

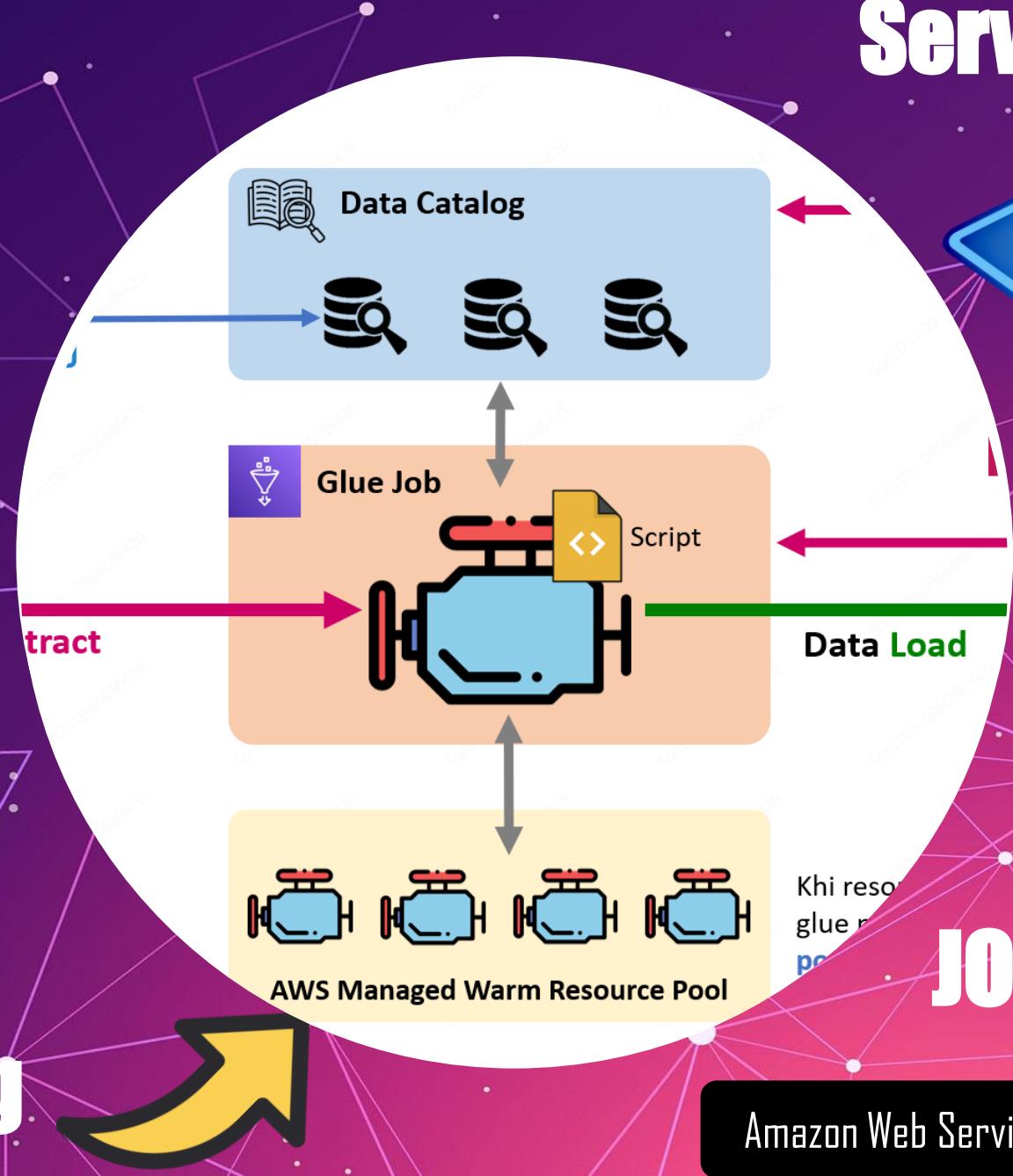


# AWS Glue



## Crawler

## Catalog



# Serverless

## Jobs

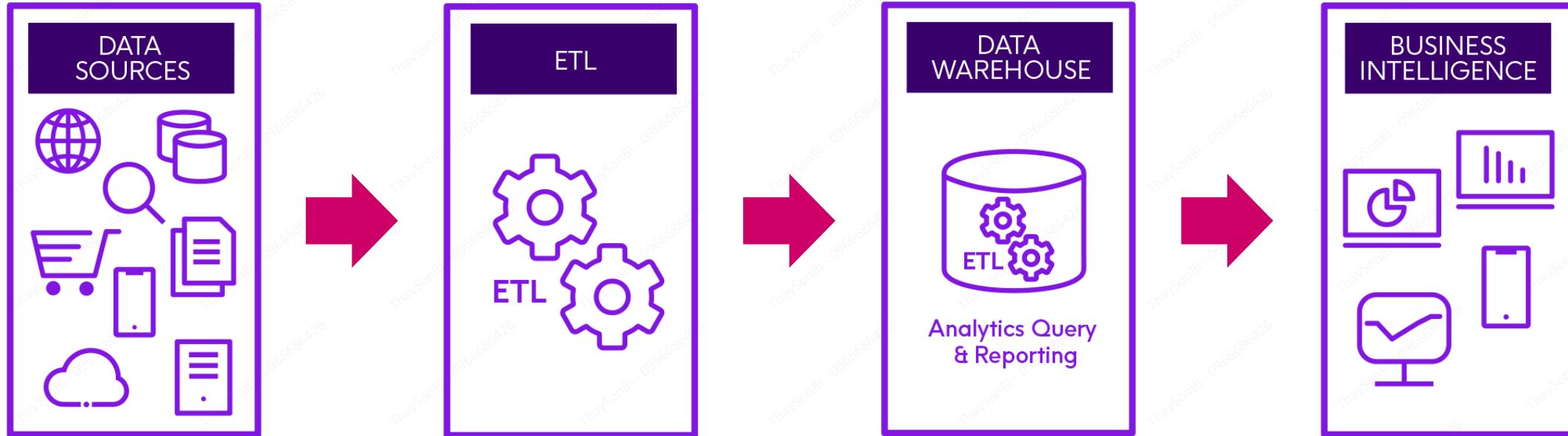
Amazon Web Service - Training



ANALYTICS  
SERVICES

# EXTRACT – TRANSFORM – LOAD

AWS  
Glue



S3



RDS



Kinesis



DynamoDB



ORACLE  
FUSION MIDDLEWARE  
Data Integrator



Microsoft  
SQL Server  
Integration Services



Glue



RDS



Redshift



QuickSight



Power BI



Excel



## Serverless ETL (Extract, Transform và Load)

... với **datapipeline** (thực hiện ETL) và user server (**EMR**)

