



DATA INTEGRATION & FEDERATION

# Analytics in 15

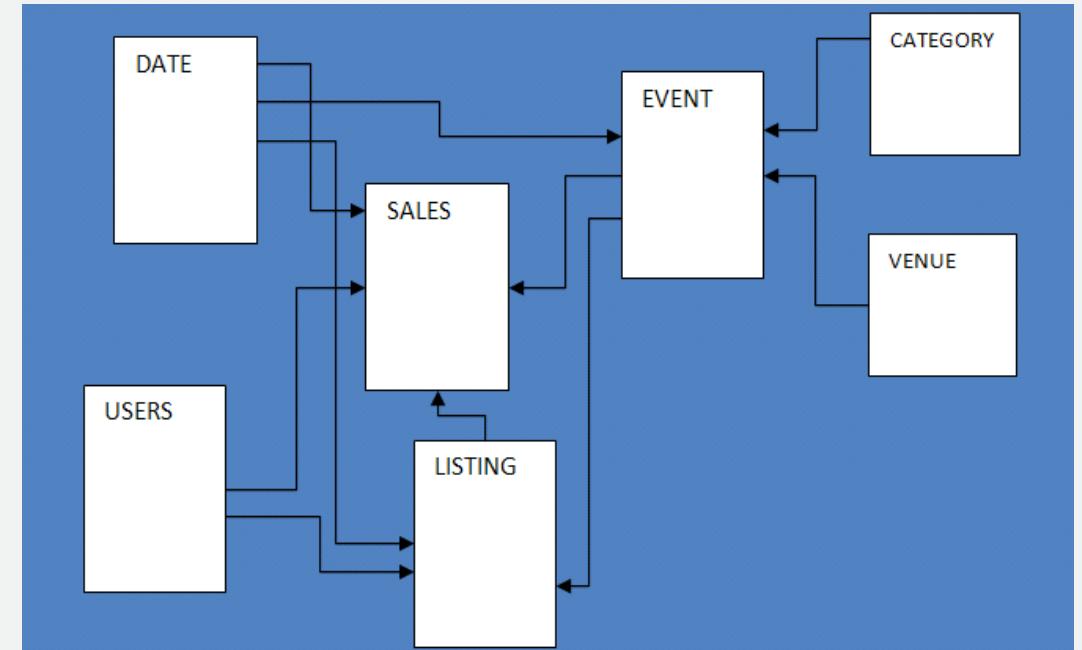
## Build No-code/Low-code Data Pipelines with Amazon Redshift

Runhong (Ronnie) Deng

Sr. Data Architect  
Amazon Web Services




# Agenda

- Existing capabilities
  - Federated queries
  - Materialized View
  - Data Sharing
- New features
  - Auto-copy (Preview)
  - Merge Command (Preview)
  - Dynamic Data Masking (Preview)



