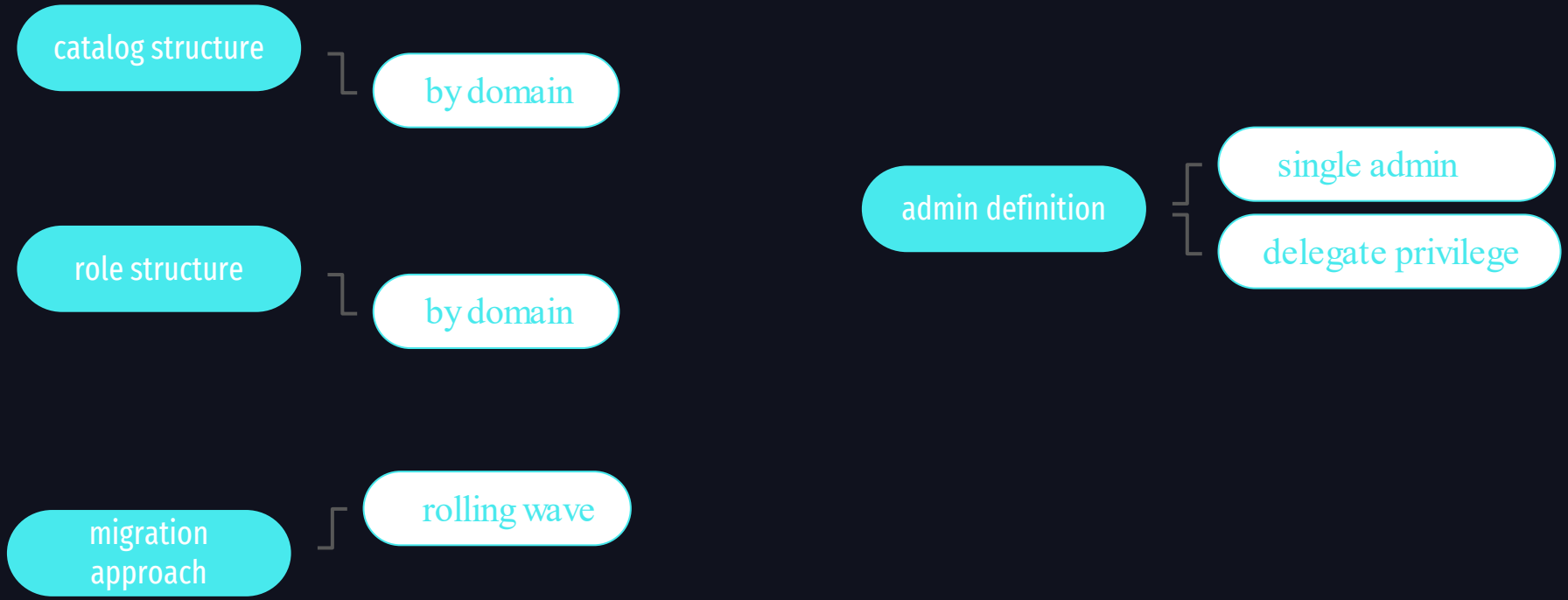
Strategy



Strategy



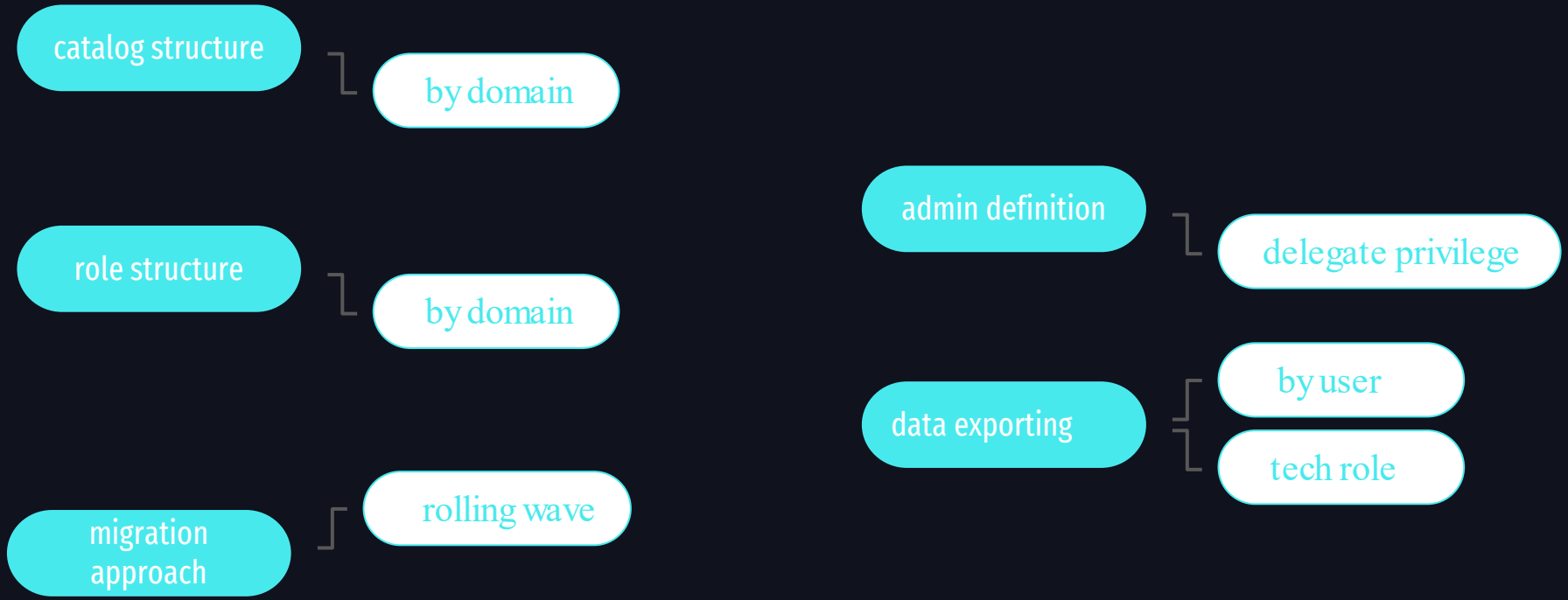
Strategy



Strategy

- catalog structure
  - by domain
- role structure
  - by domain
- migration approach
  - rolling wave
- admin definition
  - delegate privilege

