midhunmathew

Databricks System Tables

Databricks System Tables provide essential metadata and logs for managing jobs, clusters, billing, and data access within the Databricks platform. Here's a simplified, point-by-point overview of these tables and their real-world applications.

databricks

with the new
ER DIAGRAM

Get ready to dive deep into tables with:

- 🎯 Job Tracking, Execution, Data Lineage and Access
- 🎯 Query Execution History, Marketplace and API Events
- 🎯 Billing, Usage, Cluster and Compute Management
- 🎯 Sample Queries & Entity Relationship Diagram

LinkedIn /mathew-midhun

databricks

DATABRICKS SYSTEM TABLES

midhunmathew

What are system tables?

System tables are a Databricks-hosted analytical store of your account's operational data found in the system catalog. System tables can be used for historical observability across your account.

Requirements

- To access system tables, your workspace must be enabled for Unity Catalog.
- Must be in a supported region.

Enable / Disable/ List Schemas

CLI QUERIES

Bash

```
databricks system-schemas enable METASTORE_ID SCHEMA_NAME
```

Bash

```
databricks system-schemas disable METASTORE_ID SCHEMA_NAME
```

Bash

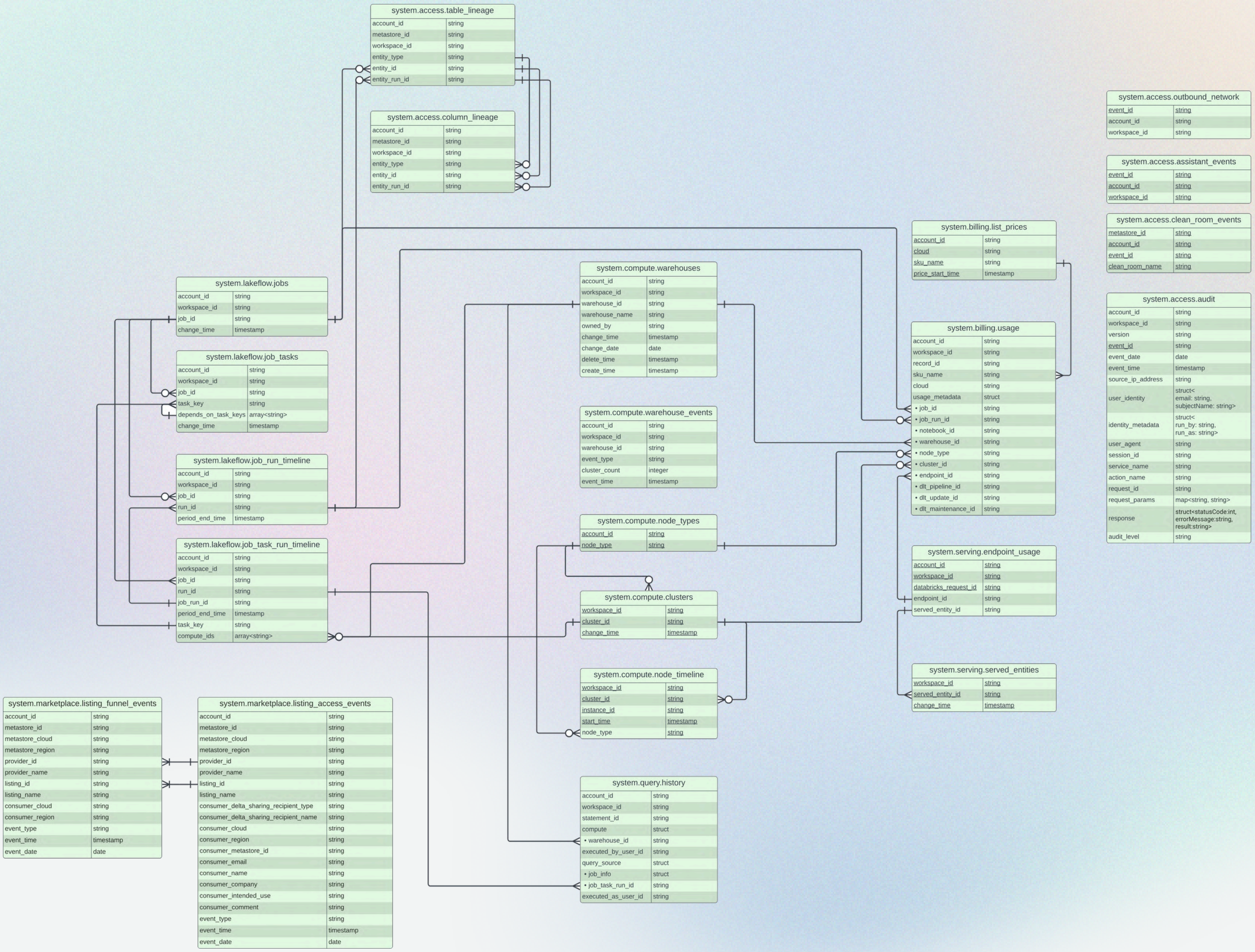
```
databricks system-schemas list METASTORE_ID
```