**Di chuyển** và **biến đổi** dữ liệu giữa **Source** và **Destination**

**Crawls** dữ liệu nguồn và tạo ra **AWS Glue Data Catalog**

Data **Source: Storage**: S3, RDS, JDBC and Dynamo DB

Data **Source: Streams**: Kinesis Data Stream & Apache Kafka

Data **Targets**: S3, RDS, JDBC Databases



**Liên tục cập nhật metadata của các nguồn dữ liệu**

**One Catalog cho một region và cho một account**

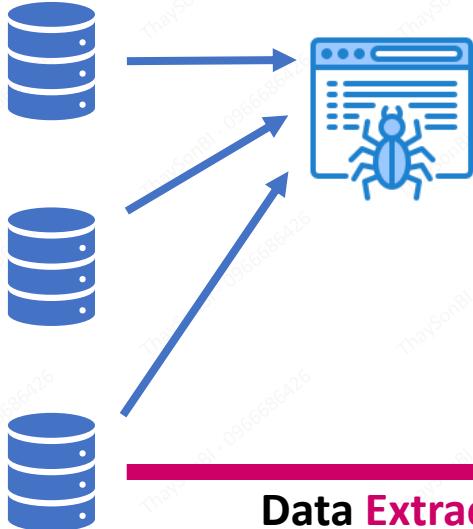
**Avoid data silos..**

Athena, Redshift, EMR, LakeFormation use **Data Catalog**

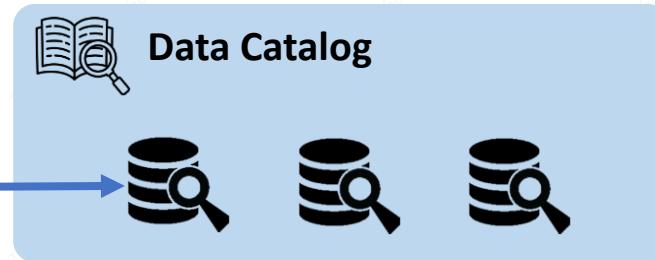
...bằng cách cài đặt **Crawler** cho data source



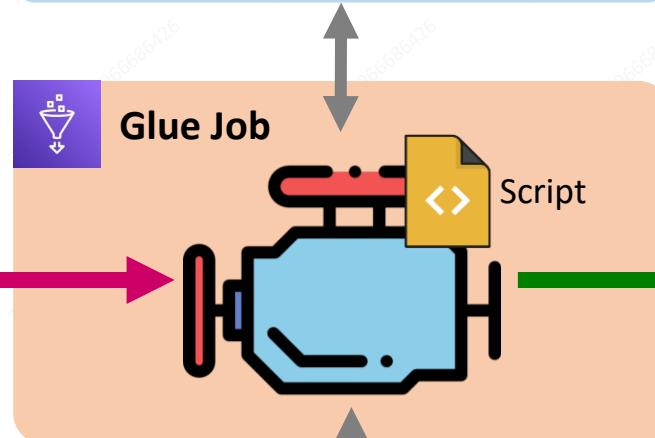
Crawler kết nối với **data source** để xác định **schema** và tạo ra **metadata** trong **data catalog**



Có thể truy cập và quản trị thông qua **AWS Console**



**Glue jobs** có thể được khởi động **manual** hoặc qua **EventBridge**



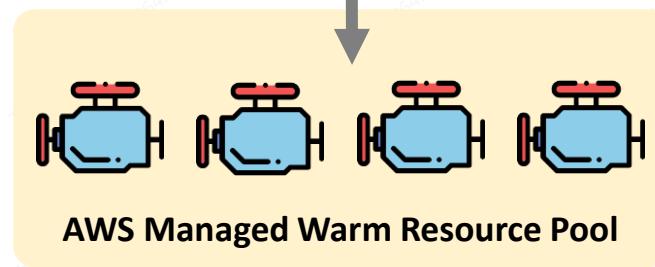
Data Load



Hỗ trợ các database đích **S3, RDS** và **JDBC** tương thích



Data source hỗ trợ bởi Glue bao gồm **S3, RDS, DynamoDB, Kinesis Data Stream** và **Kafka**



AWS Managed Warm Resource Pool

Khi resource được yêu cầu, glue phân phối từ **AWS warm pool** để thực hiện quá trình **ETL**