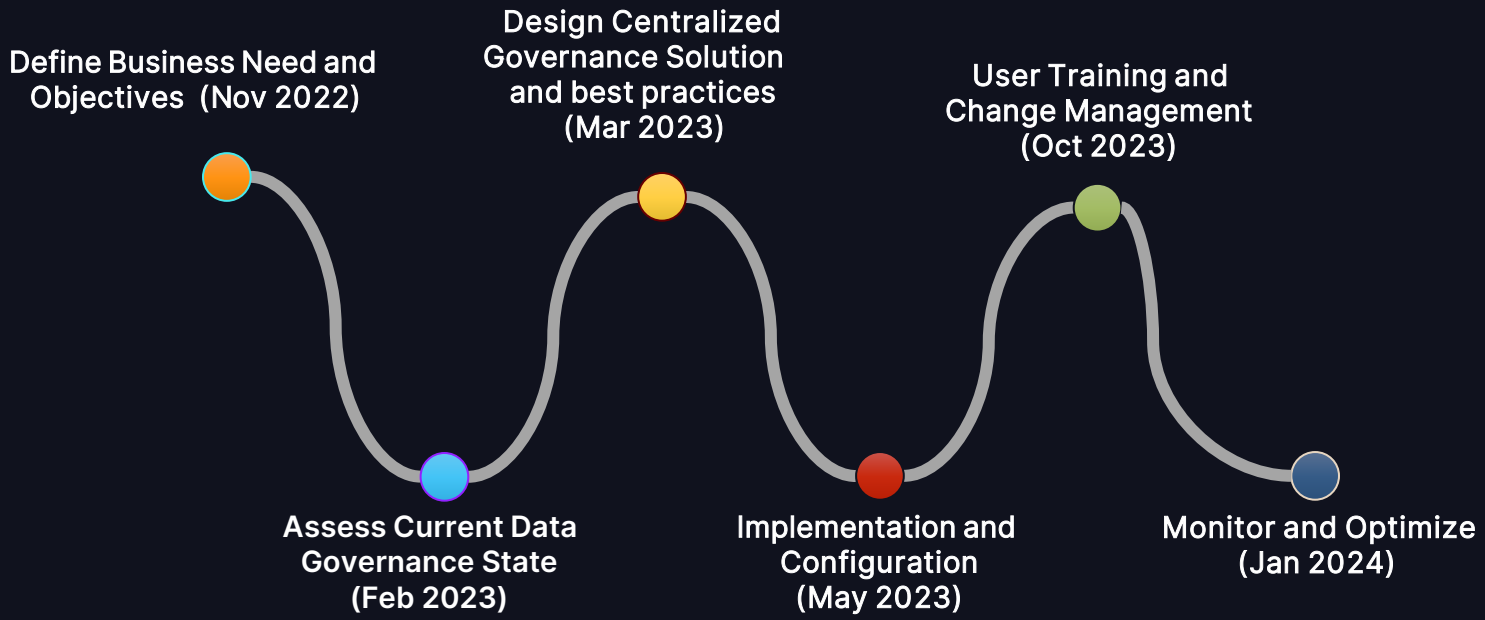
Strategy



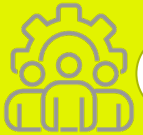
Strategy

- catalog structure → by domain
- role structure → by domain
- migration approach → rolling wave
- admin definition → delegate privilege
- data exporting → by user

