



DATA INTEGRATION & FEDERATION

# Analytics in 15

## Accelerated ETL for Amazon Redshift with AWS Glue

Brian Dolan

Sr. Go-to-Market Specialist, AWS Glue  
Amazon Web Services

# Agenda

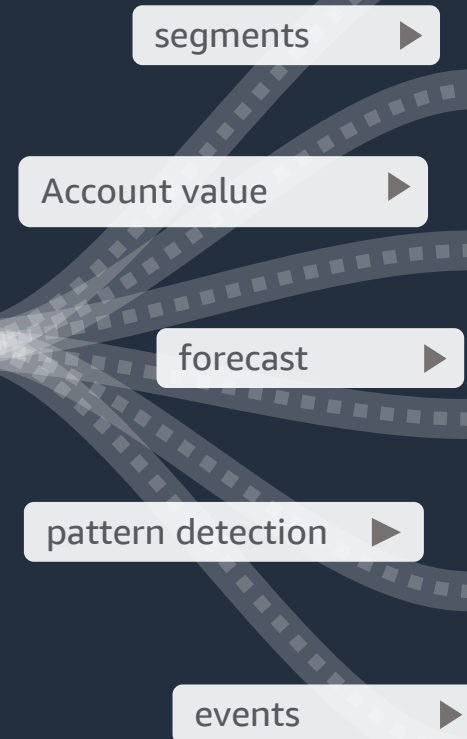
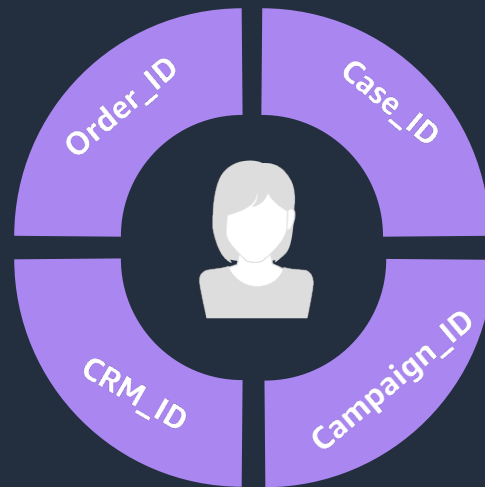
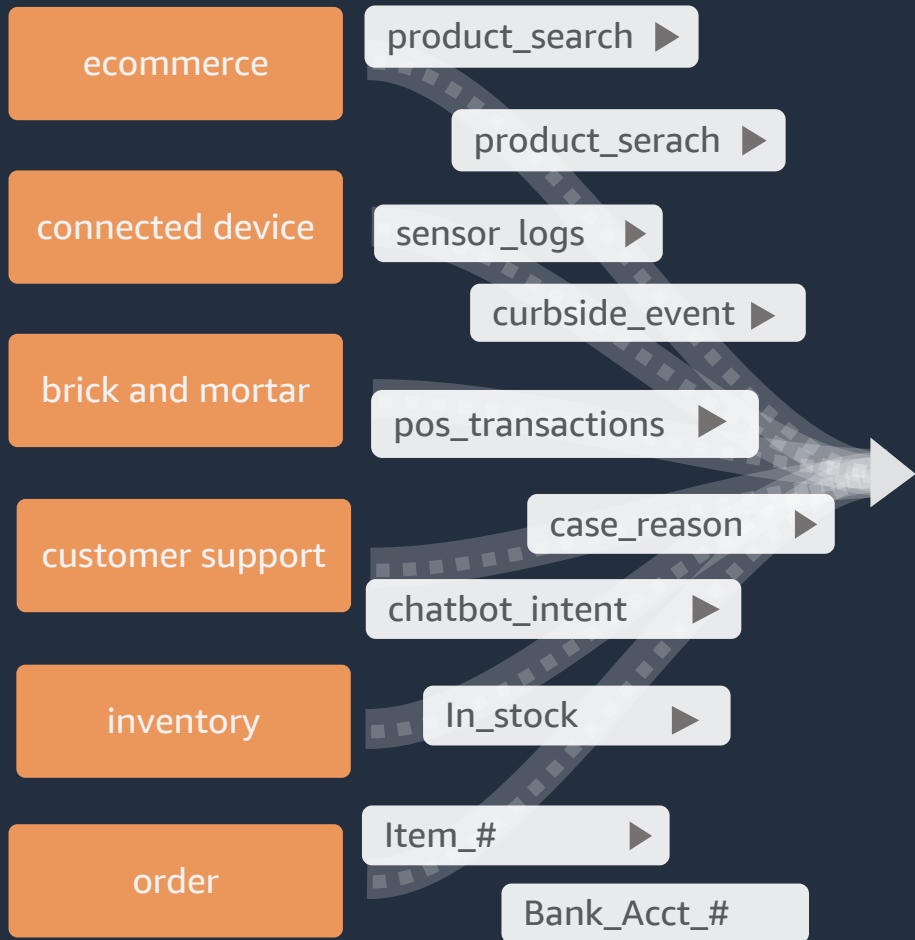
- Amazon Redshift and AWS Glue
- Redshift ETL with Glue Studio
- Demo
- Recap

