



AWS Builder Academy

AWS Modern Data Architecture

June 15, 2023
At AWS Bangkok Office



Wi-fi: Guest
Password: BrokenWires@@2019

Meet our team!



Chatchai K. (Bas)
Solutions Architect Manager



Pongpat S.(Tae)
Solutions Architect



Patiphan P. (Boy)
Solutions Architect



Nonthapat K. (Tontan)
Associate Solutions Architect

What you will learn/do today?

- Good enough theory, and opportunities to hand-on with Modern Data on AWS.
- Come talk to us if you have any plan on AWS. 😊

Agenda - AWS Modern Data Architecture

Time	Topic
09:00 – 09:15	Registration
09:15 – 09:45	Modern Data Architecture Overview
09:55 – 10:25	Scalable Data Lakes
10:25 – 10:40	Break
10:40 – 11:00	LAB: Setting Up the Data Lake - Mandatory
11:00 – 11:45	Seamless Data Movement
11:45 – 12:00	LAB: Change Data Capture (CDC) From Relational Databases
12:00 – 13:00	Break
13:00 – 13:20	LAB: Streaming Data Ingestion
13:20 – 13:35	LAB: Fined Grained Access Control - Column
13:35 – 14:35	Purpose Built Analytics Services
14:35 – 14:50	Break
14:50 – 15:20	LAB: Query with Amazon Redshift
15:20 – 15:45	LAB: Creating Federated Query to Redshift through Athena
15:45 – 16:00	LAB: Building Data Visualization with QuickSight
16:00 – 16:20	Recap / Learning Resources

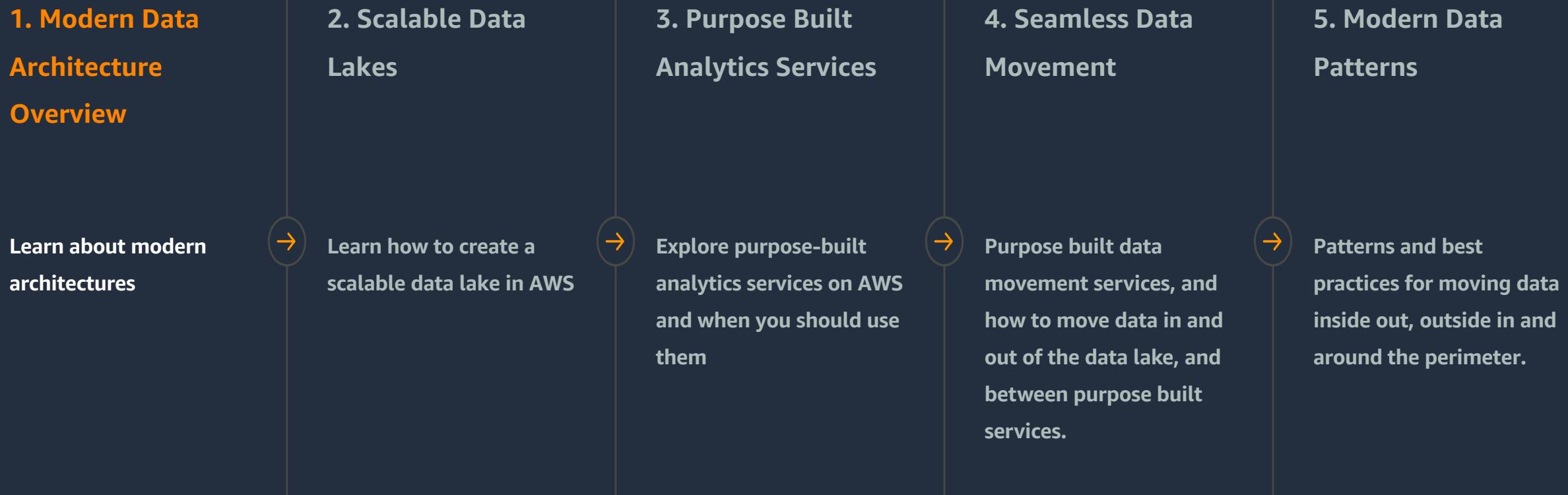


Modern Data Architecture Overview

Pongpat Sa-iam (Tae)