databricks

DATABRICKS SYSTEM TABLES

midhunmathew

Entity Relationship Diagram



LinkedIn /mathew-midhun

databricks

DATABRICKS SYSTEM TABLES

midhunmathew

Job Tracking and Execution

Analyze job performance, identify delays, and troubleshoot failed jobs.

TABLE	USAGE
system.lakeflow.jobs	Stores job metadata like **job_id** and **job_name**.
system.lakeflow.job_tasks	Tracks tasks within jobs.
system.lakeflow.job_run_timeline	Monitors when jobs start and finish.

Sample
QUERIES

Fetch job names and their recent execution status

```sql
SELECT job_id, job_name, storage, latest_run_id
FROM system.lakeflow.jobs
ORDER BY latest_run_id DESC
LIMIT 10;
```

Get a timeline of task runs from the past 7 days

```sql
SELECT task_key, status, start_time, end_time
FROM system.lakeflow.job_run_timeline
WHERE start_time >= CURRENT_TIMESTAMP() - INTERVAL '7 DAYS'
ORDER BY start_time DESC;
```

LinkedIn /mathew-midhun

databricks

DATABRICKS SYSTEM TABLES

midhunmathew

Cluster and Compute Management

Optimize cluster configurations and reduce idle time for cost savings

TABLE	USAGE
system.compute.clusters	Contains cluster-level details (**cluster_id**, **node_type**, **memory**).
system.compute.node_timeline	Tracks changes to cluster nodes over time.
system.compute.node_types	Lists available node types (useful for cost and resource planning).

Sample
QUERIES

Fetch cluster details, including cluster ID, memory, and node type

```sql
SELECT cluster_id, driver_memory_mb, worker_memory_mb, node_type
FROM system.compute.clusters
ORDER BY created_time DESC
LIMIT 10;
```

Get a list of node types and their details

```sql
SELECT node_type_id, num_cores, memory_mb
FROM system.compute.node_types
ORDER BY num_cores DESC;
```

LinkedIn /mathew-midhun

databricks

DATABRICKS SYSTEM TABLES

midhunmathew

Query Execution History

Track slow queries, troubleshoot query performance, and audit user activity

TABLE	USAGE
system.query.history	Stores executed query details such as **query_id**, **user_id**, and **warehouse_id**.

Sample
QUERIES

List the most recent queries executed, along with user information and status

```sql
SELECT query_id, user_id, warehouse_id, status, start_time
FROM system.query.history
ORDER BY start_time DESC
LIMIT 10;
```

LinkedIn /mathew-midhun

databricks

DATABRICKS SYSTEM TABLES

midhunmathew

Data Lineage and Access

Enhance data governance, ensure regulatory compliance, and track sensitive data

TABLE	USAGE
system.access.table_lineage	Tracks lineage of tables to identify data sources.
system.access.column_lineage	Tracks lineage of individual columns.
system.access.audit	Logs access events, including user IPs and workspace actions.

Sample
QUERIES

Get table lineage to see data sources used in transformations

```sql
SELECT entity_id, source_type, source_id
FROM system.access.table_lineage
ORDER BY source_type, entity_id;
```

Track user access events (IP, timestamp, and actions)

```sql
SELECT user_id, workspace_id, ip_address, access_time, operation
FROM system.access.audit
WHERE access_time >= CURRENT_TIMESTAMP() - INTERVAL '1 DAY'
ORDER BY access_time DESC;
```

LinkedIn /mathew-midhun

databricks

DATABRICKS SYSTEM TABLES

midhunmathew

Billing and Usage

Monitor and optimize cloud costs by analyzing real-time usage data.