# AWS Glue and Amazon Redshift

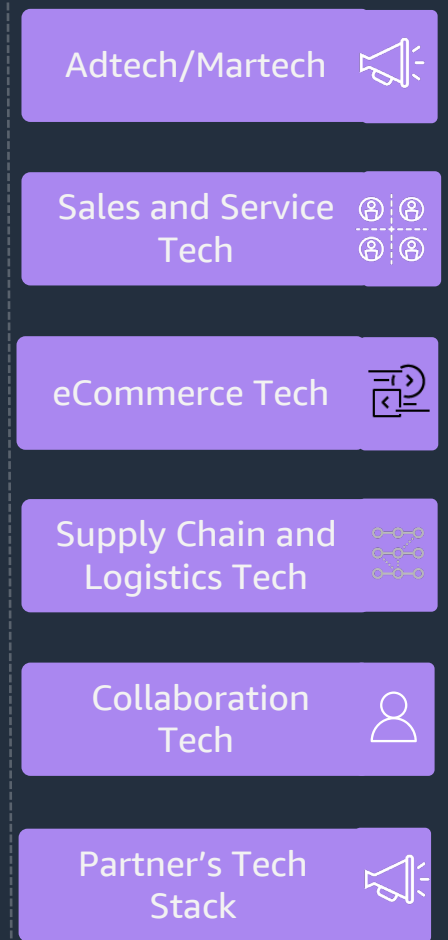


# Customer 360: Unify and Enrich Customer Data

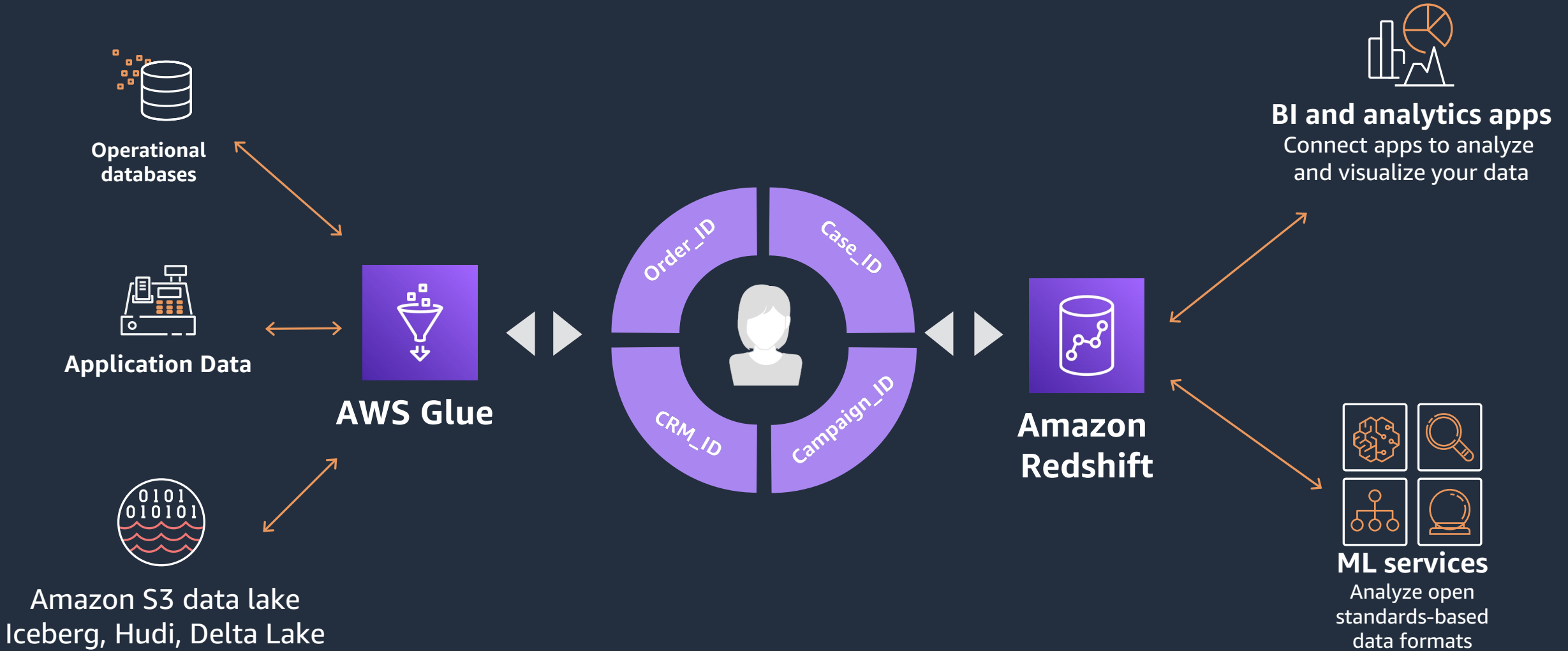
## Publishers



## Consumers



# Customer 360: Unify and Enrich Customer Data



# Redshift ETL with Glue Studio

# Glue Studio – Visual ETL



The screenshot displays the AWS Glue Studio interface for a data pipeline named 'data-pipeline'. The pipeline is visualized with three nodes: 'Data source - S3 bucket primaryData', 'Data source - S3 bucket lookupData', and 'Transform - Join joinedData'. The 'lookupData' node is connected to the 'primaryData' node, and both are connected to the 'Join' node. The right-hand side of the interface shows the 'Data source properties - S3' configuration panel, including fields for S3 URL, recursive options, data format (CSV), delimiter (Comma (,)), and escape/quote characters.

## Redshift ETL developer

Rich **visual interface**

**250+ built-in** transformations

Easy access to **90+ connectors**

Build end-to-end pipelines **without code**

# Get started quickly through direct access

**Node properties** | **Data source properties - Amazon Redshift** | Output schema | Data preview

Redshift access type

- Direct data connection - *recommended*
- Glue Data Catalog tables

Redshift connection

Choose the AWS Glue connection for Amazon Redshift, or [create a new connection](#).

redshift-demo-blog-connection

Connection: View properties | Database: dev

Redshift source

- Choose a single table  
Access data from a single Redshift table, found in a specific schema.
- Enter a custom query  
Access a custom dataset from multiple Redshift tables, found in one or more schemas.

Schema

Choose your Amazon Redshift schema.

public

Table

Search and enter the name of the source Amazon Redshift table.

Choose your Amazon Redshift table.

▶ Performance and security

▶ Custom Redshift parameters - optional

## New: DIRECT DATA CONNECTION

Browse Redshift schema and tables from Glue Studio

Auto-populated defaults for performance and security



# Customize your source using a SQL query

**Node properties** | **Data source properties - Amazon Redshift** | Output schema | Data preview

Direct data connection - *recommended*  
 Glue Data Catalog tables

**Redshift connection**  
Choose the AWS Glue connection for Amazon Redshift, or [create a new connection](#).

redshift-demo-blog-connection

Connection: View properties | Database: dev

**Redshift source**

Choose a single table  
Access data from a single Redshift table, found in a specific schema.