Solutions Architect
Amazon Web Services

Modern Data Architecture Immersion Day



1. Why Choose a Modern Data Architecture?



Challenges of data analytics at scale



Variety of sources
and data types



Performance



Increasing and
unpredictable cost



Diverse analytics needs



Manageability



Inflexible tools



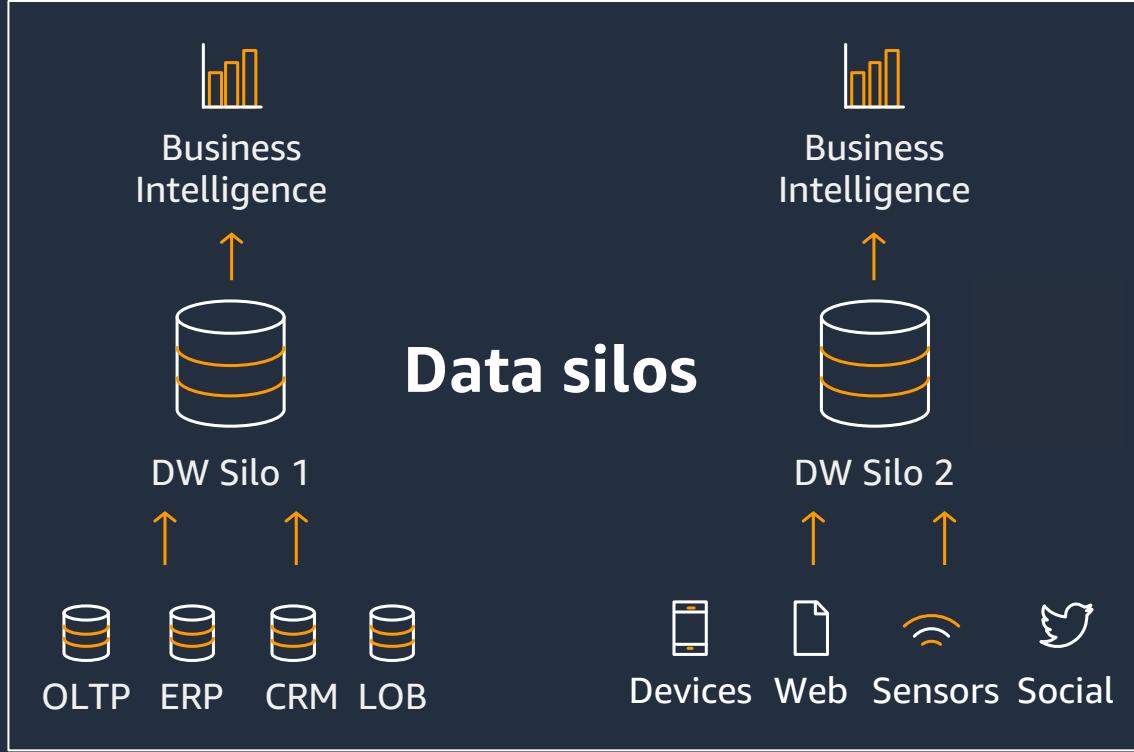
