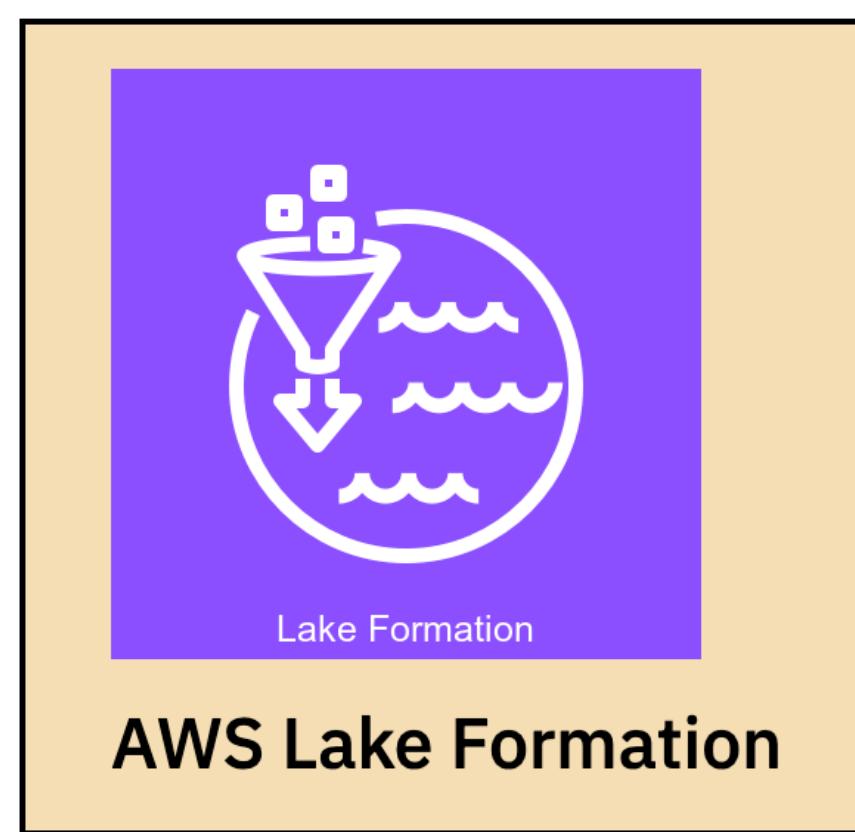




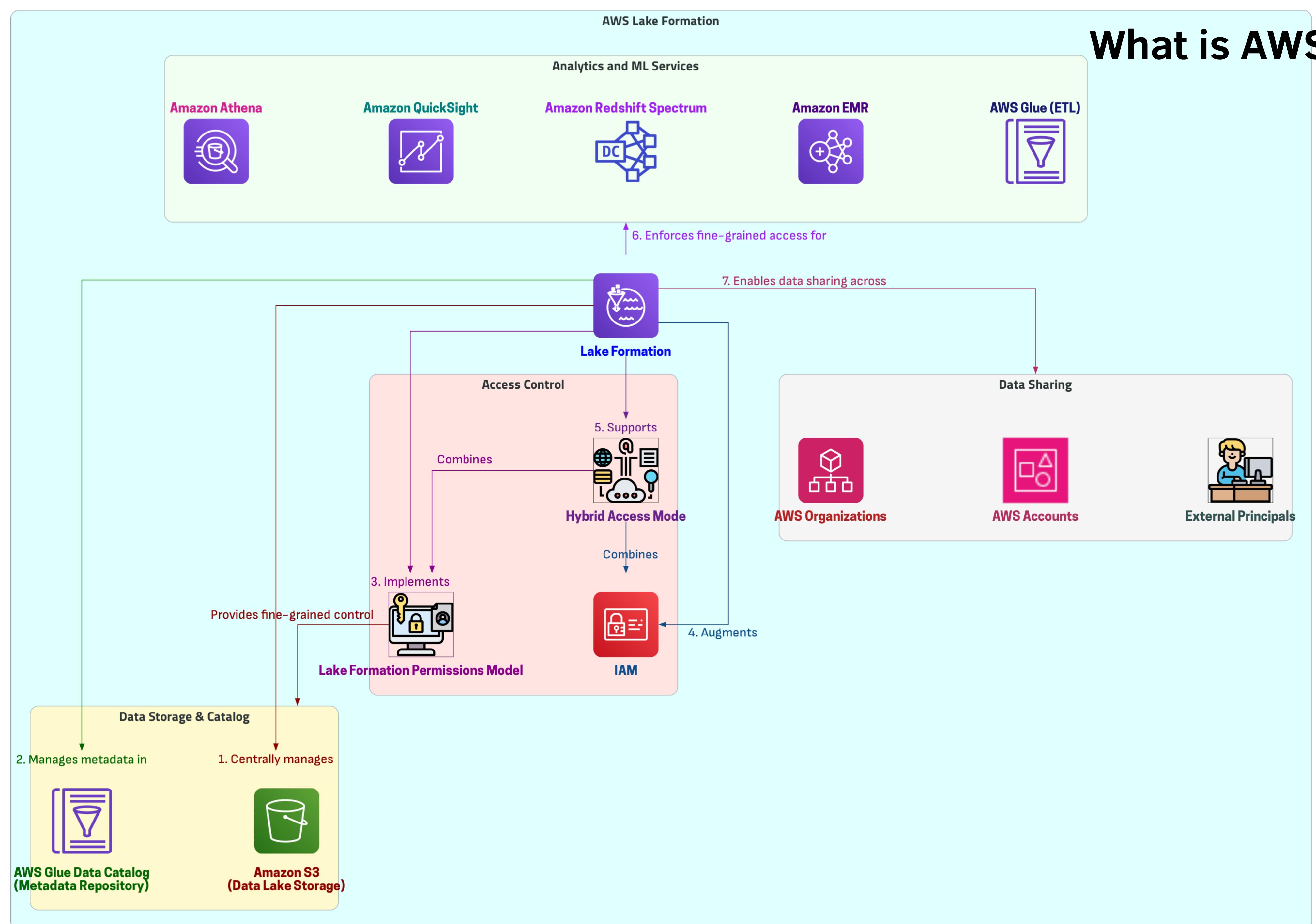
# AWS Lake Formation

# Table of Contents



- 1. What is AWS Lake Formation?
- 2. High Level Overview
- 3. Lake Formation terminology
- 4. Lake Formation features
- 5. Data Ingestion and Management
- 6. Security Management
- 7. Data Sharing
- 8. AWS Lake Formation: How it works
- 9. Cross-Account Data Sharing in AWS Lake Formation
- 10. AWS Service Integrations