# Use case

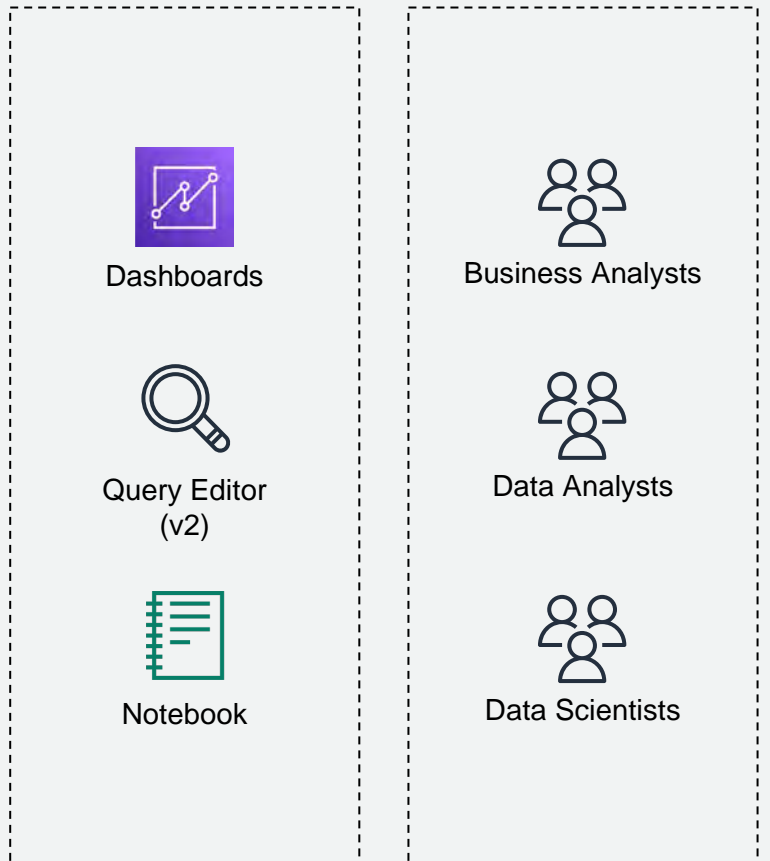
## Data Sources

-  Operational databases
-  Data lake
-  Data Streams

## Data Pipelines



## Data Consumptions



# Three existing features

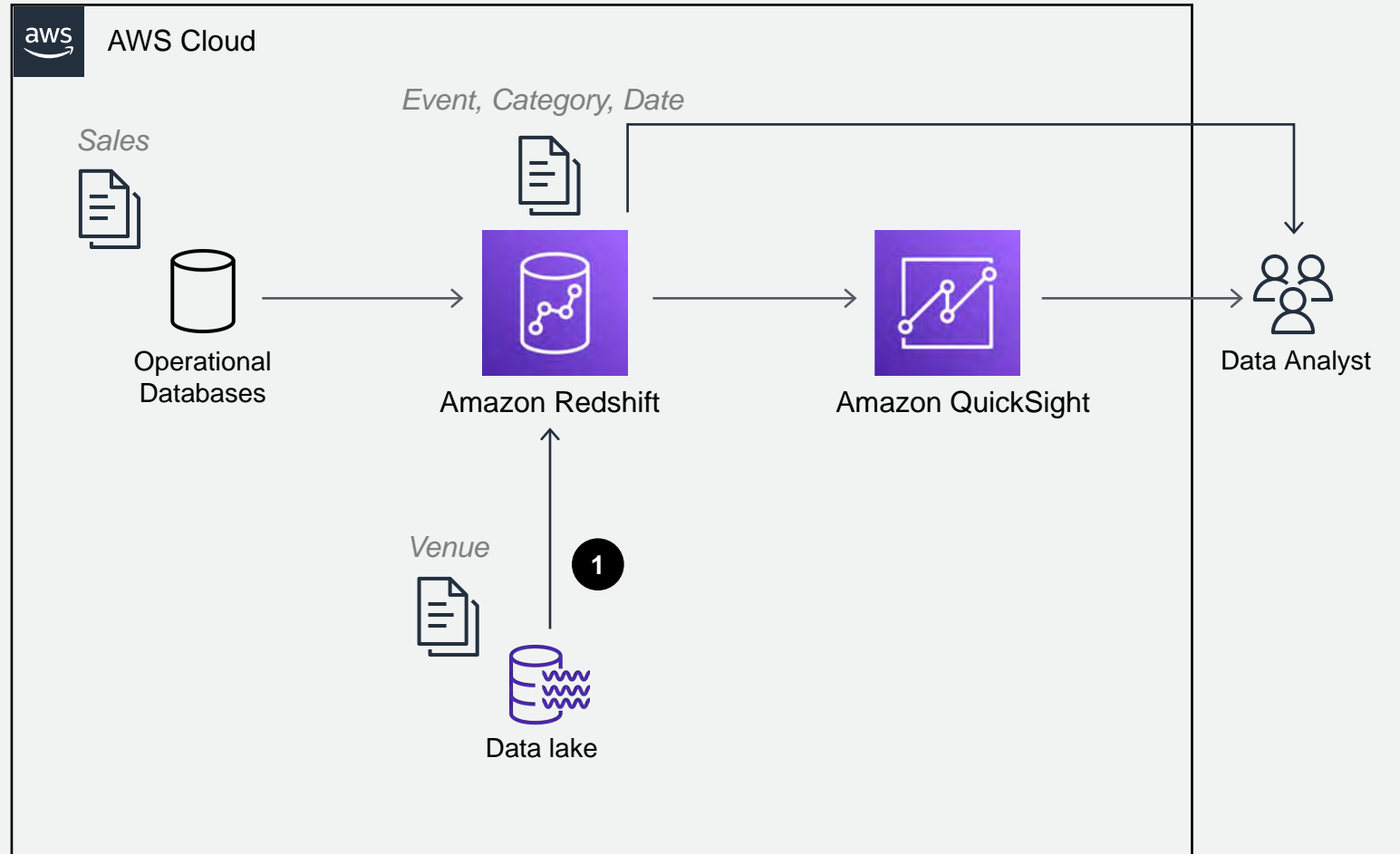
# 1 - Federated Queries

## CREATE EXTERNAL SCHEMA

```
s3_tickit_preprocessed
FROM DATA CATALOG
DATABASE
'redshift_demo_preprocessed_db'
REGION 'us-east-1'
IAM_ROLE
'arn:aws:iam::<ACCOUNT_ID>:role
/service-role/AmazonRedshift-
DemoAccessRole';
```

