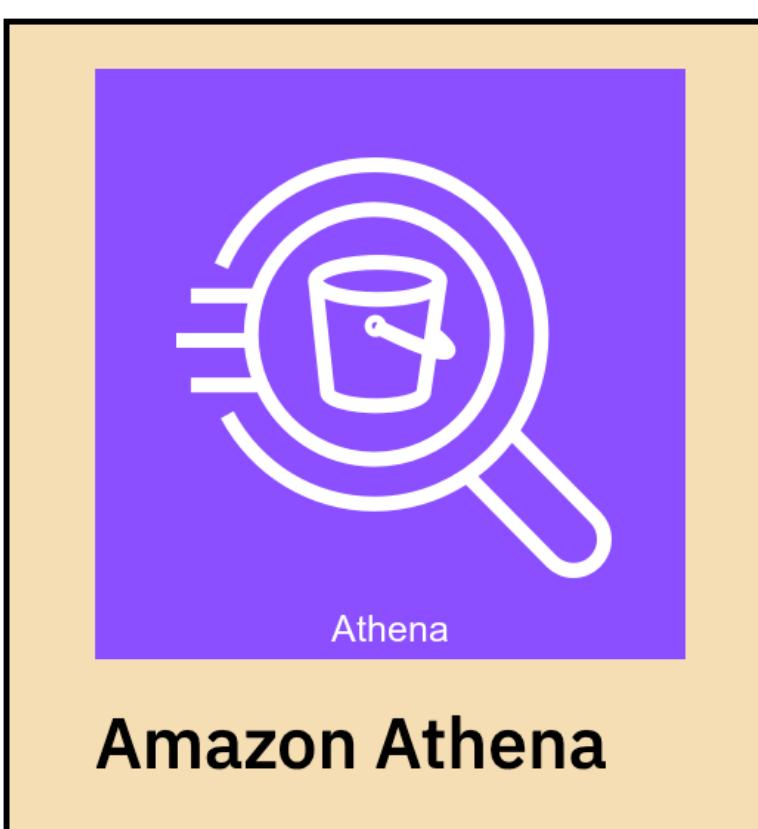


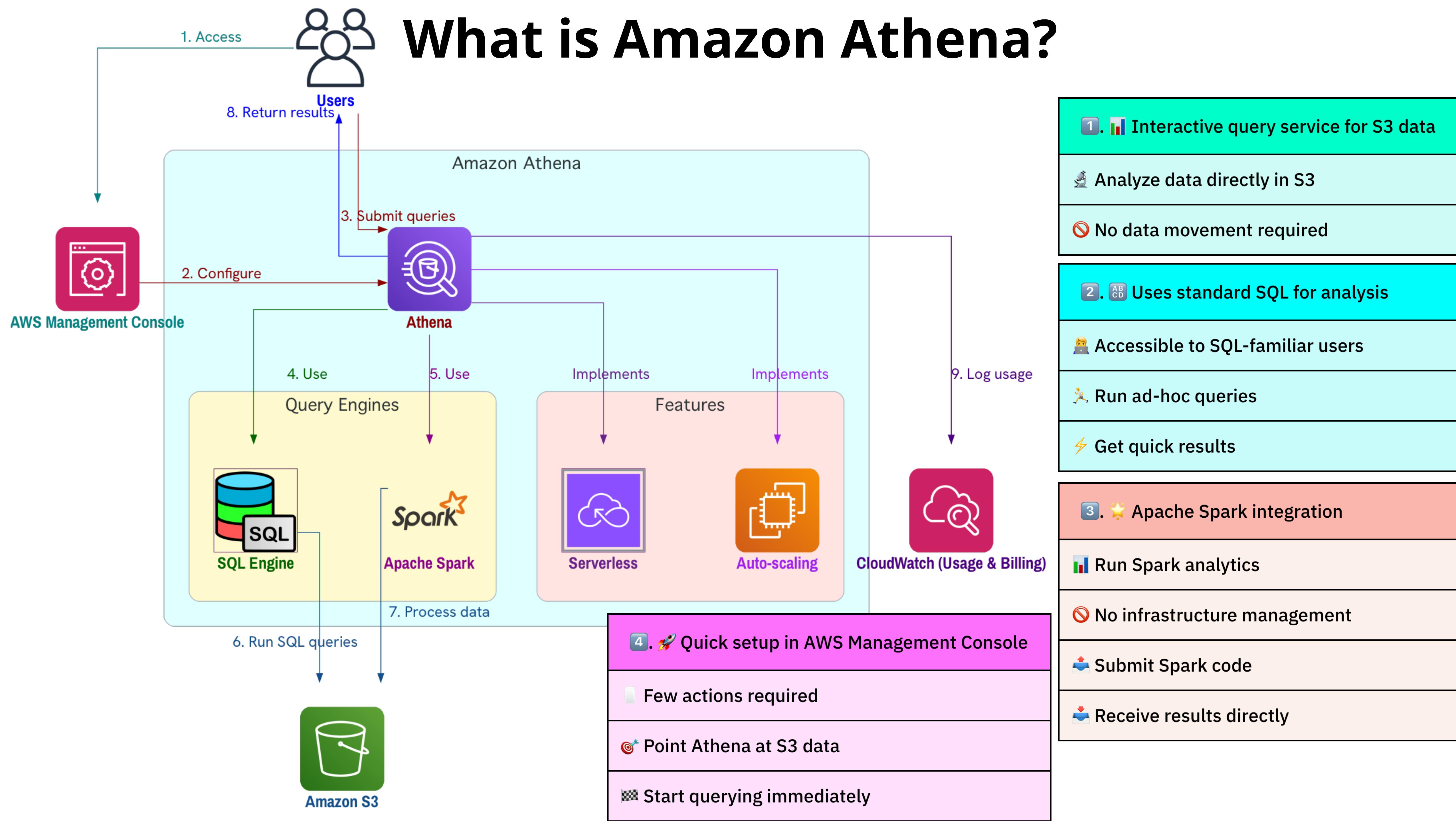
# Amazon Athena

# Table of Contents

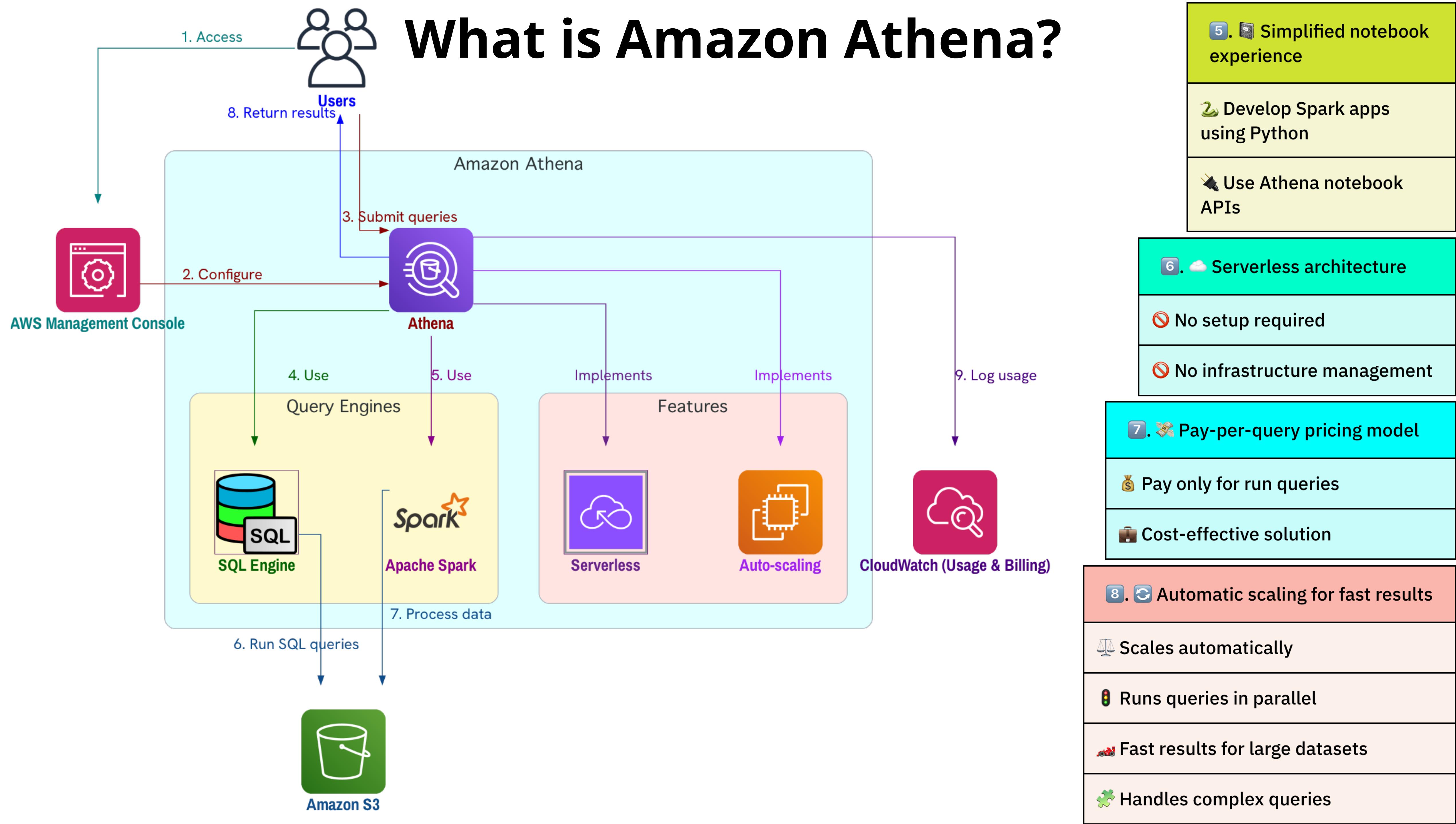


- 1. What is Amazon Athena?
- 2. When to Use Amazon Athena
- 3. When to Use Amazon EMR
- 4. When to use Amazon Redshift
- 5. AWS Service Integrations with Athena
- 6. Understanding Tables, Databases, and Data Catalogs in Athena
- 7. Table in Athena
- 8. Table Creation
- 9. AWS Glue Data Catalog
- 10. Example

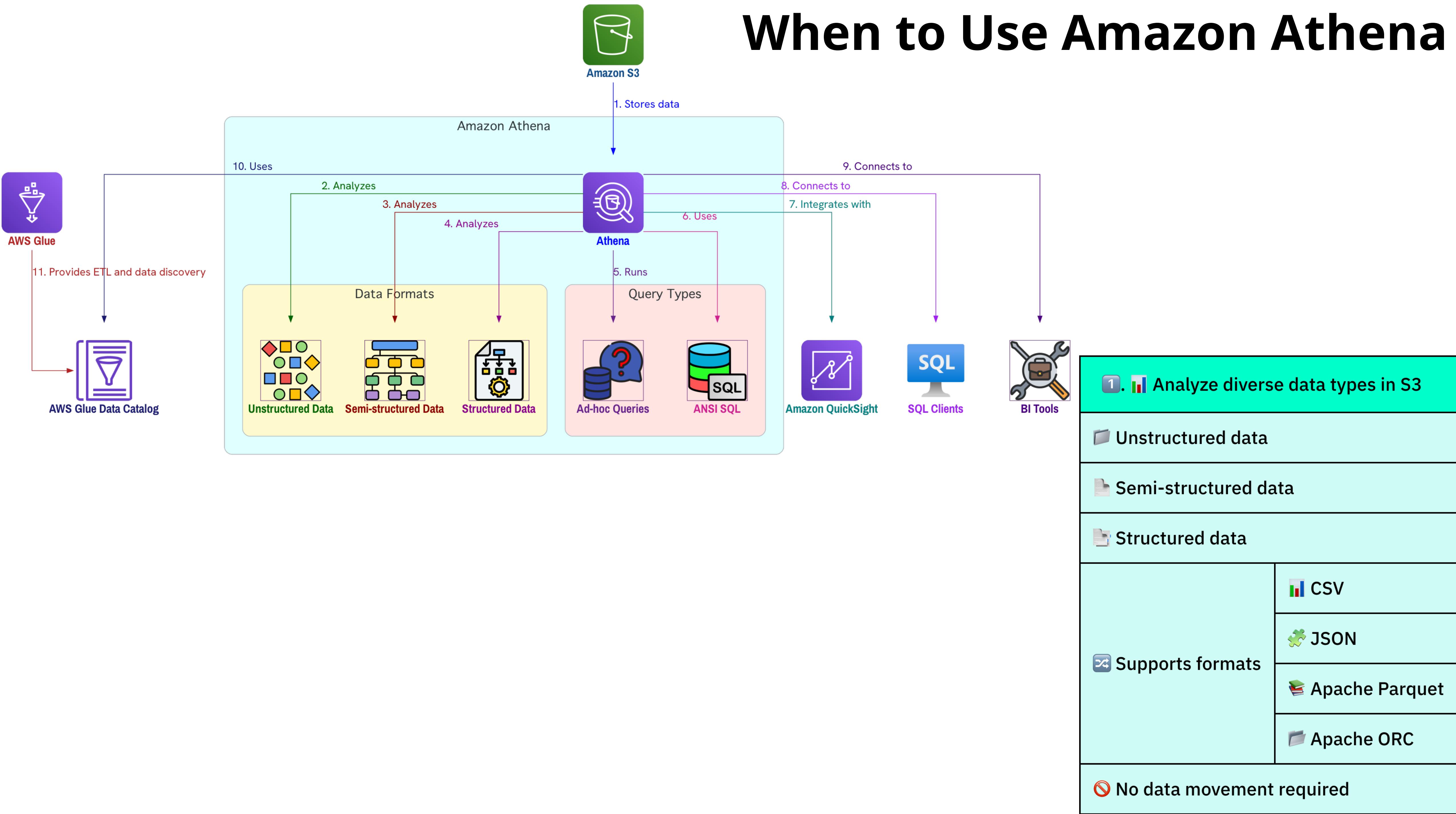
# What is Amazon Athena?