Data volume and velocity



Scalability



Security and
Compliance



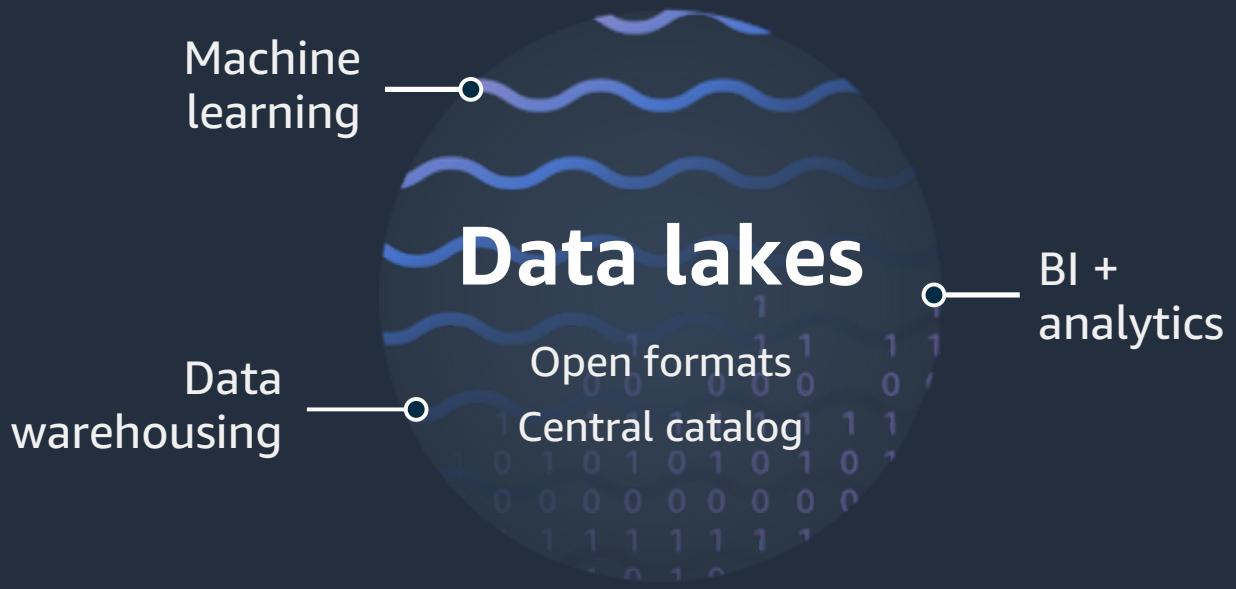
Relational vs Non-Relational

TBs-PBs Scale

Schema Defined Prior to Data Load

Operational and Ad Hoc Reporting

Large Initial Capex + \$\$K / TB/ Year



TB-EBs Scale

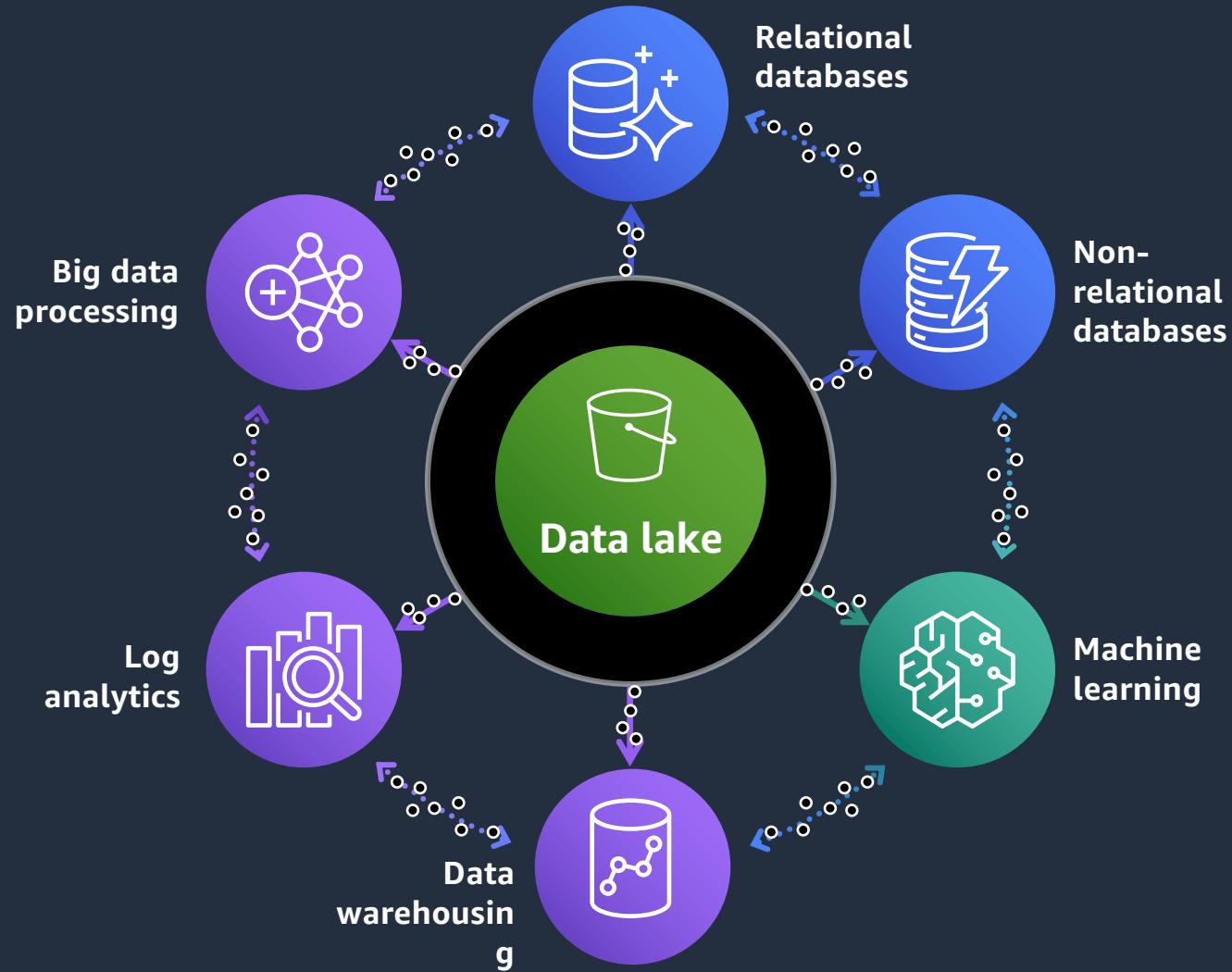
All Data in one place, a Single Source of Truth