Enter a custom query  
Access a custom dataset from multiple Redshift tables, found in one or more schemas.

**Redshift query**  
Enter the Redshift query that will feed this source code.

```
1 select venue.venueid, venue.venuename from event, venue where event.venueid = venue.venueid
```

SQL Ln 1, Col 92 Errors: 0 Warnings: 0

Infer schema | Open Redshift query editor

NEW: Custom SQL query as a source

Access Redshift views or execute joins for a dynamic data source

ELT execution of Redshift SQL

Launch Redshift query editor from Glue Studio to copy & paste SQL

# Simplify common tasks

Select existing table or create new

NEW: Choose between common target actions like APPEND or MERGE

Use Glue Studio's built in options or enter your own custom

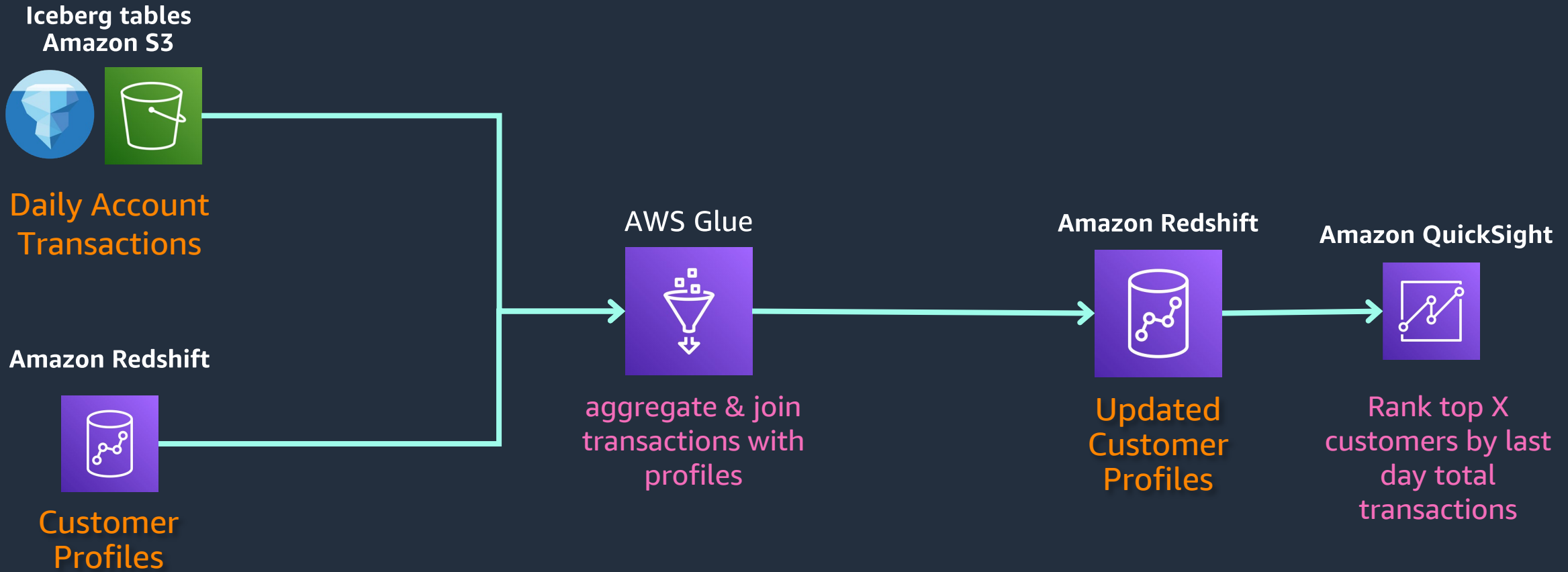
The screenshot shows the 'Data target properties - Amazon Redshift' configuration panel in AWS Glue Studio. It is divided into several sections:

- Node properties**: Includes tabs for 'Node properties', 'Data target properties - Amazon Redshift', 'Output schema', and 'Data preview'.
- Choose your Amazon Redshift schema.**: A dropdown menu with 'public' selected and a refresh button.
- Table**: A search box with 'venue' entered and a refresh button.
- Handling of data and target table**: A list of radio button options:
  - APPEND (insert) to target table: AWS Glue will append data to existing columns of the table and discard any extra columns.
  - MERGE data into target table: AWS Glue will either update or append data to the table based on a set of conditions.
  - TRUNCATE target table: Same as Append, except AWS Glue will first clear the contents of the table.
  - DROP and recreate target table: AWS Glue will delete and recreate the table with the schema from the source data.
- Choose keys and simple actions**: A button that is currently selected.
- Enter custom MERGE statement**: A button for entering a custom statement.
- Matching keys**: A dropdown menu with 'Choose one or more fields' selected.
- When matched**: A list of radio button options:
  - Update record in the table with data from source
  - Delete record in the table
- When not matched**: A list of radio button options:
  - Insert source data as a new row into the table
  - Do nothing

# Demo

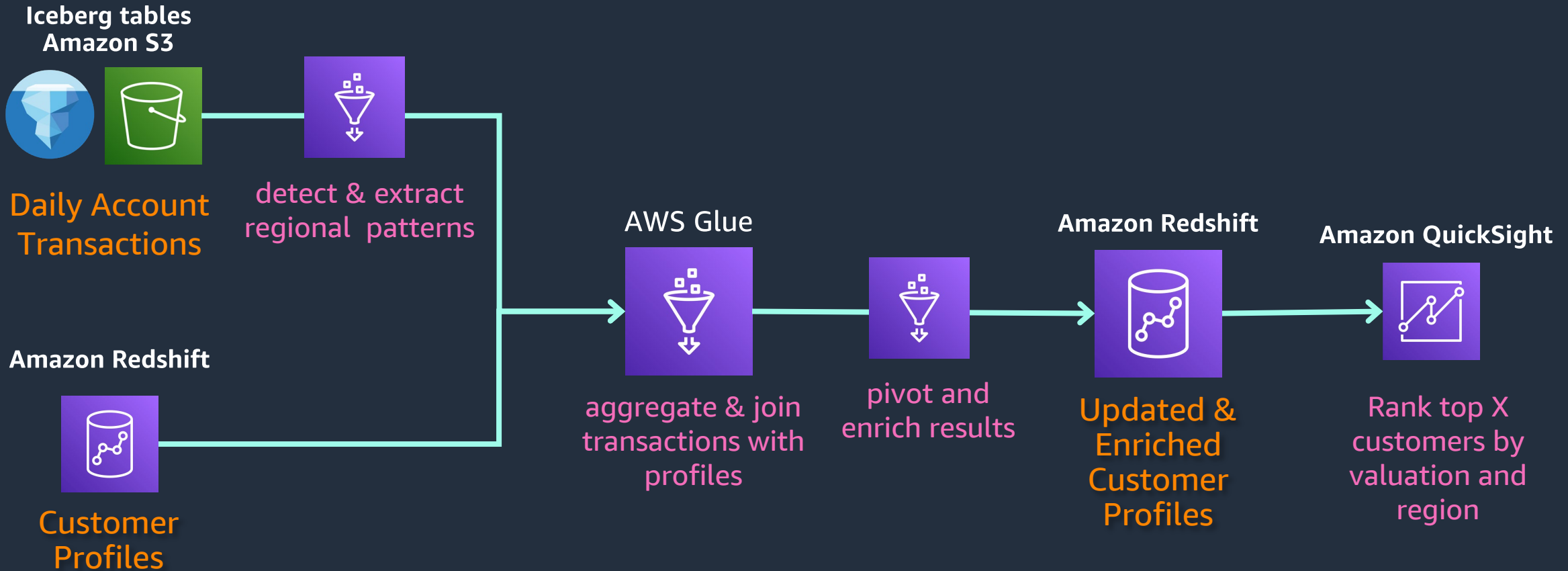


# Update and Enrich Customer Profiles



**[Video 1 – 00:00 to 03:27]**

# Update and Enrich Customer Profiles



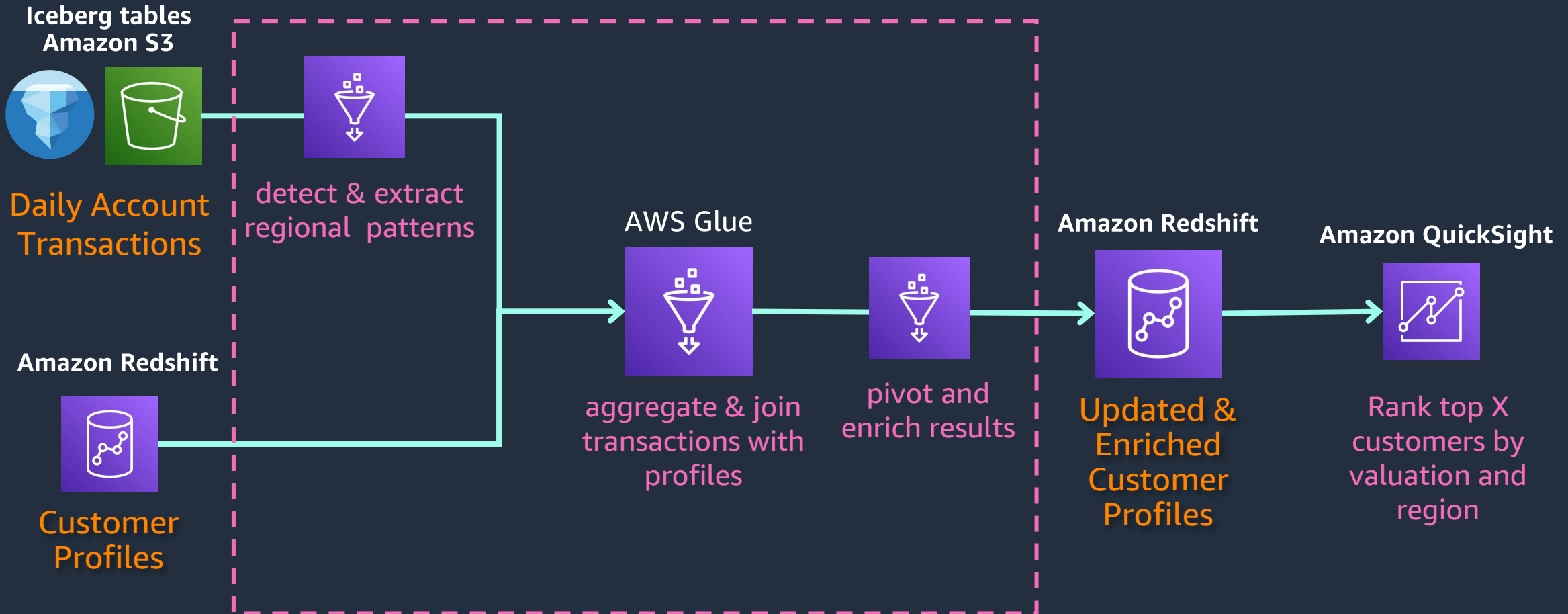
# [Video 2 - 00:00 to 02:36]

# RECAP





# Update and Enrich Customer Profiles



**AWS Glue Studio : Author, Execute, Schedule and Monitor**



# Thank you!

To learn more, please visit:

<https://aws.amazon.com/glue/>

<https://aws.amazon.com/redshift/>

<https://aws.amazon.com/blogs/big-data/exploring-new-etl-and-elt-capabilities-for-amazon-redshift-from-the-aws-glue-studio-visual-editor/>



# Thank you!

To learn more, please visit:

<https://aws.amazon.com/big-data/datalakes-and-analytics/>