## GRANT USAGE ON SCHEMA

```
s3_tickit_preprocessed
TO ROLE dev_tickit_ro;
```



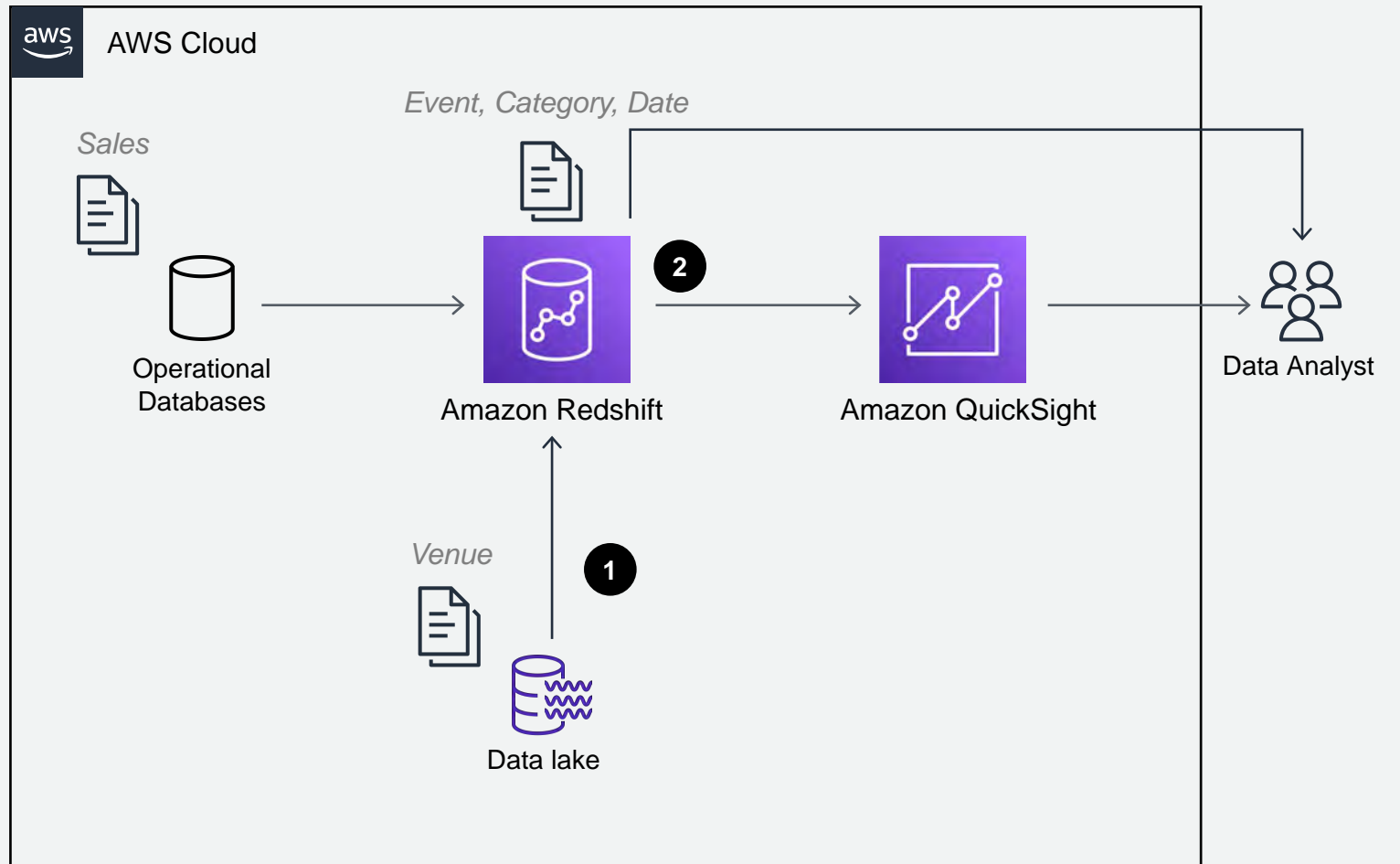
## 2 - Materialized View

### CREATE MATERIALIZED VIEW

```
"dev"."ticket".mv_event_details
```

```
AUTO REFRESH YES
```

```
AS SELECT e.eventid, e.eventname,  
c.catname,  
v.col1 as venueid, v.col2 as  
venuecity, v.col3 as venuestate,  
v.col4 as venueseats  
FROM "dev"."ticket"."event" as e,  
"dev"."ticket"."category" as c,  
"dev"."s3_ticket_preprocessed"."venu  
e" v  
WHERE c.catid=e.catid  
and e.venueid=v.col0
```



