
AWS
re:Invent
NOV. 28 – DEC. 2, 2022 | LAS VEGAS, NV

ANT220

Using Amazon AppFlow to break down data silos for analytics and ML

Kamen Sharlandjiev

Sr. Analytics Specialist SA
WWSO Analytics

Gaurav Gupta

Sr. Product Manager - Technical
Amazon AppFlow

Ray Jang

Principal Product Manager - Technical
Amazon AppFlow

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Agenda

Introduction to Amazon AppFlow

Demo

Breaking down data silos for Analytics and ML

Simplifying data integration for analytics and ML

Summary

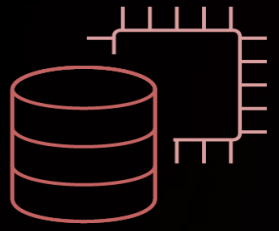
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon AppFlow

Fully managed integration service that enables you to securely transfer data between SaaS applications like Salesforce, SAP, Zendesk, Slack, and ServiceNow, and AWS services like Amazon S3 and Amazon Redshift; perform simple data transformations; and integrate these applications with AWS services, in just a few clicks

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Top integration needs of customers



Batch processing of data



Event-driven processing of data



Robust set of connectors to data sources



Big data analytics & machine learning



Data synchronization across apps & services



Secure & private data access and exchange

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

3 key challenges our customers face today

Data is stored in silos



- Data is stuck in a variety of sources (on premises/SaaS/AWS)
- Managing Servers on premises, Amazon EC2 or multiple cloud & SaaS applications

Unlocking value of data



- Need for gathering insights from variety of sources
- It is challenging to extract useful information when data cannot be analyzed collectively

Keeping data in sync is complicated and expensive



- Third-party software is expensive
- Customers need to manage infrastructure, pay for idle time and provision for excess capacity

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Key Benefits with Amazon AppFlow



No / Low Code

Build data pipeline in a few clicks using console or through APIs – easy to use for business users and developers.



Cost Savings

With zero licensing fee, Amazon AppFlow offers significant cost-savings over building connectors in-house or using other third-party services



Speed & Agility

Create data flows in minutes – easy to reconfigure to meet the ever-changing needs of the business



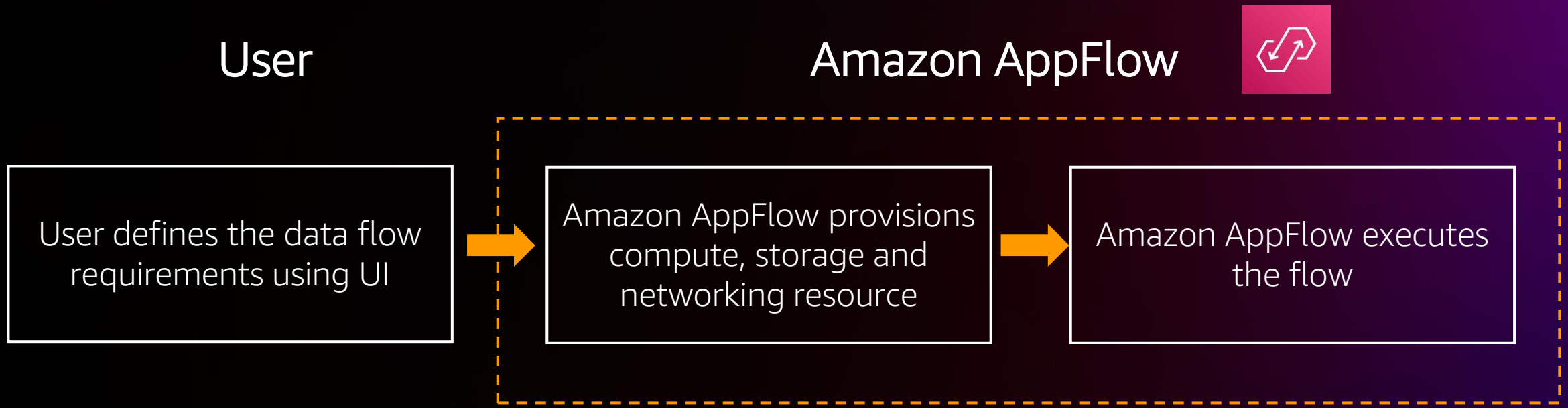
Secure & Scalable

Transfer millions of records over AWS PrivateLink with supported SaaS partners and prevent exposure of data to the public internet

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

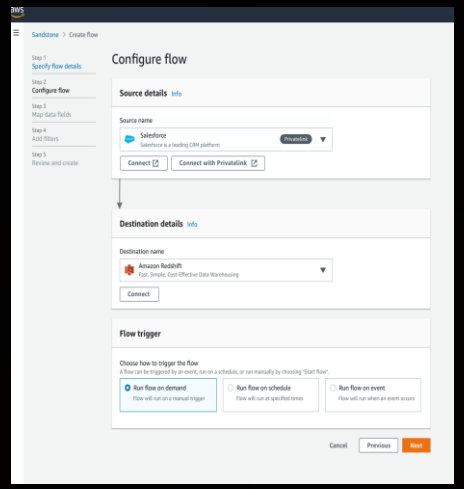
How it works?



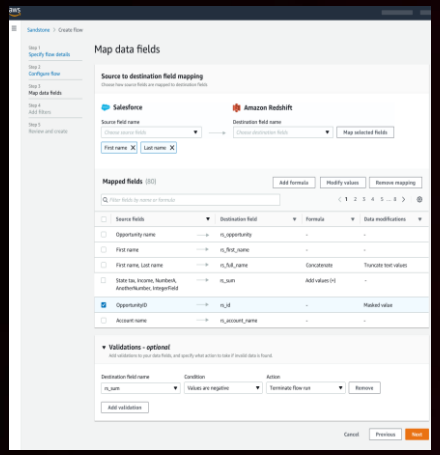
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon AppFlow allows you to create a flow in minutes!

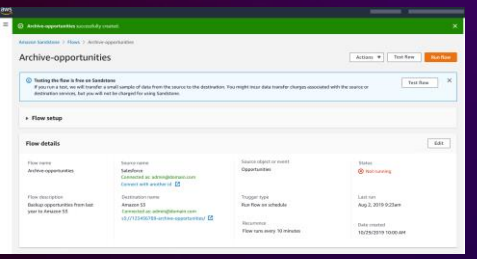
Select source and destination; login → Specify flow trigger → Map fields from source to destination → Add filters, validation, transformation → Activate your flow or run it at the click of a button



- Event Based
- Scheduled (1 minute minimum)
- On-demand (run flow immediately)



- Filter records
- Mask sensitive fields
- Combine fields to create new ones
- Validate records
- Truncate fields



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Why is it important to talk about data silos?