Relational and Non-Relational Data

Decouples Storage (low cost) and Compute

Schema on Read

Diverse Analytical Engines



Key Pillars of the Modern Data Architecture

Scalable Data Lake

Purpose-built Data Services

Seamless Data Movement

Unified Governance

Performant and Cost Effective

2. Principles

Scalable Data Lake

Purpose-built Data Services

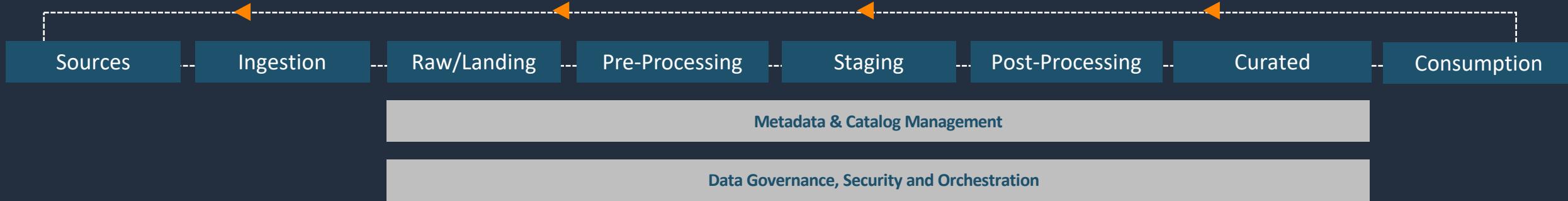
Seamless Data Movement

Unified Governance

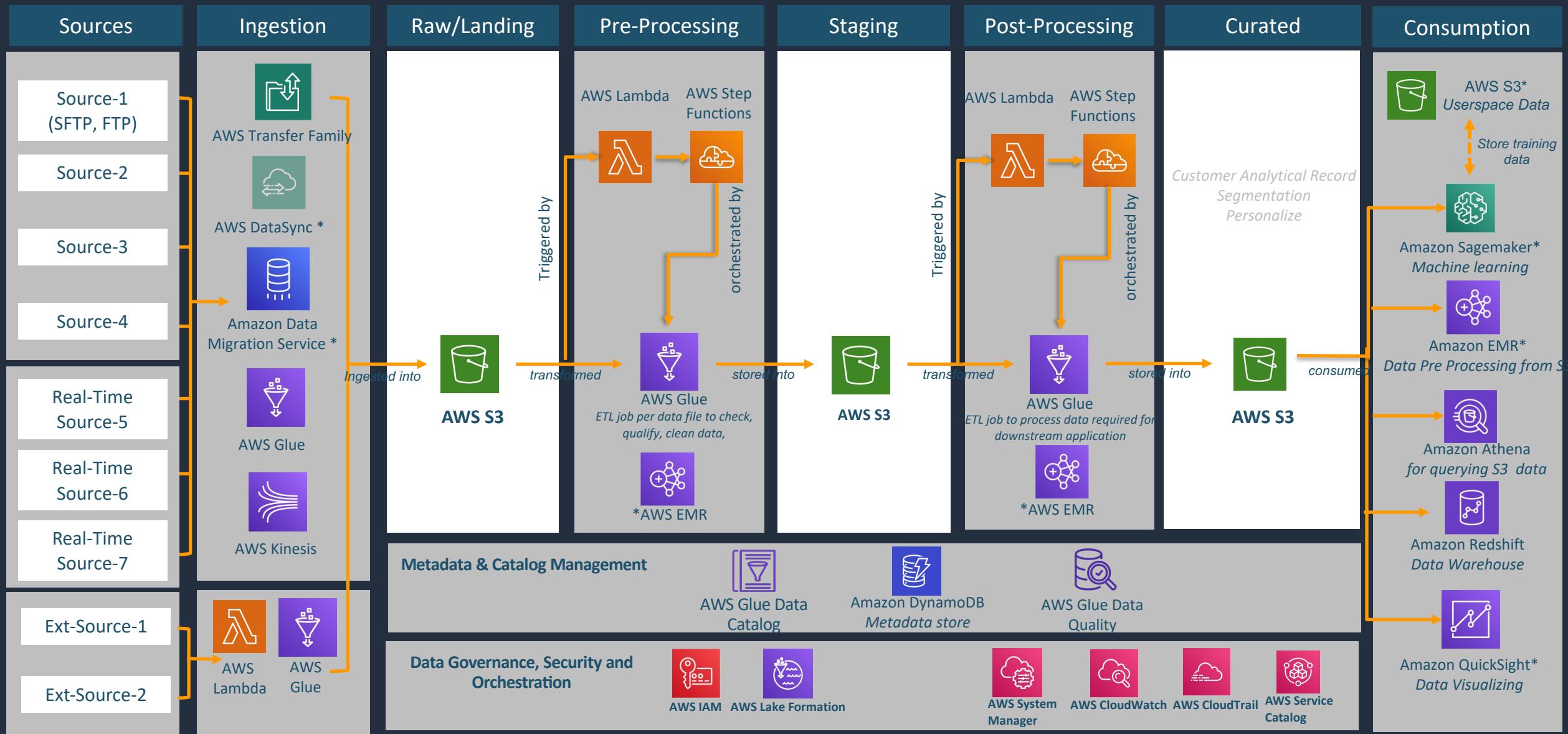
Performant and Cost Effective



Data Lake Technology Architecture Design



Data Lake Technology Architecture Design



* The components are not part of the scope.

Scalable Data Lakes



Variety of sources
and data types



Data volume and velocity



Increasing and
unpredictable cost



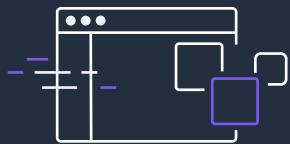
A data lake is a **centralized repository** that allows you to store all your **structured and unstructured data** at any scale



Variety of sources
and data types



Diverse analytics needs



Performance



Inflexible tools



Data Scientists



Data Engineers



Data Analysts



Data Owners

Purpose Built Data Services

Choose the right service

- Never compromise on performance, scale, or cost

Best of breed

- Specialized, best of breed analytics services are designed to be the best at specific tasks.

Democratization

- allowing different personas to use the right tool for the job
- removing gate keepers

Lab Time! Please go to:
go.jansat.co/md-lab



Wi-fi: Guest
Password: BrokenWires@@2019





Sign in

Choose a preferred sign-in method

Email one-time password (OTP)

Enter your personal or corporate email to receive a one-time password

Login with Amazon

Login with your Amazon.com retail account

Amazon employee

Login with your Amazon Corporate account. Only for Amazon Employees.

One-time email passcode

Send a passcode to the email below.

Email

jansat@amazon.com

Back

Send passcode

[Get help signing in](#)

One-time email passcode

We sent a passcode to jansat@amazon.com. You should receive it within 5 minutes.

Passcode (9-digit) [Resend passcode](#)

[Back](#)[Sign in](#)

[Get help signing in](#)

Description

Container Immersion Day for AWS Builder Academy

Terms and Conditions

Read and accept before joining the event

1. By using AWS Workshop Studio for the relevant event, you agree to the [AWS Event Terms and Conditions](#) and the [AWS Acceptable Use Policy](#). You acknowledge and agree that are using an AWS-owned account that you can only access for the duration of the relevant event. If you find residual resources or materials in the AWS-owned account, you will make us aware and cease use of the account. AWS reserves the right to terminate the account and delete the contents at any time.
2. You will not: (a) process or run any operation on any data other than test data sets or lab-approved materials by AWS, and (b) copy, import, export or otherwise create derivate works of materials provided by AWS, including but not limited to, data sets.
3. AWS is under no obligation to enable the transmission of your materials through AWS Workshop Studio and may, in its discretion, edit, block, refuse to post, or remove your materials at any time.
4. Your use of AWS Workshop Studio will comply with these terms and all applicable laws, and your access to AWS Workshop Studio will immediately and automatically terminate if you do not comply with any of these terms or conditions.

I agree with the Terms and Conditions

Cancel

Previous

Join event

▼ AWS account access

[Open AWS console
\(us-east-1\) !\[\]\(4658fc881287bc22b537ed0517e70445_img.jpg\)](#)

[Get AWS CLI credentials](#)

Lab Stream 0

AWS Modern Data Architecture Immersion Day X

AWS Modern Data Architecture Immersion Day >
Stream 0. Setting up the Data Lake (Mandatory)

Stream 0. Setting up the Data Lake (Mandatory)

This exercise creates the foundational data lake, in which we will securely store, analyze, and visualize data. We start by storing encrypted sample test data in our Amazon Simple Storage Service S3 (<https://aws.amazon.com/s3/>) data lake. We use AWS Lake Formation (<https://aws.amazon.com/lake-formation/>) to govern the data and create a metadata catalog. We will use (<https://aws.amazon.com/Glue>) AWS Glue to cleanse and transform the data, and finally query the data using Amazon Athena (<https://aws.amazon.com/athena/>).