# 3 - Data Sharing

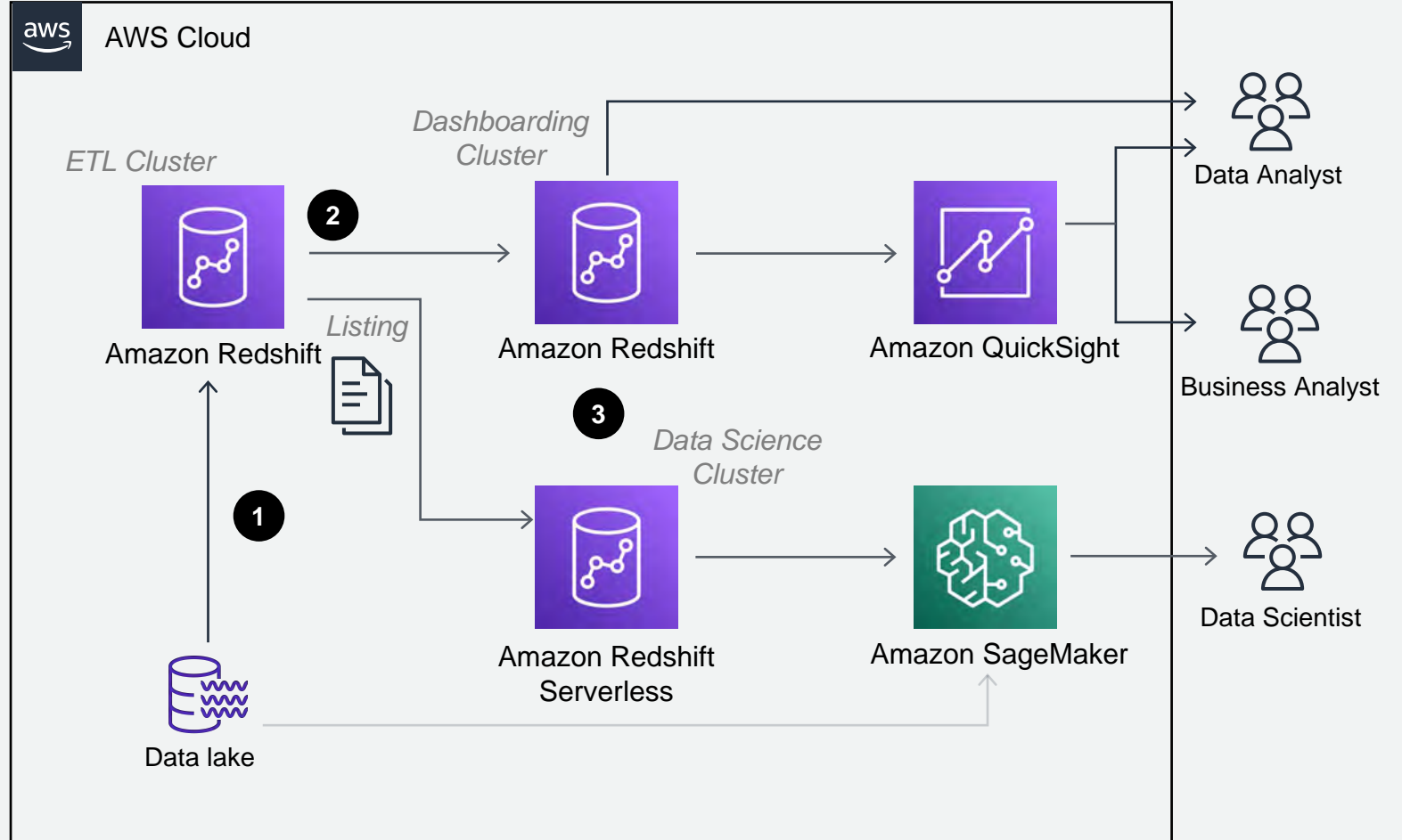
```
CREATE DATASHARE
listing_share SET
PUBLICACCESSIBLE TRUE;

ALTER DATASHARE listing_share
ADD SCHEMA public;

ALTER DATASHARE listing_share
ADD TABLE public.listing;

GRANT USAGE ON DATASHARE
listing_share to NAMESPACE
'<<consumer namespace>>';

-- Run on consumer to get
<<consumer namespace>>
SELECT current_namespace;
```