Easily

Step by Step

Hands On

# Crawler CSV from S3

Explain

ĐỀ MÔ

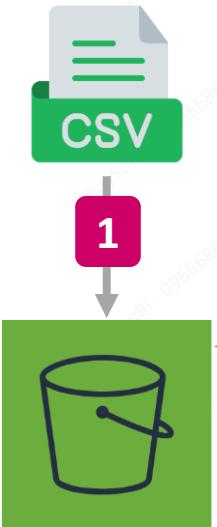
Practice



Amazon Web Service - Training



quiz\_0909.csv



2.1



role-glue-full-access = administrator

Glue



Crawler

Data Catalog

Ddbname = tsbi

2.2

Load schema of file to database

**bucket/folder/partition/file\_data**  
**02-report/quiz/year=2023/quiz\_0909.csv**



# AWS Athena



Analytics

## Query

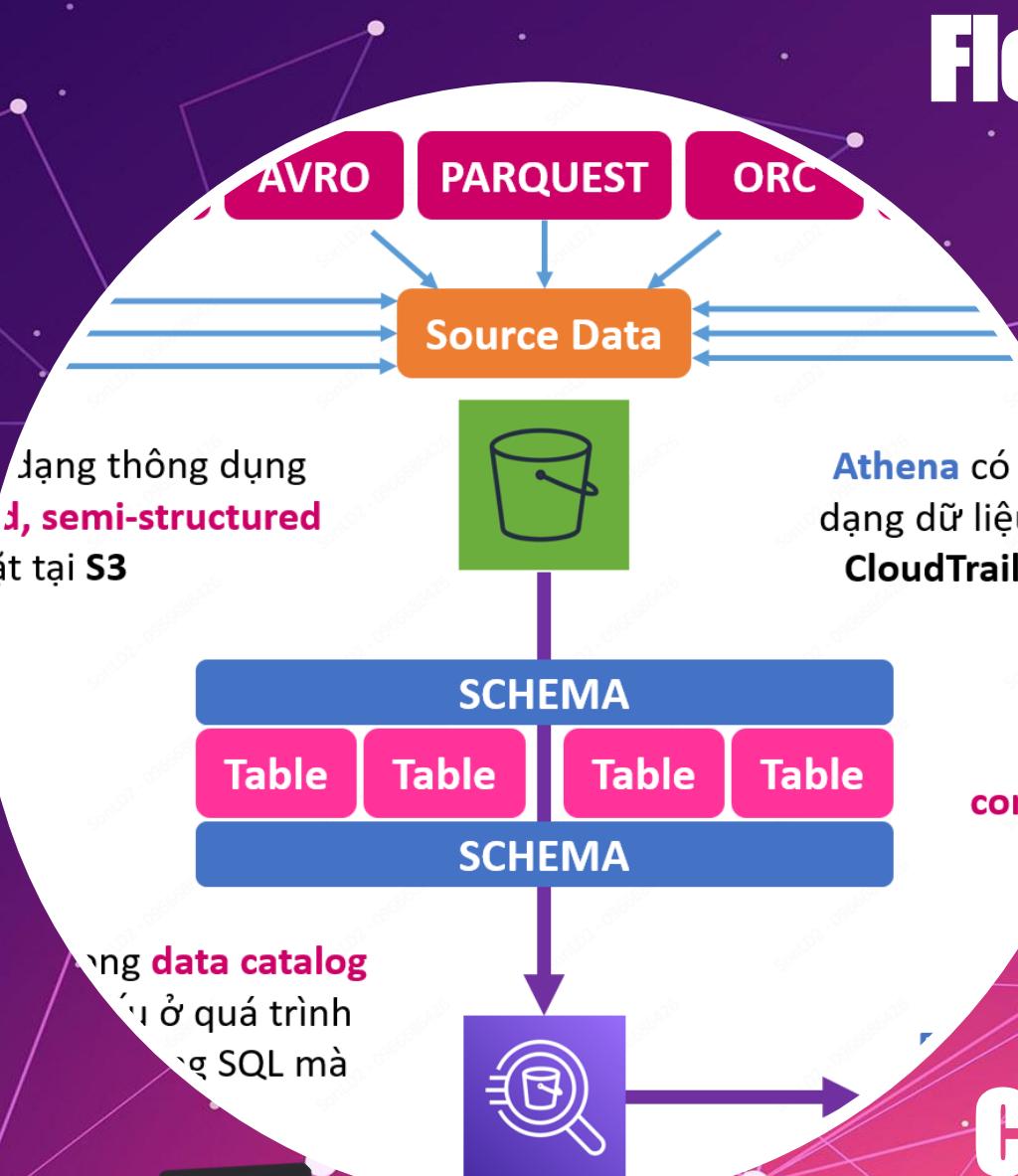


dạng thông dụng  
**semi-structured**  
tại S3

## Serverless



trong **data catalog**  
ở quá trình  
SQL mà



# Flexibility

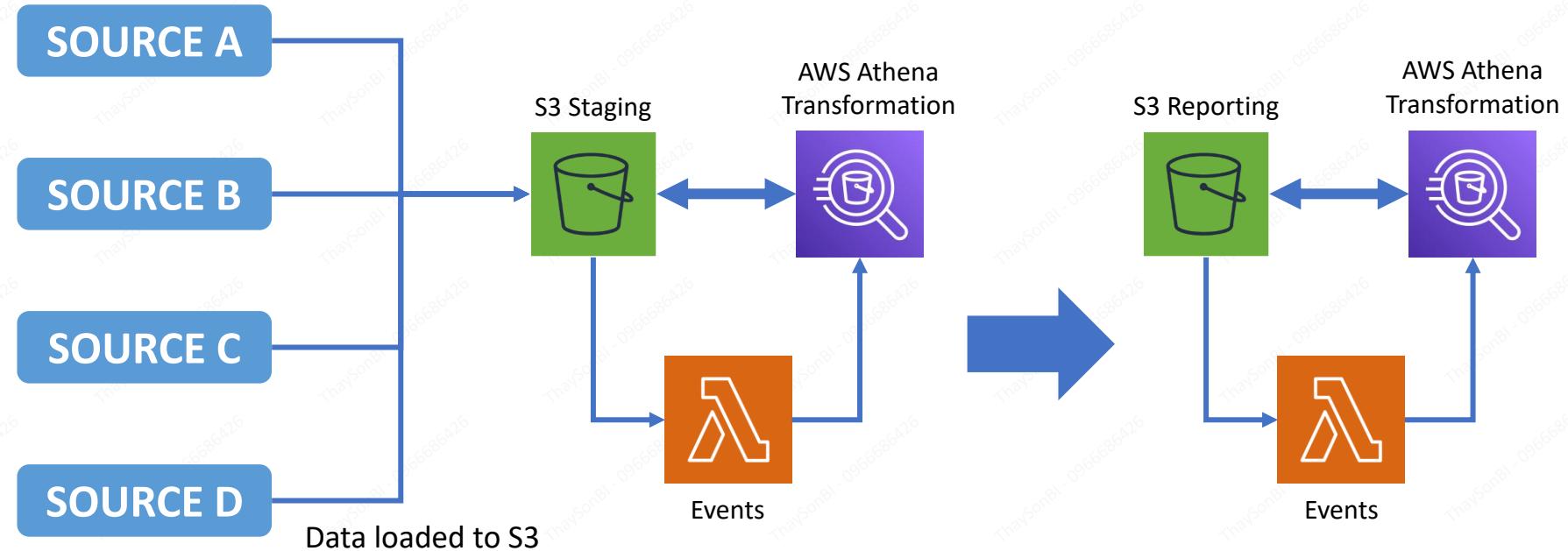
Athena có thể  
dạng dữ liệu hoặc  
CloudTrail, AW

Chi  
consu

Amazon Web Service - Training



Athena là một dịch vụ truy vấn trực tiếp làm quá trình **phân tích dữ liệu** trên S3 trở nên đơn giản và dễ thực hiện hơn với **SQL cơ bản**.