timelines



Migration



Access control

Design and implement data security rules for access control on eds

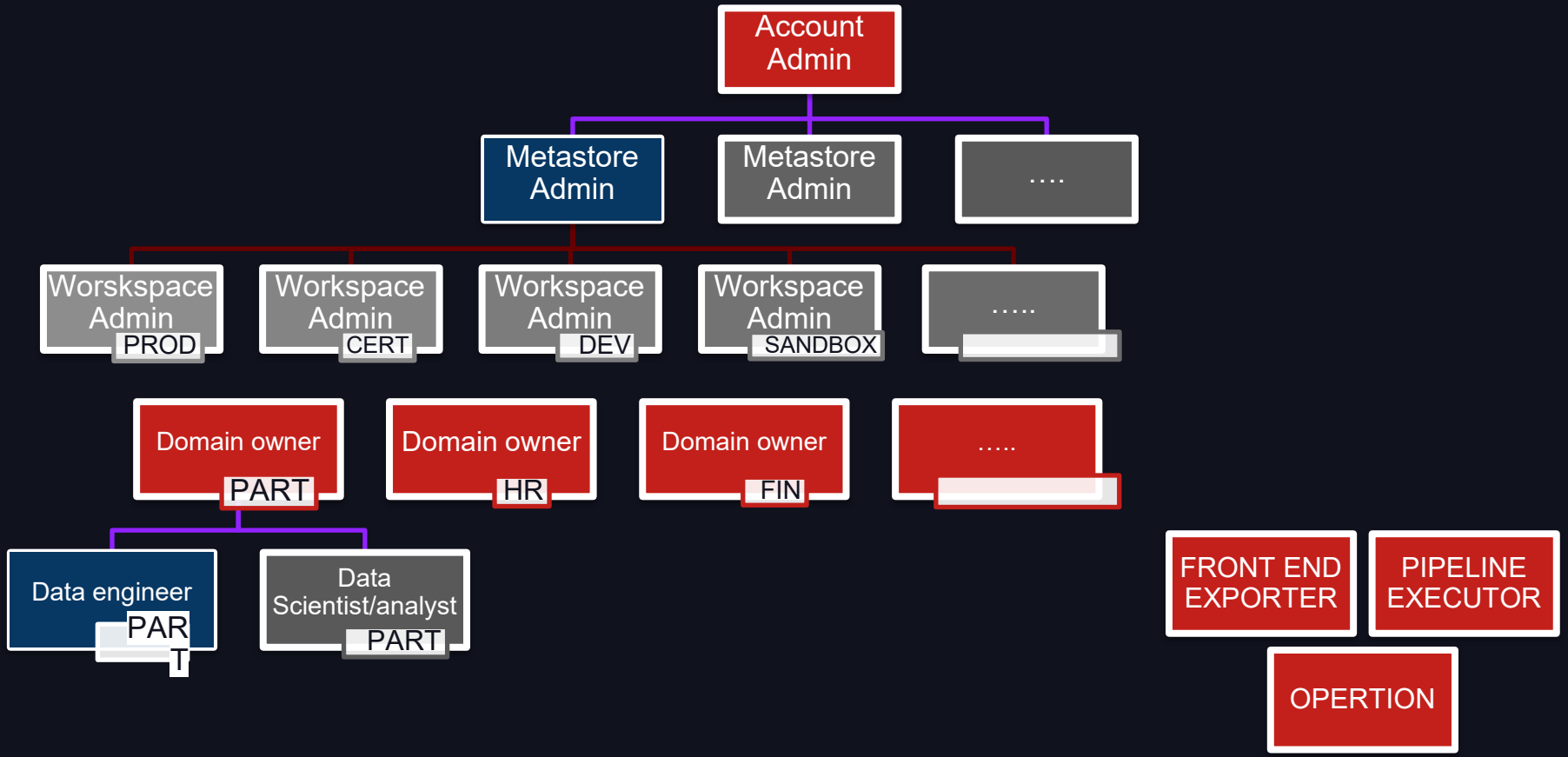
Data masking

How CNH can use data masking in databricks to enforce access control on data

Data auditing

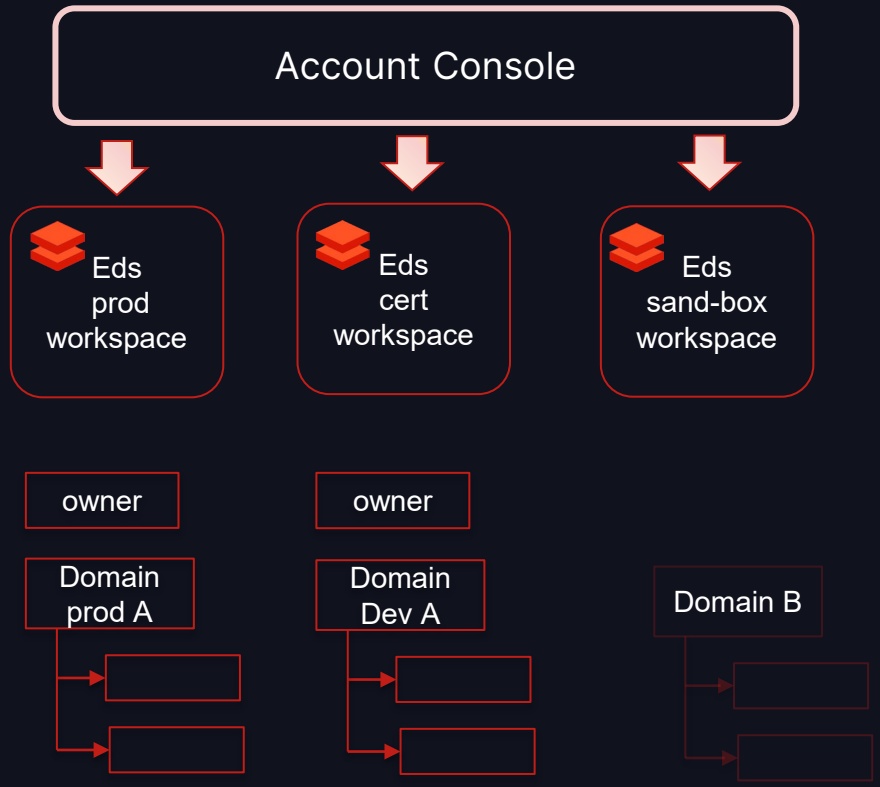
Vital component in ensuring data integrity and compliance