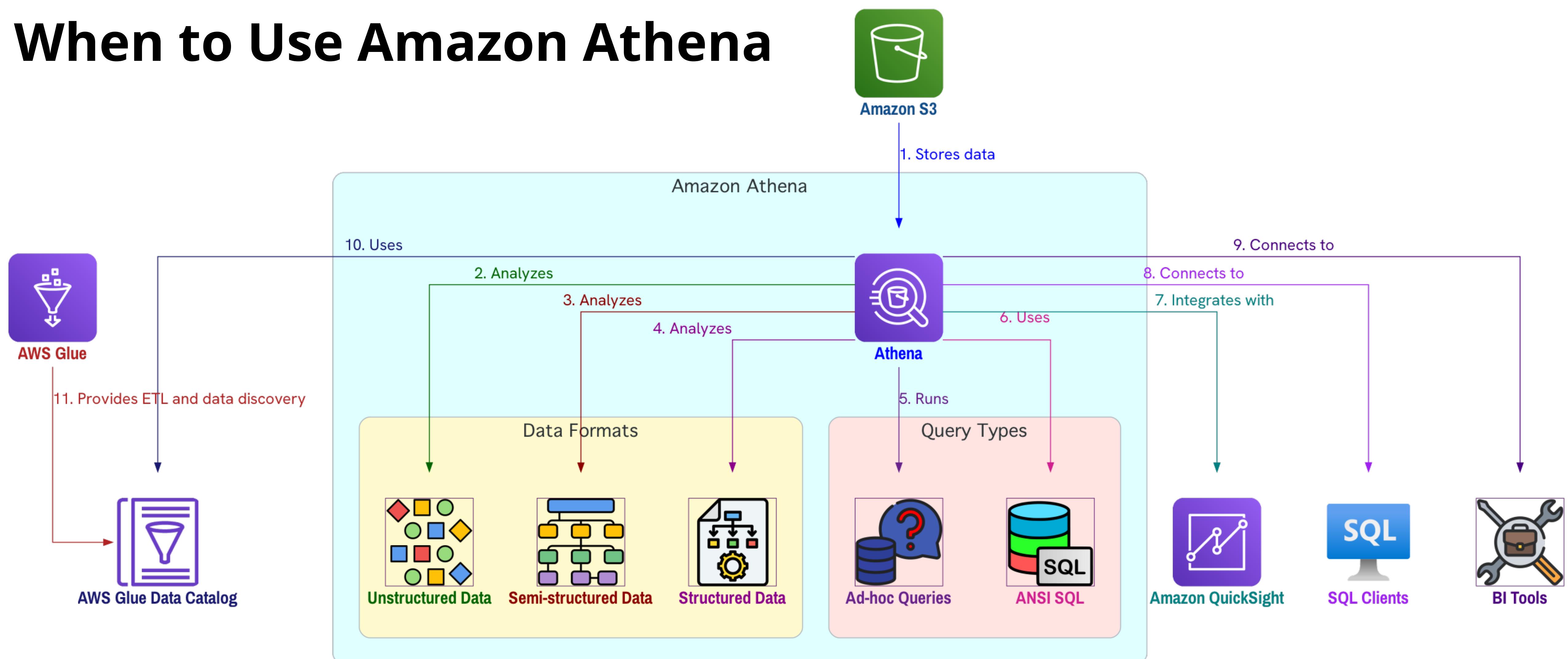
# What is Amazon Athena?



# When to Use Amazon Athena



# When to Use Amazon Athena



## 2. Execute ad-hoc ANSI SQL queries

Quick insights

🚫 No data aggregation

🚫 No data loading

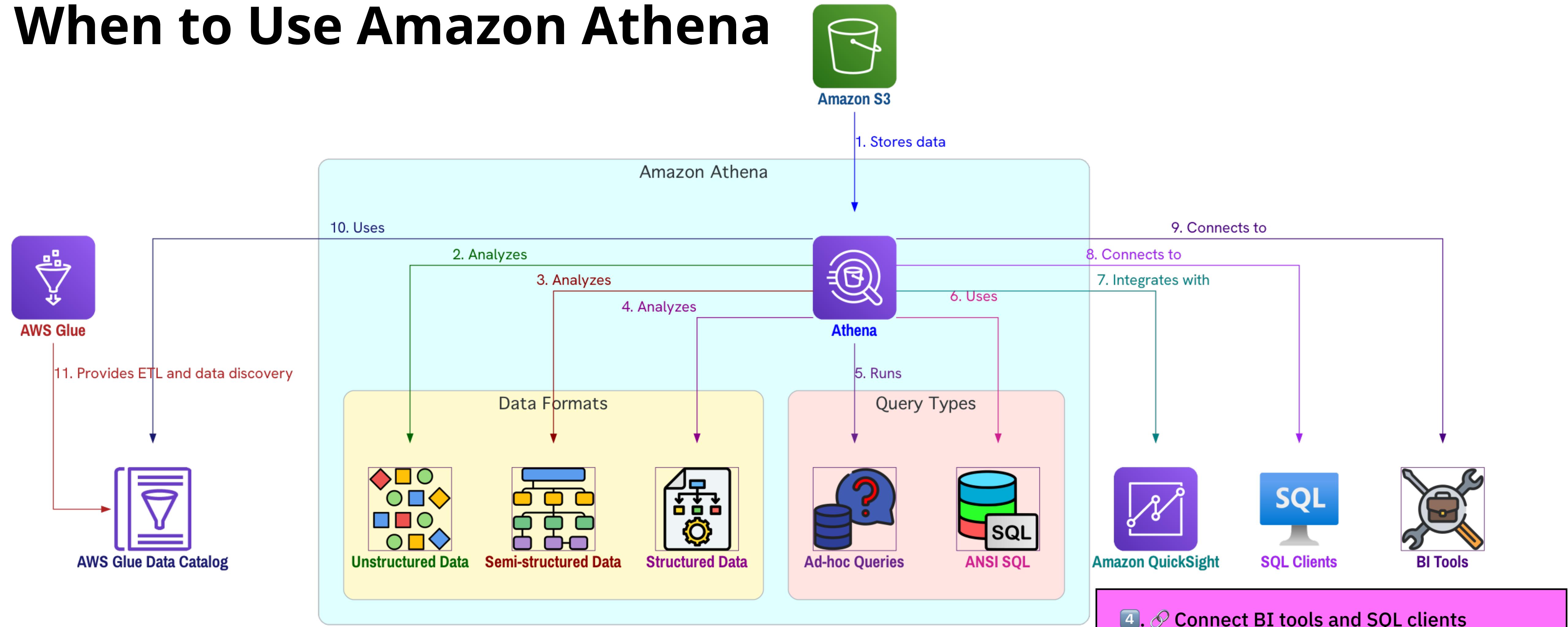
## 3. Visualize with QuickSight integration