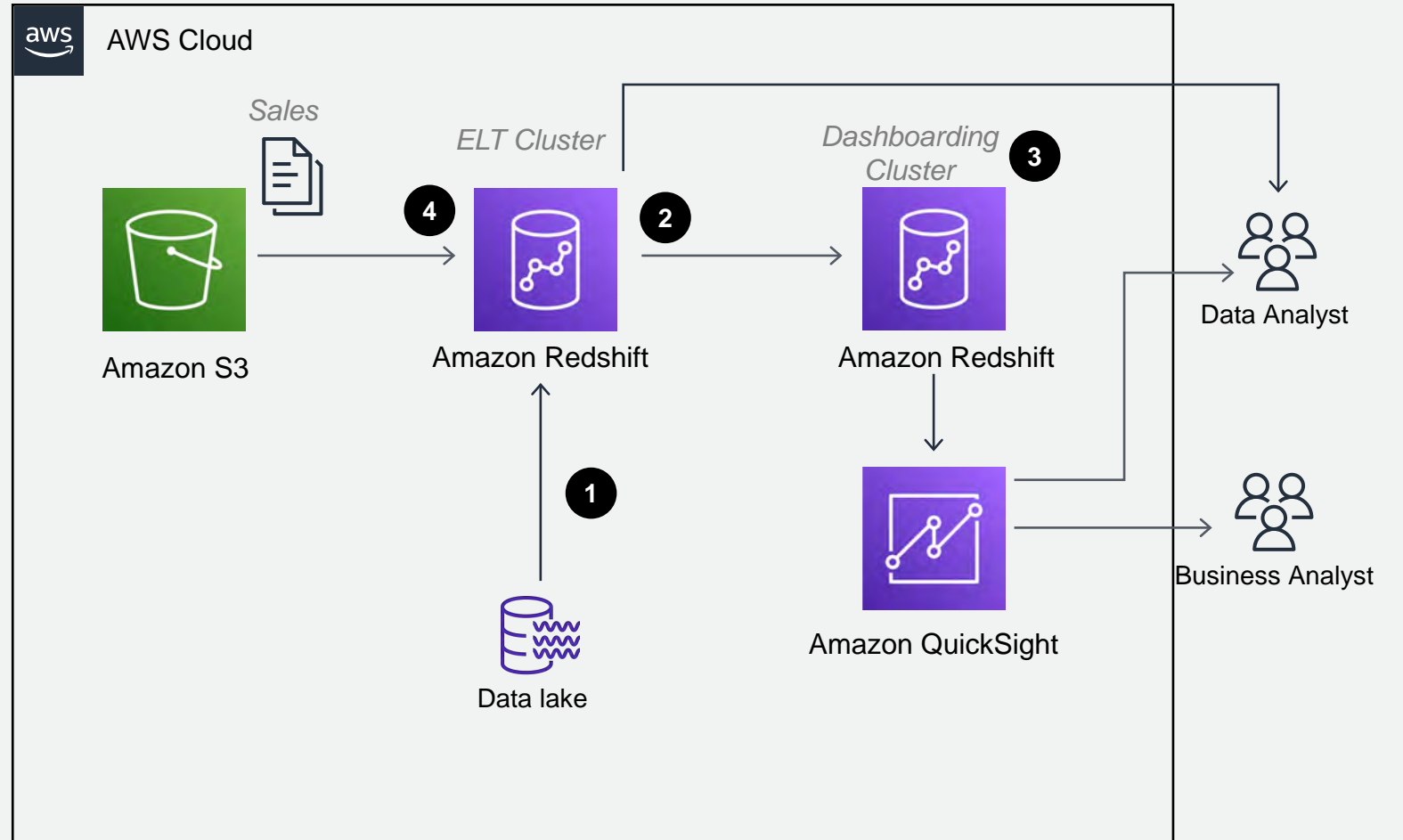
# Three new features (Preview)



# 4 - Auto-copy (Preview)

Simplify data loading from Amazon S3 into Amazon Redshift.