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Customers' data requirements are changing



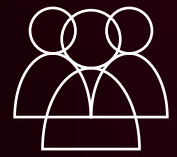
Growing exponentially



From new sources



Increasingly diverse



Used by many people

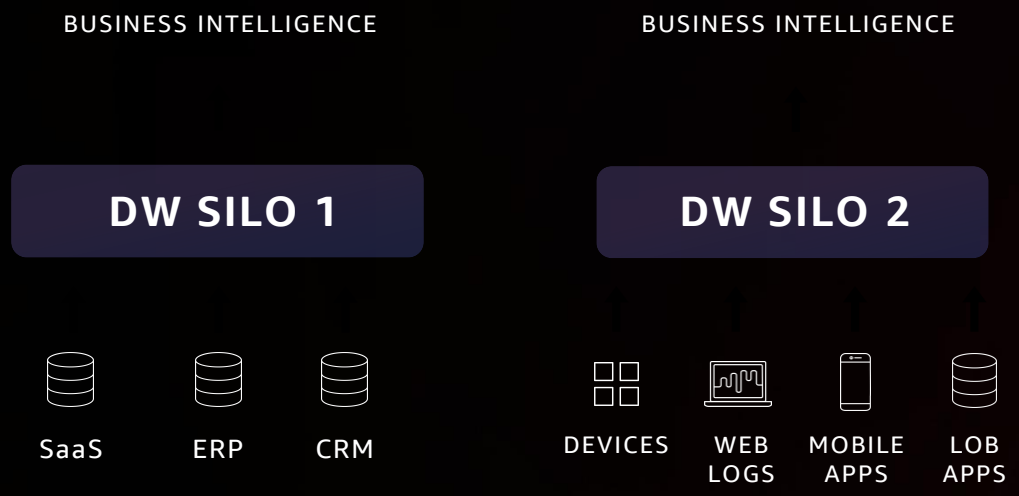


Analyzed with ML and by other applications