🔗 Seamless integration

🎨 Transform query results

📊 Compelling visuals

# When to Use Amazon Athena



## 4. Connect BI tools and SQL clients

JDBC driver

ODBC driver

Detailed reports

Data exploration

# When to Use Amazon Athena



Amazon S3

1. Stores data

Amazon Athena

10. Uses

2. Analyzes

3. Analyzes

4. Analyzes

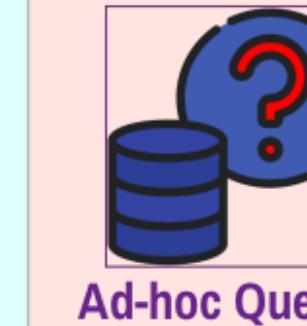


Athena

6. Uses

5. Runs

Query Types



Ad-hoc Queries

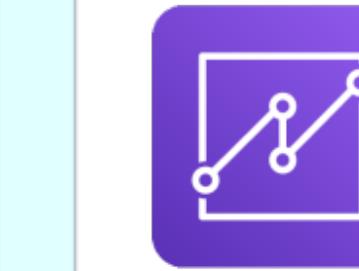


ANSI SQL

9. Connects to

8. Connects to

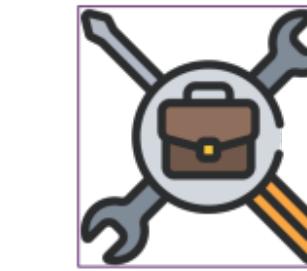
7. Integrates with



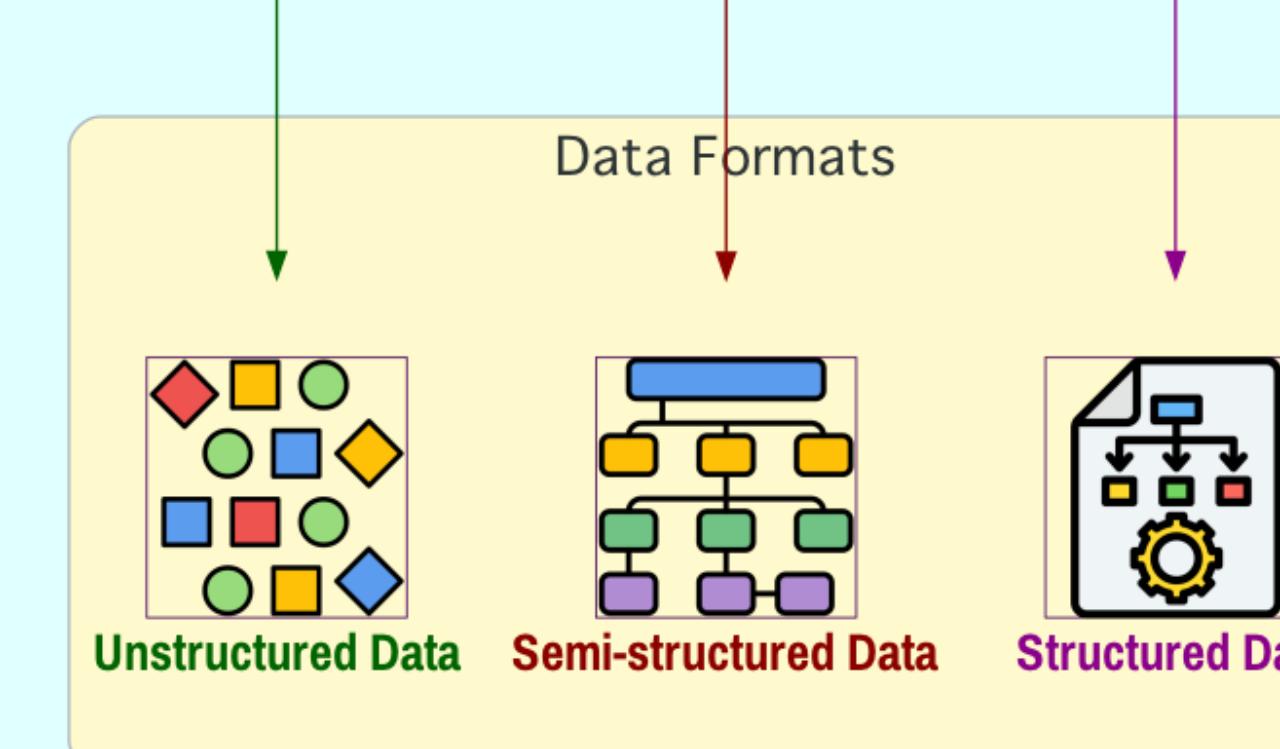
Amazon QuickSight



SQL Clients