## Serverless Interactive Query Service

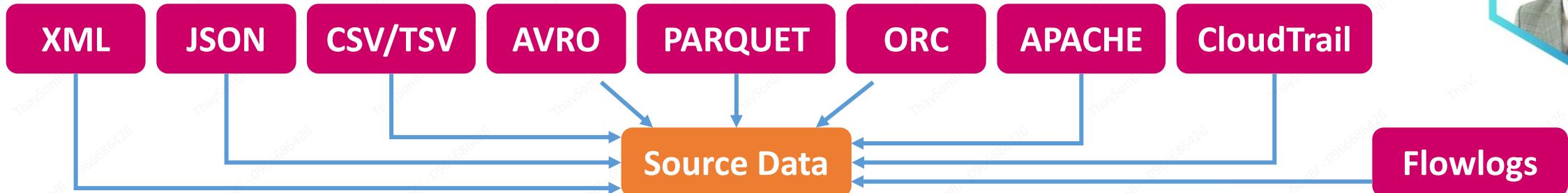
Truy vấn **Ad-hoc** và chỉ tính phí dựa trên **data consumed**

Dữ liệu được giữ **nguyên bản không thay đổi** trên S3

Structured data, unstructured, semi-structured và **Paralell**

**Schema** dịch dữ liệu, xử lý truy vấn **trong khi đọc**

Kết quả truy vấn có thể gửi cho **các dịch vụ khác**



Athena hỗ trợ các định dạng thông dụng **structured, unstructured, semi-structured** với nguồn đặt tại S3

Athena có thể đọc trực tiếp nhiều định dạng dữ liệu hoặc **AWS data format** như CloudTrail, AWS Flowlogs và ELB Logs

Table được xác định trước trong **data catalog** sau đó dữ liệu được tham chiếu ở quá trình **read**. Cho phép truy vấn dưới dạng SQL mà không **transform** dữ liệu nguồn

Chi phí dựa trên **data consumed** trong khi query

Dữ liệu đầu ra có thể được gửi cho các **công cụ visualization**



**Queries** không phải **loading/transforming**

**Serverless querying** - **Chi phí** được biết **chính xác**

Truy vấn đa dạng các loại **AWS logs** ...Flow logs, CloudTrail, ELB Logs, cost report...