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

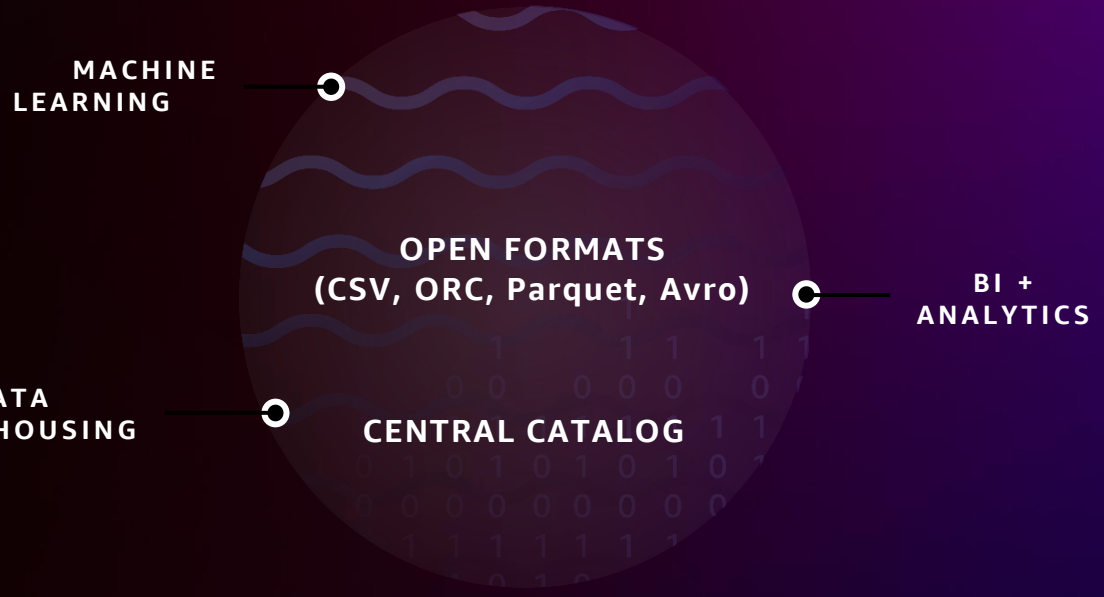
Breaking down data silos

Data silos



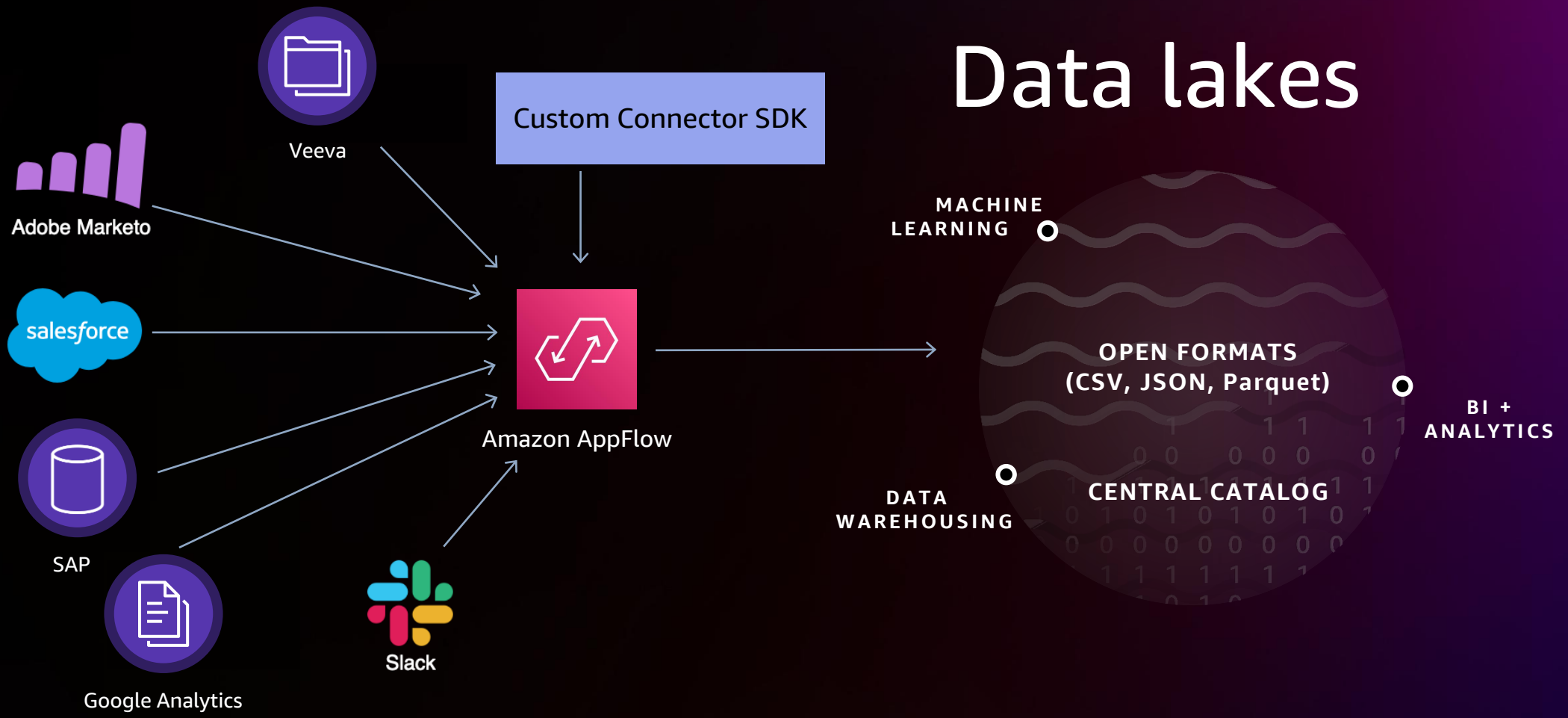
to

Data lakes



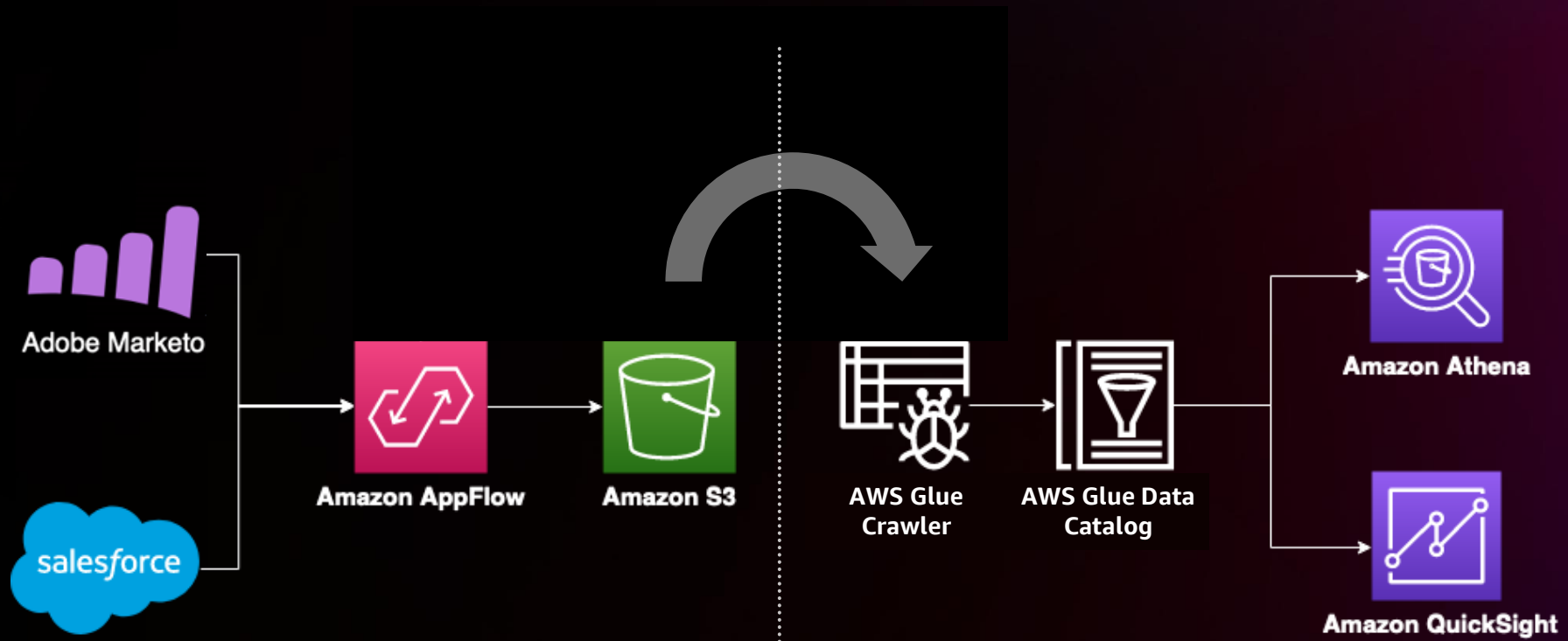
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Breaking data silos