BI Tools



Data Formats

Unstructured Data

Semi-structured Data

Structured Data

Query Types

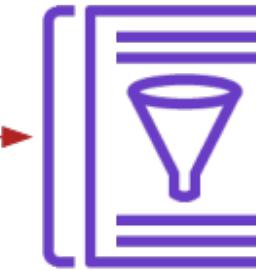
Ad-hoc Queries

ANSI SQL



AWS Glue

11. Provides ETL and data discovery



AWS Glue Data Catalog

10. Uses

## 5. Leverage AWS Glue Data Catalog

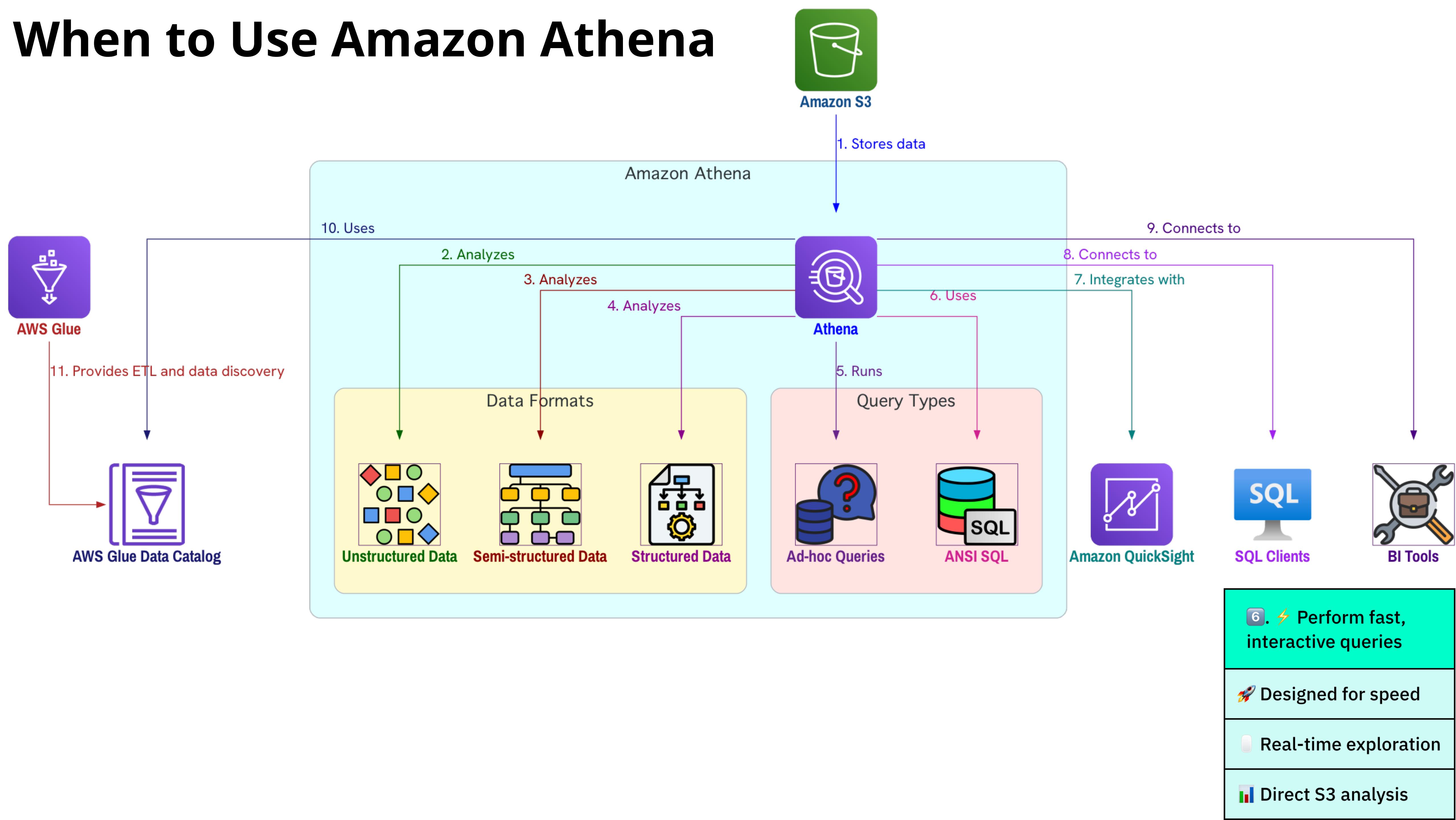
Persistent metadata store

Simplify data management

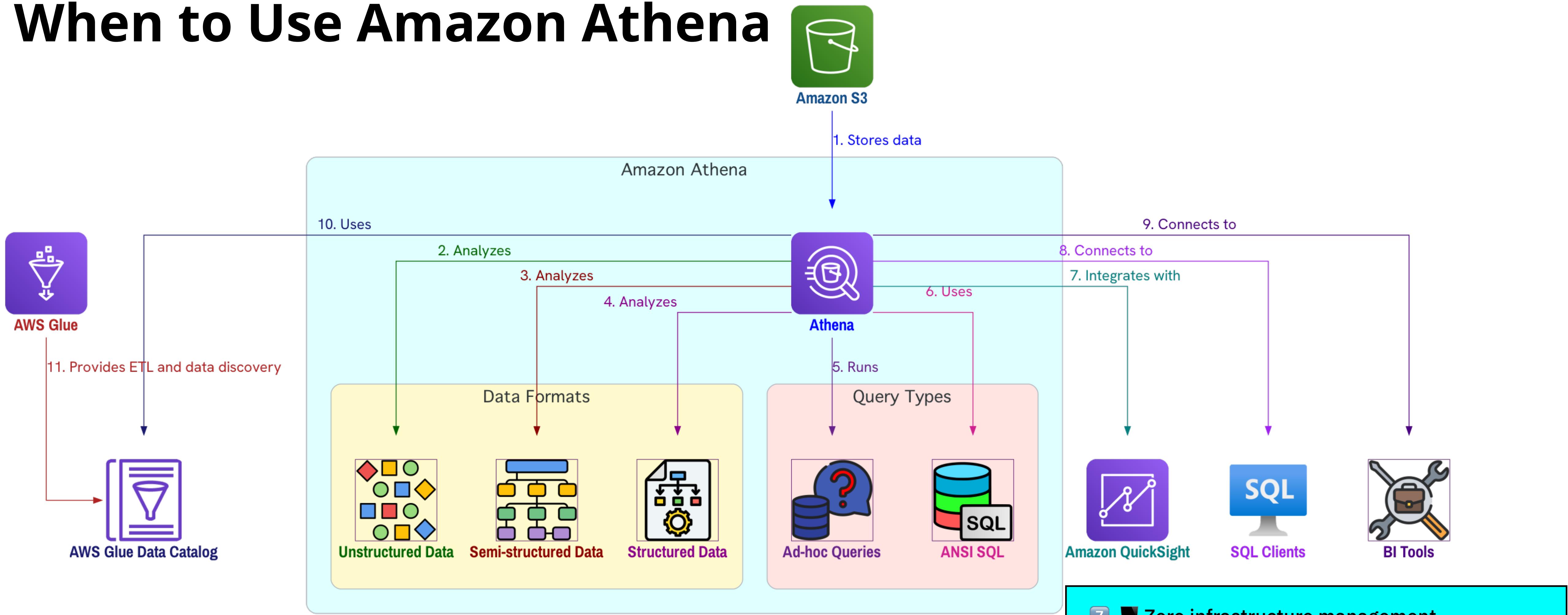
Enhance data discovery

AWS ecosystem integration

# When to Use Amazon Athena



# When to Use Amazon Athena



## 7. Zero infrastructure management

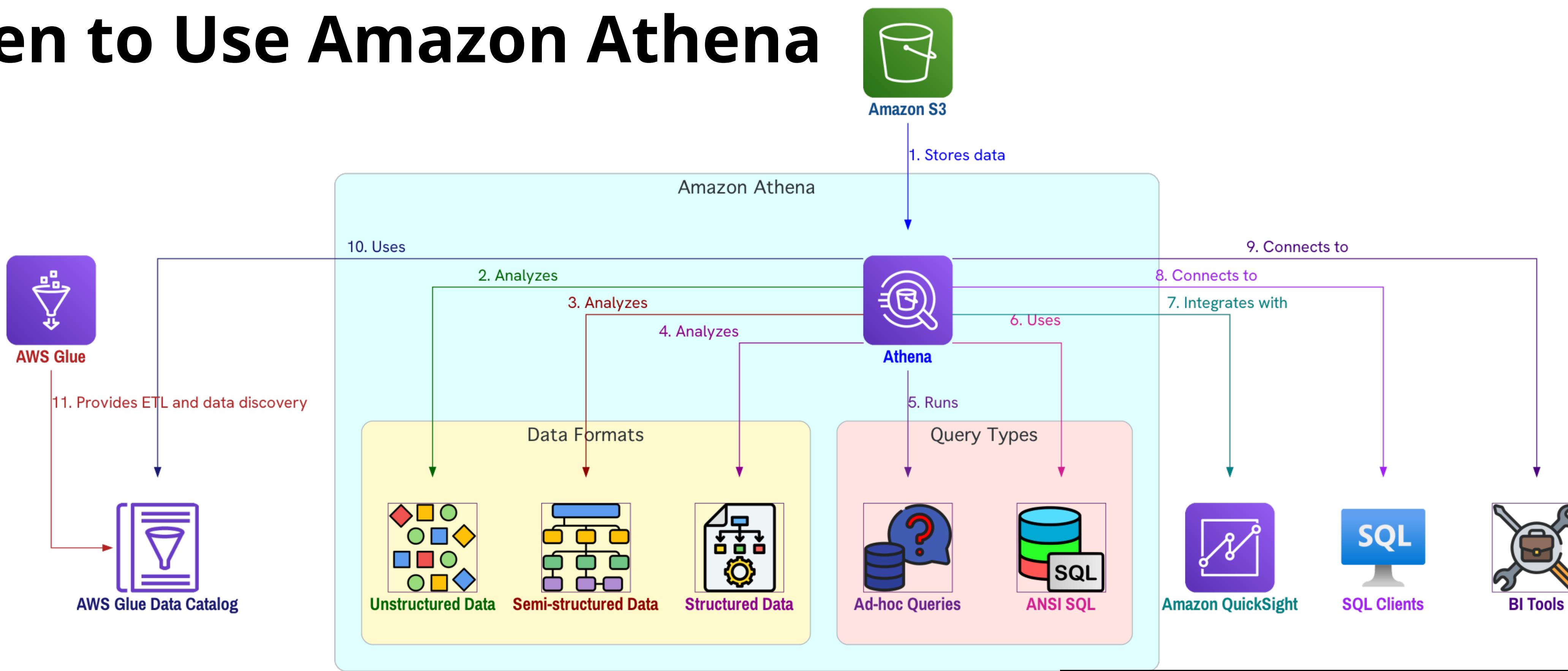
☁ Serverless nature

🔍 Ad hoc SQL queries