# What is AWS Lake Formation?



1. ⚡ Purpose: Govern, secure, and share data

Centrally manage data lakes

Analytics and machine learning

Handles data in Amazon S3

Metadata in AWS Glue Data Catalog

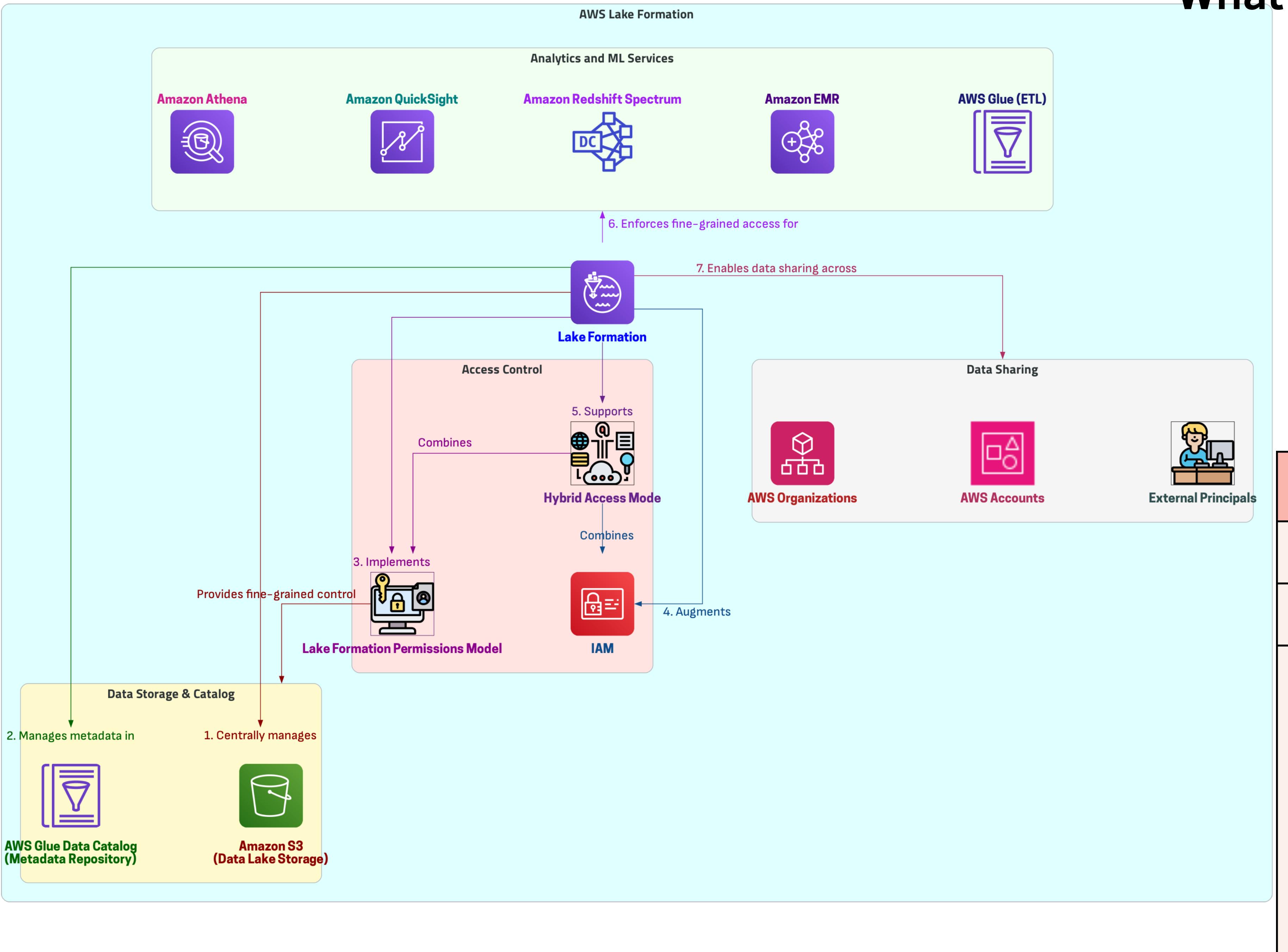
2. 🔑 Fine-grained access control

Column, row, cell-level control

Enhanced security

Compliance with regulations

# What is AWS Lake Formation?



## 3. Permissions model

👉 Works with IAM permissions

⌚ Grant/revoke mechanism

🔍 Amazon Athena

📈 QuickSight

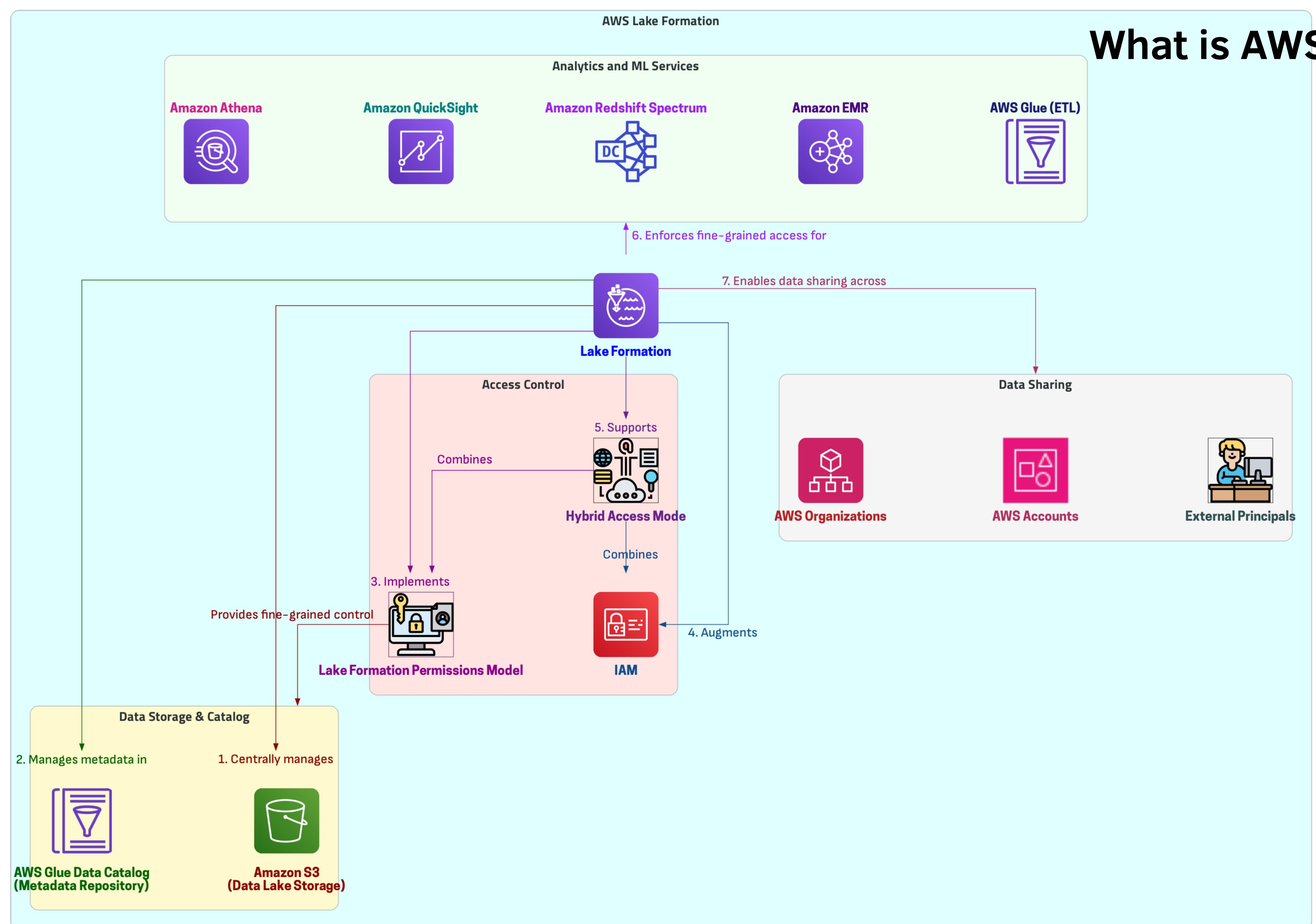
⭐ Redshift Spectrum

🐘 EMR

🔧 AWS Glue

🔗 Applies across AWS services

# What is AWS Lake Formation?



4. Hybrid access mode

Combines Lake Formation and IAM

Gradual adoption

Focus on specific use cases

5. Cross-account data sharing

Share across AWS accounts

Share with organizations

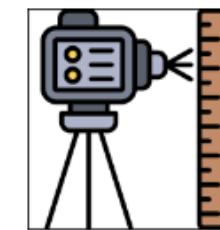
Share with IAM principals

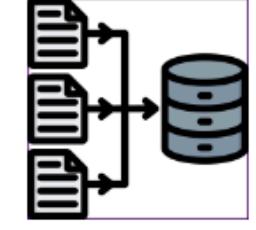
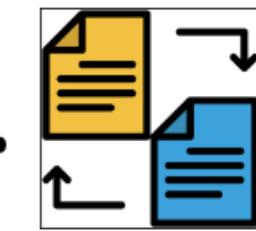
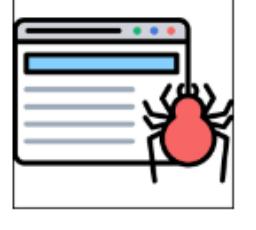
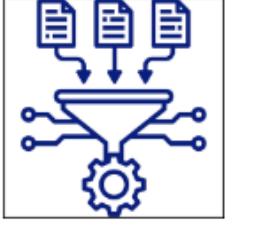
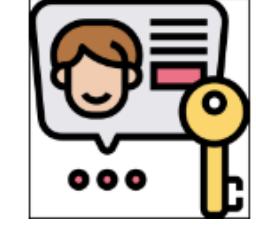
Maintain fine-grained control

Collaboration and data exchange



# High Level Overview



- 1.**  **Identify existing data stores:** Locate data in Amazon S3 or relational and NoSQL databases.  

- 2.**  **Move data:** Transfer the identified data into your data lake.  

- 3.**  **Crawl and catalog:** Process the data to create a comprehensive catalog.  

- 4.**  **Prepare data:** Clean and organize the data for analytics purposes.  

- 5.**  **Provide secure access:** Grant users self-service access to the data through their preferred analytics services.  


# Lake Formation terminology

<b>1. Data lake</b>	<b>2. Data access</b>	<b>3. Hybrid access mode</b>	<b>4. Blueprint</b>
Persistent data in Amazon S3	Secure and granular access	Uses Lake Formation and IAM/S3 permissions	Data management template
Managed by Lake Formation	Grant/revoke permissions model	Allows selective onboarding	Facilitates data ingestion
Stores:  Structured and unstructured data	Augments IAM policies	Focuses on one use case at a time	Predefined source types
Raw and transformed data	Compatible with AWS analytics services		Creates workflows
<b>5. Workflow</b>	<b>6. Data Catalog</b>	<b>7. Data lake administrator</b>	<b>8. Principal</b>
Container for AWS Glue components	Persistent metadata store	Can grant any permission	IAM user or role
Created in Lake Formation, executed in AWS Glue	Managed service for metadata	First user of the Data Catalog	Active Directory user
Visible as directed acyclic graph (DAG)	Stores databases and tables	Manages granular permissions	
On-demand or scheduled execution	Hierarchical permissions		