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

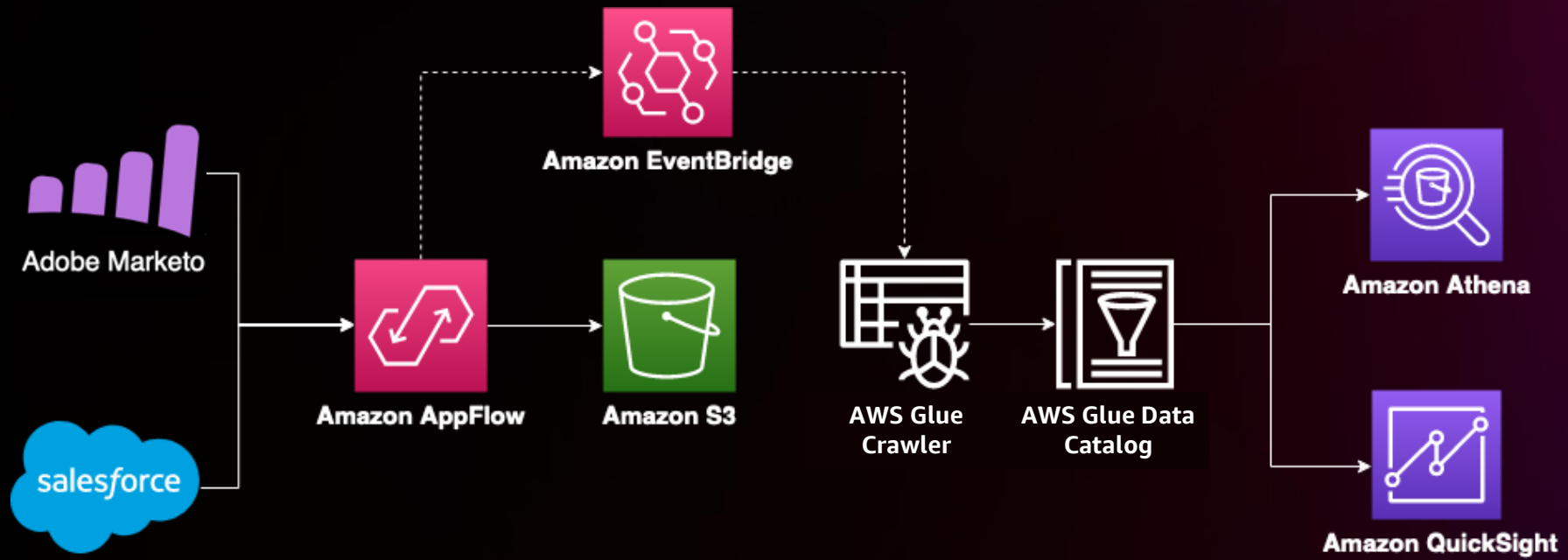
The manual boundary



- Manually create
- Manually run
- Manually correct schema

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Why integration matters



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

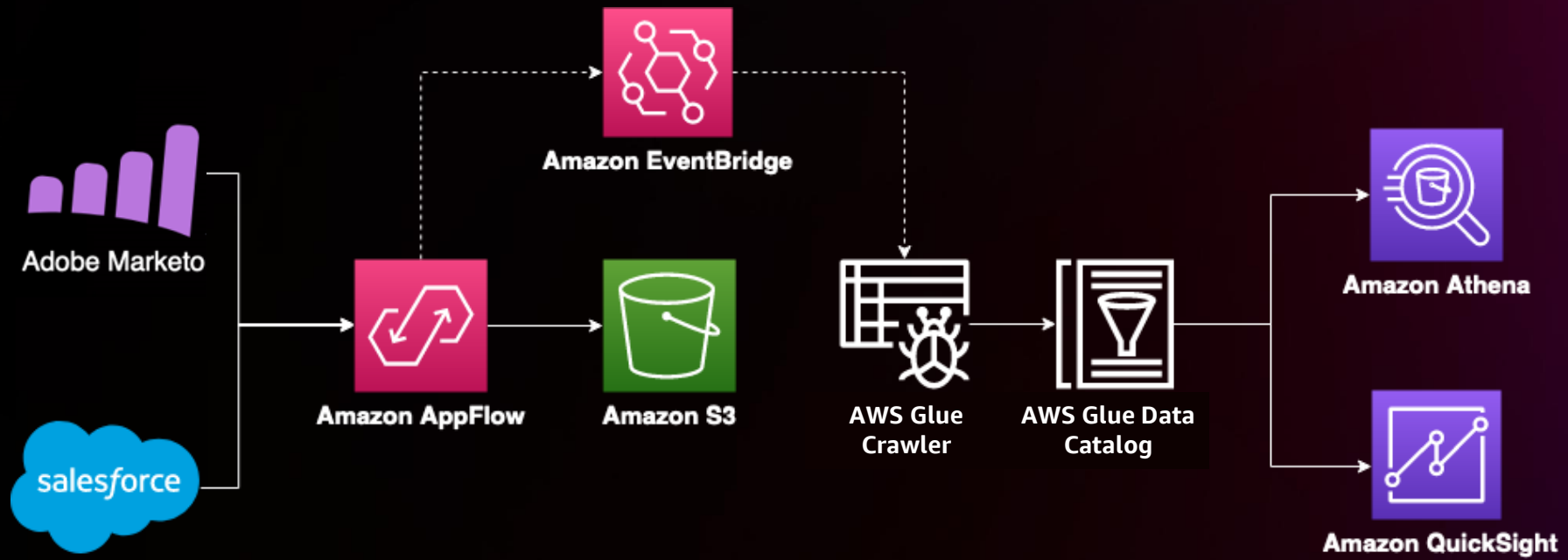
Amazon AppFlow

Simplifying data integration for AWS analytics and ML

Ray Jang
Principal Product Manager - Technical
Amazon AppFlow

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

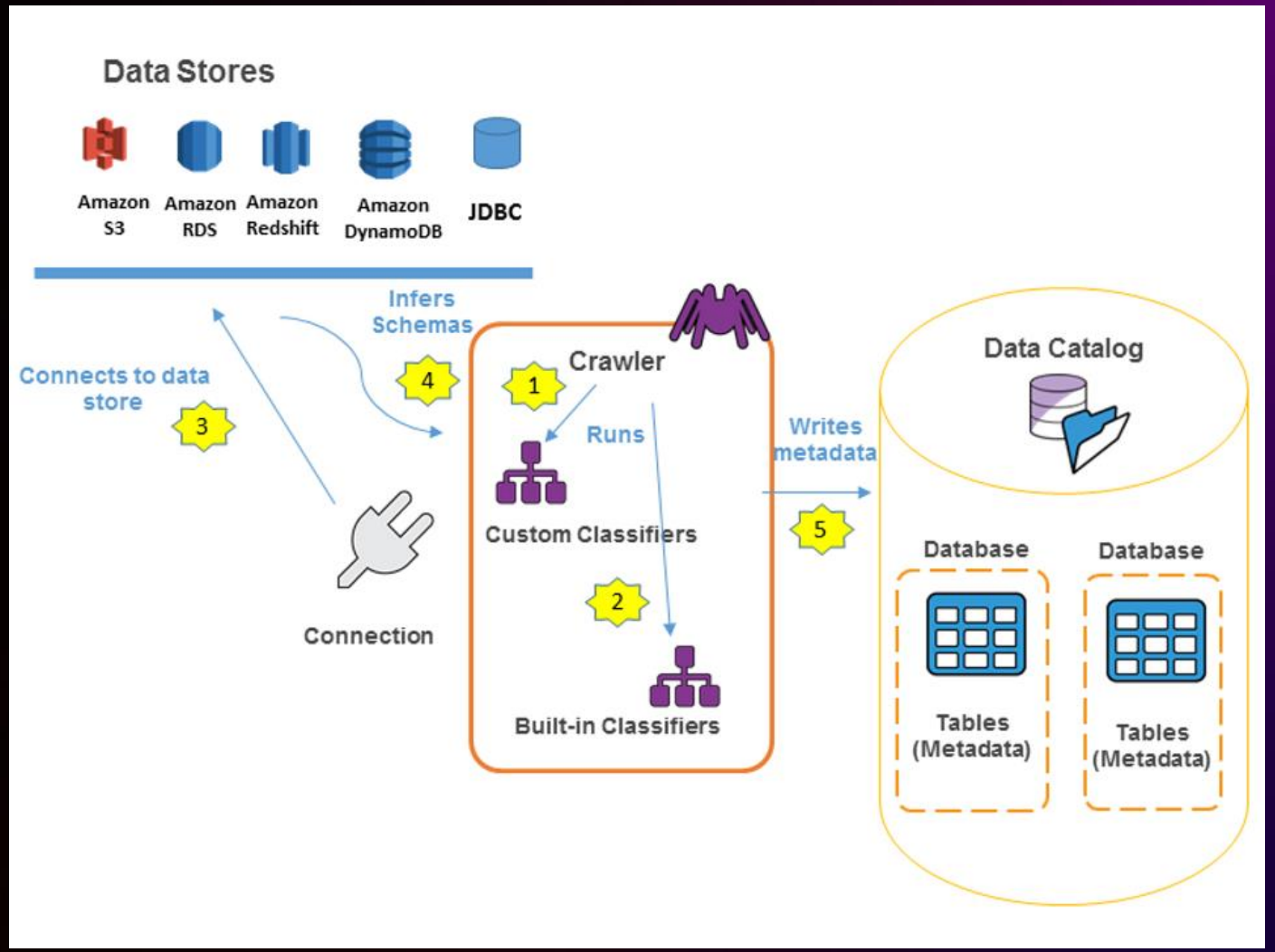
Simplifying cataloging and sharing across AWS