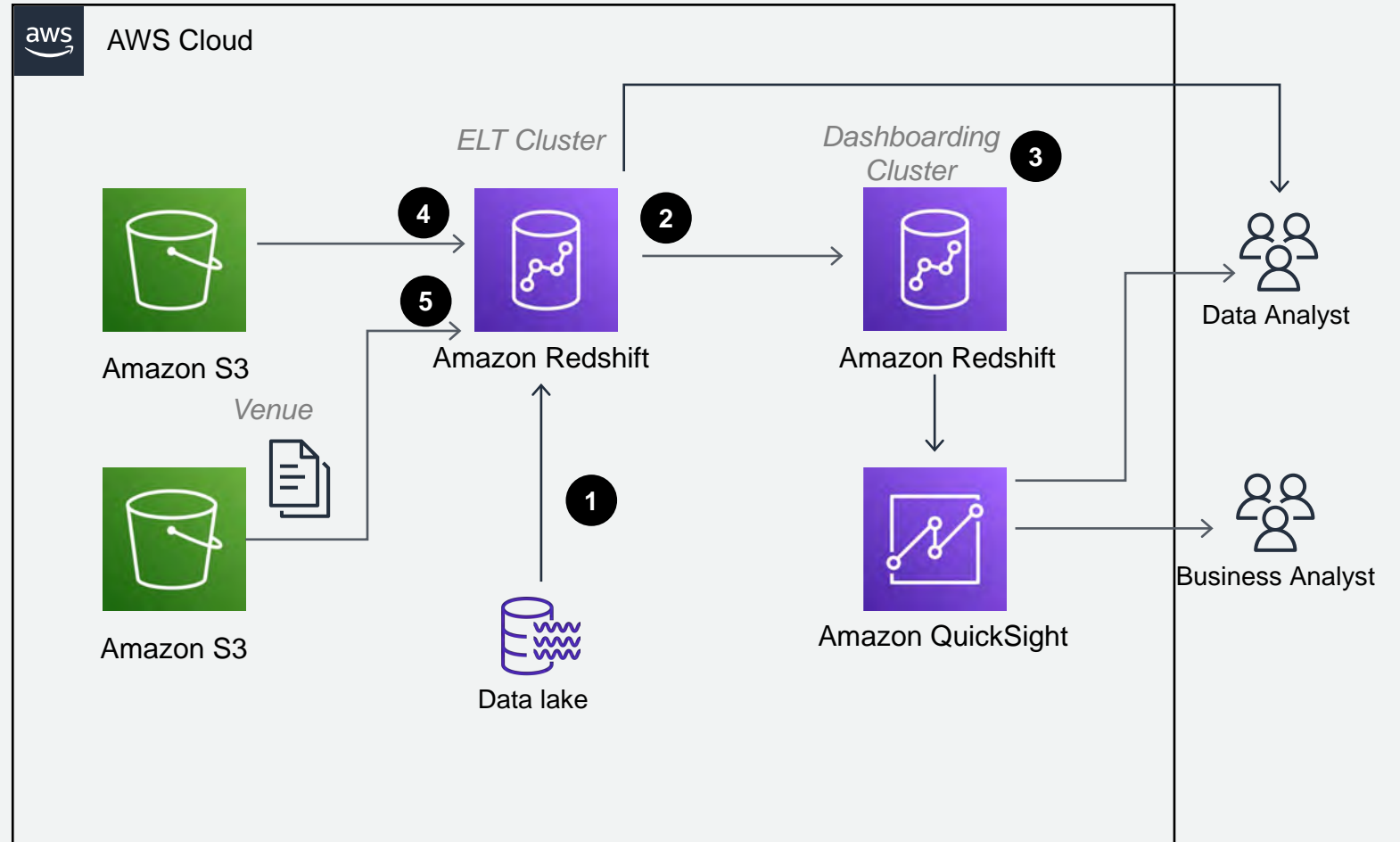
Continuous file ingestion without the need for additional tools or custom solutions.



# 5 - Merge (Preview)

Conditionally merges rows from a source table into a target table.

Traditionally, this can only be achieved by using insert, update or delete statement separately.