Access control



Access control

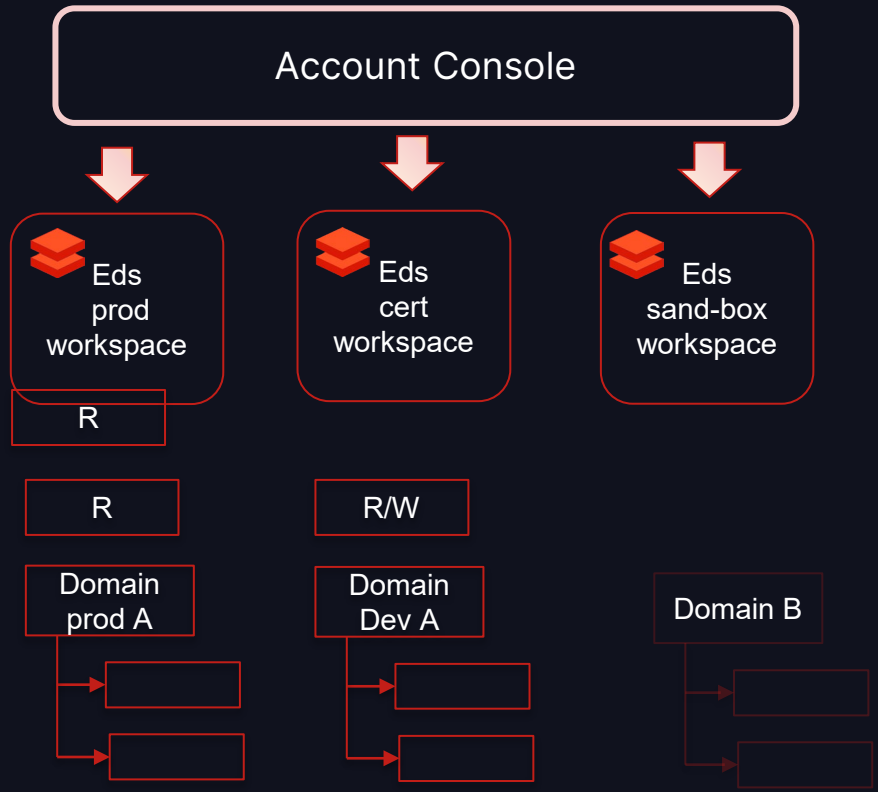
Domain owner



Who	Domain admin, for data access managment.
Workspace access	All Workspaces
Data Access	Domain-specific R/W access, e.g. Finance only. Or multi-domain, e.g. all EDS domains.
Databricks permission	Create and manage domain-specific objects (catalogs, schema, etc.) Create or drop, grant privileges on, and change ownership of catalogs and other data objects within a specific domain.

Access control

Data Engineer



Who	ETL Developers
Workspace access	Only selected domain-specific workspaces
Data Access	Domain-specific access, e.g. Finance only Or multi-domain, e.g. all EDS domains. R/W access in **Dev**, read-only in **Prod**.
Databricks permission	**Dev** Create clusters and jobs using existing cluster policies (defined by the DE admin). Attach notebooks to clusters. Create notebooks, SQL queries, dashboards. **Prod** No permissions to create clusters or jobs. Can attach notebooks to existing clusters.