TABLE	USAGE
system.billing.usage	Tracks cloud usage, including clusters and SQL warehouses.
system.billing.sku_prices	Stores SKU pricing details for cost tracking.

Sample
QUERIES

Get cluster usage information for the last month

```sql
SELECT cluster_id, warehouse_id, usage_date, total_dbu_consumed
FROM system.billing.usage
WHERE usage_date >= CURRENT_DATE - 30
ORDER BY usage_date DESC;
```

Fetch SKU pricing details

```sql
SELECT sku_name, unit_price_usd, change_time
FROM system.billing.sku_prices
ORDER BY change_time DESC;
```

LinkedIn /mathew-midhun

databricks

DATABRICKS SYSTEM TABLES

midhunmathew

Marketplace and API Events

Analyze marketplace service performance and track API activity.

TABLE	USAGE
system.marketplace.listing_funnel_events	Tracks usage for external marketplace integrations.
system.marketplace.listing_access_events	Logs API usage for marketplace services.

Sample
QUERIES

Get recent marketplace listing access events

```sql
SELECT consumer_cloud, event_time, provider_id, listing_id
FROM system.marketplace.listing_access_events
ORDER BY event_time DESC
LIMIT 10;
```

Analyze API usage trends based on funnel events

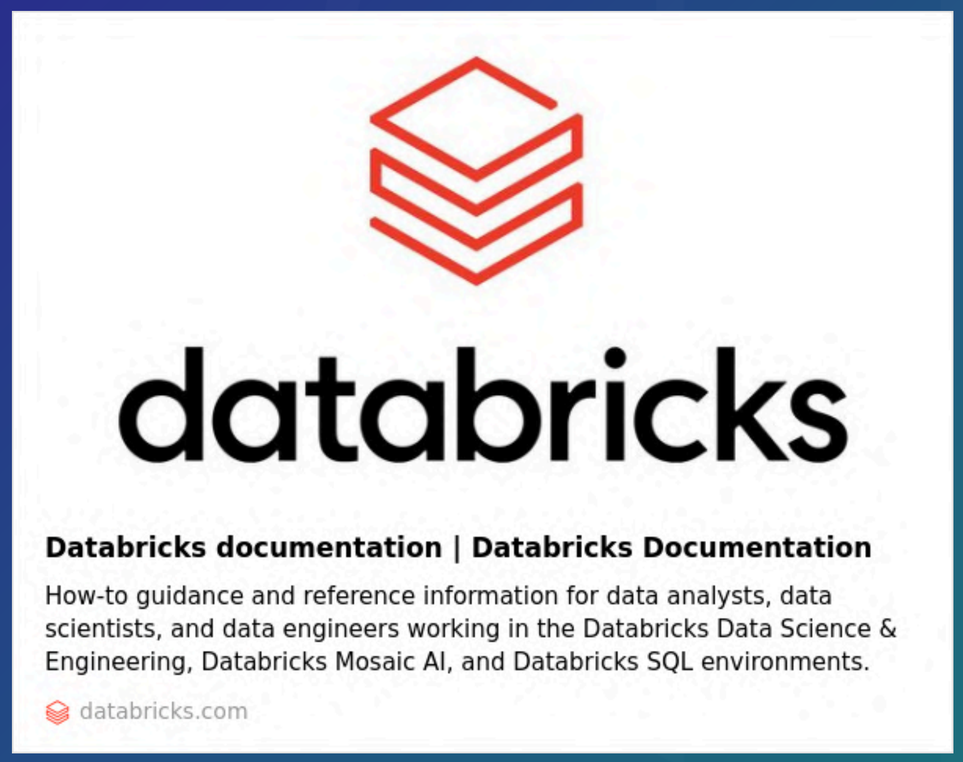
```sql
SELECT listing_id, event_type, COUNT(*) AS event_count
FROM system.marketplace.listing_funnel_events
GROUP BY listing_id, event_type
ORDER BY event_count DESC;
```

LinkedIn /mathew-midhun

midhunmathew

Your Databricks Journey Continues...

Stay ahead with databricks documentation & learning resources!



docs.databricks.com

learn.microsoft.com

/en-us/azure/databricks/admin/system-tables/

LinkedIn /mathew-midhun