# Lake Formation features

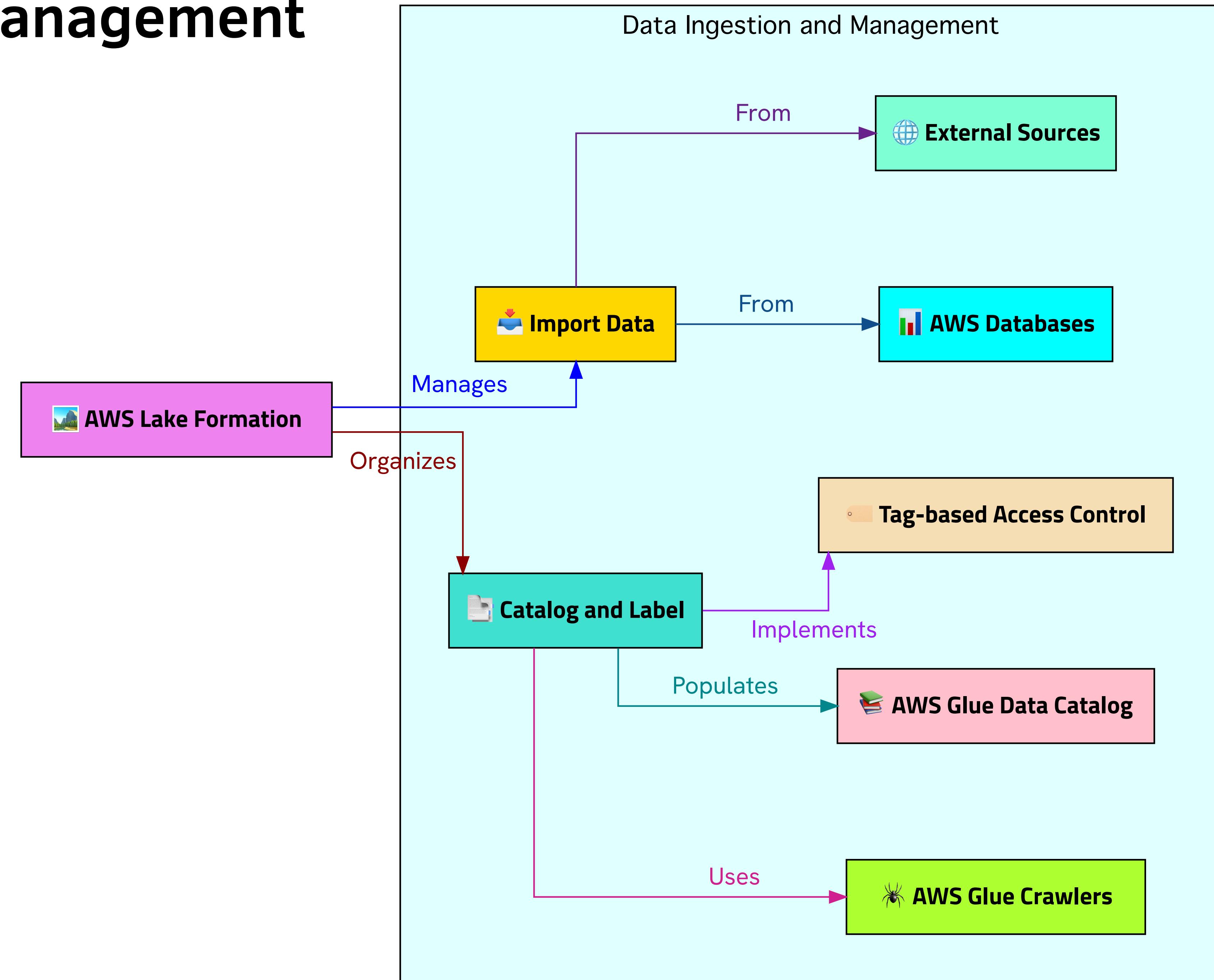
1 Data ingestion and management

2 Security management

3 Data sharing

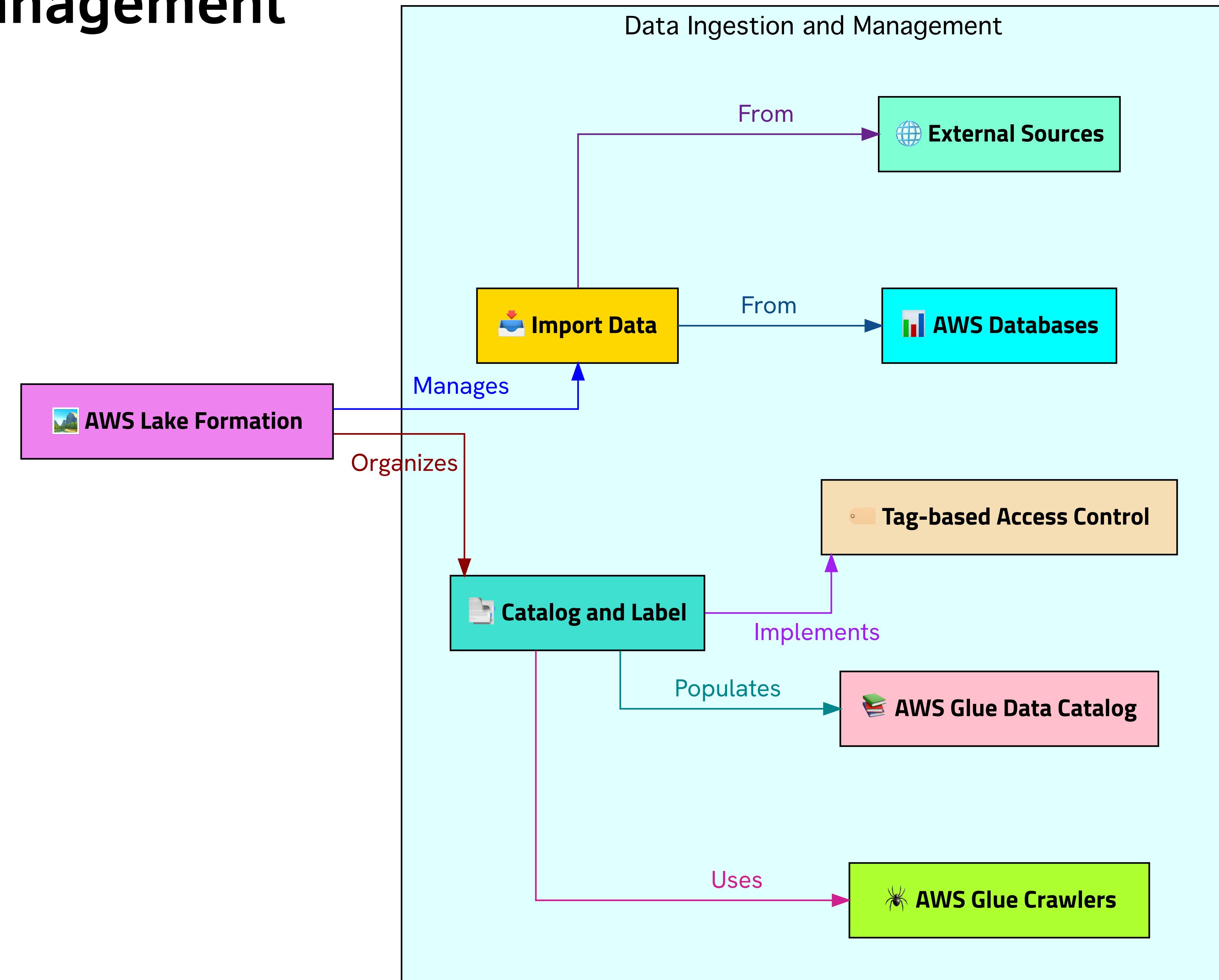
# Data Ingestion and Management

1. Import Data from AWS Databases	
Specify location and credentials	
Reads data and metadata	
Imports to data lake	
Records metadata in central catalog	
Supported databases	<ul style="list-style-type: none"><li>MySQL</li><li>PostgreSQL</li><li>SQL Server</li><li>MariaDB</li><li>Oracle</li></ul>
Running in	<ul style="list-style-type: none"><li>Amazon RDS</li><li>Amazon EC2</li></ul>
Loading types	<ul style="list-style-type: none"><li>Bulk</li><li>+ Incremental</li></ul>



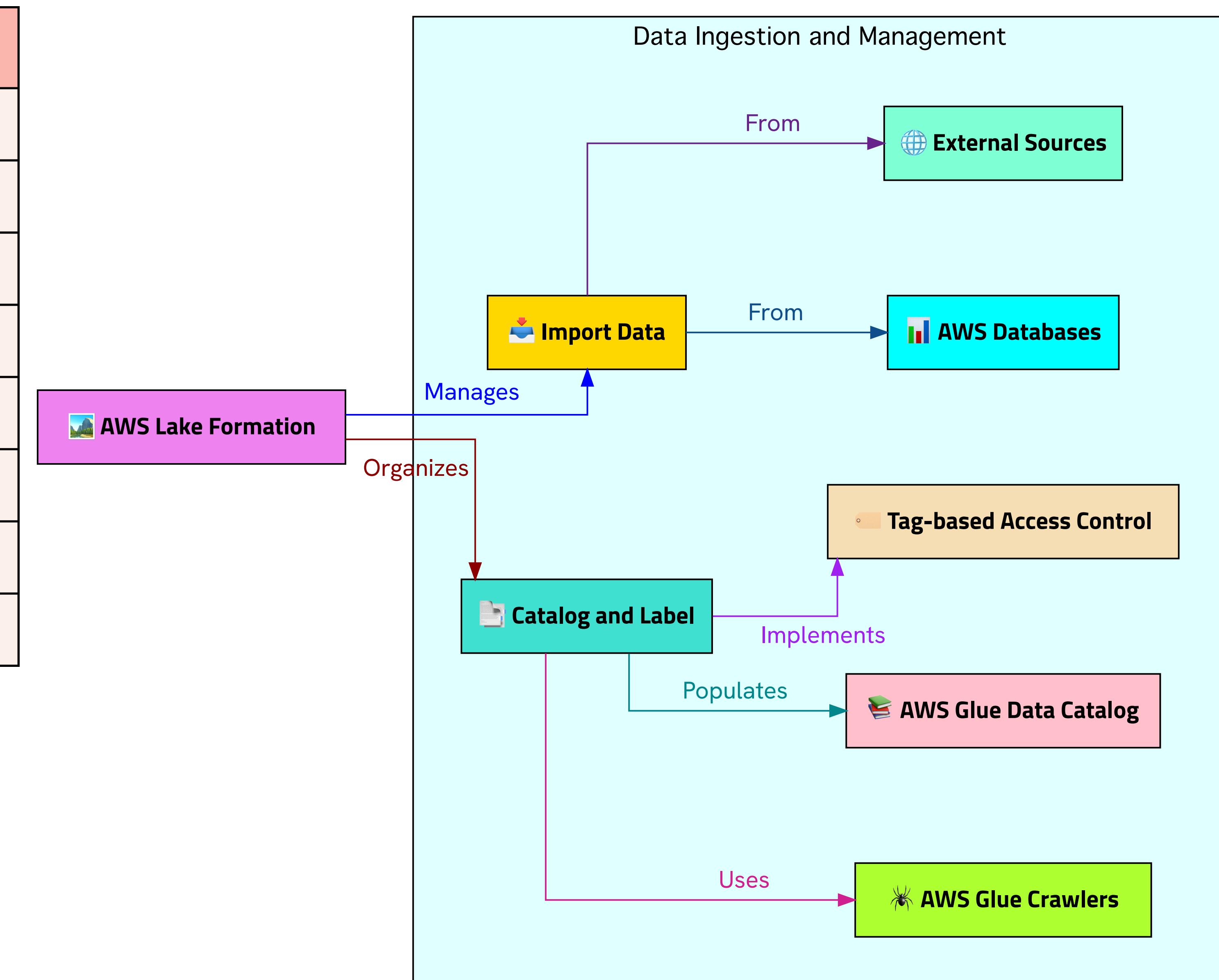
# Data Ingestion and Management

2. Import Data from External Sources	
On-premises databases	JDBC connection
Identify target sources	
Provide access credentials	
Loads data into data lake	
Custom ETL jobs	AWS Glue
	For other database types



# Data Ingestion and Management

3. Catalog and Label Data	
AWS Glue crawlers	Read data in Amazon S3
	Extract database schema
	Extract table schema
AWS Glue Data Catalog	Store extracted data
	Searchable catalog
Tag-based access control (TBAC)	Manage permissions
	Databases Tables Columns



## 1. Define and Manage Access

Controls: Single management point,  
 Security policies ( Database, Table, Column, Row, Cell levels),  
 User types ( IAM, Federated),   
Fine-grained controls for Redshift Spectrum, Athena, Glue ETL, EMR Spark

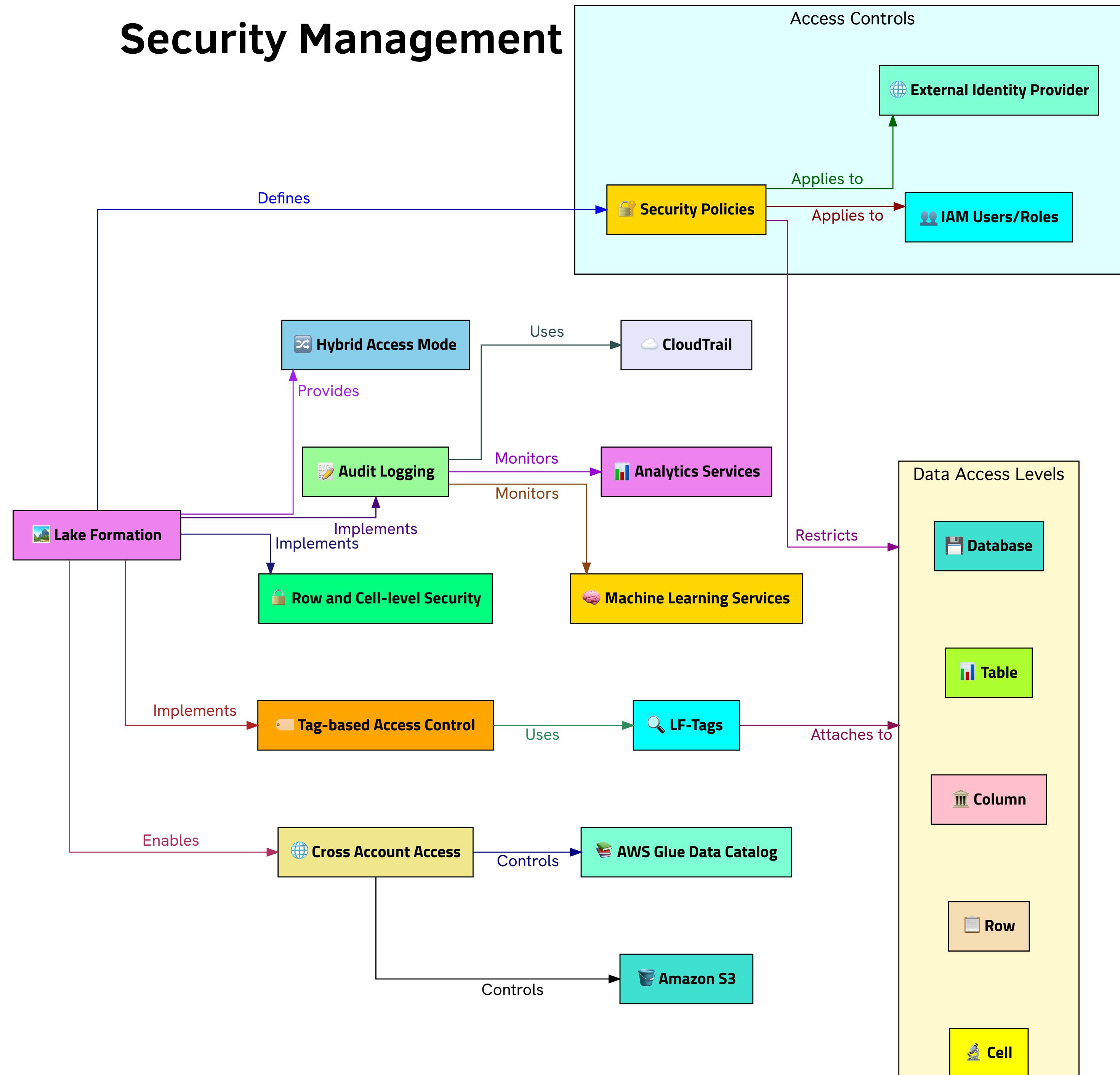
## 2. Hybrid Access Mode:

Selective enablement, Incremental implementation, Coexistence with existing policies

## 3. Implement Audit Logging:

CloudTrail integration, Access monitoring, Compliance verification, Audit capabilities ( User/role, Data access, Services, Timestamp)

# Security Management



4 . 🔒 Row and Cell-level Security: 🔍 Data filters, 🛡️ Sensitive data protection, 🔒 PII safeguarding

5 . 🏷️ Tag-based Access Control:

