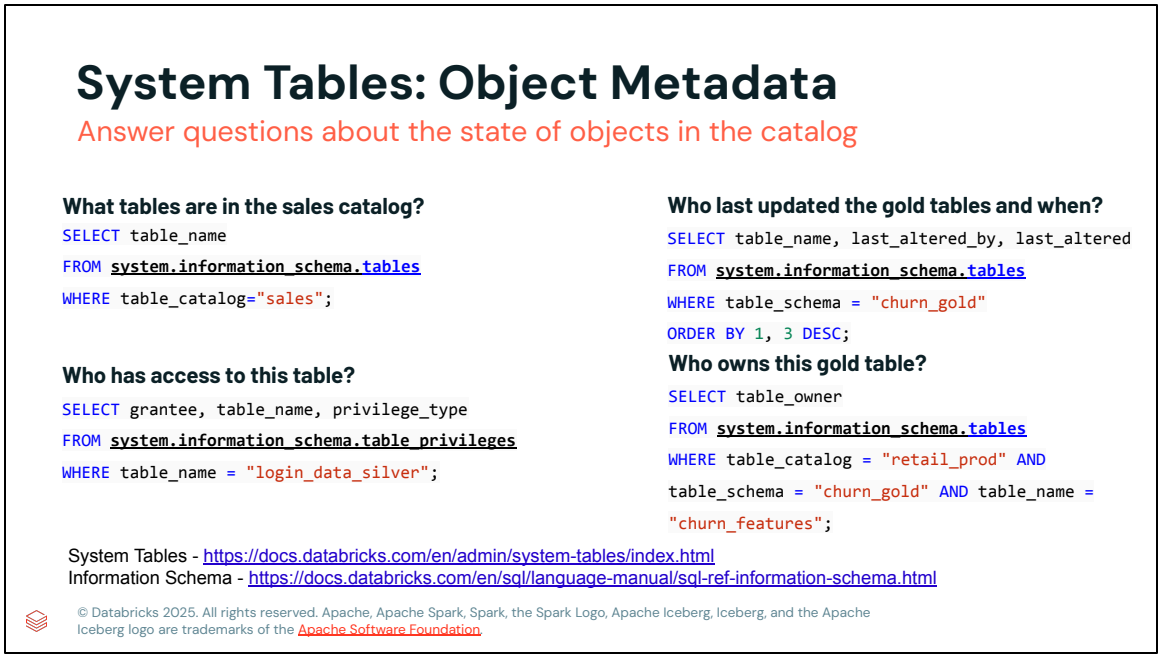
databricks

Unity Catalog

LECTURE

# Audit Your Data

© Databricks 2025. All rights reserved. Apache, Apache Spark, Spark, the Spark Logo, Apache Iceberg, Iceberg, and the Apache Iceberg logo are trademarks of the Apache Software Foundation.




- Unity Catalog now has extensive system tables that can help you answer a lot of important questions about your lakehouse, which makes it more transparent and easier to manage and monitor.
- These are found in the "system" catalog and starting with object metadata you may use "information_schema" schema for you to perform analytics on your operational data so you can answer questions around ownership, inventory, data access, tags and so on.

For example:

- What tables are in a specific catalog?
- who has access to a specific table?
- who did the last update to a gold table and when?
- Who has ownership of a particular table and more

System Tables - https://docs.databricks.com/en/admin/system-tables/index.html
Information Schema - https://docs.databricks.com/en/sql/language-manual/sql-ref-information-schema.html

# System Tables: Billing Logs

Understand cost allocation across your data estate

**What is the daily trend in DBU consumption?**

```
SELECT usage_date as `Date`, sum(usage_quantity) as `DBUs Consumed`
  FROM system.billing.usage
GROUP BY usage_date
ORDER BY usage_date ASC;
```

**How many DBUs of each SKU have been used so far this month?**

```
SELECT sku_name as `SKU`, sum(usage_quantity) as `DBUs`
  FROM system.billing.usage
WHERE
  month(usage_date) = month(CURRENT_DATE)
GROUP BY sku
ORDER BY `DBUs` DESC;
```

**Which 10 users consumed the most DBUs?**

```
SELECT identity_metadata.run_as as `User`,
sum(usage_quantity) as `DBUs`
  FROM system.billing.usage
GROUP BY identity_metadata.run_as
ORDER BY `DBUs` DESC
LIMIT 10;
```

**Which Jobs consumed the most DBUs?**

```
SELECT usage_metadata.job_id as `Job ID`,
sum(usage_quantity) as `DBUs`
  FROM system.billing.usage
GROUP BY `Job ID`;
```