🚫 No cluster management

🔧 Simplified analysis process

# When to Use Amazon Athena



8. 💰 Pay only for queries you run

💰 Cost-effective pricing

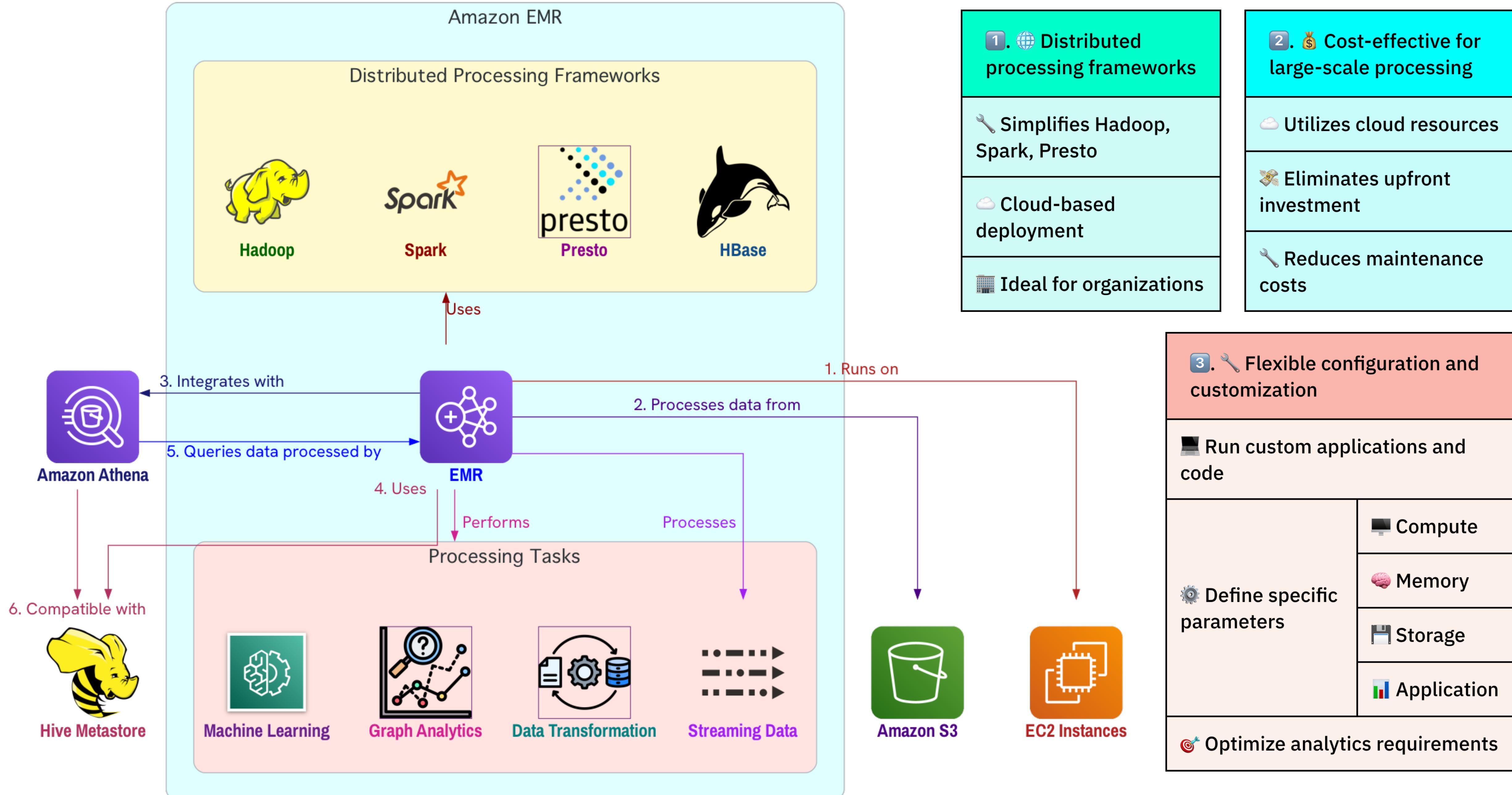
12 34 Pay per query

📊 Suitable for occasional use

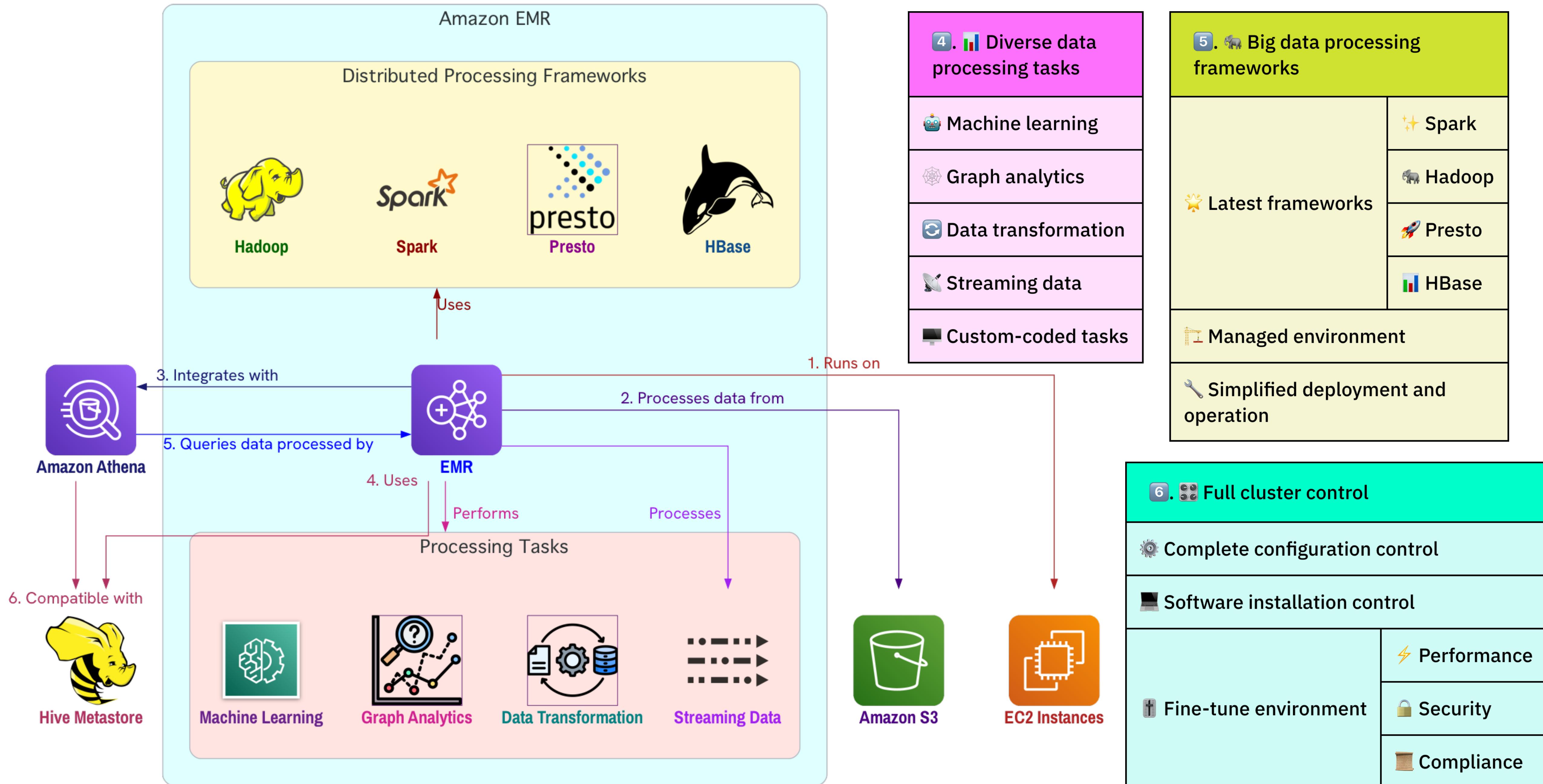
📈 Efficient for frequent analysis

🚫 No dedicated infrastructure

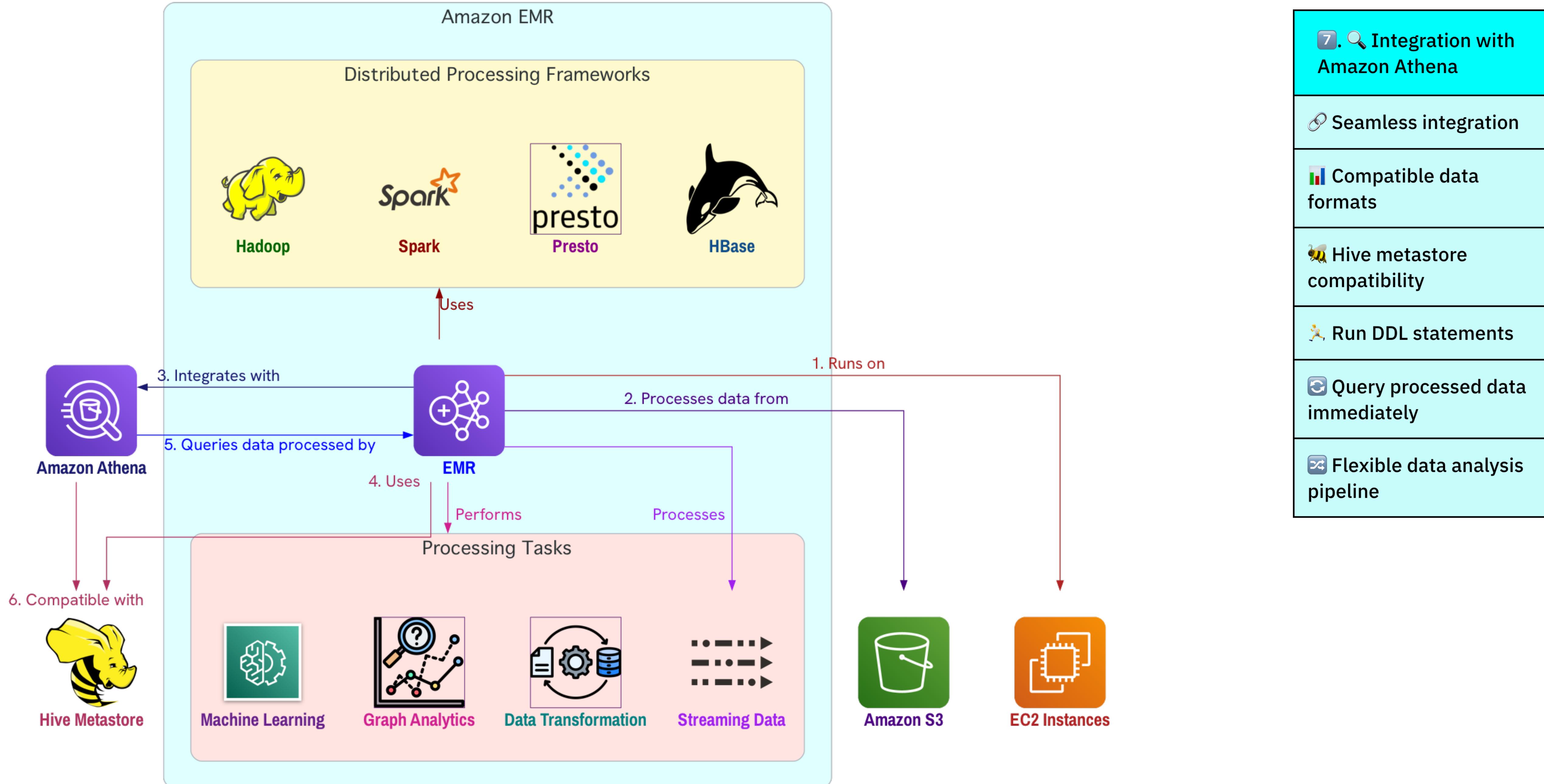
# When to Use Amazon EMR



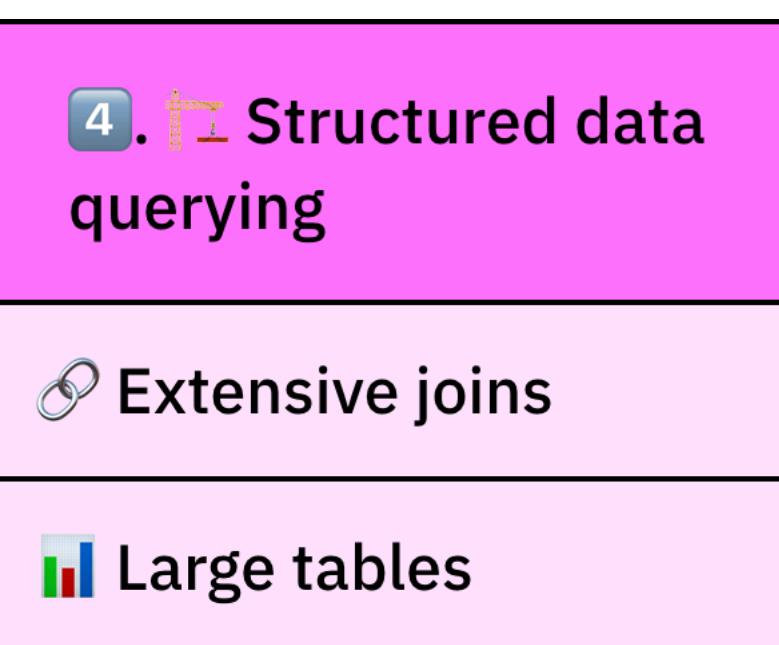
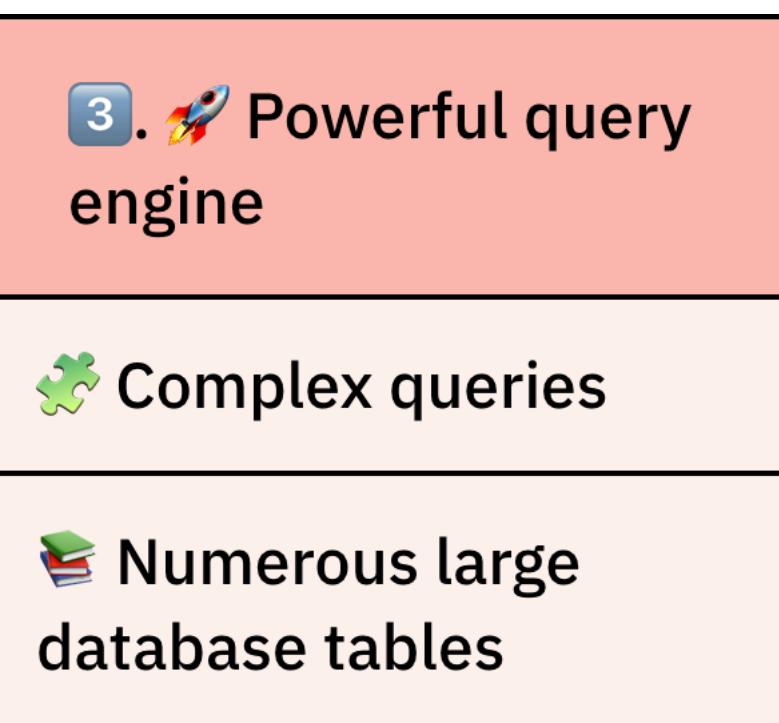
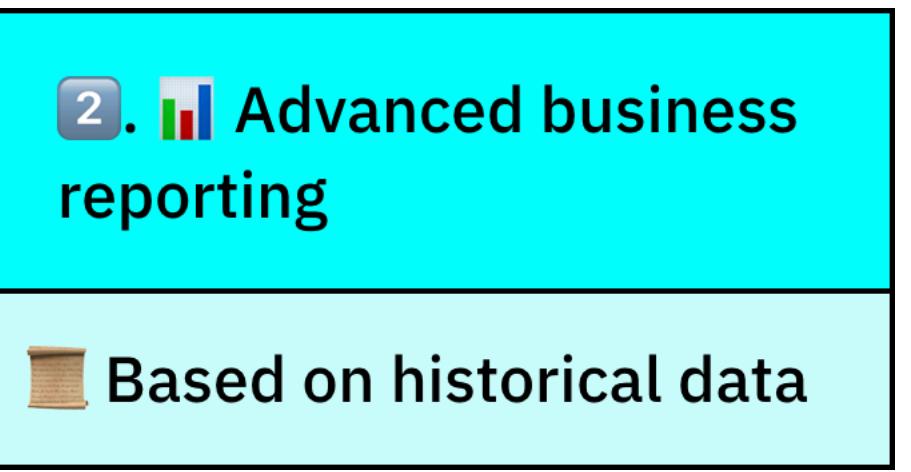
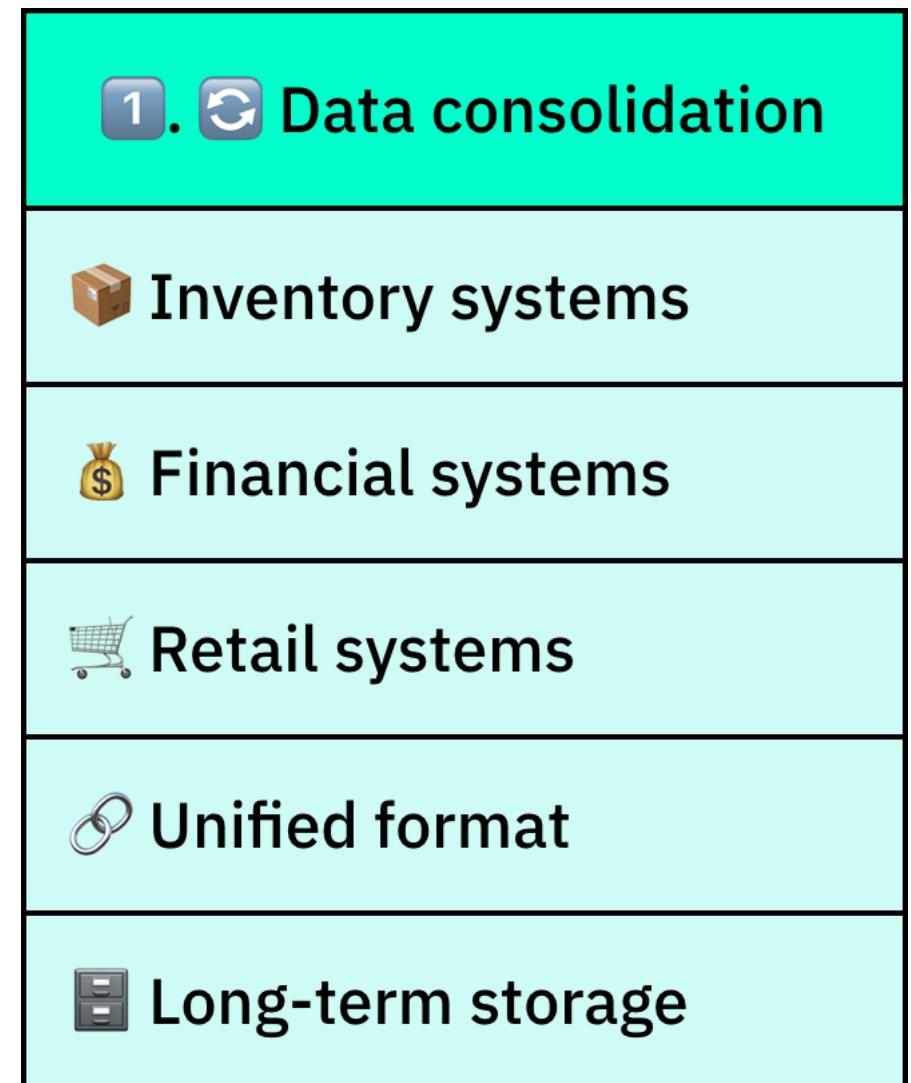
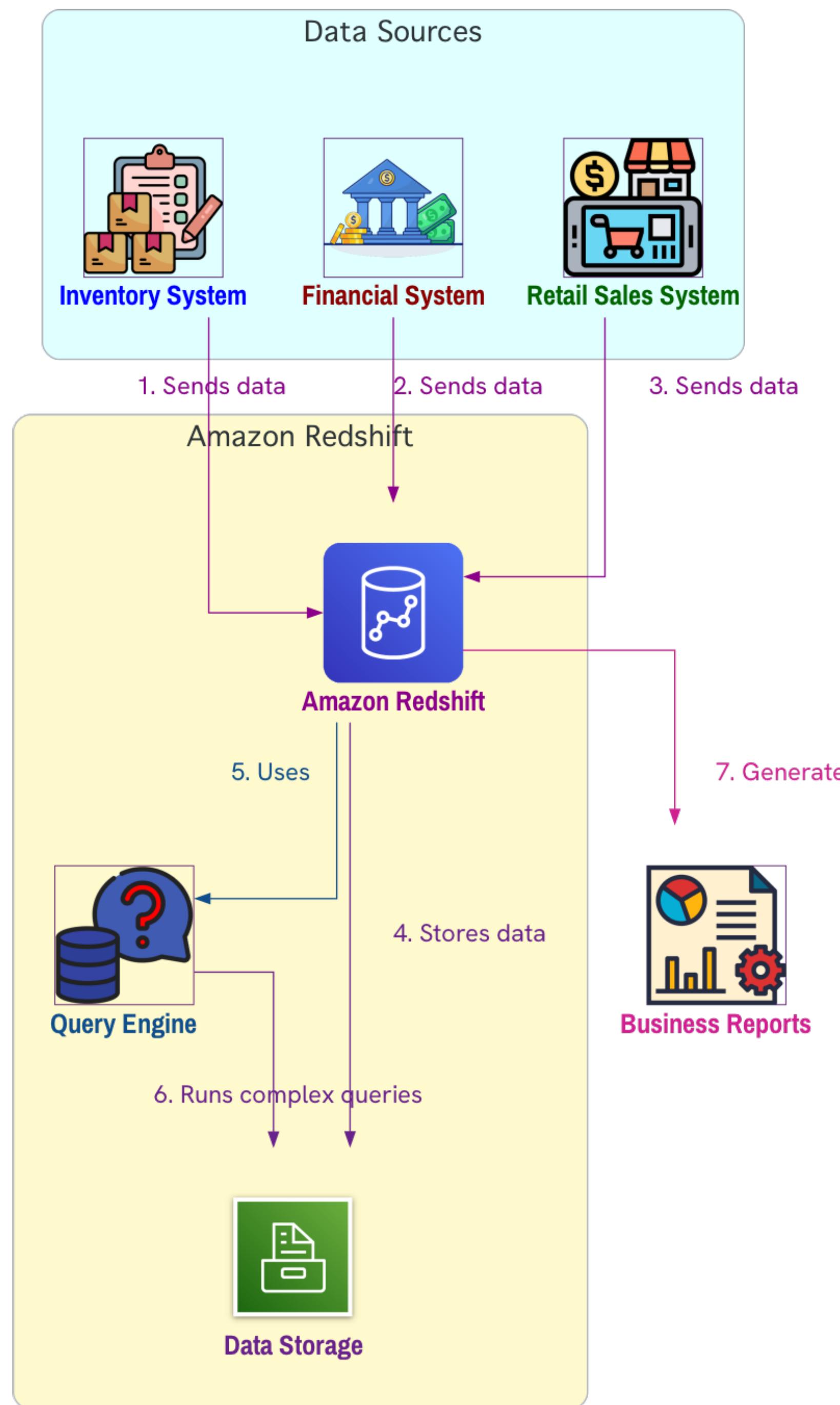
# When to Use Amazon EMR

