Data governance in **Databricks** ensures that data is secure, compliant, discoverable, and usable across the organization. Below are the key **techniques and features** used to implement **data governance** effectively in Databricks:

🔐 1. Access Control (Unity Catalog + ACLs)

- **Unity Catalog** is the cornerstone of modern governance in Databricks.
    - Enables centralized data access control across **workspaces**.
    - Supports **fine-grained permissions** at catalog, schema, table, view, function, and column level.
- **Role-Based Access Control (RBAC)**: Assign roles (e.g., data_reader, data_engineer) with predefined permissions.
- **Table Access Control (TAC)**: For older workspaces without Unity Catalog, TAC can still be used to control SQL-based data access.

🔍 2. Data Lineage

- **Unity Catalog automatically tracks lineage** across:
    - Notebooks
    - Jobs
    - Pipelines (Delta Live Tables)
- You can **view end-to-end lineage** of datasets — where the data came from and how it's being transformed.
- Helps in **impact analysis**, compliance audits, and debugging.

🧪 3. Data Quality & Validation

- **Delta Live Tables (DLT)**:
    - Built-in **expectations** allow you to enforce data quality rules.
    - Example: expect(col("age") > 0, "non_negative_age")
- **Great Expectations**:
    - Can be integrated for more complex and customizable data validation logic.

🔄 4. Schema Enforcement & Evolution

- **Delta Lake** ensures:
    - **Schema enforcement**: Prevents bad data from corrupting your tables.
    - **Schema evolution**: Allows automatic column addition (controlled when needed).
- Enables governance on schema changes and avoids unexpected data corruption.

🗃️ 5. Data Cataloging & Discovery

- Unity Catalog acts as a **centralized metastore**:
    - Supports **tags**, **descriptions**, and **column-level documentation**.
    - Integrated with **Databricks Search** for data discovery.
- Integration with tools like **Collibra** or **Alation** for enterprise cataloging.

📊 6. Audit Logging & Monitoring

- **Audit Logs**: Track user activity, such as who accessed or modified data.
    - Integrated with **Azure Monitor**, **AWS CloudTrail**, or **SIEMs**.
- **Unity Catalog** logs changes to access controls and schema changes.

👥 7. Data Ownership & Stewardship

- Define **data owners and stewards** for each dataset or domain.
- Assign responsibility for:
    - Metadata management
    - Data quality monitoring
    - Access approval workflows

🔁 8. Versioning & Time Travel

- **Delta Lake time travel** allows querying previous versions of data.
- Useful for:
    - Audits
    - Data recovery
    - Reproducibility of ML/BI workflows

🔐 9. Row-Level & Column-Level Security

- **Unity Catalog** supports:
    - **Column-level access control**: Mask or restrict access to specific columns.

    - **Row-level filters** (Public Preview): Enforce access policies dynamically at query time.
- Common use case: Hide sensitive PII based on user roles.

🌐 10. Compliance & Sensitive Data Tagging

- Use **tags or custom attributes** to label columns (e.g., PII, Financial, HIPAA).
- Enables **policy enforcement** and simplifies data compliance (e.g., GDPR, CCPA).
- Combine with **Data Loss Prevention (DLP)** tools or **Azure Purview** integration.

🧱 11. Governance Across Workspaces (Multi-Tenant Architecture)

- Unity Catalog supports **multiple workspaces sharing the same governance model**.
- Helps enforce consistent governance in multi-tenant or multi-team setups.