🏷️ LF-Tags (📝 Define, 🔗 Attach),

🔄 Controlled access sharing, 📊

Scalable data governance, 🔎

Text-based metadata search

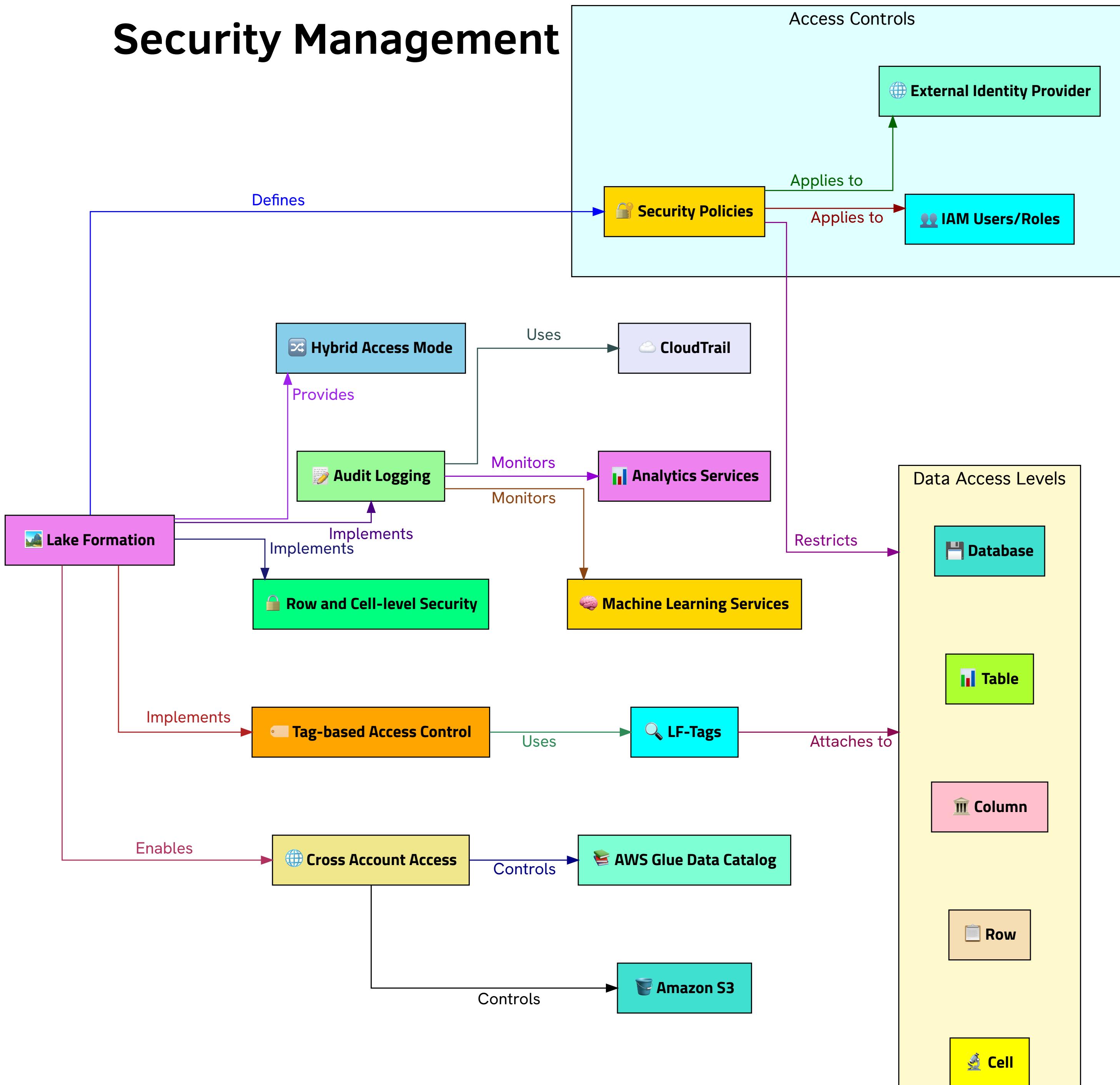
6 . 🌐 Cross Account Access: 🏢

Multi-account data lake

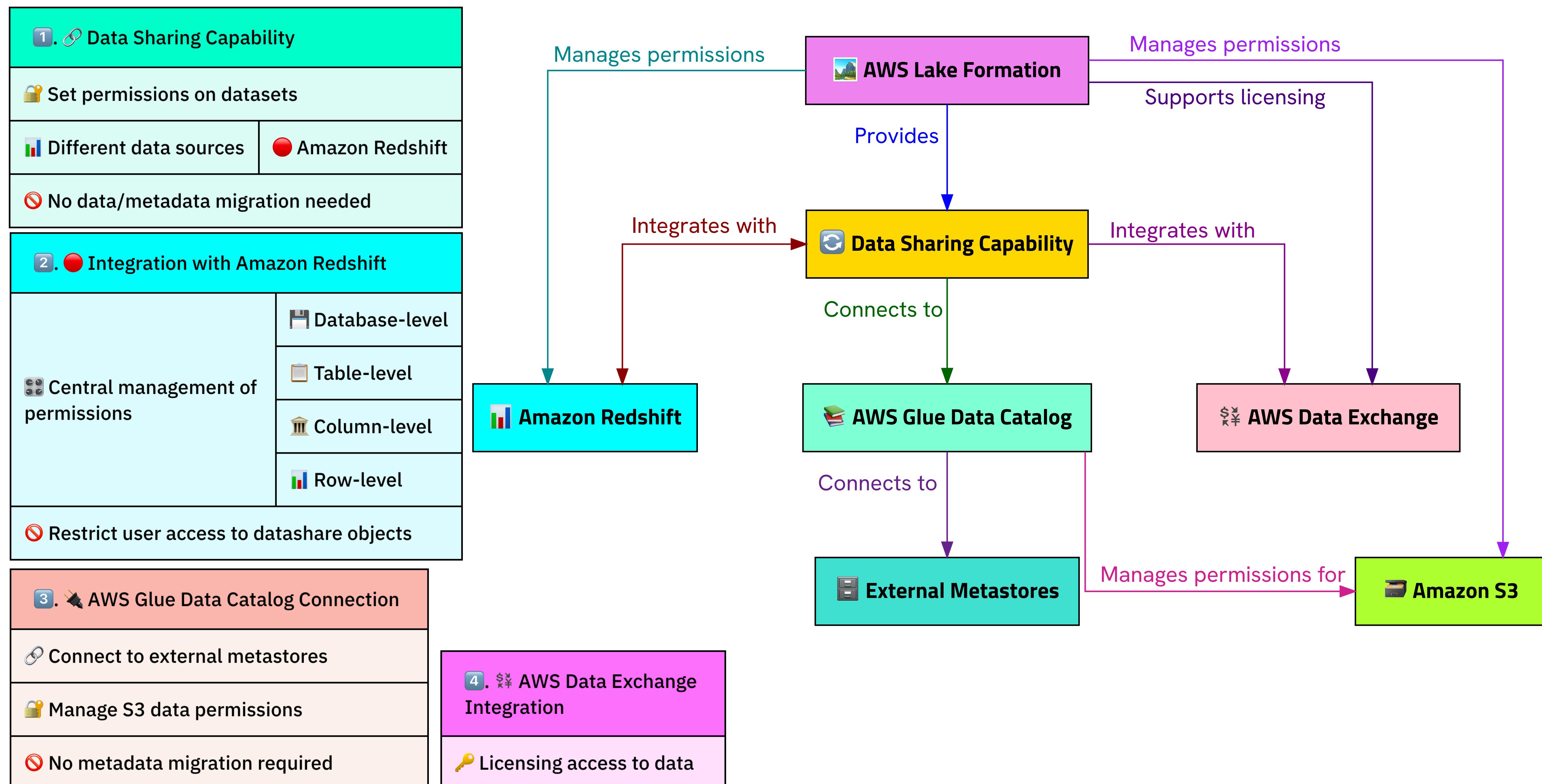
management, 🔍 Centralized

approach, 🔬 Fine-grained access control (📚 Data Catalog, 📦 S3 locations)