**Athena federated query** kết hợp với **AWS lambda** để truy vấn tới các data source khác

Tích hợp với **AWS Glue Data Catalog**  
... và **Web Server Logs**



Easily

Step by Step

Hands On

# Read S3 File With Athena

ĐỀ MÔ

Explain

Practice



Amazon Web Service - Training

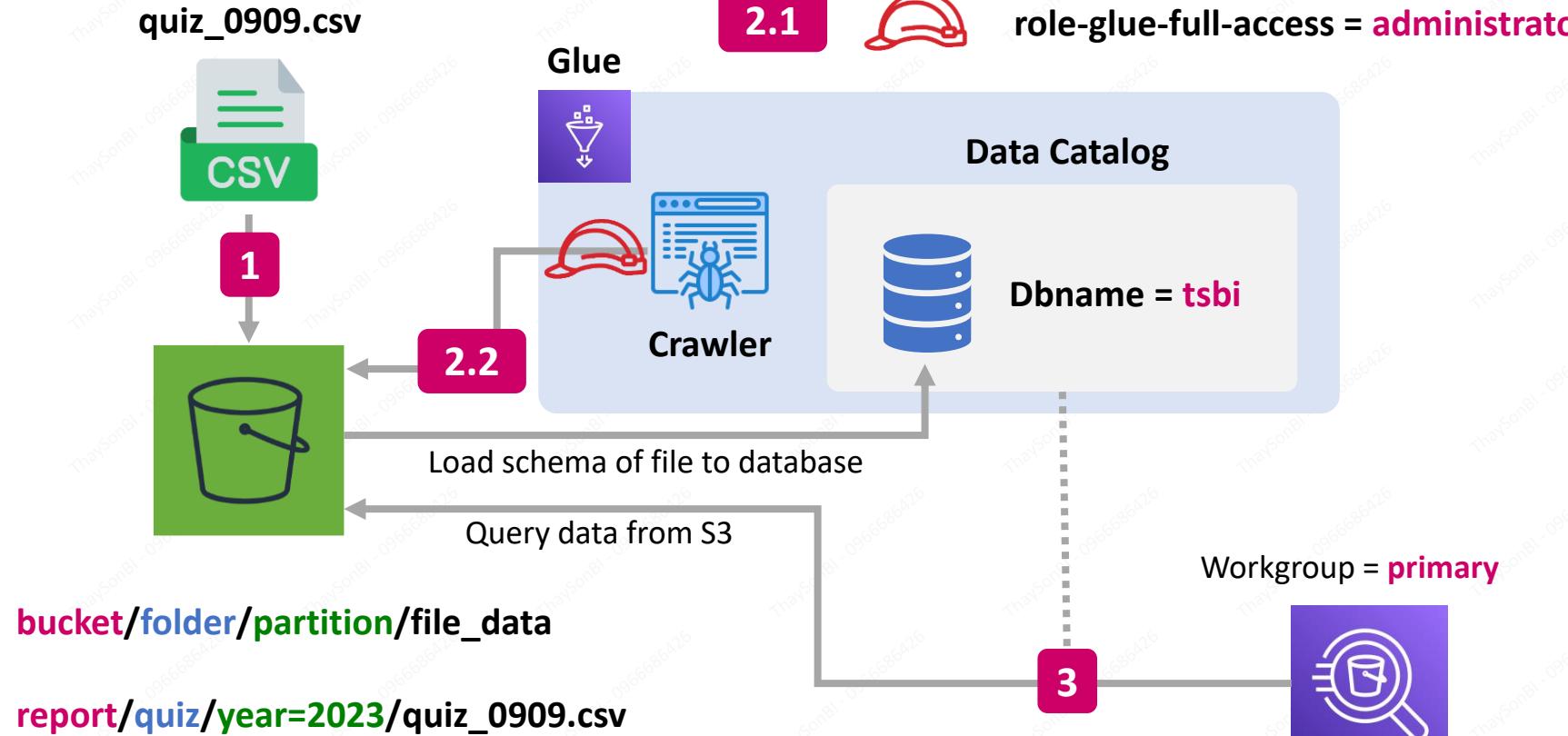


ANALYTICS  
SERVICES

# DEMO



AWS  
Athena





# AWS QuickSight

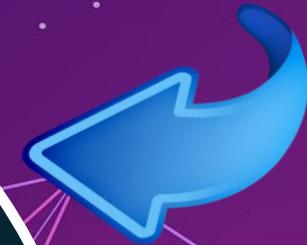
Analytics



# Dashboard

# Report

Easily



## Business Summary Dashboard

Segment

All

Category

All

All

### Amazon QuickSight

Sales Trend

Last 90 Days

YoY

MoM

\$3,762K

20.35%

↑

-19.45%

↓

1



Mobility

Amazon Web Service - Training