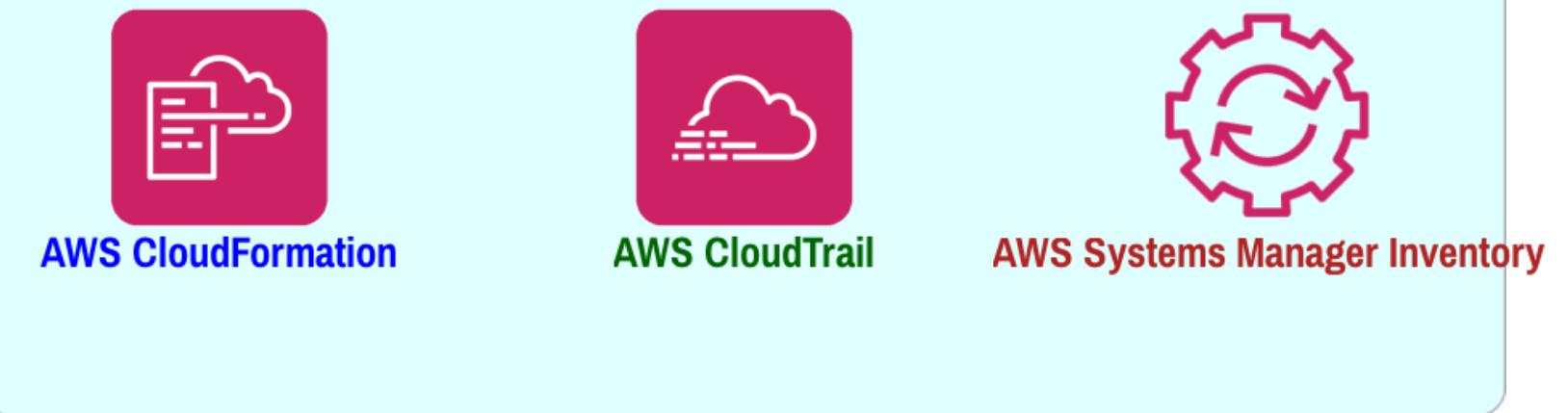


# When to Use Amazon EMR

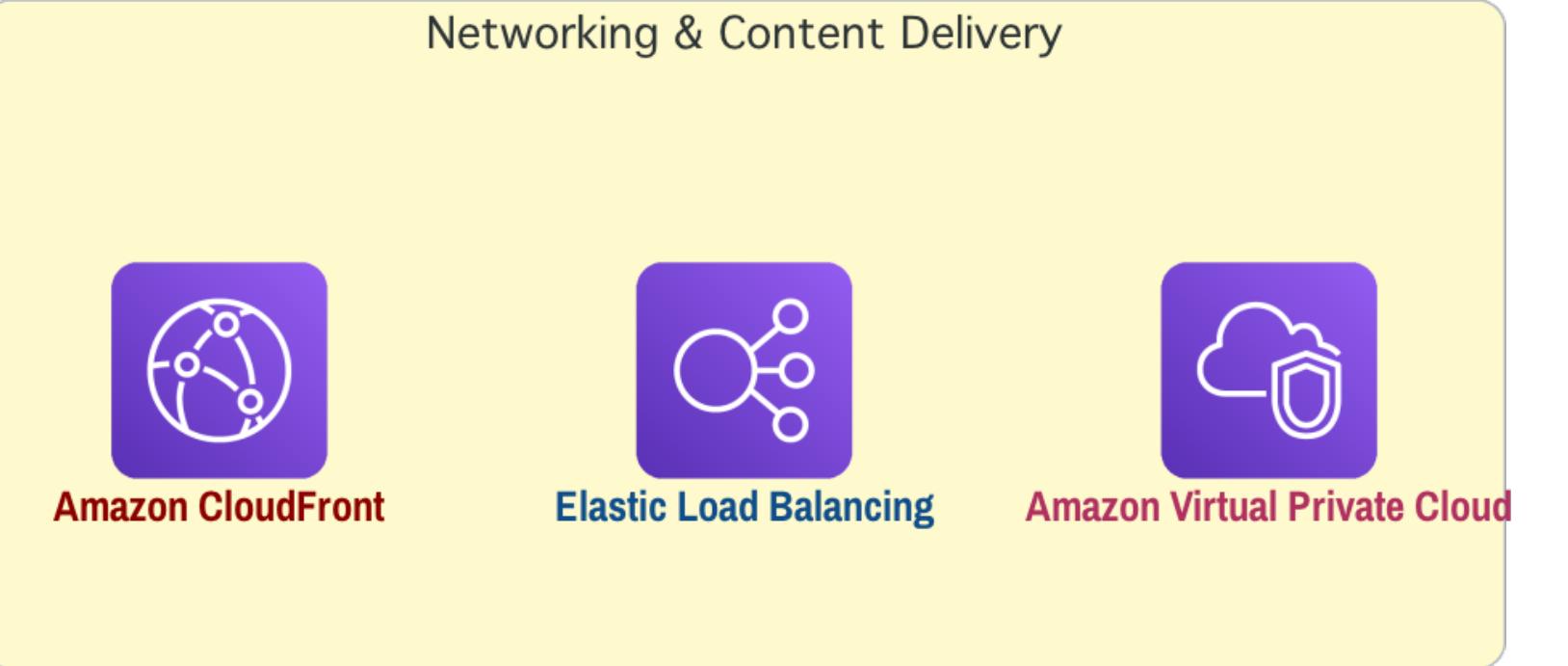


# When to use Amazon Redshift

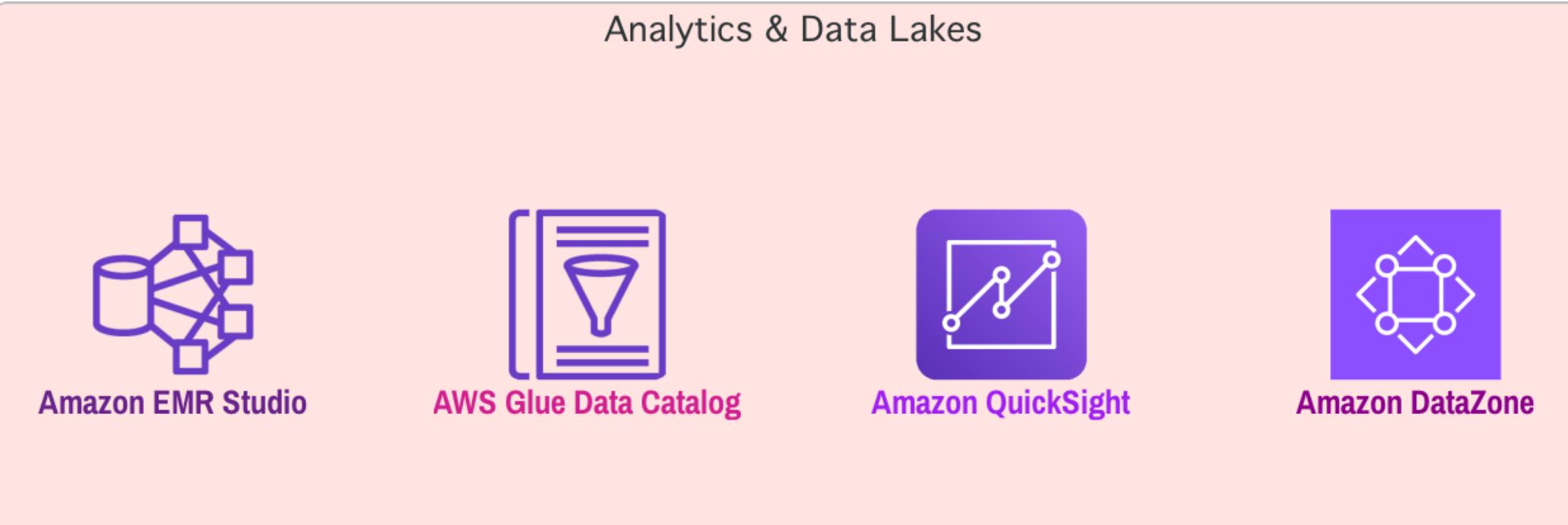




## Networking &amp; Content Delivery



## Analytics &amp; Data Lakes

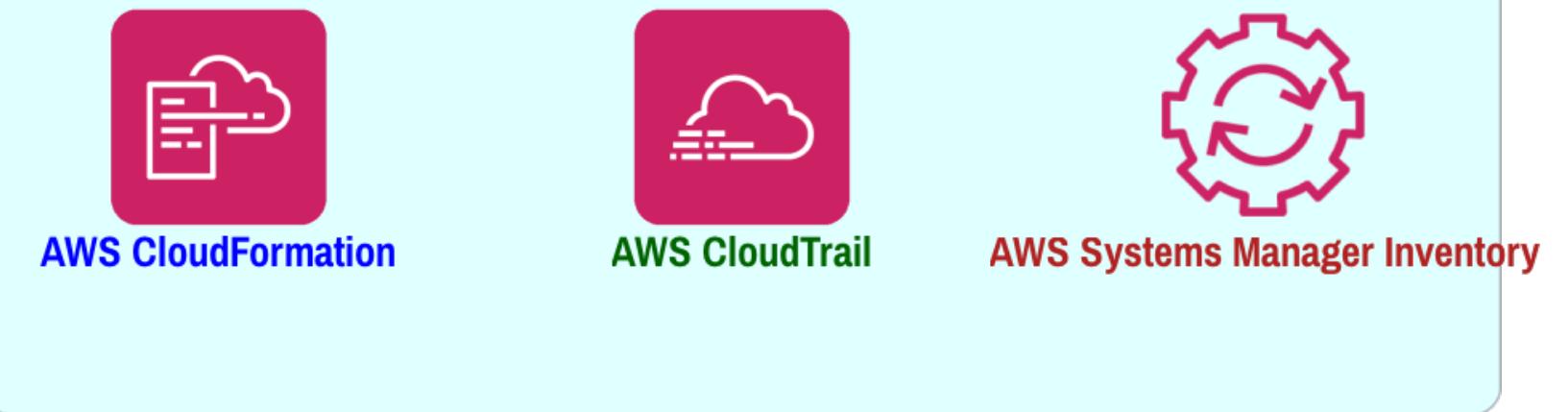


## Storage &amp; Security

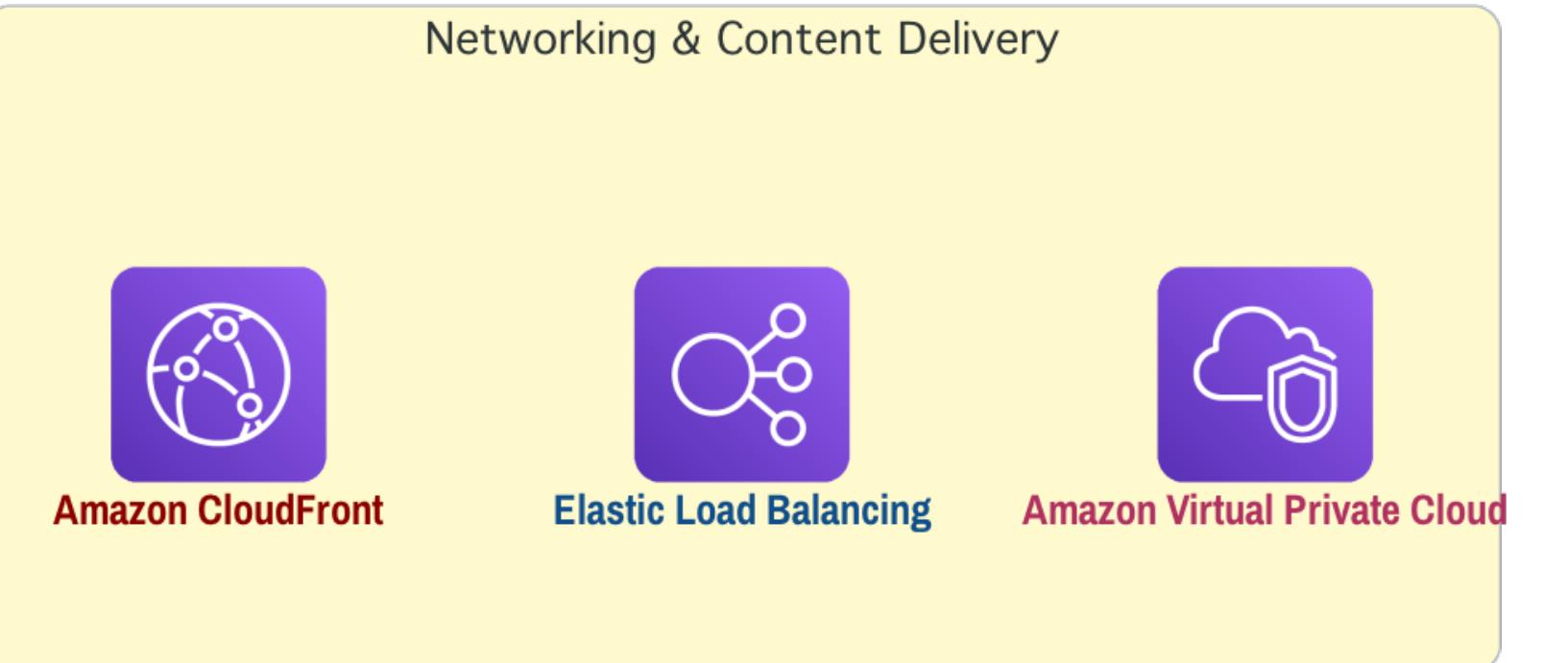


# AWS service integrations with Athena

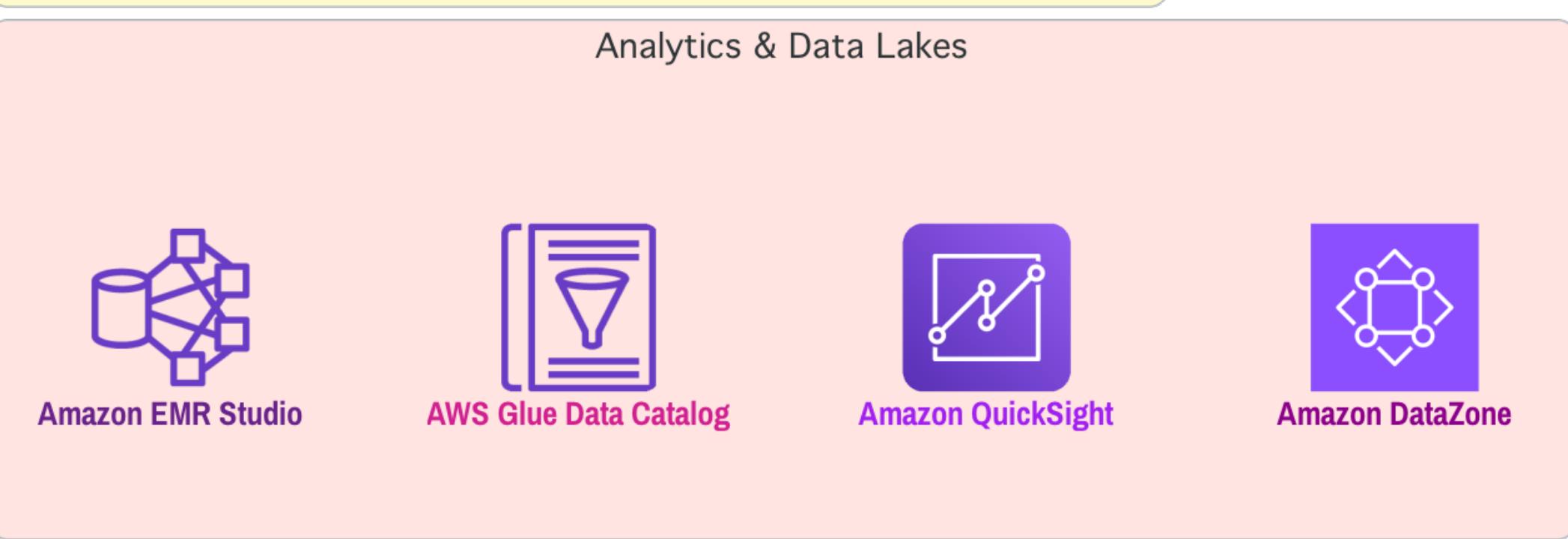
- 1 . 🚧 AWS CloudFormation:** Create Athena Objects, Infrastructure as code
- 2 . 🔎 Amazon CloudFront:** Query CloudFront logs, Analyze content delivery
- 3 . 🕵️ AWS CloudTrail:** Enhance AWS service activity analysis, Audit and security insights
- 4 . 🌐 Amazon DataZone:** Simplify AWS analytics services, Cross-organizational data sharing, Build business use case groupings
- 5 . ⚡ Elastic Load Balancing:** Query Application Load Balancer logs, Analyze traffic, latency, data transfer
- 6 . 🖥️ Amazon EMR:** Interactive query development, SQL analytics alongside other workloads
- 7 . 🔗 AWS Glue:** Metadata store for S3 data, Central metadata management, ETL and data discovery integration



## Networking &amp; Content Delivery



## Analytics &amp; Data Lakes



## Storage &amp; Security



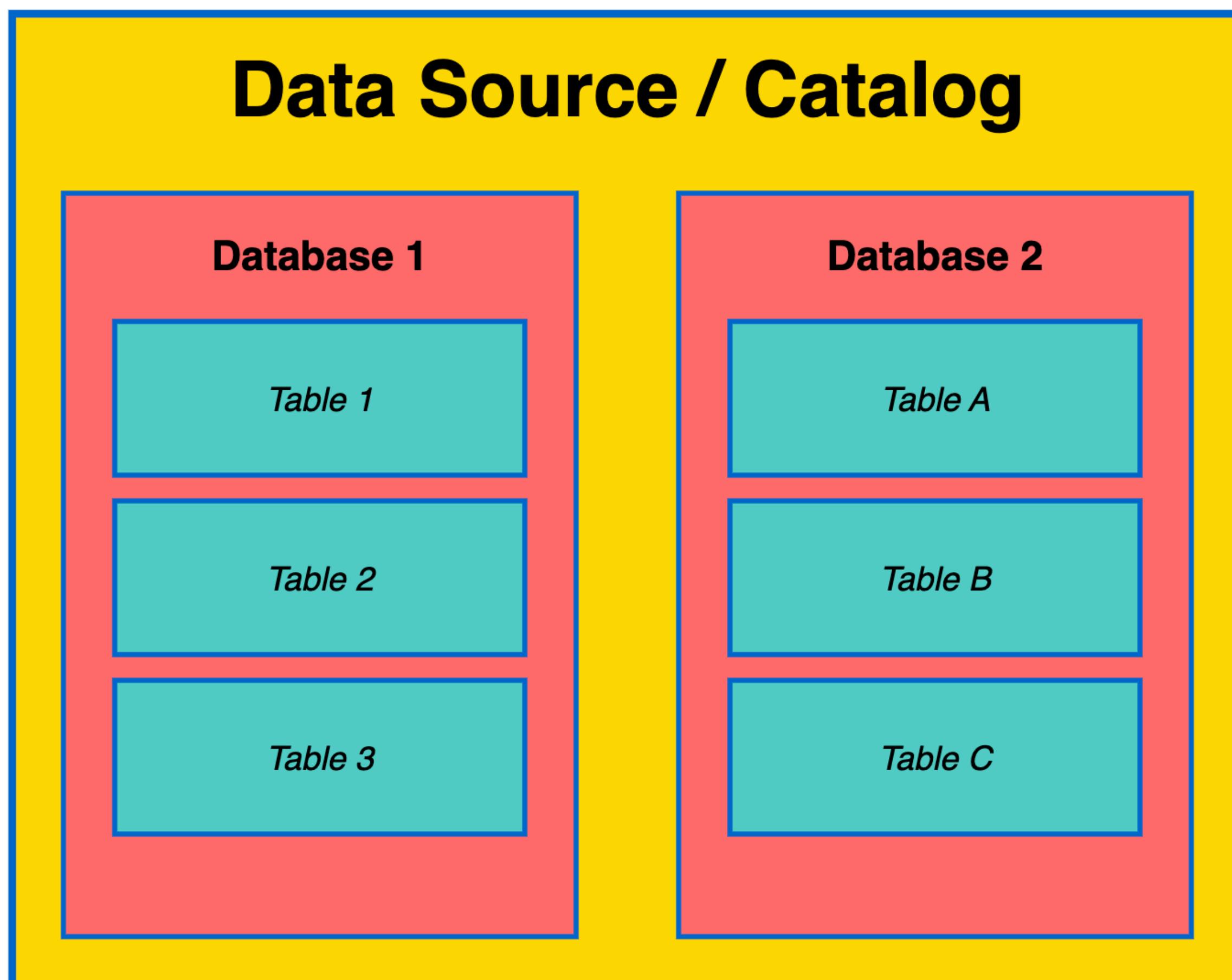
# AWS service integrations with Athena

- 8 . 🔒 **AWS IAM:** 🔑 Athena API actions in IAM policies, 🛡️ Secure access control
- 9 . 📈 **Amazon QuickSight:** 🖥️ Data visualization and reporting, 🔎 BI tool integration, ⚡ JDBC/ODBC connectivity
- 10 . 📁 **Amazon S3 Inventory:** 📊 Query S3 inventory with SQL, 🔒 Audit replication and encryption, 📋 Compliance reporting
- 11 . 🔄 **AWS Step Functions:** 🛡️ Manage Athena operations, 🔎 Control query execution, 🗓️ Schedule queries, 🌈 Data lake interaction
- 12 . 📊 **AWS Systems Manager Inventory:** 🌎 Query across regions and accounts, 🔎 Inventory data analysis
- 13 . 🌐 **Amazon VPC:** 🕵️ Analyze VPC flow logs, 🛡️ Detect threats and risks, 📊 Network traffic pattern analysis

# Understanding Tables, Databases, and Data Catalogs in Athena

---

In Athena, **catalogs**, **databases**, and **tables** are containers for the metadata definitions that define a schema for underlying source data.