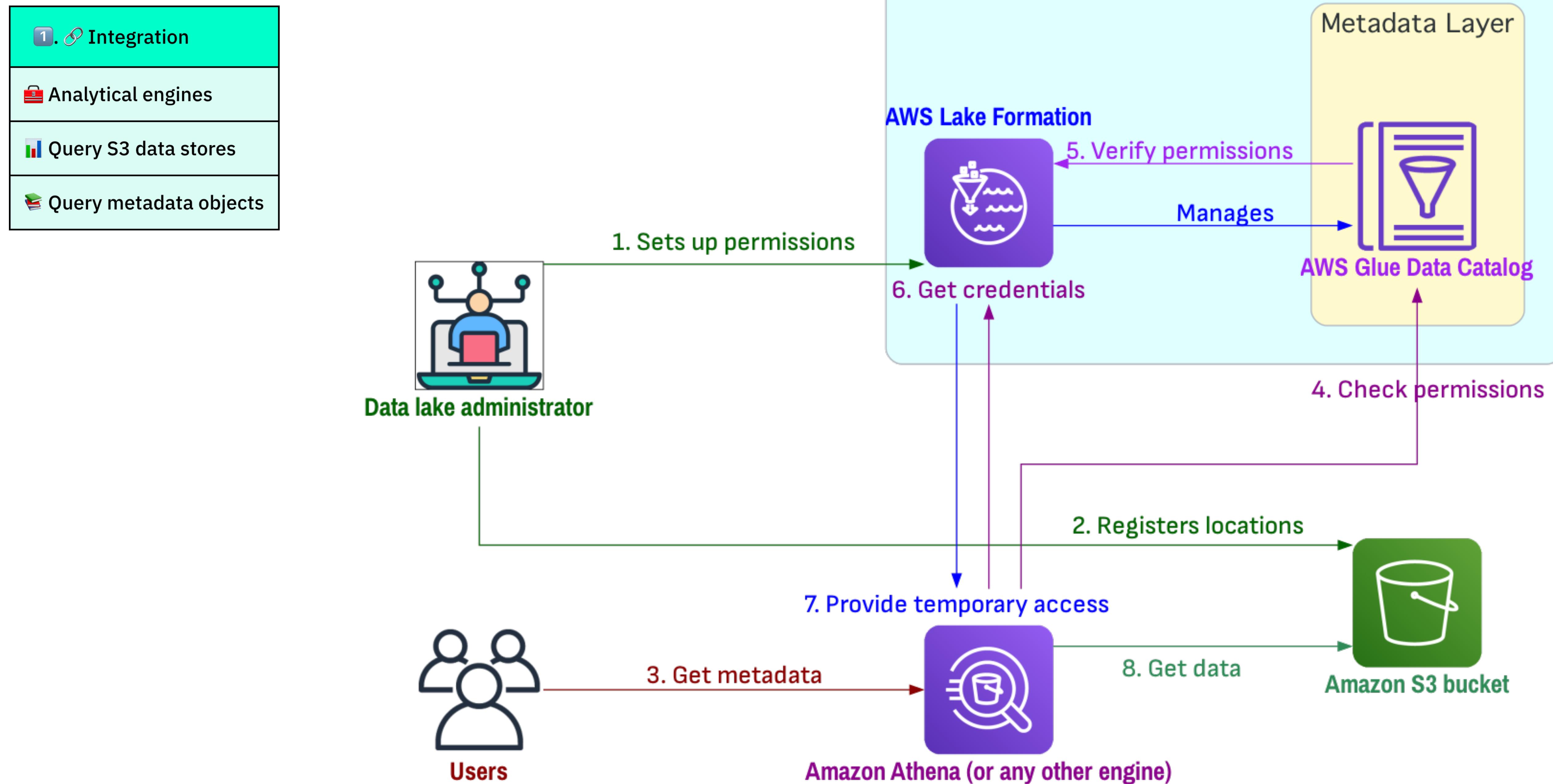
## Security Management



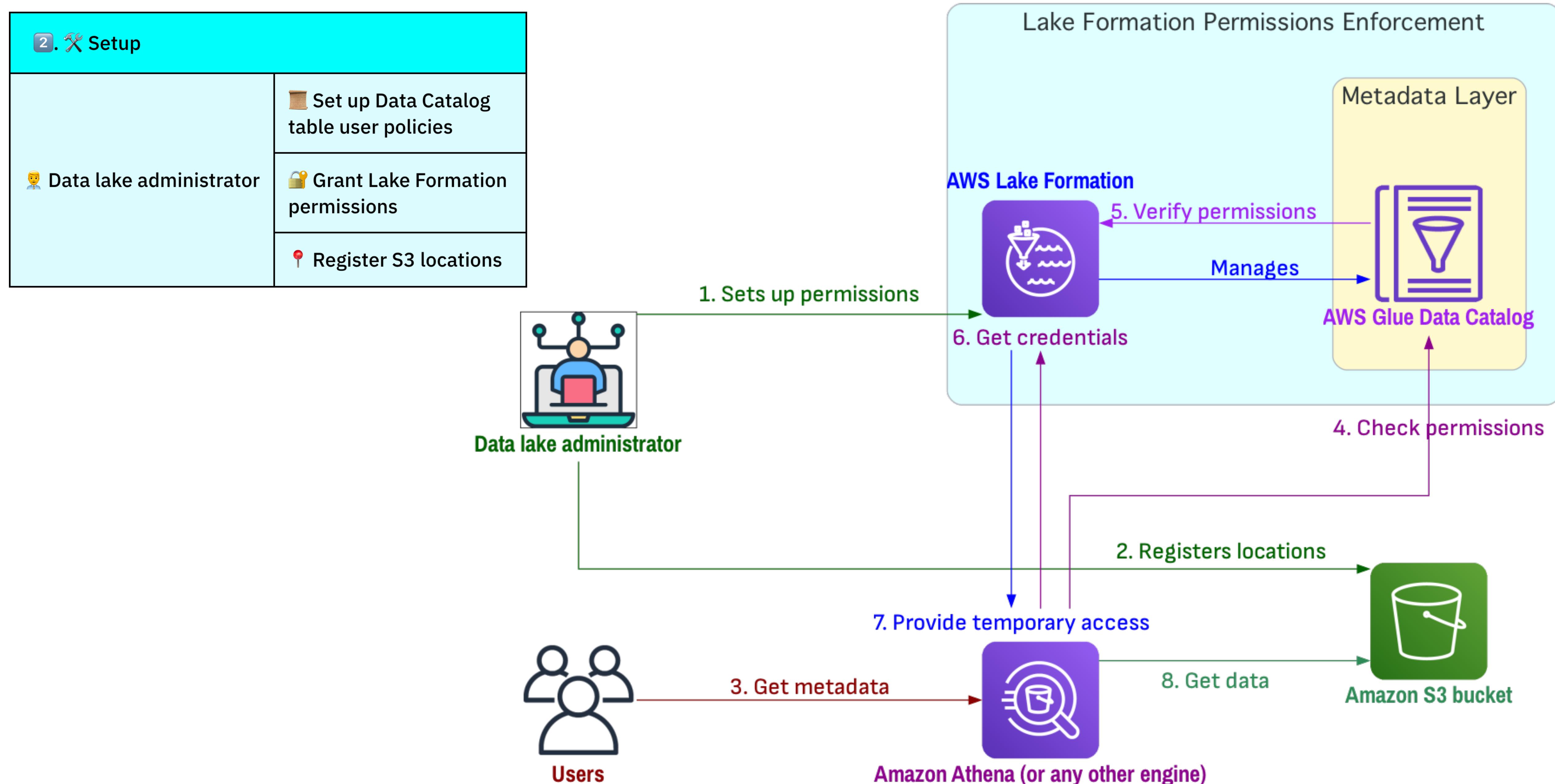
# Data Sharing



# AWS Lake Formation: How it works

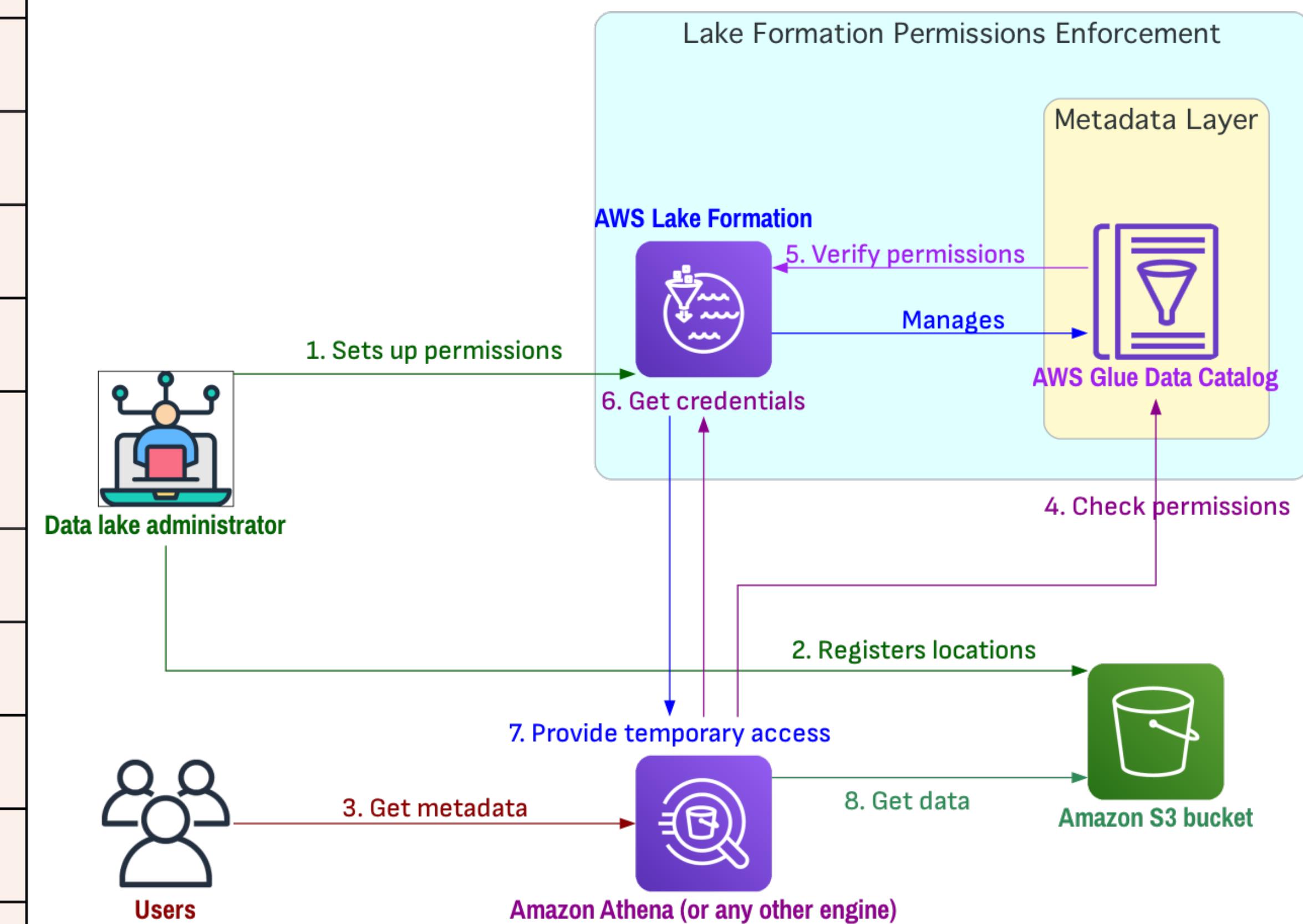


# AWS Lake Formation: How it works

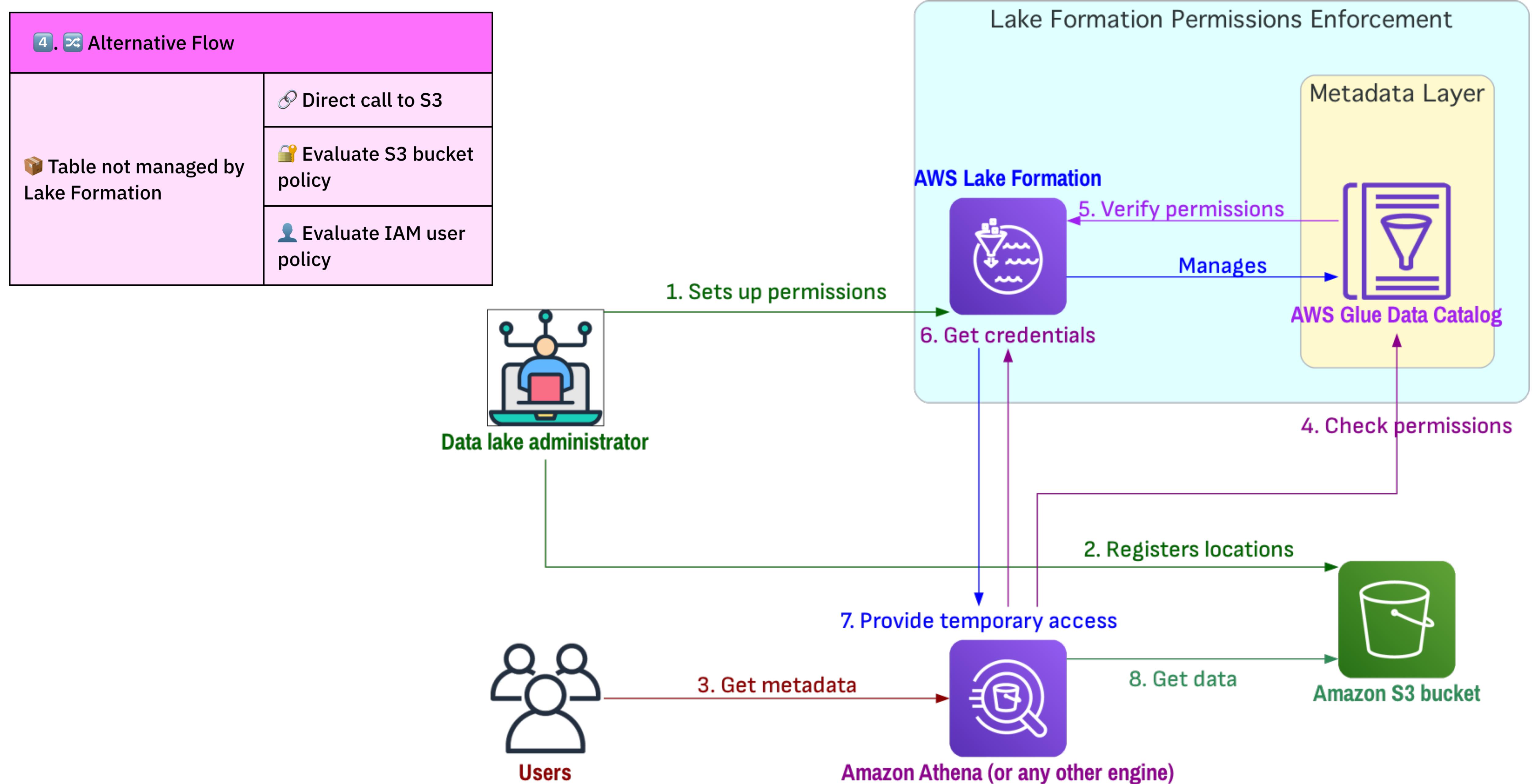


# AWS Lake Formation: How it works

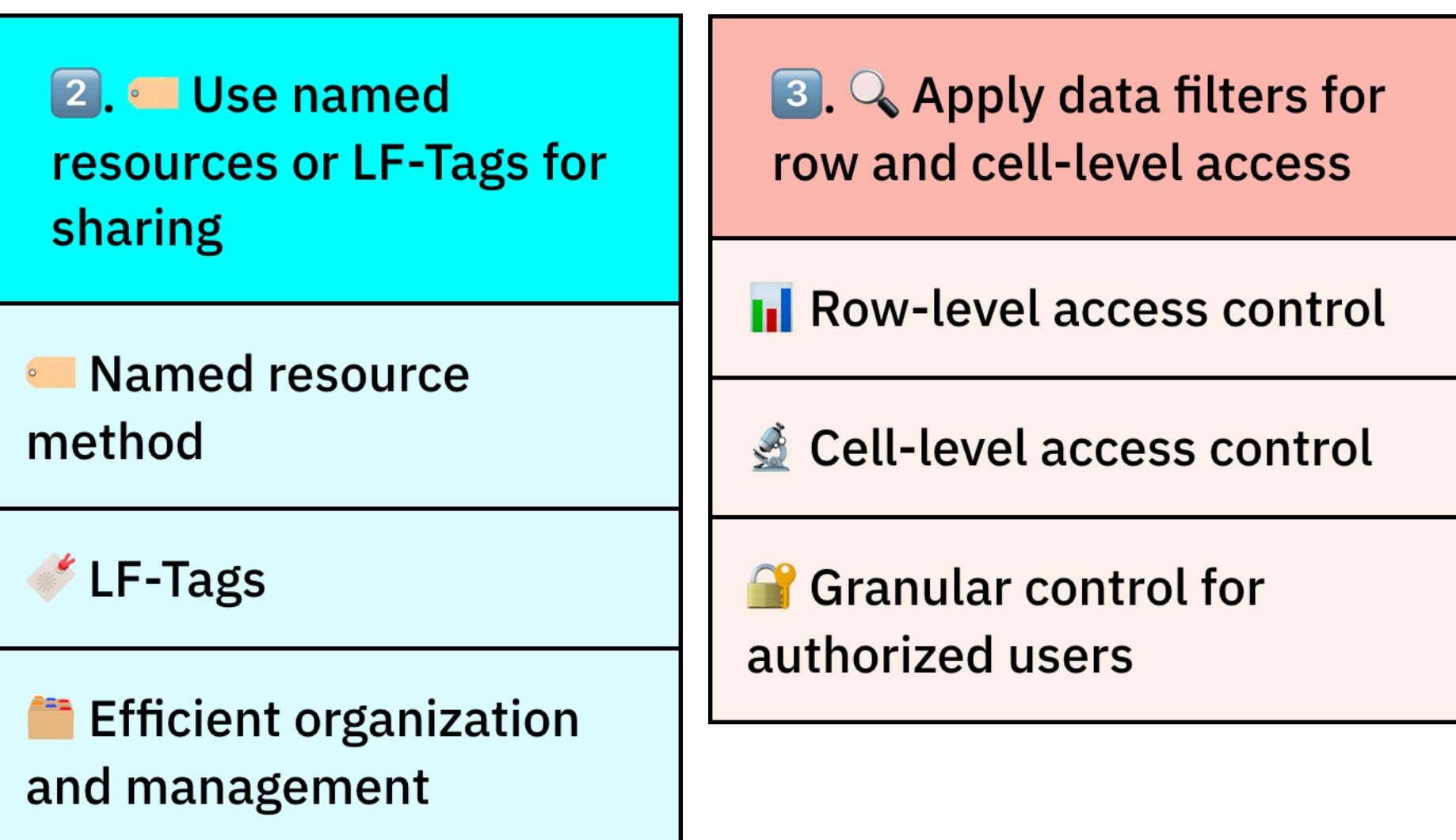
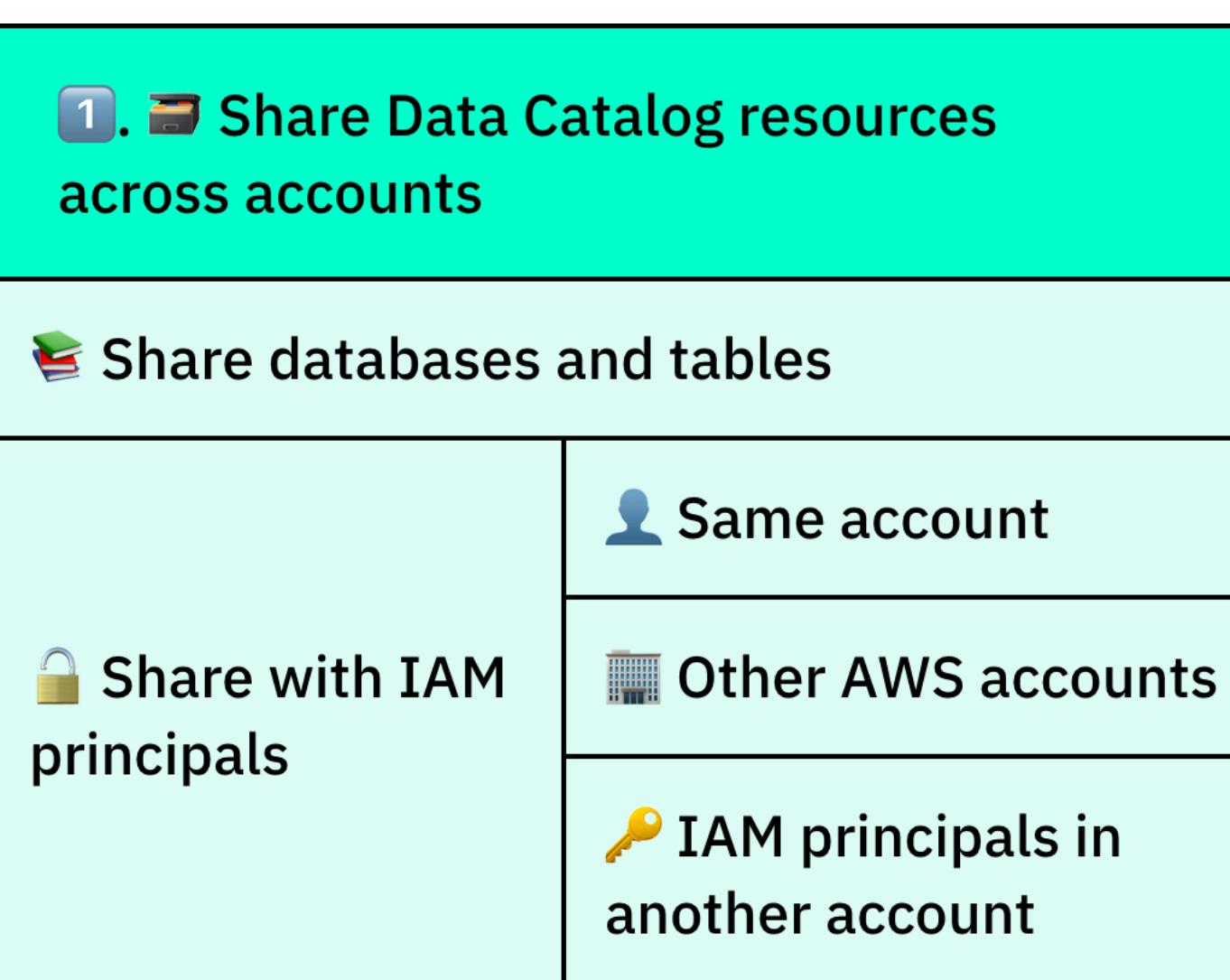
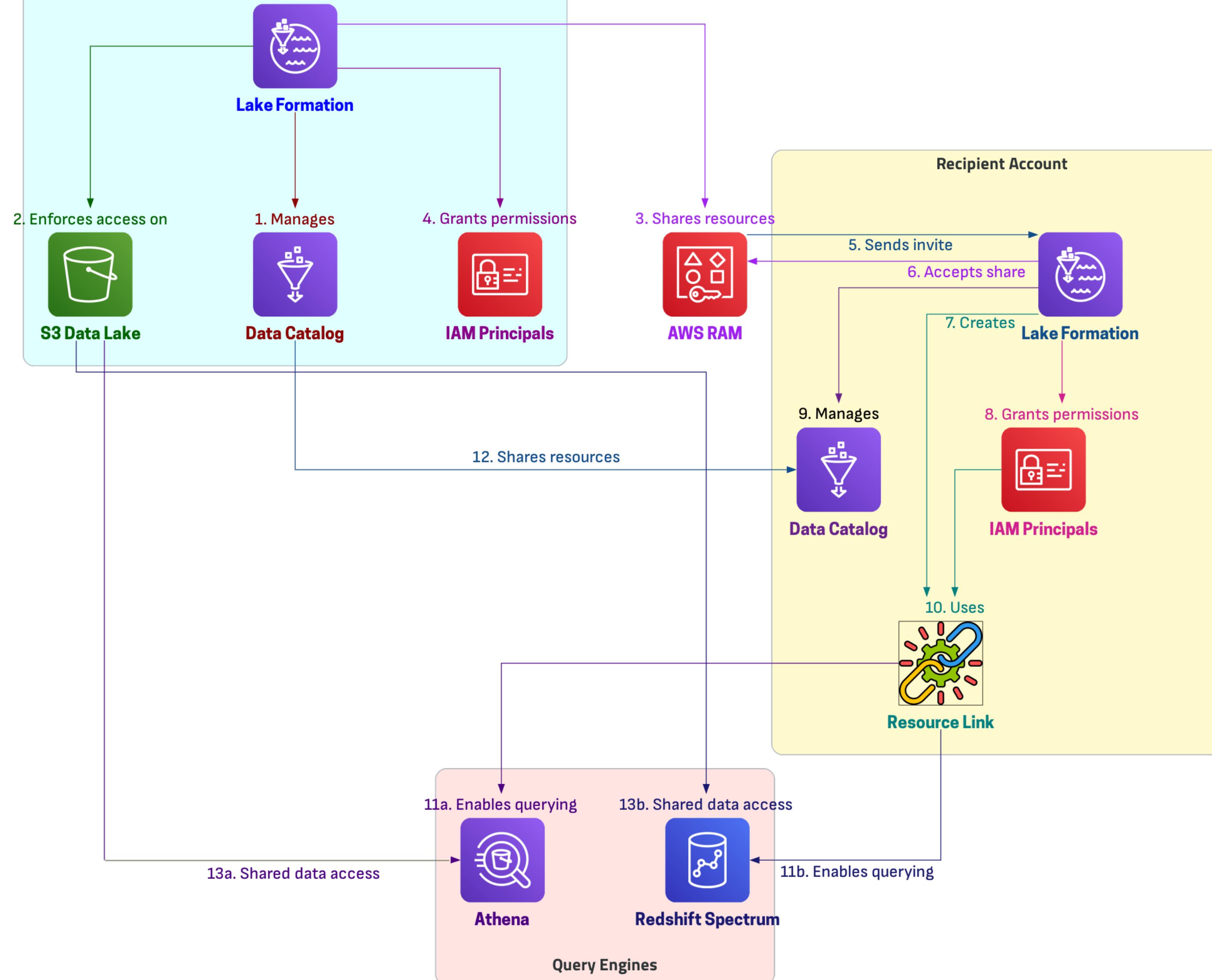
3. ⏪ Workflow Steps	
1. 🛡️ Get metadata	User submits query/ETL script
	🔍 Engine identifies requested table
	📥 Request metadata from Data Catalog
2. 🔒 Check permissions	📊 Data Catalog checks with Lake Formation
	✓ If authorized, return allowed metadata
	>Data Catalog informs engine of table management
3. 💳 Get credentials	🔒 Engine requests temporary access
	📈 Lake Formation grants temporary access
	📁 Engine fetches data from S3
4. 📈 Get data	🔍 Perform necessary filtering
	📊 Return results to user