**Business Analytics & Business Intelligence service**

**Serverless**, trực quan hóa và phân tích Ad-hoc

**Khám phá và tích hợp** với các AWS Data source kể cả với các data source bên ngoài AWS khác

Ví dụ: **Dashboard** hoặc các biểu đồ trực quan

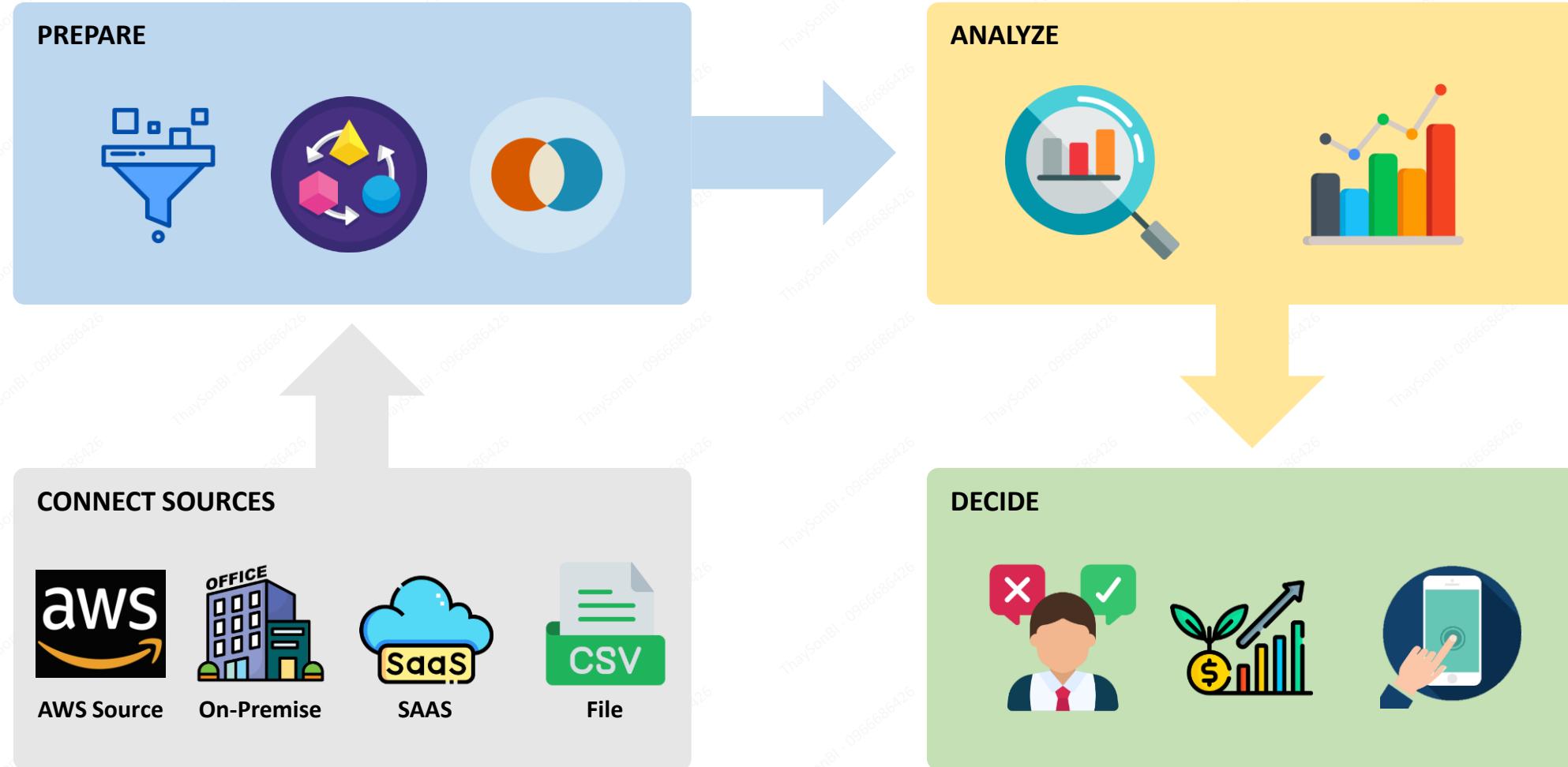


ANALYTICS  
SERVICES

# QUICKSIGHT WORK FLOW



AWS  
*QuickSight*





ANALYTICS  
SERVICES

# QUICKSIGHT SOURCE



AWS  
*QuickSight*



Athena, Aurora, Redshift, Redshift Spectrum

S3, AWS IOT

Jira, GitHub, Twitter, SaleForce

Microsoft SQL Server, MySQL, Postgre SQL

Apache Spark, Snowflake, Presto, Teradata



Easily

Step by Step

Hands On

# QuickSight Setup

Explain



Practice

Amazon Web Service - Training





# Create Basic Report

Step by Step

Hands On

Explain



Easily

Practice

Amazon Web Service - Training





ANALYTICS  
SERVICES

# PRACTICE ARCHITECTURE



AWS  
*QuickSight*