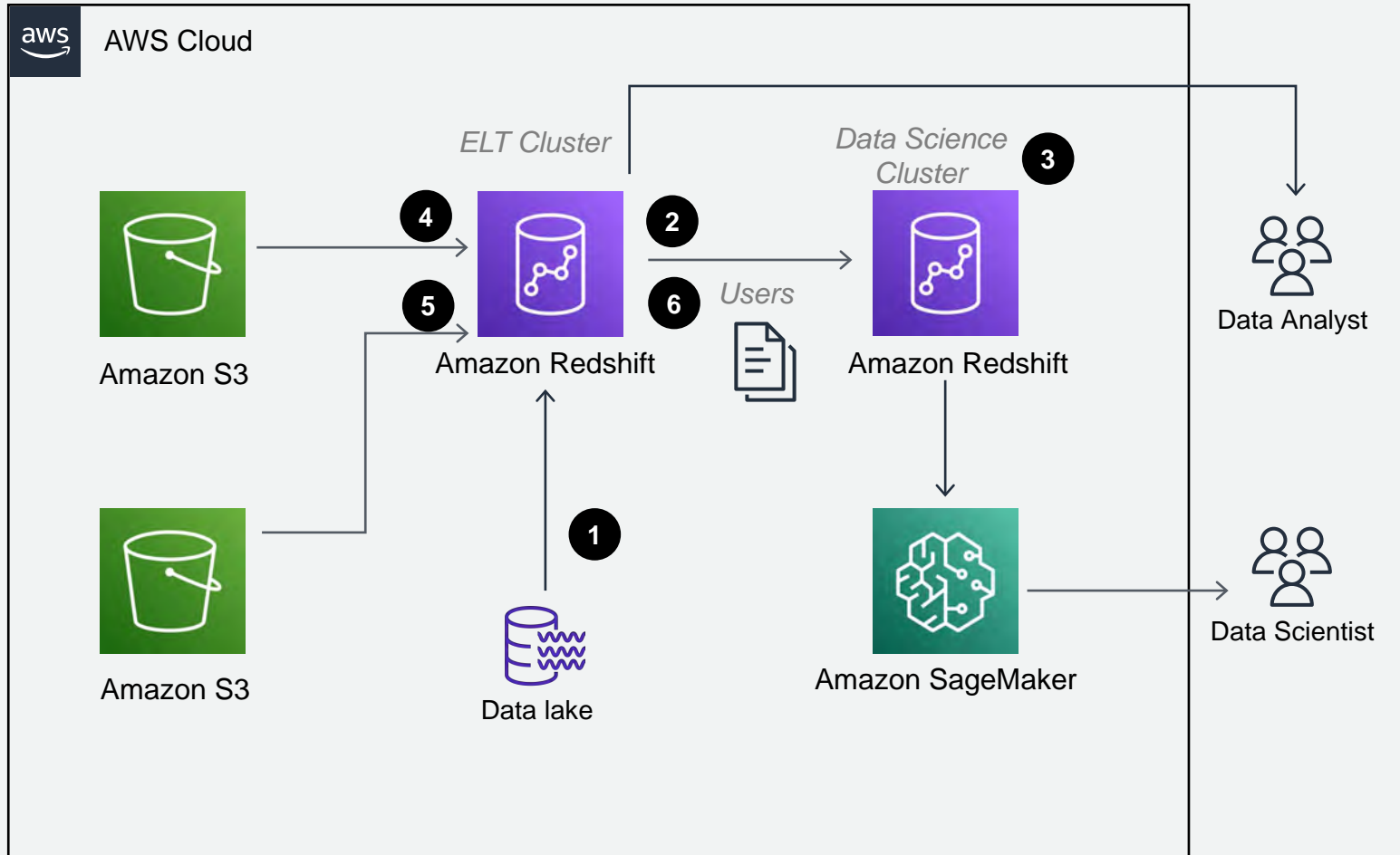


# 6 - Dynamic Data Masking (Preview)

Protect sensitive data in your data warehouse.

Manipulate how Amazon Redshift shows sensitive data to the user at query time without transforming it in the database.

Control access to data through masking policies that apply custom obfuscation rules to a given user or role.