Migration

Access control

Design and implement data security rules for access control on Databricks

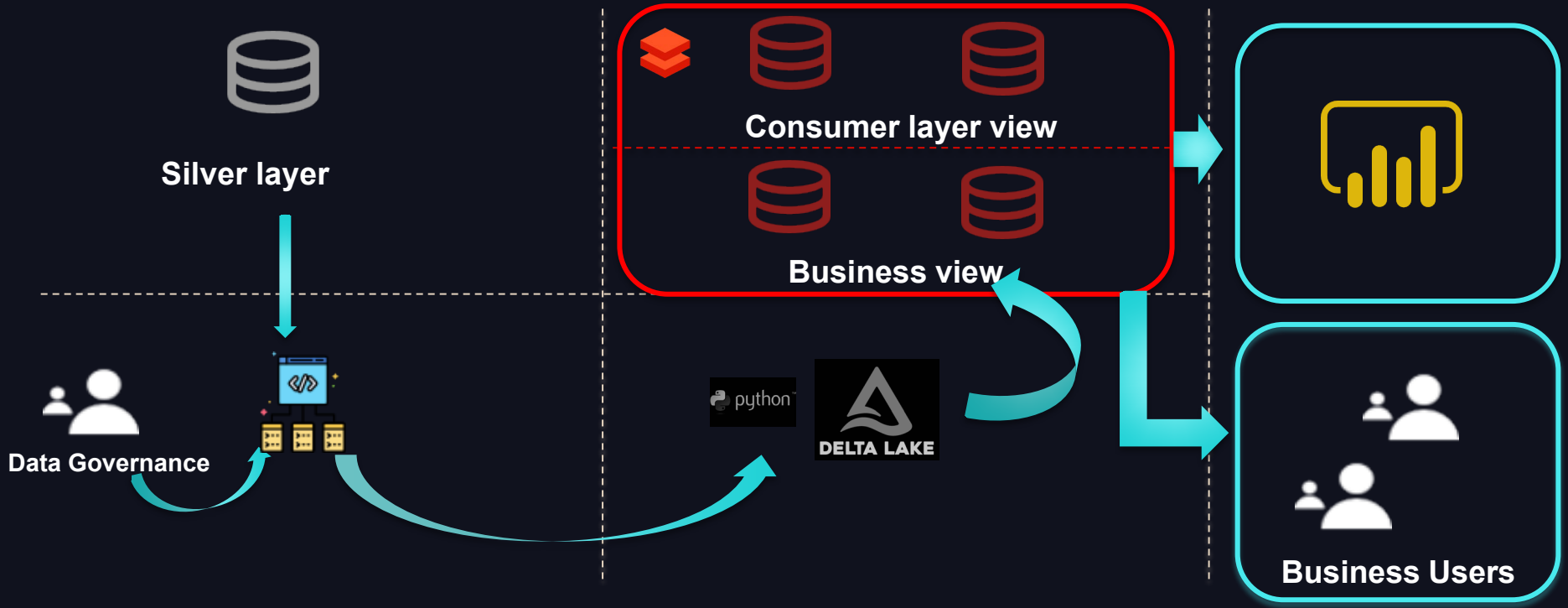
Data masking

How CNH can use data masking in databricks to enforce access control on data

Data auditing

Vital component in ensuring data integrity and compliance

Data Masking



Demo

role managed table

Table

	table_name	permission
1	readacted_customer	EDS_OWNER_PROD
2	vw_part_masked_row	eds_read_only_part_pr...
3	vw_part_masked_colu...	EDS_OWNER_PROD

function role checking

Just now (<1s)

```
select is_member("EDS_OWNER_PROD")
```

Table

	is_member(EDS_OWNER_PROD)
1	true

Column masking

03:08 PM (<1s)

```
select * from gold_cert.master_data_private.vw_part_masked_column
```

Table

	PART_MASKED	CD_MFR	NR_PACK_QUANTITY_MASKED	.00 NR_WEIGHT_NUMBER
1	911663008	COE	0	0
2	YF12B00019F1	KOB	1	0
3	1291250015	S	0	0
4	BSU1W	IH	1	454
5	DDN993208	FOR	1	27
6	CAS1020	CAS	1	0
7	2400312R92	IH	0	0
8	77268845	KOB	1	77
9	0086305889	CUR	1	366
10	921251	ZED	1	1452
11	82985402	MEX	1	422
12	16165174	JES	5	4
13	292689	BER	1	600
14	FGAF25017DS	MKT	6	0
15	3109450R1	IH	0	0

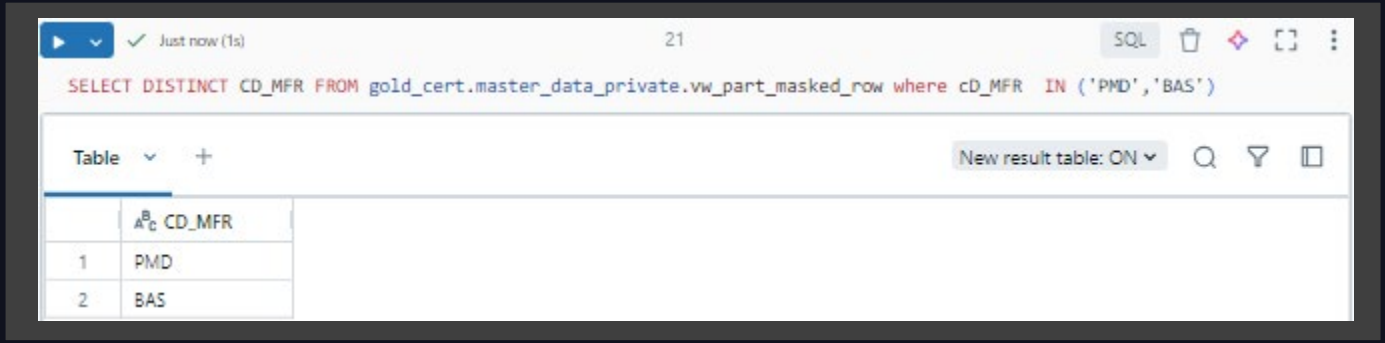
```
select * from gold_cert.master_data_private.vw_part_masked_column
```

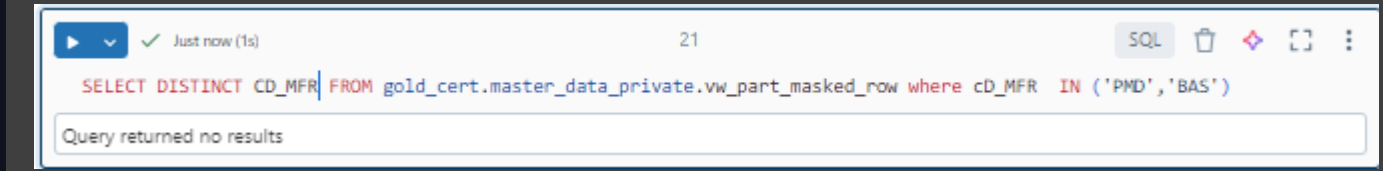
Table — New result table: ON

	PART_MASKED	CD_MFR	NR_PACK_QUANTITY_MASKED	.00 NR_WEIGHT_NUMBER
1	*xxxxxx	CAS	x	0
2	xxxxxx*	STA	x	0
3	xxxxxxx*x	F	x	0
4	xxxxxxx*xx	F	x	0
5	xxxxxxxx	BEL	x	0
6	xxxxxxxx	SM	x	80
7	xxxxxx	BER	x	1200
8	xxxxxx*x	BHB	x	23800
9	xxxxxx*x	CAS	x	40
10	xxxxxxxxx	COE	x	0
11	**xxx***	MKT	xx	0
12	xxxxxxxx	LEC	x	0
13	xxxxxxxx	HIT	x	100040
14	xxxxxxxxx	COE	x	0
15	**xx*xxxxxx	KOB	x	7