▼ Stream 0. Setting up the Data Lake (Mandatory)

- A. Configure CloudTrail
- B. Configure Lake Formation
- C. Populate the Data Catalog
- D. Transform and Prepare the



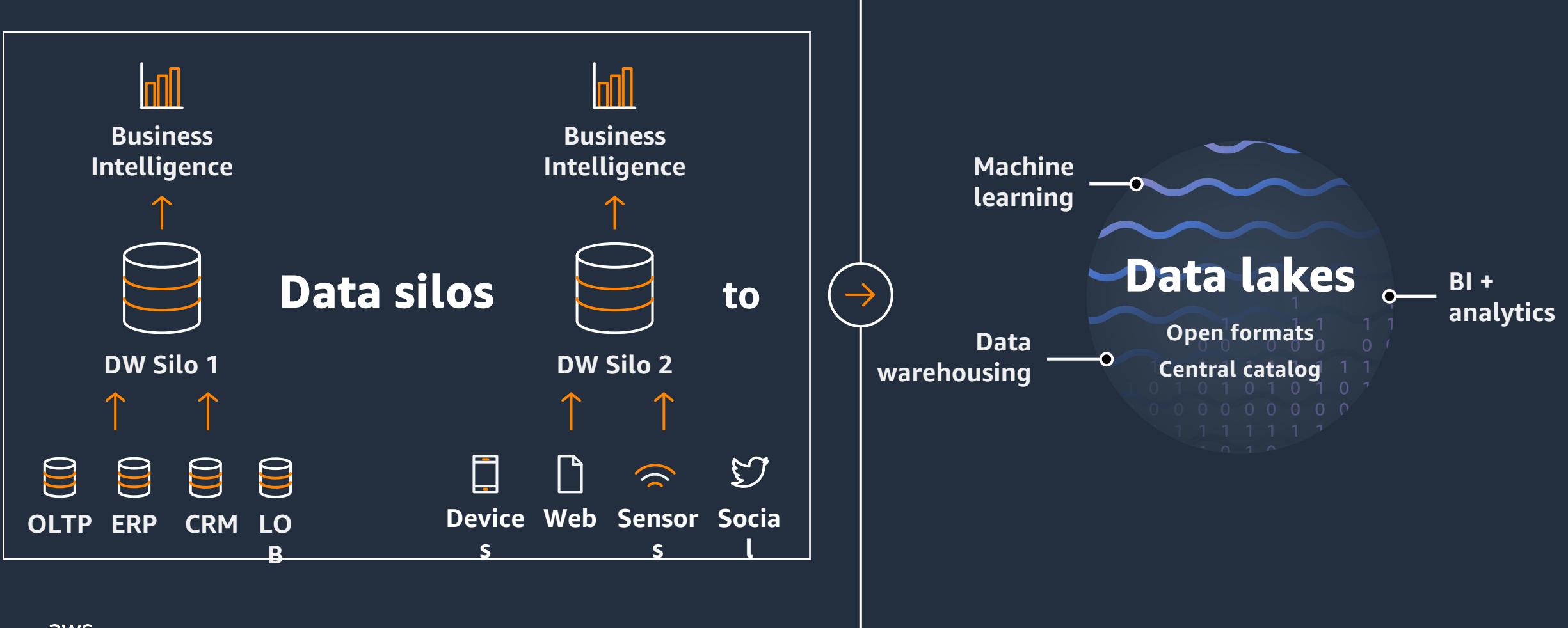


Data lakes and analytics on AWS: Turn data into insights

Pongpat Sa-iam (Tae)

Solutions Architect
Amazon Web Services

Traditional data warehousing approaches don't scale



Why choose AWS for data lakes and analytics?

1



**Easiest to build
data lakes and
analytics**

2



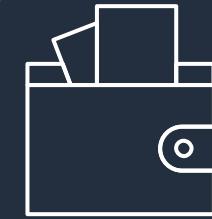
**Most secure
infrastructure for
analytics**

3



**Most
comprehensive
and open**

4



**Most scalable
and cost
effective**

1. Easiest to build data lakes and analytics



- A single storage layer (S3) for all analytics and ML
 - A service to build secure data lakes in days
 - Deep integration across analytics & infrastructure
 - (including federated queries)
-

**The fastest way to go from zero to insights,
covering all data for all users**



2. Most secure infrastructure for analytics



Customers need to have multiple levels of security, identity and access management, encryption, and compliance to secure their data lake



Security

Amazon GuardDuty
AWS Shield
AWS WAF
Amazon Macie
VPC



Identity

AWS IAM
AWS SSO
Amazon Cloud Directory
AWS Directory Service
AWS Organizations



Encryption

AWS Certification Manager
AWS Key Management Service
Encryption at rest
Encryption in transit
Bring your own keys,
HSM support



Compliance

AWS Artifact
Amazon Inspector
Amazon Cloud HSM
Amazon Cognito
AWS CloudTrail



2. Most secure infrastructure: certifications



Global



CSA
Cloud Security
Alliance Controls



ISO 9001
Global Quality
Standard



ISO 27001
Security Management
Controls



ISO 27017
Cloud Specific
Controls



ISO 27002
Information
Security
Management
Systems



PCI DSS Level 1
Payment Card
Standards



SOC 1
Audit Controls
Report



SOC 2
Security, Availability, &
Confidentiality Report



SOC 3
General Controls
Report



© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

United States



CJIS
Criminal Justice
Information Services



FedRAMP
Government Data
Standards



FERPA
Educational
Privacy Act



ISO FFIEC
Financial Institutions
Regulation



FIPS
Government Security
Standards



FISMA
Federal Information
Security Management



GxP
Quality Guidelines
and Regulations



HIPAA
Protected Health
Information



ITAR
International Traffic
in Arms Regulation



MPAA
Protected Media
Content



NIST
National Institute of
Standards and Technology



SEC Rule 17a-4(f)
Financial Data
Standards



VPAT/Section 508
Accountability
Standards



FISC [Japan]
Financial Industry
Information Systems



IRAP [Australia]
Australian Security
Standards



ISMS
Information Security
Management System



MTCS Tier 3 [Singapore]
Multi-Tier Cloud
Security Standard



My Number Act [Japan]
Personal Information
Protection

Europe



C5 [Germany]
Operational Security
Attestation



Cyber Essentials
Plus [UK]
Cyber Threat
Protection



G-Cloud [UK]
UK Government
Standards



IT-Grundschutz
[Germany]
Baseline Protection
Methodology



© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

3. Most comprehensive and open

Data as a Service

Identify Business Risk

Business Operation and Forecasting

Data Democratization

Enable Self-service

Data Clean Room

Less ETL (Zero ETL)