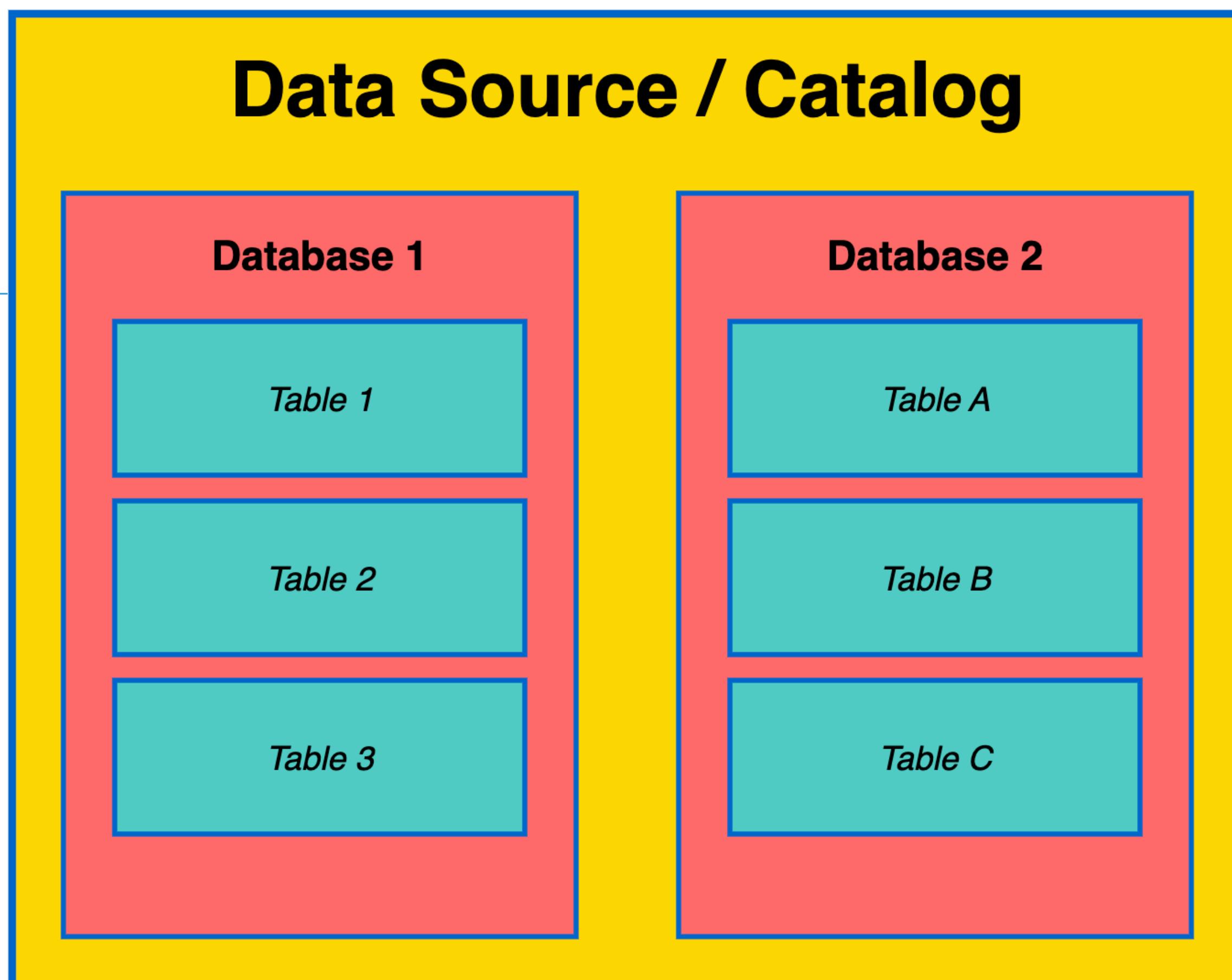
# Understanding Tables, Databases, and Data Catalogs in Athena

In Athena, **catalogs**, **databases**, and **tables** are containers for the metadata definitions that define a schema for underlying source data.

## Hierarchy of Data Objects

Athena uses the following terms to refer to hierarchies of data objects:

1. **Data source** (also referred to as a catalog): A group of databases
2. **Database** (sometimes called a schema): A group of tables
3. **Table**: Data organized as a group of rows or columns



# Tables in Athena

For each dataset you want to query in Athena, a table must exist. The table metadata includes:

- Location of data in Amazon S3
- Structure of the data (e.g., column names, data types)
- Table name

Databases are a logical grouping of tables and hold only metadata and schema information for a dataset.

## Table Metadata



Location of data in Amazon S3



Structure of the data



Table name

## Database in Athena



Logical Grouping



Table 1



Table 2



Table 3



Table 4



Metadata & Schema Only

# Tables in Athena

For each dataset you want to query in Athena, a table must exist. The table metadata includes:

- Location of data in Amazon S3
- Structure of the data (e.g., column names, data types)
- Table name

Databases are a logical grouping of tables and hold only metadata and schema information for a dataset.

## Creating Tables

Before querying data, a table must be registered in Athena. This can be done:

1. Automatically
2. Manually

### Table Creation Methods



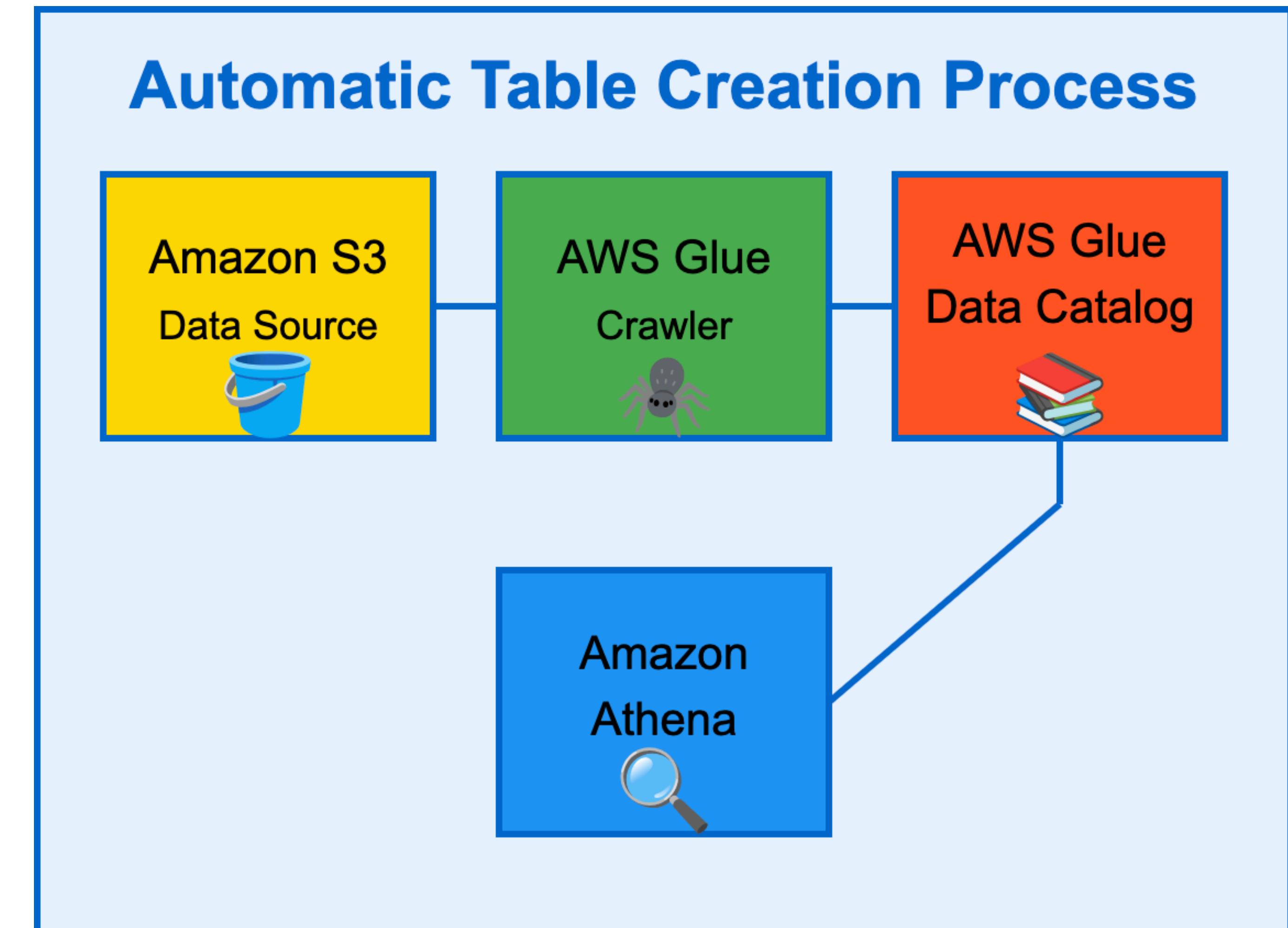
Automatic



Manual

# Automatic Table Creation

You can create a table automatically using an **AWS Glue crawler**. When AWS Glue creates a table, it registers it in its own AWS Glue Data Catalog.



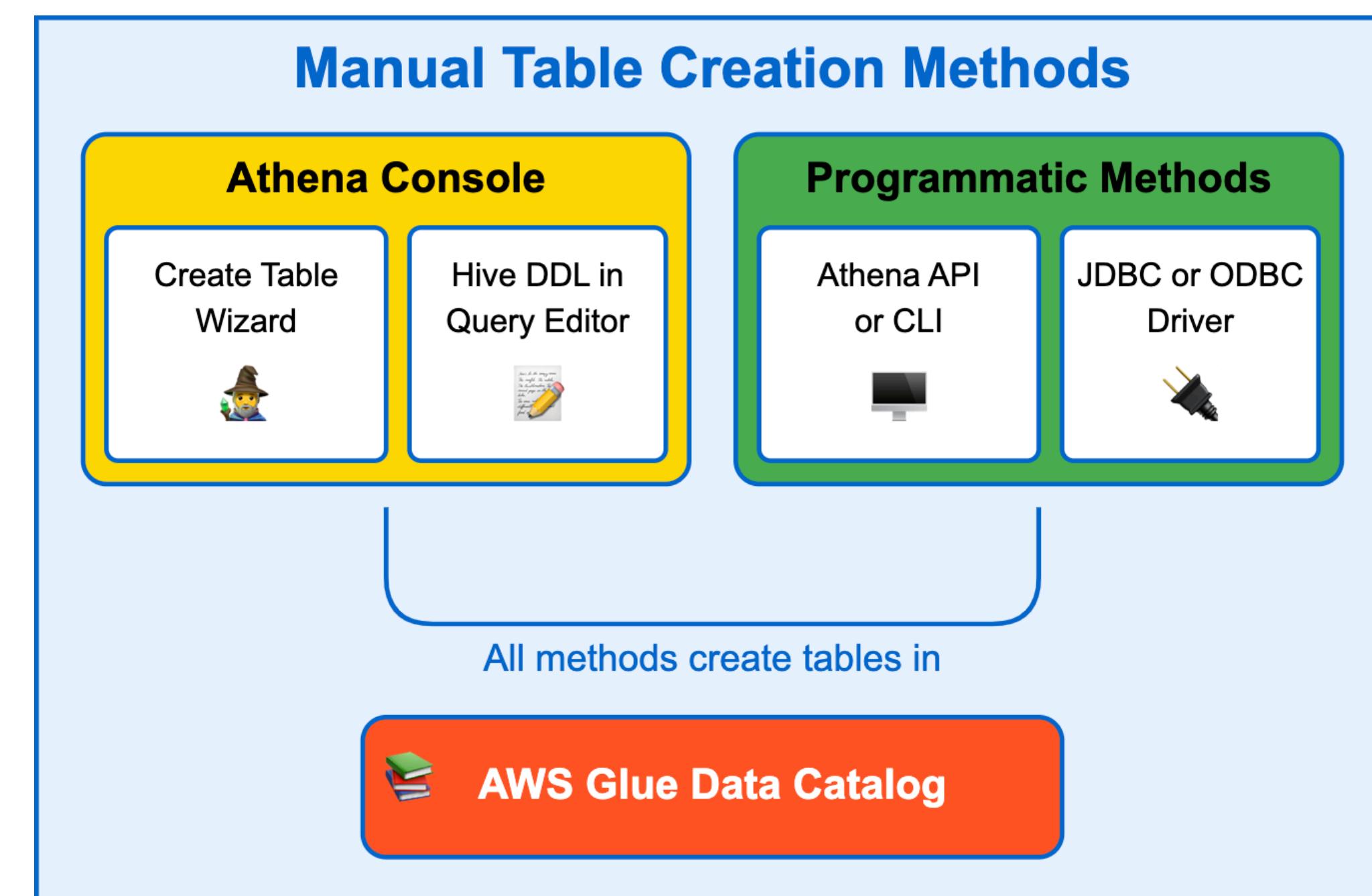
# Automatic Table Creation

You can create a table automatically using an **AWS Glue crawler**. When AWS Glue creates a table, it registers it in its own AWS Glue Data Catalog.