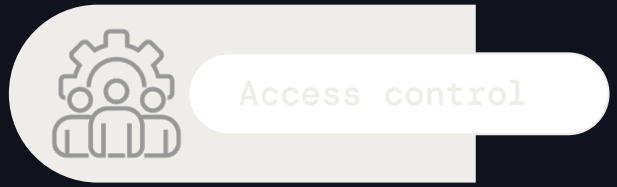
Demo

Row masking





Migration



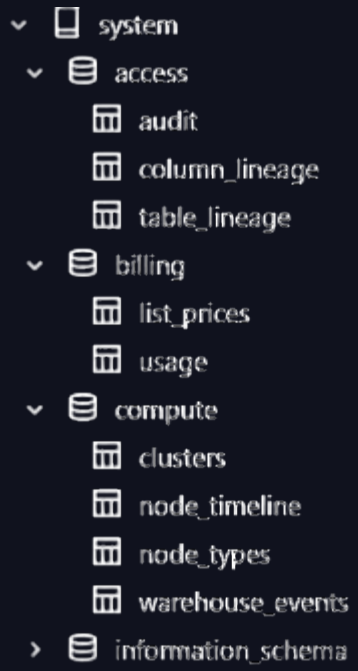
Design and implement data security rules for access control on Databricks

How CNH can use data masking in databricks to enforce access control on data

Vital component in ensuring data integrity and compliance

Centralized data auditing

System Tables



Databricks by default provides different audit tables, that are stored inside system.access, these tables are:

- Audit;
- Column_lineage;
- Table_lineage;

Dashboards

Workspace Action

Query - Workspace Access (Actions -> createTable, deleteTable, getTable, updateTables, last 7 days)

Search event_date, workspace_name, workspace_id and 2 others...

event_date	workspace_name	workspace_id	grantee	action_name
2024-04-24	EDS-CERT - euedscadb017	[redacted]	[redacted]	getTable
2024-04-24	EDS-CERT - euedscadb017	[redacted]	[redacted]	createTable
2024-04-24	SANDBOX - cnh-we-prod.eds.adb-01	[redacted]	[redacted]	getTable
2024-04-24	SANDBOX - cnh-we-prod.eds.adb-01	[redacted]	[redacted]	getTable
2024-04-24	EDS-PROD - euedspadb017	[redacted]	[redacted]	getTable
2024-04-24	EDS-CERT - euedscadb017	[redacted]	[redacted]	getTable
2024-04-24	SANDBOX - cnh-we-prod.eds.adb-01	[redacted]	[redacted]	getTable
2024-04-24	SANDBOX - cnh-we-prod.eds.adb-01	[redacted]	[redacted]	getTable
2024-04-23	EDS-PROD - euedspadb017	[redacted]	[redacted]	getTable
2024-04-24	EDS-PROD - euedspadb017	[redacted]	[redacted]	getTable
2024-04-24	SANDBOX - cnh-we-prod.eds.adb-01	[redacted]	[redacted]	updateTables

Dashboards

Group monitoring

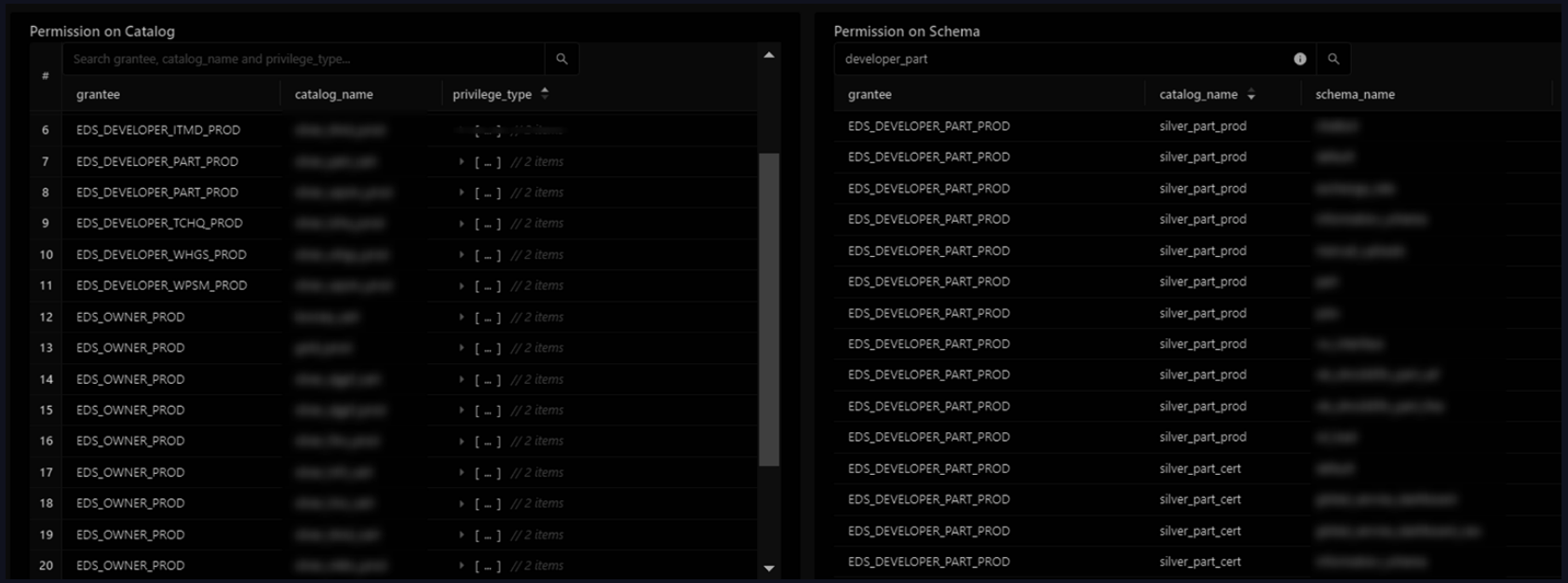
Query - Group and Users (Batch Result)