Data, visualization, engagement & Machine learning



Data Exploration



Business Dashboards



Digital User Engagement / C360



Predictive Analytics and MLOps

Database & Analytics

Data Warehousing

Big Data Processing

Interactive Query

Operational Databases

Real time Analytics



Data Lake Storage



Governance, Security & Management



Data Catalog & ETL

Data movement

Migration & Streaming Services

Data Sources

On-premises Data sources

The AWS analytics portfolio

Data, visualization, engagement, & machine learning

NEW



Data Exchange



Quicksight



Pinpoint



SageMaker



Comprehend



Lex



Polly



Rekognition



Translate
+ many more

Analytics



Redshift



EMR (Spark & Hadoop)



AWS Glue
(Spark & Python)



Athena



Elasticsearch Service



Kinesis Data Analytics

Data lake infrastructure & management



S3/Glacier



Lake Formation



AWS Glue

Data movement

Database Migration Service | Snowball | Snowmobile | Kinesis Data Streams | Kinesis Data Firehose | Managed Streaming for Apache Kafka



No compromises on performance, scale, and cost



Amazon
Redshift

3x better price performance than other cloud data warehouses

Automated performance tuning and near-linear scaling



Amazon
EMR

Optimized runtimes that provide the best price-performance

1.7x faster than standard Apache Spark; **2.6x faster** than standard Presto



Amazon
OpenSearch
Service

UltraWarm storage tier **reduces costs by 90%**

Store **6x more log data** without increasing costs



Amazon
S3

Amazon S3 Select retrieves a subset of data leading to queries that run **up to 400% faster**

Amazon S3 intelligent tiering **saves up to 70% on storage costs** for data lakes



With Graviton instance, customers **save 25.7%** for typical workloads

3. Open standards, formats, and Apache open source



- Flink
- Ganglia
- Hbase
- HCatalog
- HDFS
- Hive
- Hudi
- Java
- JupyterHub
- Kafka
- Livy

Mahout
MapReduce
MxNET
MySQL
Oozie
ORC
Parquet
Phoenix
Pig
Presto
Python

PyTorch
R
Scala
Spark
Sqoop
SQL
TensorFlow
Tez
YARN
Zeppelin
Zookeeper

4. Most scalable, cost-effective, high-performance infrastructure for analytics



**On-demand,
Reserved, and
Spot instances
to reduce costs**



**100 Gbps
bandwidth
network
interfaces for
performance**



**Industry leading
choice of 200+
instance types to
meet workload
needs**



**Five highly
available
storage tiers
and intelligent
tiering**

4. Most scalable, cost-effective infrastructure for analytics



Some examples of advanced capabilities in analytics services



EMR

Autoscaling

57% less than on-premises per IDC report



Redshift

Less than 1/10th of the cost of traditional, on-premises solutions



Athena & QuickSight

Serverless pay only for what is used

Pricing per session for visualization



MSK

Get started for under \$2.50/day for fully managed Apache Kafka



Overview of Data Movement services



Most ways to move data to the data lake

Data movement

Professional services and partners to help migration



Data movement from your on-premises datacenters



Data movement from real-time sources

Synchronizing data across environments

Data movement from on-premises datacenters

- Dedicated network connection – AWS Direct Connect
- Secure appliances – AWS Snowball
- Ruggedized shipping containers – AWS Snowmobile
- Database migration – AWS DMS
- Gateway that lets applications write to the cloud

Data movement from real-time sources

- Connect devices to AWS
- Real-time data streams – Kinesis & MSK
- Real-time video streams – Kinesis Video Analytics



AWS storage services

AWS storage services

Object, file, and block storage



Amazon Simple Storage Service (S3)

Object storage with industry-leading scalability, availability, and security for you to store and retrieve any amount of data from anywhere.



Amazon Elastic File System (EFS)

A simple, serverless, elastic, set-and-forget file system for you to share file data without managing storage.

FSx

Amazon FSx

Fully managed, cost-effective file storage offering the capabilities and performance of popular commercial and open-source file systems.



Amazon Elastic Block Store (EBS)

Easy to use, high-performance block storage service for both throughput and transaction-intensive workloads at any scale.



Amazon File Cache

High-speed cache for datasets stored anywhere, accelerate cloud bursting workloads.

Your choice of object storage classes



S3 Standard

S3 Intelligent-Tiering

S3 Standard-IA

S3 One Zone-IA

S3 Glacier
Instant Retrieval

S3 Glacier
Flexible Retrieval

S3 Glacier
Deep Archive

Frequent ←

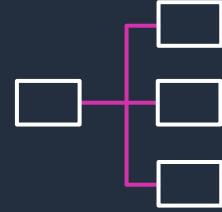
Access Frequency

→ *Infrequent*

<ul style="list-style-type: none">• Active, frequently accessed data• Milliseconds access• ≥ 3 AZ• ≥ 3 AZ• \$0.023/GB	<ul style="list-style-type: none">• Data with changing access patterns• Milliseconds access• ≥ 3 AZ• \$0.023 to \$0.0125/GB (\$0.004 to \$0.00099/GB Archive)• No retrieval fees• Monitoring fee per Obj.• Min storage duration• Min object size	<ul style="list-style-type: none">• Infrequently accessed data• Milliseconds access• ≥ 3 AZ• \$0.0125/GB• Retrieval fee per GB• Min storage duration• Min object size	<ul style="list-style-type: none">• Re-creatable, less accessed data• Milliseconds access• 1 AZ• \$0.0100/GB• Retrieval fee per GB• Min storage duration• Min object size	<ul style="list-style-type: none">• Archive data instant retrieval• Milliseconds access• ≥ 3 AZ• \$0.0040/GB• Retrieval fee per GB• Min storage duration• Min object size	<ul style="list-style-type: none">• Archive data• Select minutes or hours• ≥ 3 AZ• \$0.0036/GB – (\$4.10/TB)• Retrieval fee per GB• Min storage duration• Min object size	<ul style="list-style-type: none">• Archive data• Select 12 or 48 hours• ≥ 3 AZ• \$0.00099/GB – (\$1.01/TB)• Retrieval fee per GB• Min storage duration• Min object size
--	--	--	--	--	--	---