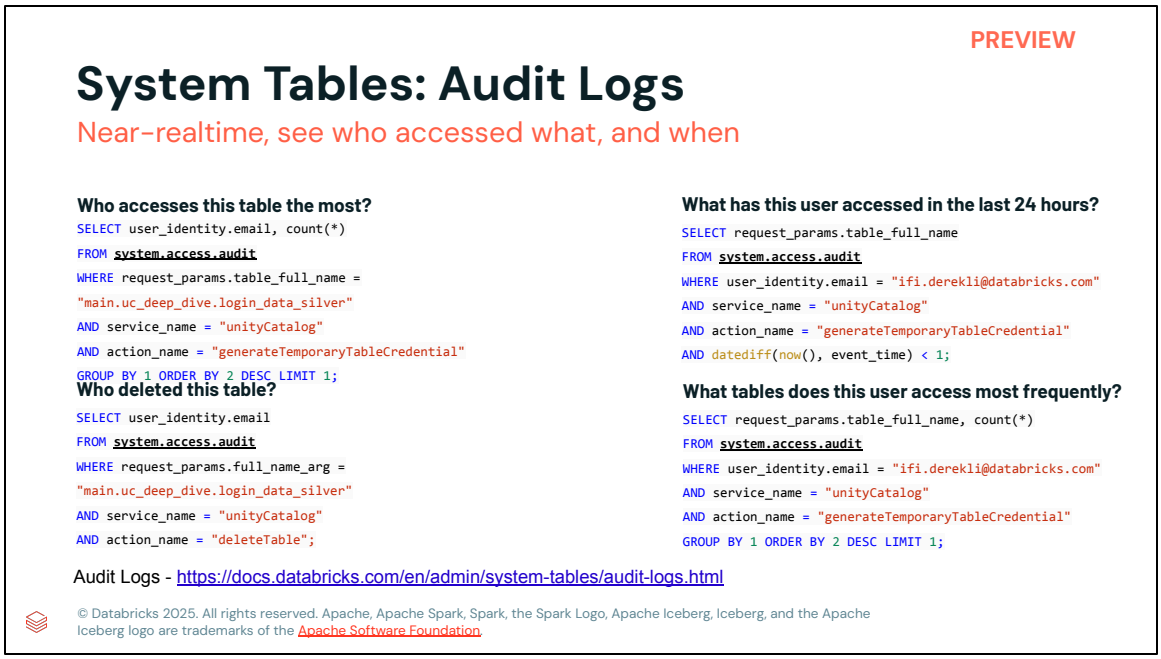
System Billing Schema - https://docs.databricks.com/en/admin/system-tables/billing.html

© Databricks 2025. All rights reserved. Apache, Apache Spark, Spark, the Spark Logo, Apache Iceberg, Iceberg, and the Apache Iceberg logo are trademarks of the Apache Software Foundation.

You can also leverage within the System catalog the Billing logs via the “billing” schema and the table “Usage” to be able to answer questions like:

- What is the daily trend in DBU Consumption?
- How many DBUs of each SKU have been used for this month?
- Which 10 users consumed the most DBUs?
- Which Jobs consumed the most DBUs?

System Billing Schema - https://docs.databricks.com/en/admin/system-tables/billing.html

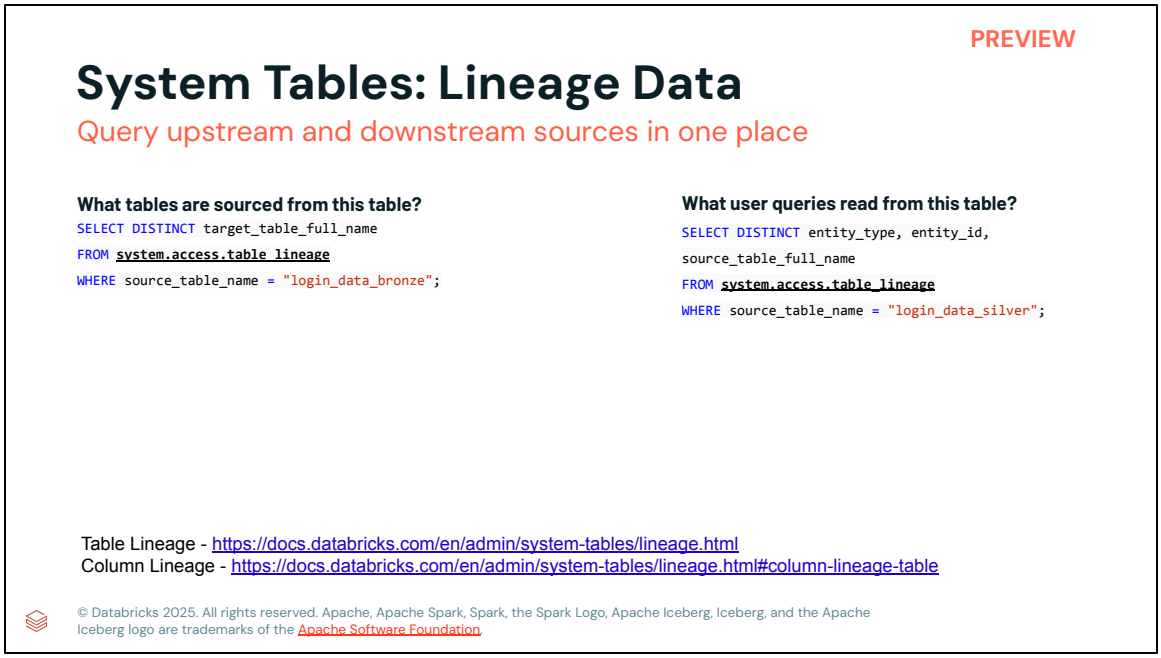



You can also leverage “Audit Logs” found in the “Access” schema and “Audit” table to answer questions like:

- Which table has the most access?
- Who deleted a specific table?
- What has a particular user accessed in the last 24 hours?
- What tables does a specific user access most frequently

Here are some samples, and be aware that this feature is still in Public Preview, so please check with your workspace administrator to see if this feature is enabled.

Audit Logs - https://docs.databricks.com/en/admin/system-tables/audit-logs.html




What about Lineage Data? Here, are two options: First on Table level and second in “Column” level, for that you can leverage the “Access” schema and “table_lineage” table to answer questions like:

- What tables are sourced from a specific table?
- What user queries read from a specific table?

This feature is in preview mode, so please check with your admin.

Table Lineage - https://docs.databricks.com/en/admin/system-tables/lineage.html
Column Lineage - https://docs.databricks.com/en/admin/system-tables/lineage.html#column-lineage-table