## Manual Table Creation

To create a table manually, you can:

- Use the Athena console to run the Create Table Wizard
- Use the Athena console to write Hive DDL statements in the Query Editor
- Use the Athena API or CLI to run a SQL query string with DDL statements
- Use the Athena JDBC or ODBC driver



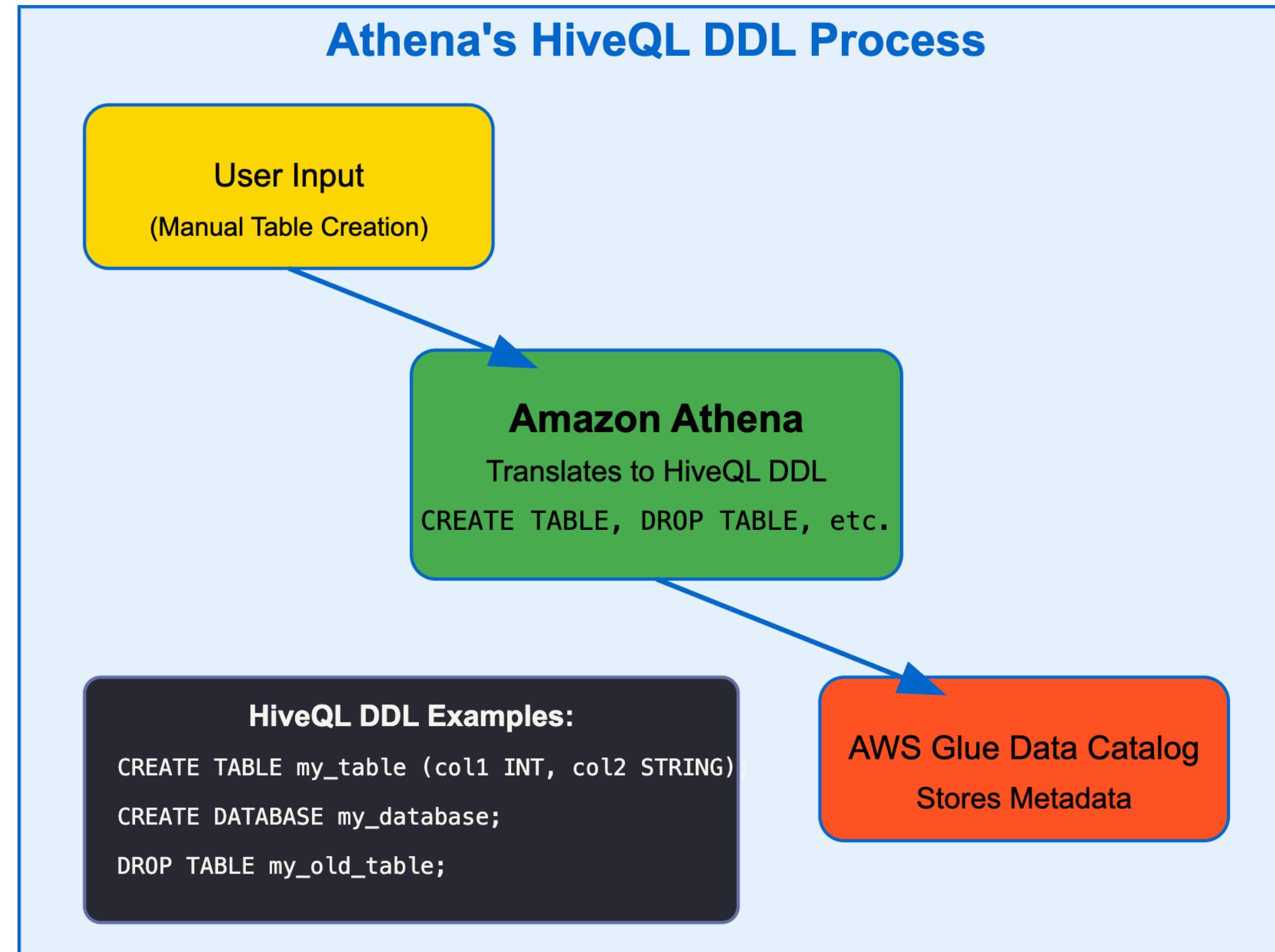
# Automatic Table Creation

You can create a table automatically using an **AWS Glue crawler**. When AWS Glue creates a table, it registers it in its own AWS Glue Data Catalog.

## Manual Table Creation

To create a table manually, you can:

- Use the Athena console to run the Create Table Wizard
- Use the Athena console to write Hive DDL statements in the Query Editor
- Use the Athena API or CLI to run a SQL query string with DDL statements
- Use the Athena JDBC or ODBC driver



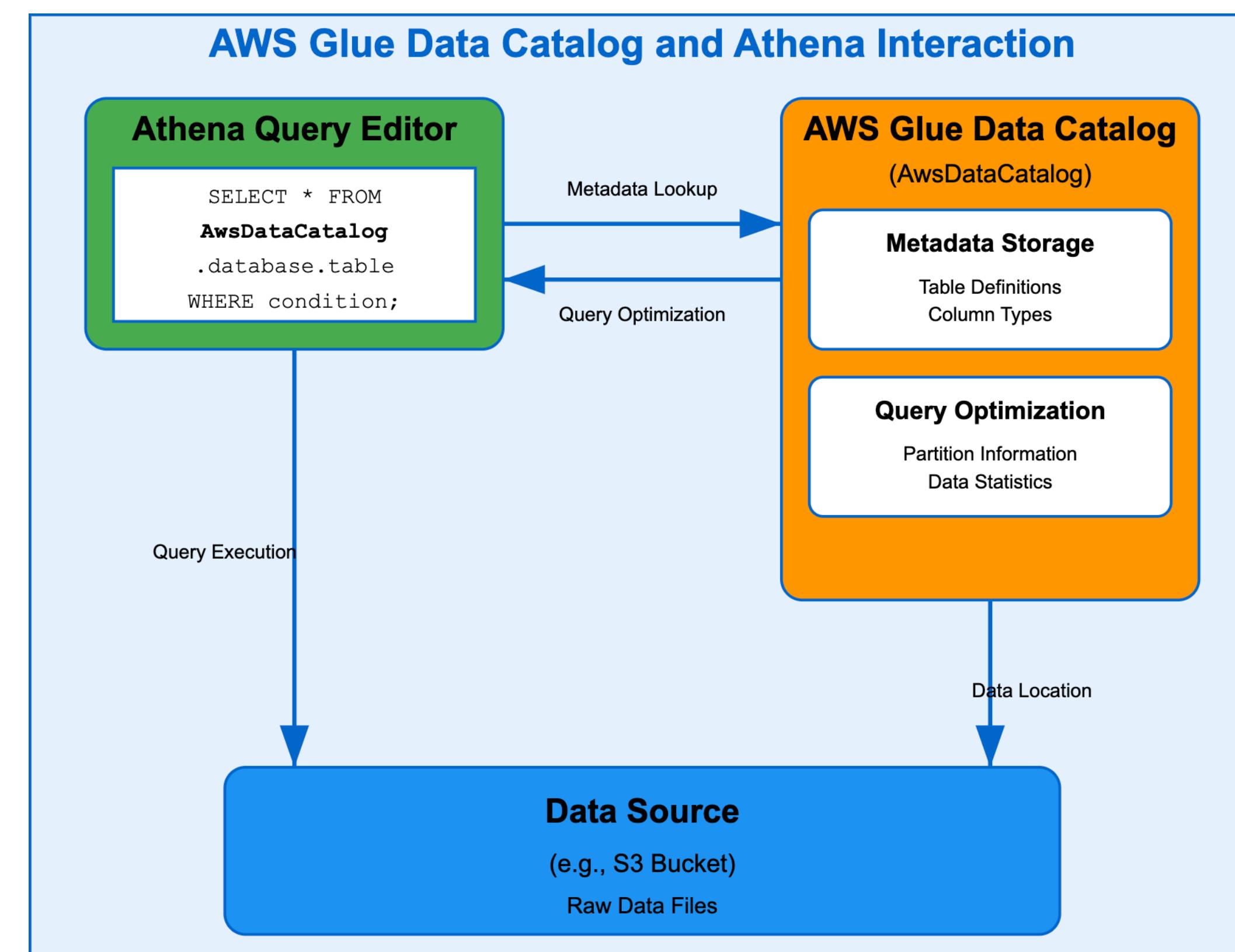
When creating tables and databases manually, Athena uses HiveQL data definition language (DDL) statements such as `CREATE TABLE`, `CREATE DATABASE`, and `DROP TABLE` under the hood to create tables and databases in the AWS Glue Data Catalog.

# AWS Glue Data Catalog

Athena uses the AWS Glue Data Catalog to:

- Store and retrieve metadata
- Use it when running queries to analyze the underlying dataset

In the Athena query editor, this catalog (or data source) is referred to with the label `AwsDataCatalog`.



# AWS Glue Data Catalog

Athena uses the AWS Glue Data Catalog to:

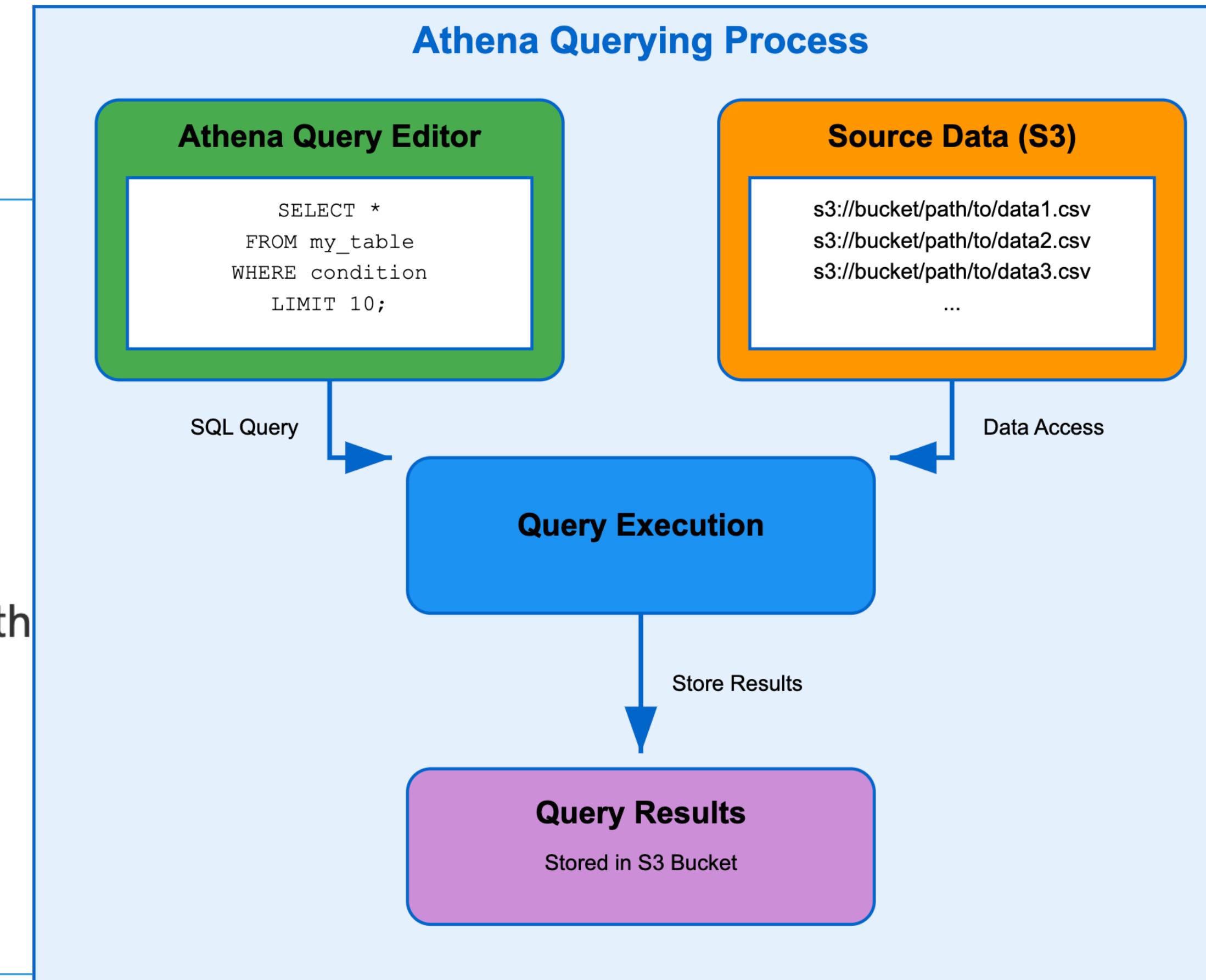
- Store and retrieve metadata
- Use it when running queries to analyze the underlying dataset

In the Athena query editor, this catalog (or data source) is referred to with

## Querying Data

After creating a table:

1. Use SQL **SELECT** statements to query the table, including getting specific file locations for your source data
2. Query results are stored in Amazon S3 in the specified query result location



# Example

Editor    Recent queries    Saved queries    Settings    Workgroup  
primary

**Data**

Data source: AwsDataCatalog

Database: default

Tables and views:

- some\_table
- some\_table2

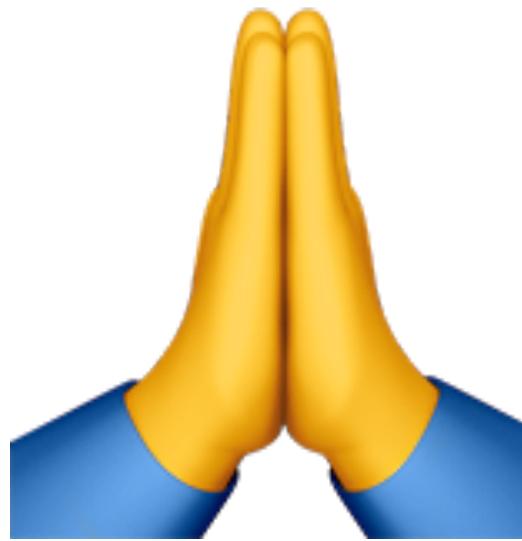
Query:

```
SELECT * FROM "awsdatacatalog"."default"."some_table" limit 10;
```

Run    Explain    Cancel    Clear    Create ▾

✓ Completed Time in queue: 240 ms Run time: 6.535 sec Data scanned: 0.91 KB

#	id	data	category
1	1	a	A
2	3	d	d1
3	4	e	e1
4	4	f	f1
5	2	b	b1



**Thanks  
for  
Watching**