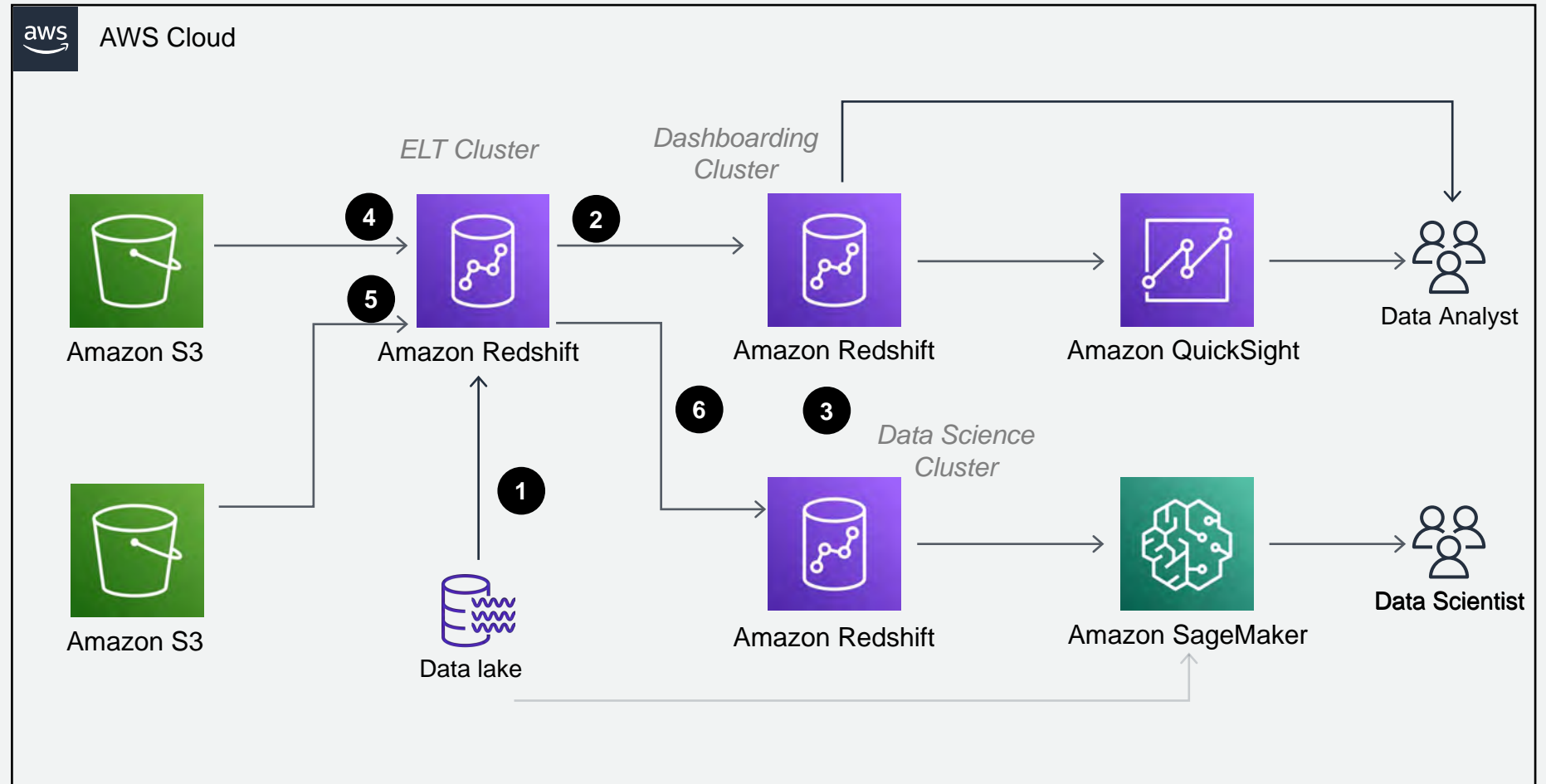


# Demo



# Summary

- Ingestion:**
  1. Federated Queries
  4. Auto-copy (Preview)
- Processing**
  2. Materialized View
  5. Merge (Preview)
- Servicing**
  3. Data Sharing
  6. Dynamic Data Masking (Preview)
- Many more  
... ..



# Additional Resources

## Ingestion

- <https://docs.aws.amazon.com/redshift/latest/dg/federated-overview.html>
- [https://docs.amazonaws.cn/en\\_us/redshift/latest/dg/loading-data-copy-job.html](https://docs.amazonaws.cn/en_us/redshift/latest/dg/loading-data-copy-job.html)

## Processing

- <https://docs.aws.amazon.com/redshift/latest/dg/materialized-view-overview.html>
- [https://docs.aws.amazon.com/redshift/latest/dg/r\\_MERGE.html](https://docs.aws.amazon.com/redshift/latest/dg/r_MERGE.html)
- <https://aws.amazon.com/blogs/aws/new-amazon-redshift-integration-with-apache-spark/>

## Servicing

- <https://docs.aws.amazon.com/redshift/latest/dg/datashare-overview.html>
- [https://docs.aws.amazon.com/redshift/latest/dg/t\\_ddm.html](https://docs.aws.amazon.com/redshift/latest/dg/t_ddm.html)
- <https://aws.amazon.com/about-aws/whats-new/2022/11/amazon-redshift-data-sharing-centralized-access-control-lake-formation-preview/>



# Thank you!

To learn more, please visit:

- ✓ Amazon Redshift webpage

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

- ✓ Amazon Redshift new feature page

<https://aws.amazon.com/redshift/whats-new/>

- ✓ Analytics webpage

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