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Using the AWS Glue Crawler

1. Classify your data
2. Verify the classifier
3. Connect to data store
4. Create inferred schema
5. Write metadata
6. Verify schema
7. Verify data types
8. Debug and modify jobs
9. Maintain jobs for your data transfers

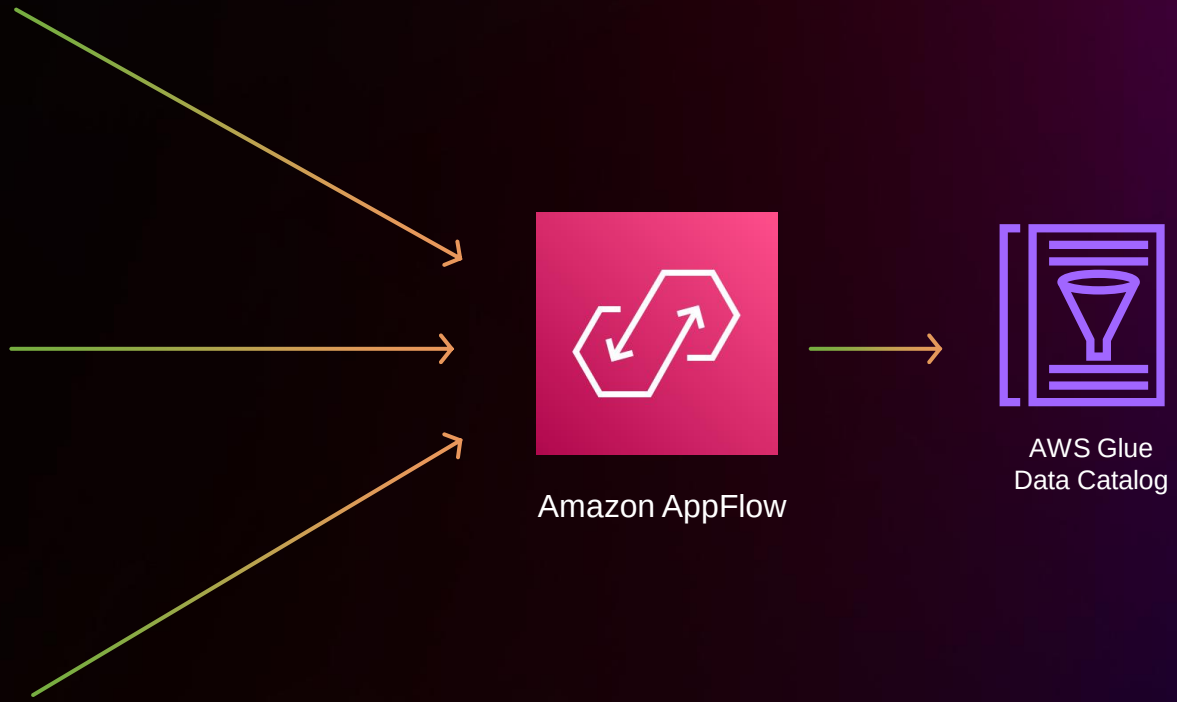


© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon AppFlow now catalogs your data in 3 clicks

NEW

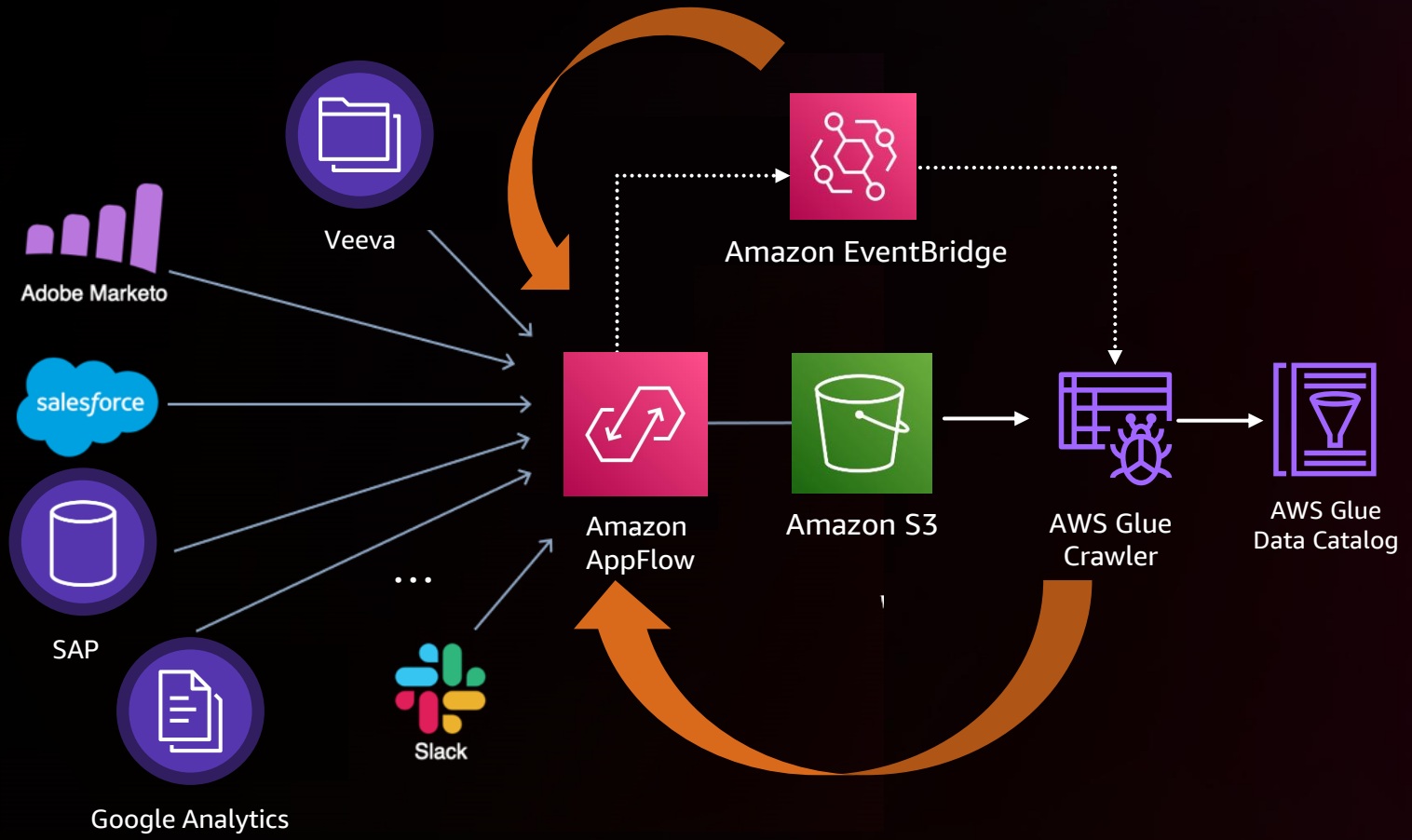
1. Classify your data
2. Verify the classifier
3. Connect to data store
4. Create inferred schema
5. Write metadata
6. Verify schema
7. Verify data types
8. Debug and modify jobs
9. Maintain jobs for your data transfers



aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon AppFlow now catalogs your data in 3 clicks

