**Databricks Data Management and Goverance Issues with Mitigations**

**Databricks Data Management & Governance Issues Comparison Table with Mitigations**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| # | Data Governance Issue | One-Liner Description | Where It Typically Arises | Governance Areas Most Affected | Mitigation Strategies |
| 1 | Lack of Centralized Catalog | Data assets are scattered without a single source of truth, making discovery and governance difficult. | Data onboarding, multi-team environments | Tables, Views, Schemas | Use Unity Catalog to centralize metadata, permissions, and lineage tracking. |
| 2 | Incomplete Data Lineage | Teams can’t trace the flow and transformation of data across pipelines and tables. | ETL workflows, Delta table updates | Tables, Pipelines, Views | Enable lineage tracking in Unity Catalog; document transformations in notebooks and pipelines. |
| 3 | Weak Access Controls | Permissions are inconsistently applied, creating risks of unauthorized access. | Table access, sharing data externally | Unity Catalog, SQL Warehouses | Use RBAC for fine-grained controls; audit permissions regularly; follow least privilege principles. |
| 4 | Inconsistent Data Classification | Sensitive and regulated data isn’t clearly tagged or categorized. | Schema onboarding, data ingestion | Tables, Columns | Implement column-level tagging for sensitivity, compliance, and business domains. |
| 5 | No Data Retention Policies | Data is kept indefinitely without clear lifecycle management, creating compliance risks. | Table lifecycle management | Tables, Backups | Define retention and purging policies; automate enforcement with scheduled jobs. |
| 6 | Unvalidated Data Quality | Bad data enters downstream systems due to lack of validation checks. | Ingestion pipelines, ETL processes | Bronze, Silver, Gold Layers | Integrate data quality validations and alerts into ingestion and transformation workflows. |
| 7 | Limited Audit and Monitoring | Data access and modification activities are not consistently logged and monitored. | Production environments | Tables, Jobs, Warehouses | Enable audit logs; export to secure storage or SIEM; review logs periodically. |
| 8 | Poor Metadata Documentation | Data assets lack clear descriptions, owners, and business context, leading to misuse. | Schema creation, onboarding | Unity Catalog, Tables | Require metadata documentation on creation; maintain a governance wiki or data dictionary. |
| 9 | Siloed Data Stewardship | No clear accountability for data domains or quality, creating gaps in ownership. | Cross-functional teams | Governance, Data Quality | Assign data stewards for domains; publish stewardship responsibilities and escalation paths. |
| 10 | Inconsistent Naming Conventions | Tables, schemas, and assets have inconsistent names, causing confusion across teams. | Data modeling, schema setup | Unity Catalog, SQL Warehouses | Define naming conventions; enforce them with automated validations and reviews. |

**Quick Reference**

* **Unity Catalog:** Central governance layer for metadata, permissions, and lineage.
* **RBAC:** Role-Based Access Control for granular permissions.
* **Lineage Tracking:** Visibility into data flow and transformations.
* **Data Stewardship:** Assignment of owners accountable for data quality and compliance.
* **Data Classification:** Tagging data assets based on sensitivity and business relevance.

**Example Mitigation Actions and Configurations**

**Enable Unity Catalog:**

* In the admin console, set up Unity Catalog with workspaces attached.
* Configure catalogs and schemas centrally.

**Grant Fine-Grained Permissions:**

sql

CopyEdit

GRANT SELECT ON TABLE main.sales.transactions TO `finance\_analysts`;

**Tag Sensitive Columns:**

sql

CopyEdit

ALTER TABLE main.hr.employee\_data

SET TAGS (

'sensitivity' = 'high',

'compliance' = 'GDPR',

'owner' = 'hr\_data\_steward'

);

**Define Retention Policies:**

sql

CopyEdit

DELETE FROM main.logs

WHERE event\_timestamp < current\_timestamp() - INTERVAL 90 DAYS;

**Document Metadata:**

sql

CopyEdit

COMMENT ON TABLE main.marketing.campaign\_performance IS 'Daily campaign metrics for performance tracking.';

**Enable Audit Logs:**

* Configure workspace audit logs to export events to secure cloud storage for compliance tracking.

**Establish Naming Conventions:**

* Example format:

php-template

CopyEdit

<domain>\_<subject>\_<layer>

e.g., sales\_orders\_bronze

**Assign Data Stewards:**

* Maintain a registry table:

sql

CopyEdit

CREATE TABLE governance.data\_steward\_registry (

table\_name STRING,

steward STRING,

domain STRING,

last\_reviewed DATE

);