Search grantee, directGroup and user_name...

#	grantee	directGroup	user_name
1	EDS_READONLY_HRIS_CLUSTER_PROD	true	[redacted]
2	EDS_READONLY_ITMD_CLUSTER_PROD	true	[redacted]
3	EDS_READONLY_HRIS_CLUSTER_PROD	true	[redacted]
4	EDS_READONLY_SCMP_CLUSTER_PROD	true	[redacted]
5	EDS_READONLY_SCMP_CLUSTER_PROD	true	[redacted]
6	EDS_READONLY_PART_CLUSTER_PROD	true	[redacted]
7	EDS_READONLY_WHGS_CLUSTER_PROD	true	[redacted]
8	EDS_READONLY_HRIS_CLUSTER_PROD	true	[redacted]
9	EDS_READONLY_PART_CLUSTER_PROD	true	[redacted]
10	EDS_DEVELOPER_HRIS_PROD	true	[redacted]
11	EDS_READONLY_HRFI_CLUSTER_PROD	true	[redacted]
12	EDS_DEVELOPER_MKTC_PROD	true	[redacted]
13	EDS_DEVELOPER_DGPF_PROD	true	[redacted]
14	EDS_READONLY_ITMD_CLUSTER_PROD	true	[redacted]

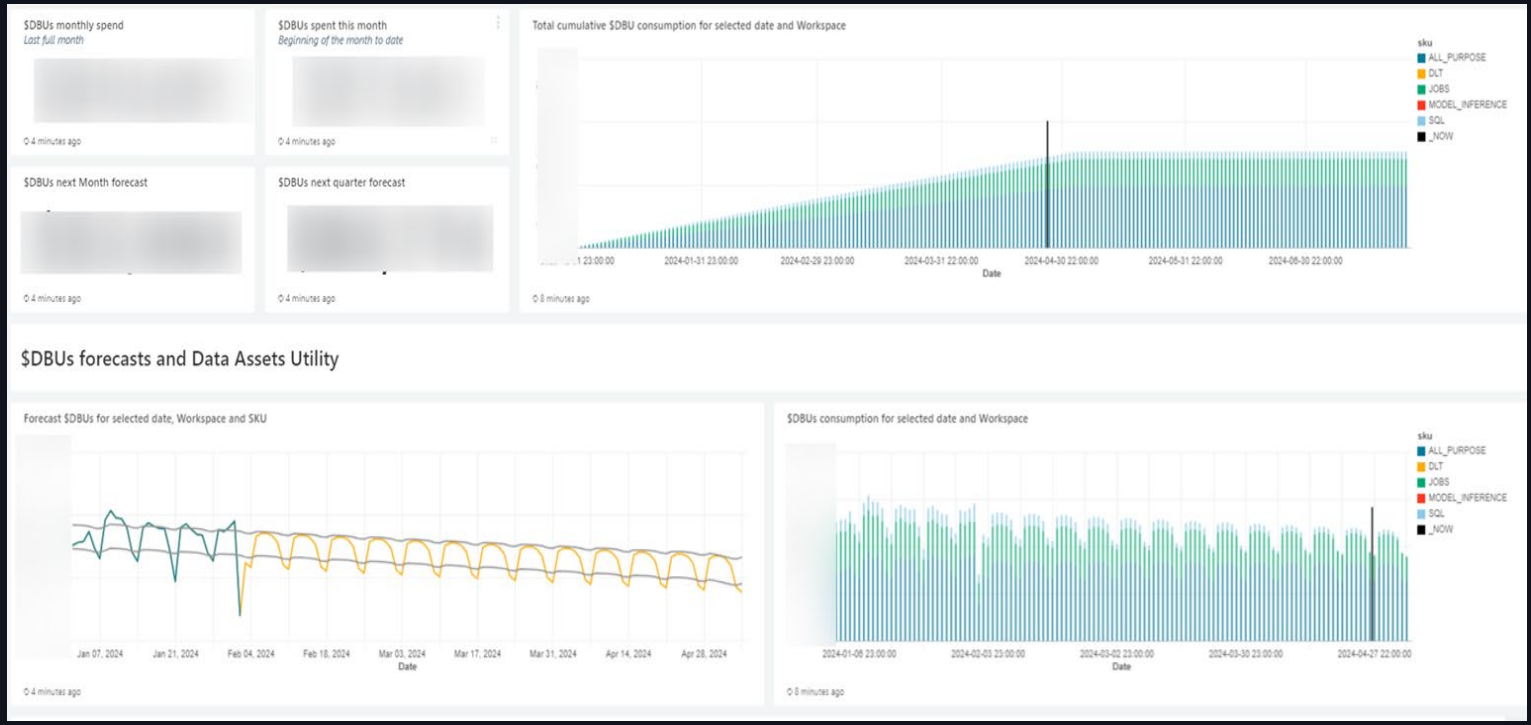
Dashboards

Permission overview

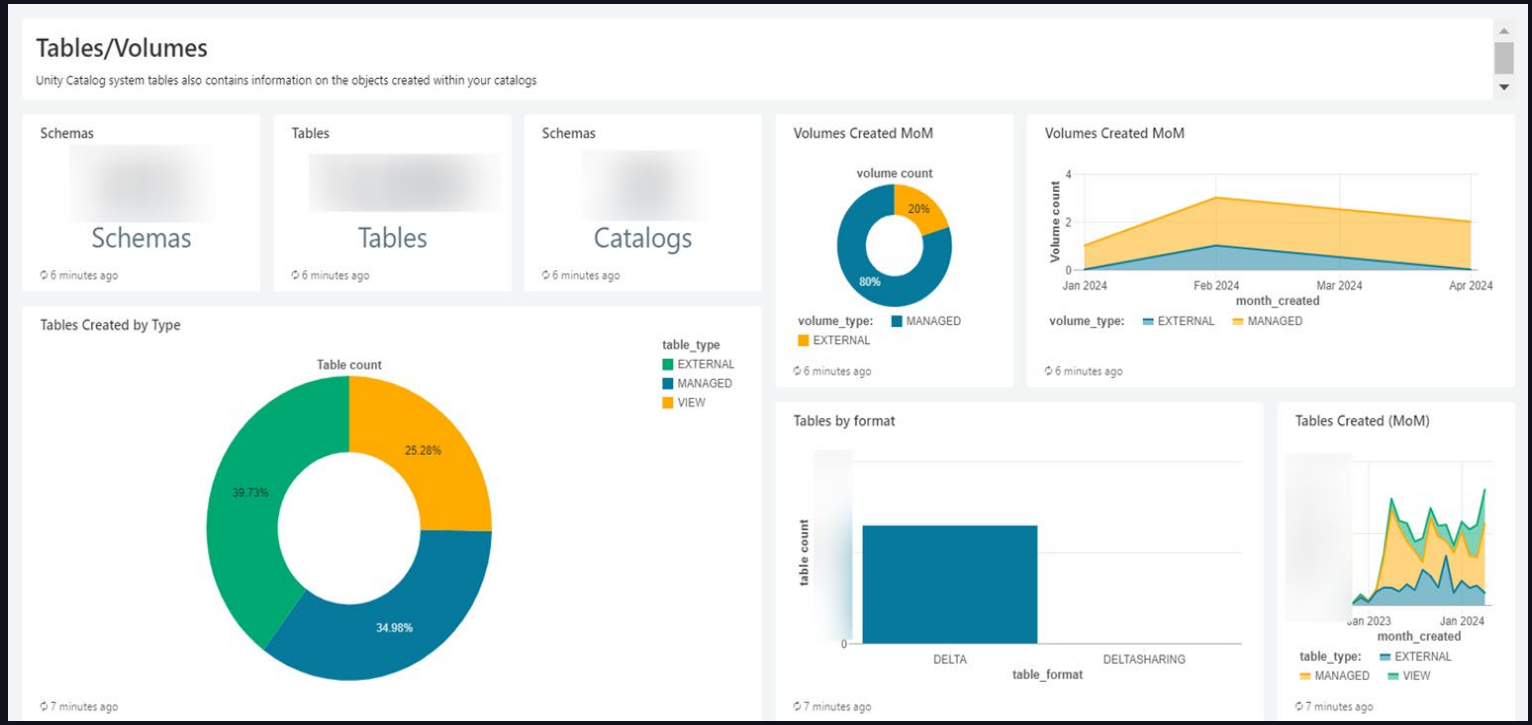


Dashboards

consumption analysis



Dashboards



Rewind & Rethink

First setup

- define a migration strategy
- design and develop technical role
- define what and who should be monitored
- define cluster policies

Per Domain

- perform usage data analysis
- create catalog and volumes
- create domain roles
- create clusters
- migrate tables inside unity

after migration

- implement audit dashboards
- revoke hive_metastores data
- revoke account user privileges
- train business users

Progress and Lessons Learned

Acceleration

- data democratization
- data sharing
- security access

Progress and Lessons Learned

Acceleration

- data democratization
- data sharing
- security access

Benefit

- lineage capability
- cost control
- user orientation

Progress and Lessons Learned

Acceleration

- data democratization
- data sharing
- security access

Benefit

- lineage capability
- cost control
- data catalog
- user orientation

Unlock Potentials

- self service analysis
- comprehensive governance
- usage monitoring

Acknowledgment

- Liang Sun
- Arianna Bonani
- Nataliya Zayeva
- Enrico Bacchetta

- Nik Green
- Cristiano Vineis
- Dario Zaffagnin

- Sergio Schena
- Luca Bolognesi
- Davide Veneziano
- Eddy Amarouche

Thanks for the attention