NEW



Amazon AppFlow now:

- Automates cataloging of your data
- Removes requirement to run a crawler
- Triggers Data Catalog table creation/update after every Flow run
- Removes requirement for Amazon EventBridge

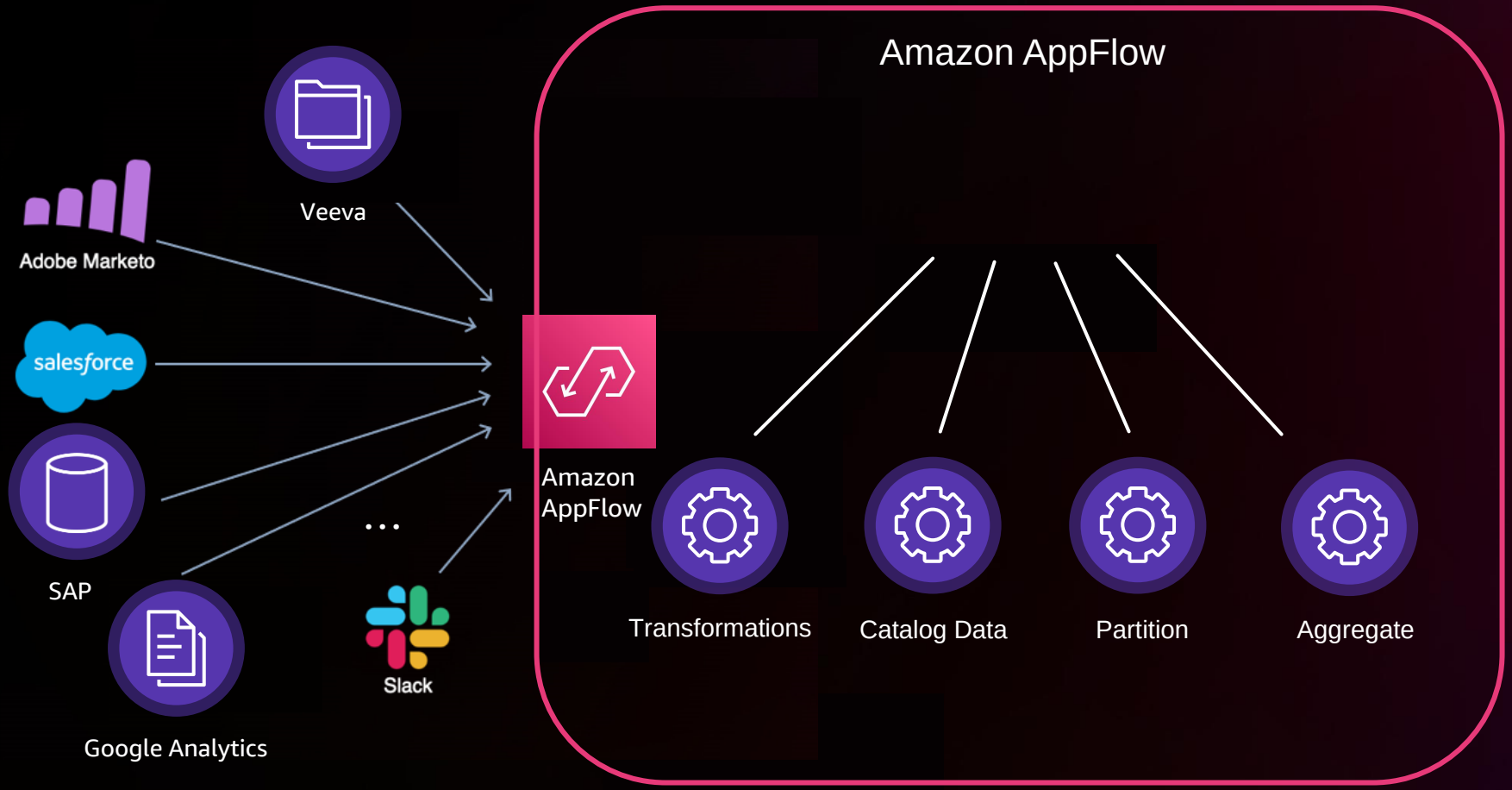
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Analyze data across AWS analytics, ML and BI



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon AppFlow now supports partitioning, aggregation



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What is partitioning?

- Divides your table based on column values
- Partitions act as virtual columns
- ***Improves query performance*** by reducing the amount of data scanned
- Amazon AppFlow supports Hive partitioning:

```
BillingCountry=France/
BillingCountry=Canada/
BillingCountry=USA/
        BillingState=KS/
        BillingState=NC/
                BillingCity=Burlington/
```

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Partitioning performance and cost benefits

Non-Partitioned Query		
Data Scanned	**Runtime***	**Cost (Athena)**
74100 MB	4.8 sec	$0.36

Partitioned Query		
Data Scanned	**Runtime***	**Cost (Athena)**
30 MB	0.7 sec	$0.0001

Partitioned vs. non-partitioned

- 85% faster performance
- 99% lower cost

* Reference: https://aws.amazon.com/blogs/big-data/top-10-performance-tuning-tips-for-amazon-athena/

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Aggregation and performance benefits

Queries run more efficiently when data scanning is parallelized

- Small files impact performance due to overhead
- Large files reduce parallelism

Up to 90% performance improvement when records are aggregated into the optimal file size