S3 Management Features



Organize

S3 Tagging

S3 Prefixes

S3 Versioning



Monitor

CloudWatch

CloudTrail

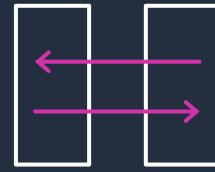
S3 Event Notifications

S3 Inventory

S3 Glacier Restore
Notifications

S3 Storage Lens

AWS Config



Replicate & Tier

S3 Lifecycle

**S3 Storage Class
Analysis**

S3 Intelligent-Tiering

Cross-Region
Replication

Replication Time
Control (RTC)



Modify

S3 Event Notifications
+ Lambda

S3 Batch Operations

S3 Object Lock

S3 Object Lambda



Data migration

Data migration



AWS DataSync

Online data transfer service that optimizes network bandwidth and accelerates data movement between on-premises storage and AWS storage.



AWS Snow Family

Offline data transfer devices with built-in security and logistics features for simplified data migration.



Transfer data between on premises and AWS

Transfer data between AWS storage services

Transfer data between AWS and other locations

On premises



Shared File System or Object Storage

Shared File Systems,
Object Storage,
or Hadoop Clusters

NFS
SMB
S3 API
HDFS



AWS DataSync Agent

The agent is deployed as a VM and connects to your shared file system or object storage



Amazon S3 on Outposts

The DataSync agent runs on EC2 on the Outpost



AWS Snowcone

The DataSync agent comes pre-installed on the device



AWS Direct Connect



Encrypted online transfer

DataSync performs data integrity checks in-transit and at-rest



AWS DataSync

DataSync seamlessly and securely connects to AWS Storage services to copy data and metadata to and from AWS



Amazon S3



Amazon Elastic File System



Amazon FSx for Windows File Server



Amazon FSx for Lustre



Amazon FSx for OpenZFS



Amazon FSx for NetApp ONTAP

Hybrid cloud storage and edge computing

Managed file transfer

Hybrid cloud storage and edge computing



AWS Storage Gateway

Hybrid cloud storage service that gives you on-premises access to virtually unlimited cloud storage.



AWS Snow Family

Edge compute, data collection, and data transfer services with security and end-to-end logistics for mobile and rugged deployments.