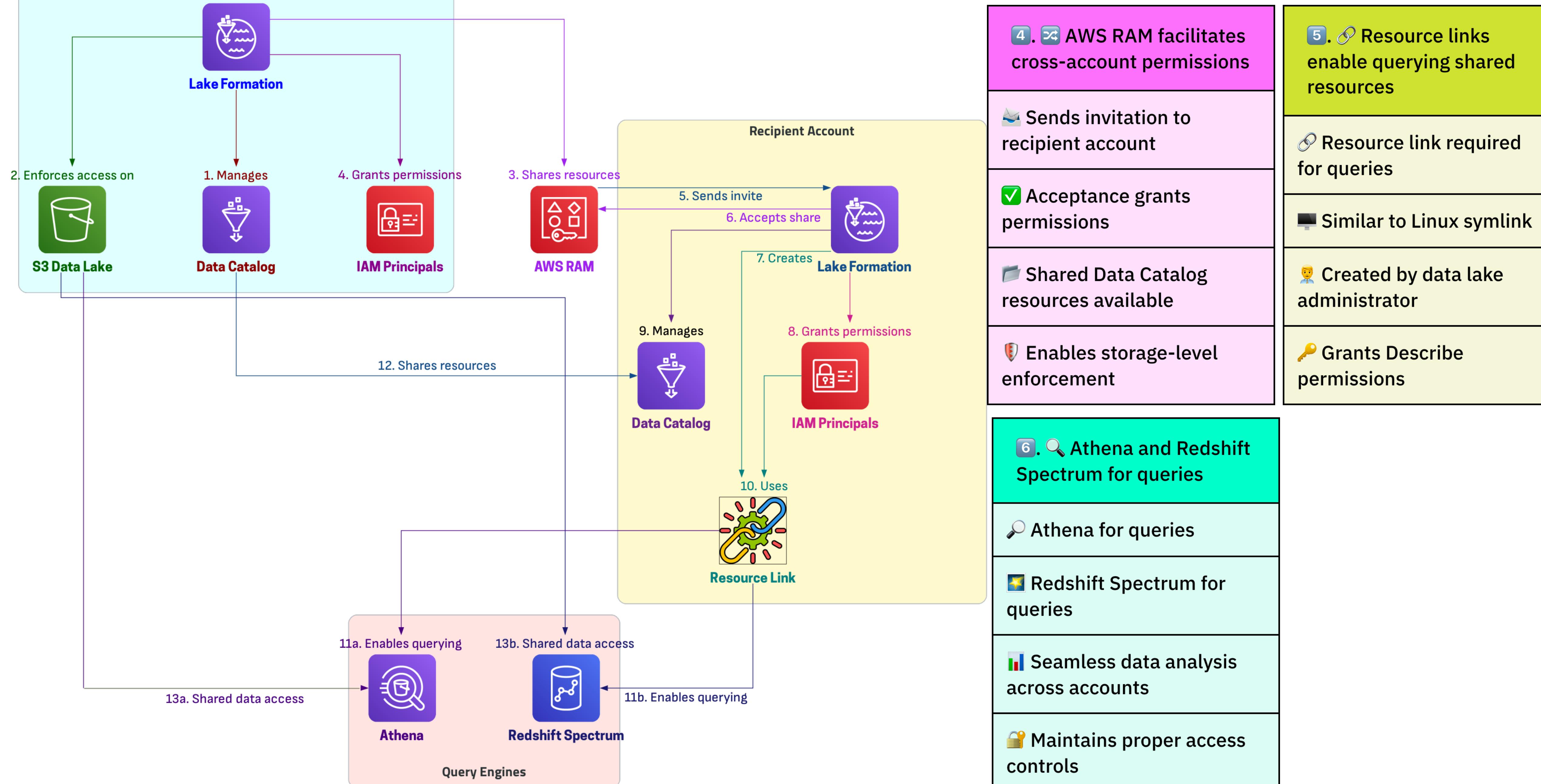
# AWS Lake Formation: How it works



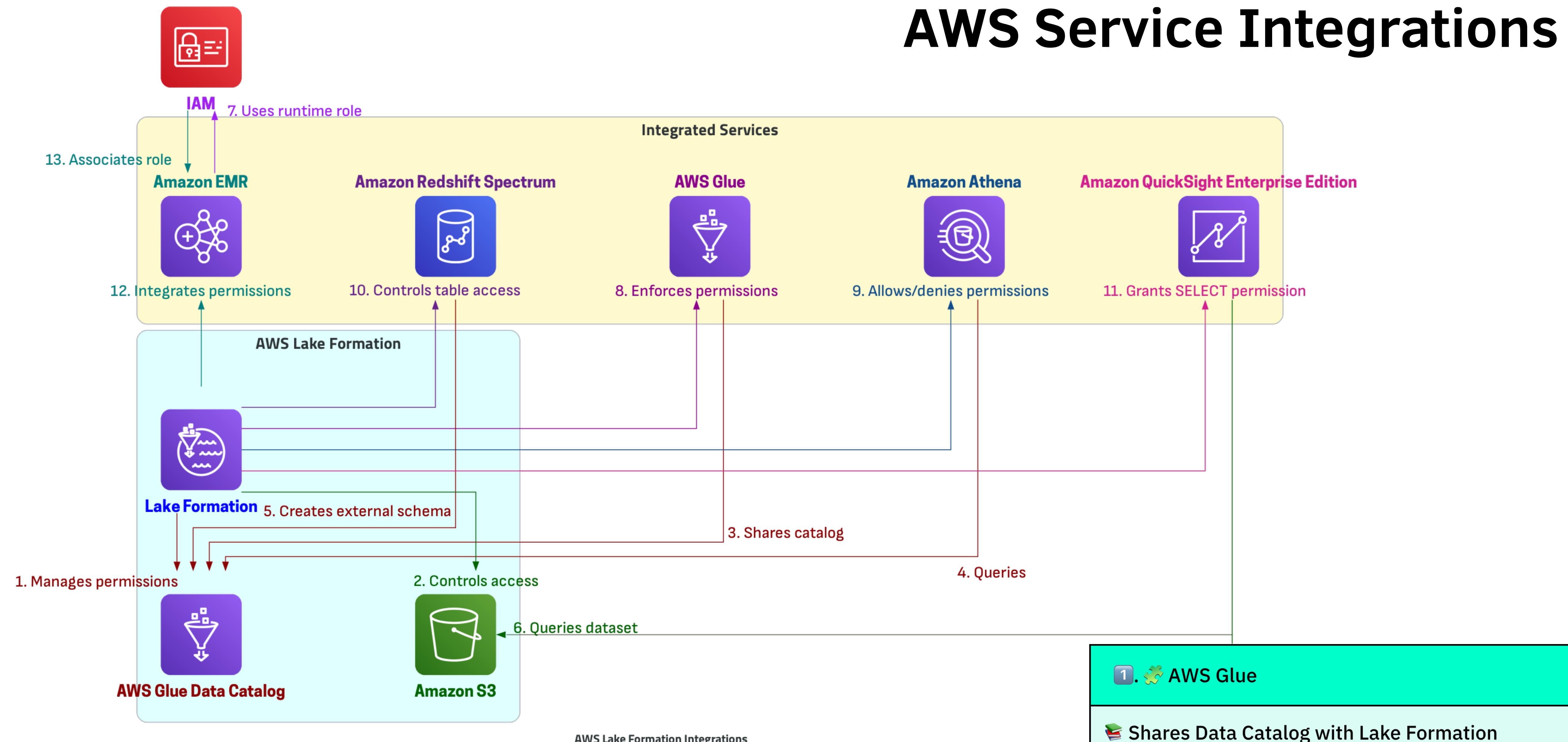
# Cross-Account Data Sharing in AWS Lake Formation