Non-aggregated query	
# of files	**Runtime**
100,000	13 seconds

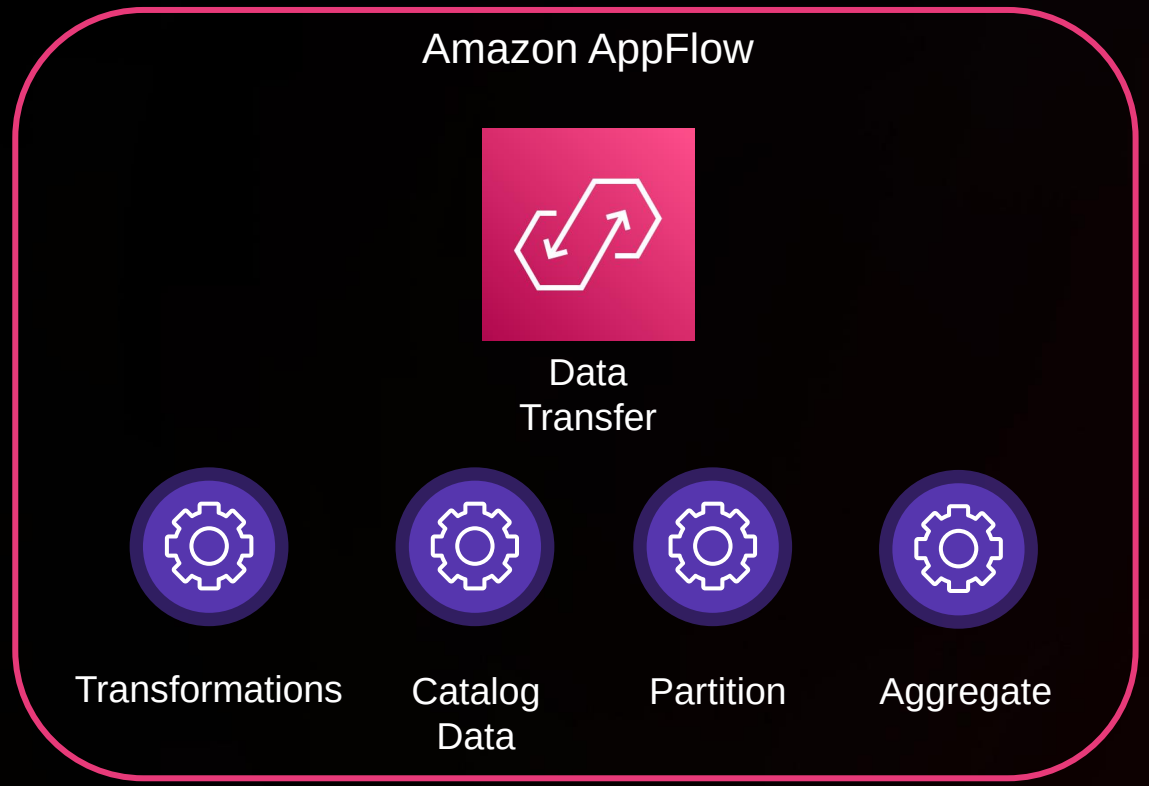
Aggregated query	
# of Files	**Runtime**
1	1.3 seconds

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon AppFlow now supports partitioning, aggregation in just a few clicks

NEW



Partitioning

- Based on column values
- Automates registration of partitions in the AWS Glue Data Catalog

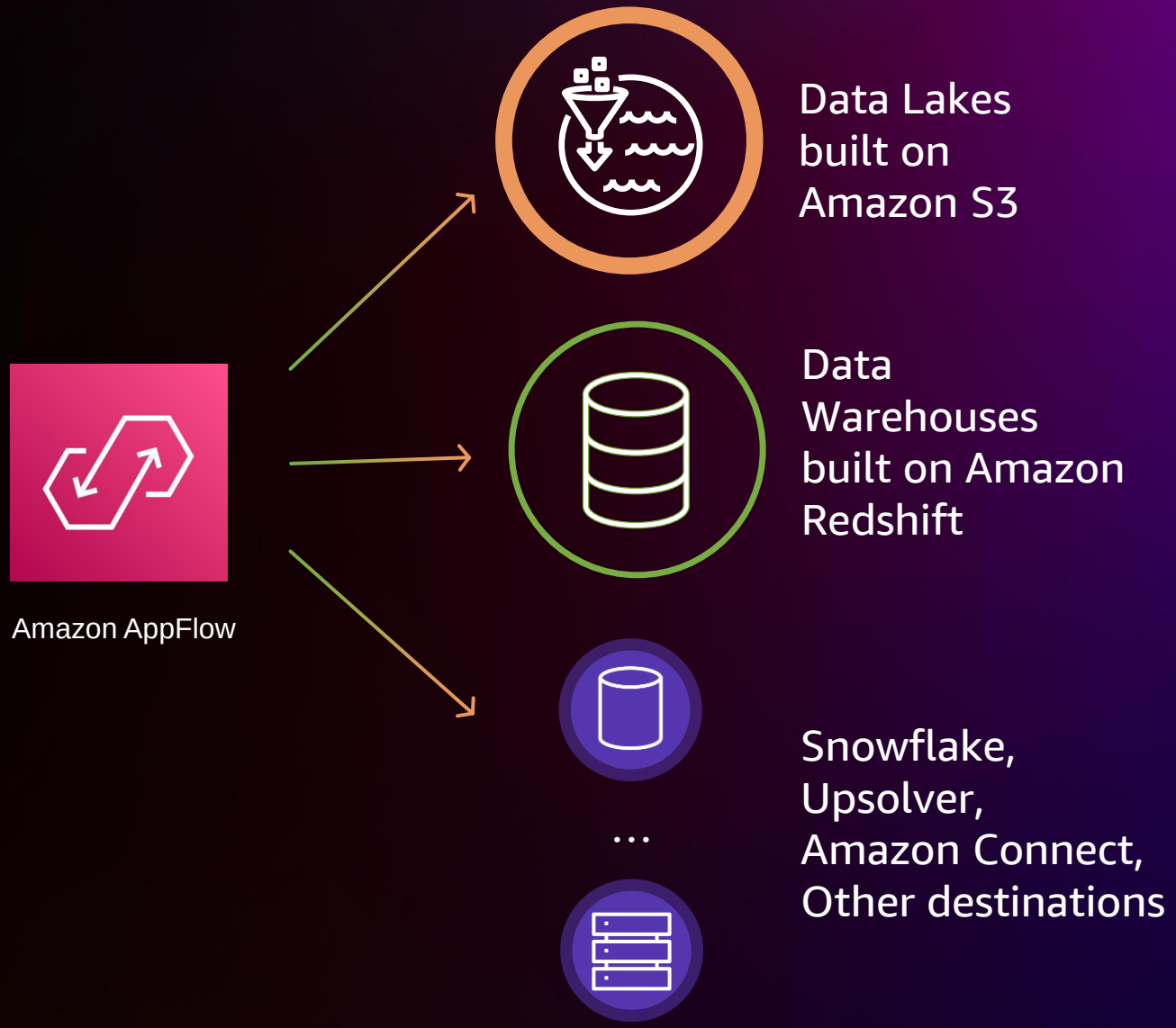
Aggregation:

- Automates aggregation of records into files to the size you specify, or
- Into one file per partition

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Customers' SaaS connectivity challenges

- Improve customers' digital experiences
- Create targeted marketing campaigns
- Analyze revenue impact



aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon AppFlow's broad library of connectors