# Cross-Account Data Sharing in AWS Lake Formation

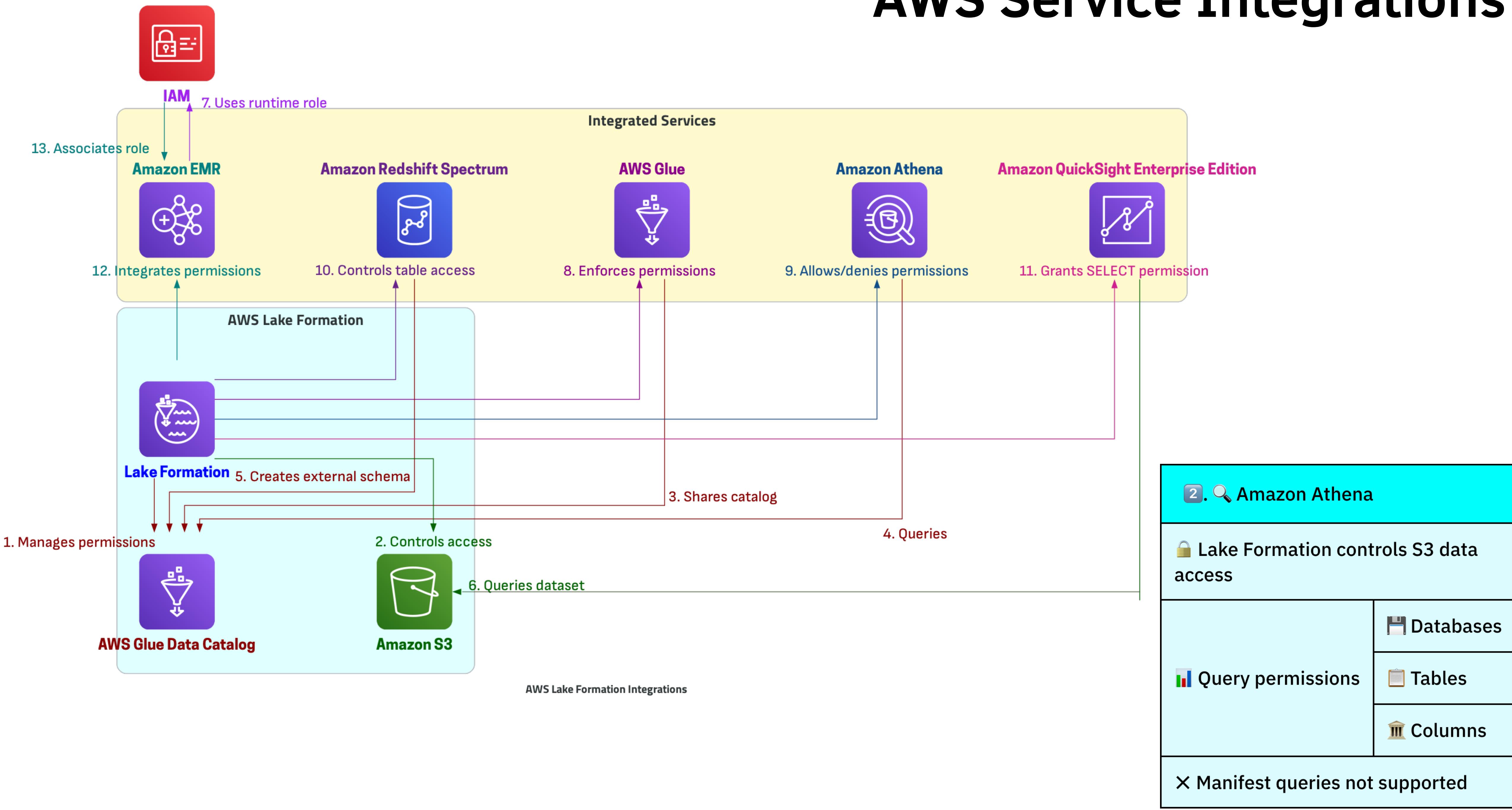


# AWS Service Integrations

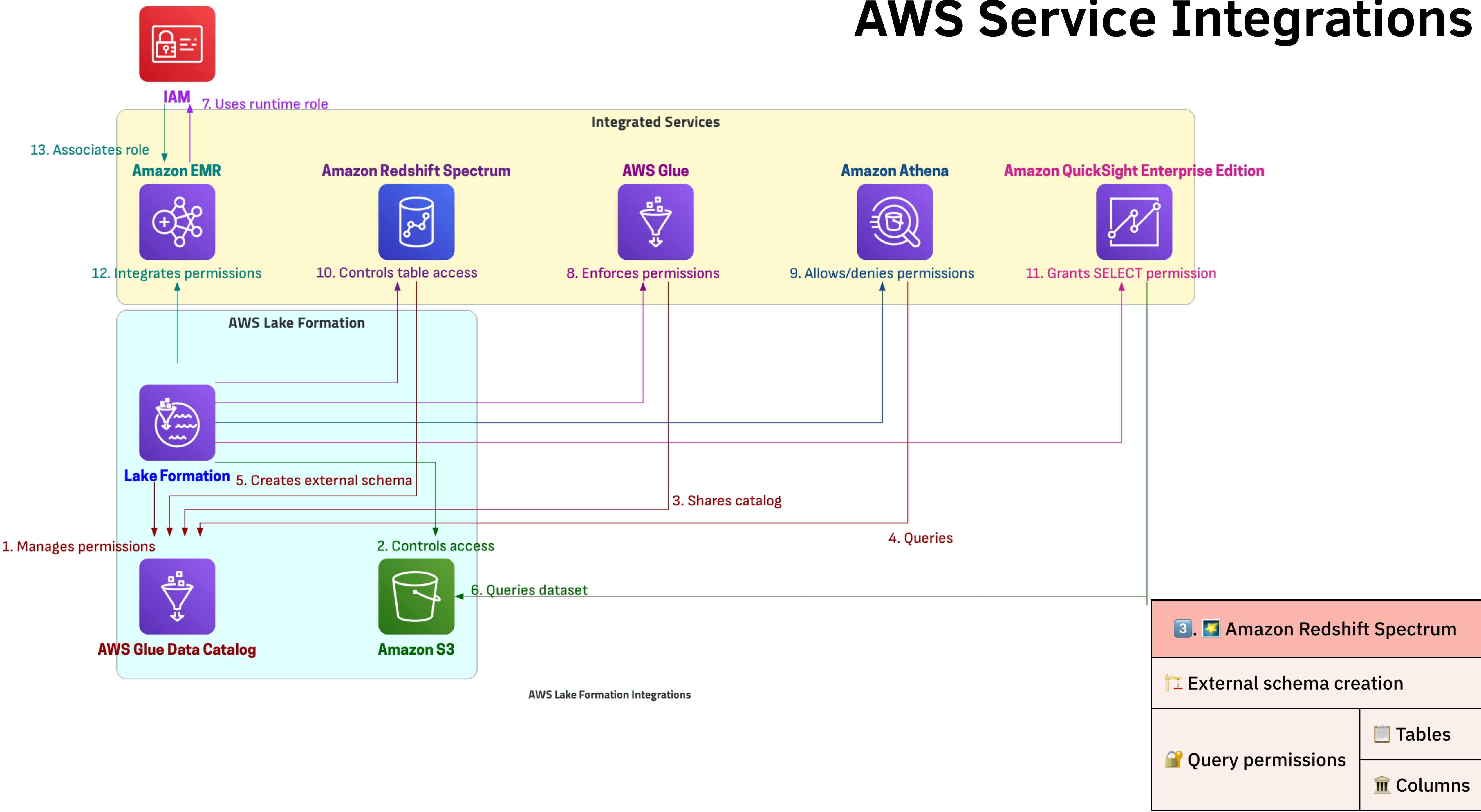


<b>1. ✨ AWS Glue</b>	
Shares Data Catalog with Lake Formation	
Access based on Lake Formation permissions	Console operations
	API operations

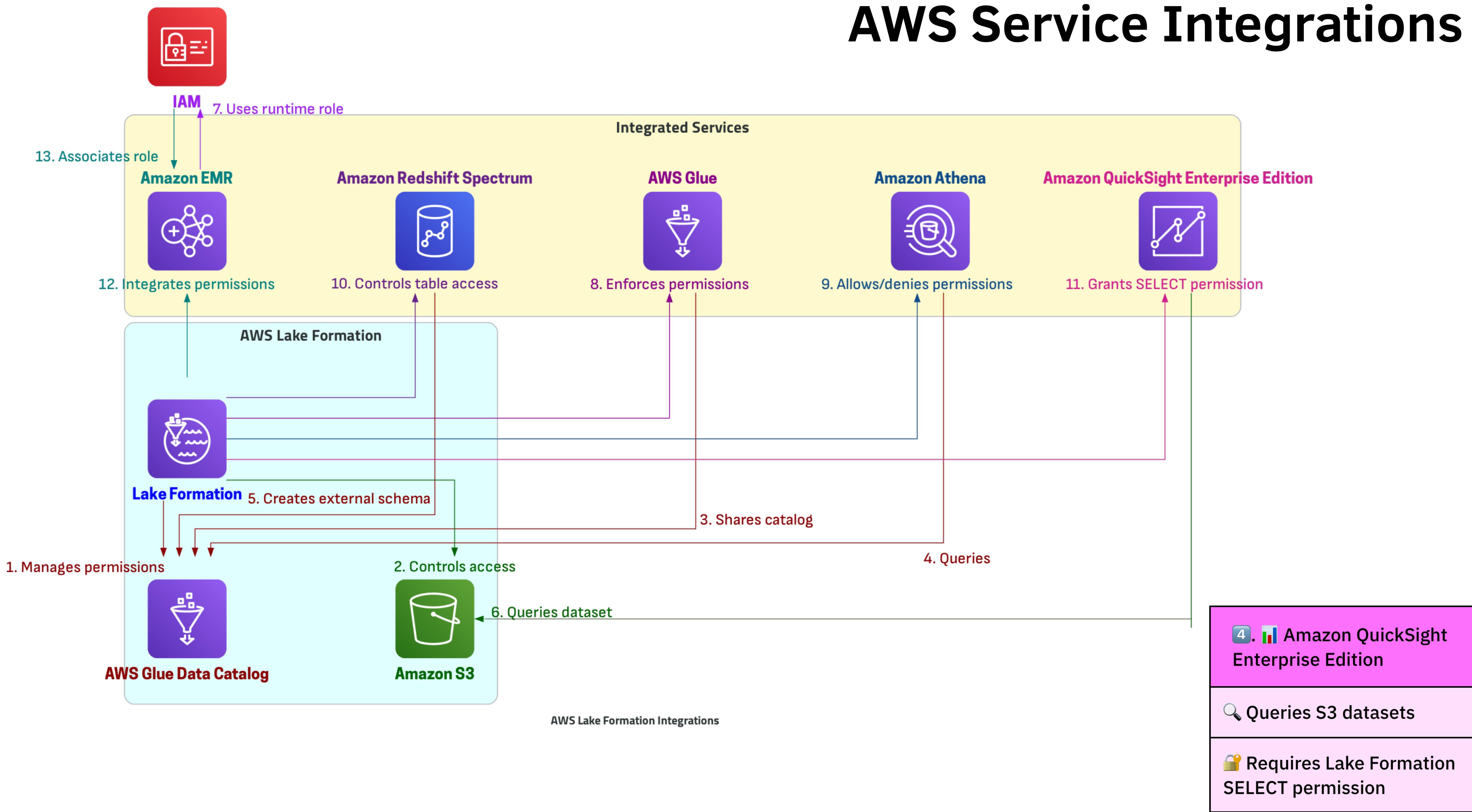
# AWS Service Integrations



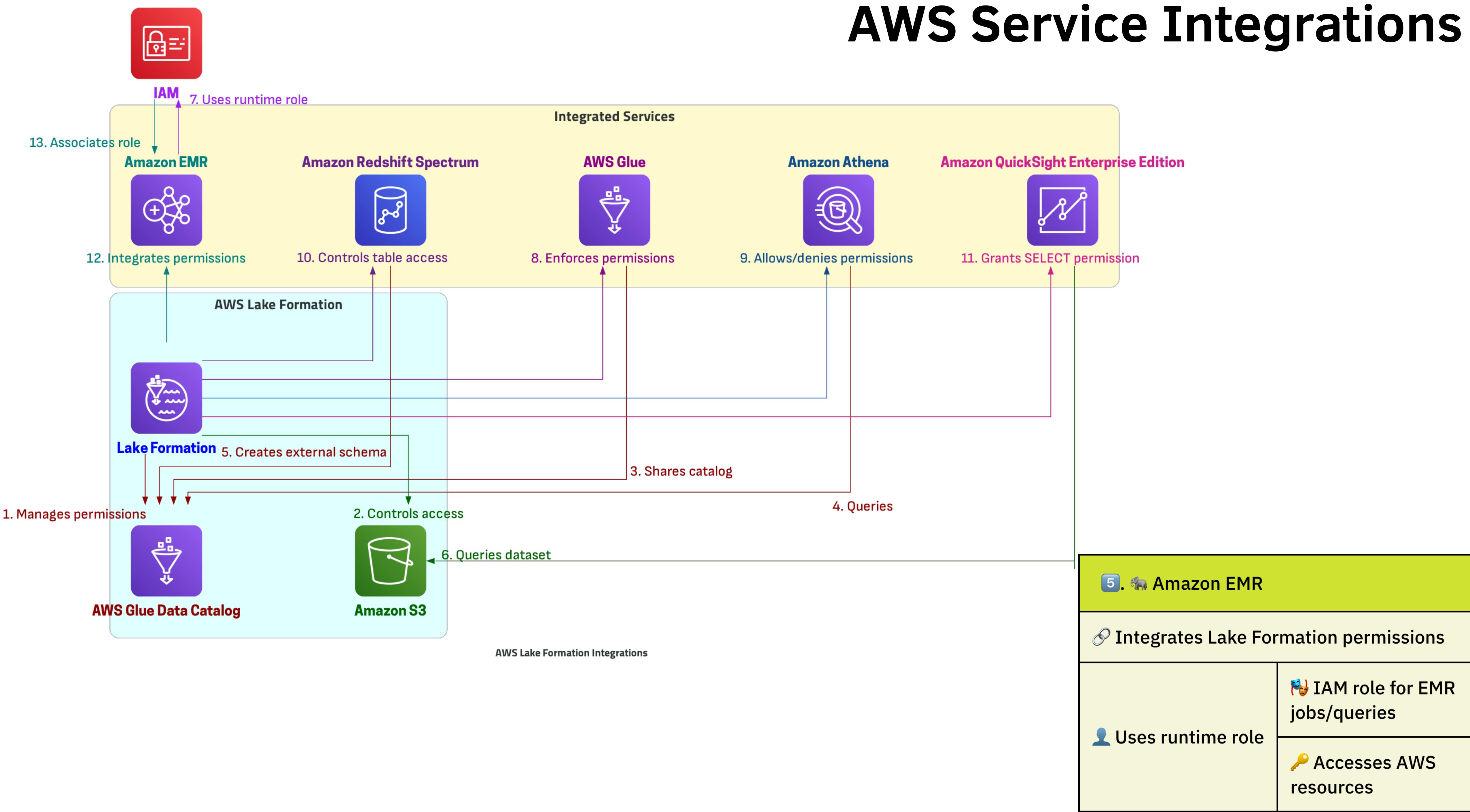
# AWS Service Integrations

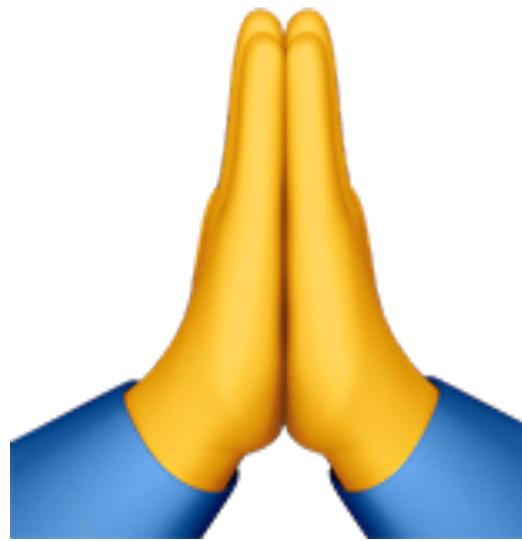


# AWS Service Integrations



# AWS Service Integrations





**Thanks  
for  
Watching**