CRM, ANALYTICS

Google Analytics

Zendesk

BUSINESS, OPERATIONS

SAP

ServiceNow

Datadog

MARKETING

Facebook Ads

Amazon AppFlow

Data Lakes built on Amazon S3

Data Warehouses built on Amazon Redshift

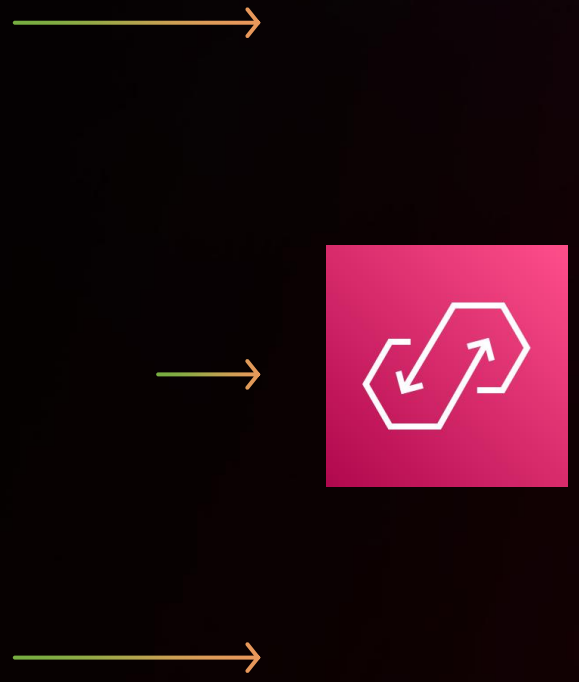
Snowflake, Upsolver, Amazon Connect, Other destinations

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Many more to be announced this week…

NEW

MANY MORE CONNECTORS TO BE ANNOUNCED THIS WEEK



Data Lakes built on Amazon S3

Data Warehouses built on Amazon Redshift

…

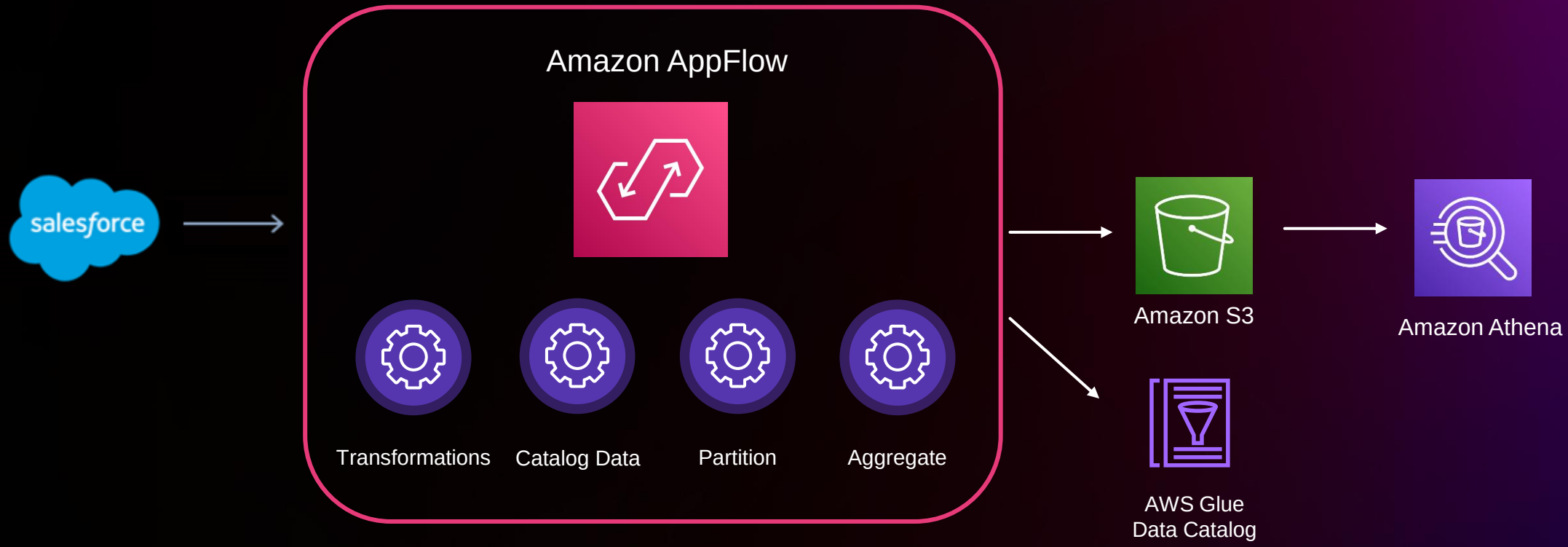
Snowflake, Upsolver, Amazon Connect, Other destinations

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Demo – Analysis of SaaS data in a few clicks

NEW



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Cloud Computing Services - Ama

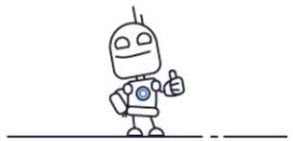
aws.amazon.com

aws | re:Invent | Products | Solutions | Pricing | Documentation | Learn | Partner Network | AWS Marketplace | Customer Enablement | Events | Explore More | Contact Us | Support | English | My Account | Sign In to the Console

Migrate to AWS and See Business Results Faster

Migrate any workload to AWS for cost savings, improved productivity, business agility and operational resilience

Learn more »



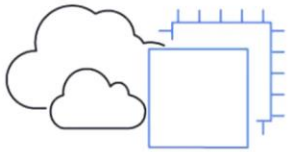
Amazon Lightsail Extended Free Tier
New customers get up to three months free on select virtual private servers



Fight Online Fraud With ML
Explore how to detect and mitigate online fraud with machine learning



Amazon DocumentDB (With MongoDB Compatibility)
Start your free trial of the database designed for enterprise JSON workloads



Longest-running Magic Quadrant Leader for CIPS
Gartner names AWS a Leader for the 12th consecutive year

AWS Customer News

Best Buy selects AWS as its strategic cloud provider to further unleash the innovative work of its technology teams and advance its mission to create world-class digital experiences for its customers. Read the press release »

Training and Certification

For Builders | For Decision Makers

For developers, data scientists, solutions architects, or anyone interested in learning how to build on AWS today







Summary

1. Access to data and AWS analytics/ML services are table stakes toward delivering better customer and business insights

2. Amazon AppFlow breaks down data silos and delivers differentiated value with a no-code, secure, and fully managed AWS service

3. New Amazon AppFlow features simplify and automate your data integration: New connectors, AWS Glue Data Catalog integration, automated data preparation

4. Try Amazon AppFlow today at https://aws.amazon.com/appflow/

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!



Please complete the session survey in the **mobile app**

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.