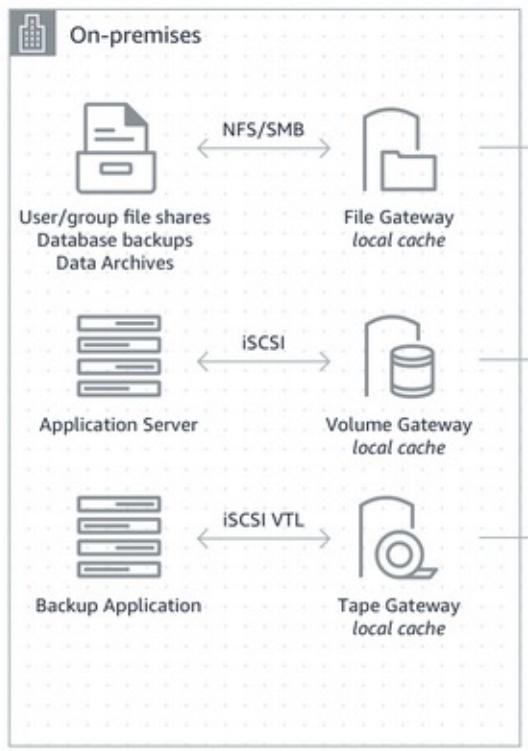
Managed file transfer



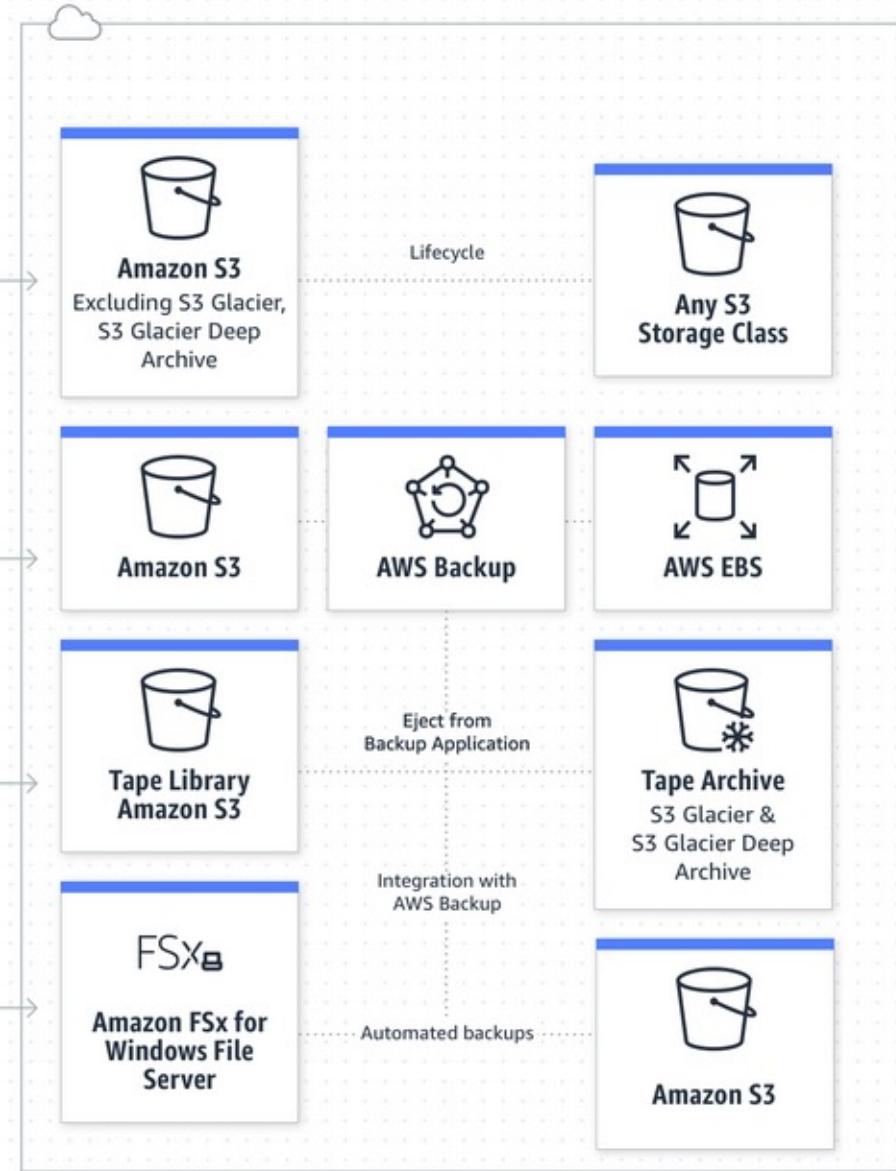
AWS Transfer Family

Simple and seamless file transfer to Amazon S3 and Amazon EFS using SFTP, FTPS, and FTP protocols.





Encryption in Transit
AWS Direct Connect
or Internet



Disaster recovery and backup

Disaster recovery and backup



AWS Elastic Disaster Recovery (DRS)

Minimize downtime and data loss with fast, reliable recovery of on-premises and cloud-based applications using affordable storage, minimal compute, and point-in-time recovery.



AWS Backup

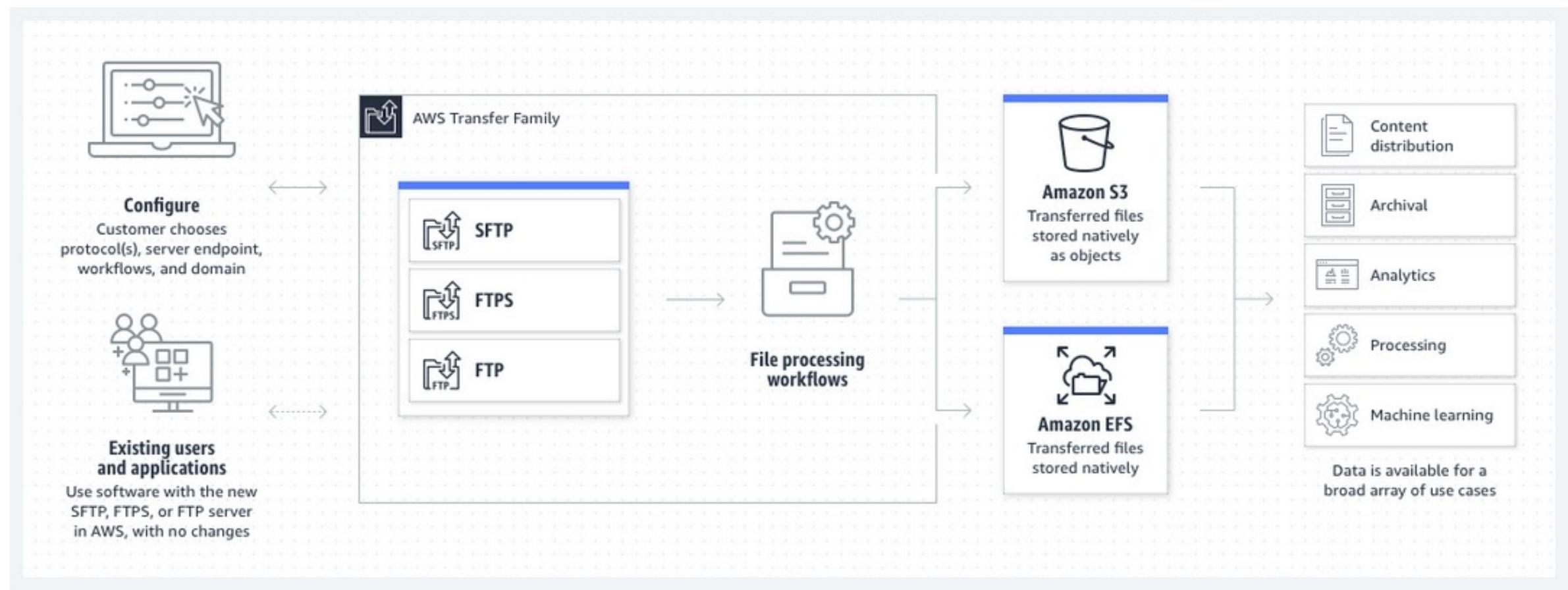
Fully managed, policy-based service to centrally manage and automate data protection, compliance, and governance for applications running on AWS.

AWS Transfer Family

SFTP, FTPS, and FTP

Applicability Statement 2 (AS2)

Seamlessly migrate, automate, and monitor your file transfer workflows into and out of Amazon S3 and Amazon EFS using the SFTP, FTPS, and FTP protocols.

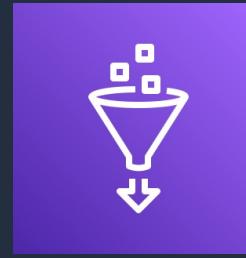


Seamless Data Movement

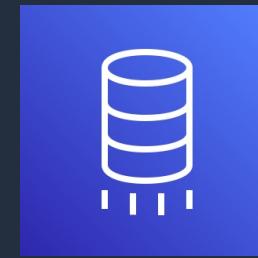
Ensure that data can easily get to wherever it's needed, with the right controls, to enable analysis and insights.



Amazon Kinesis



AWS Glue



AWS Database
Migration Service

Serverless ETL and data integration with AWS Glue

Data lake infrastructure & management

Serverless provisioning, configuration, and scaling to run your ETL jobs on Apache Spark and Python

Pay only for the resources used for jobs

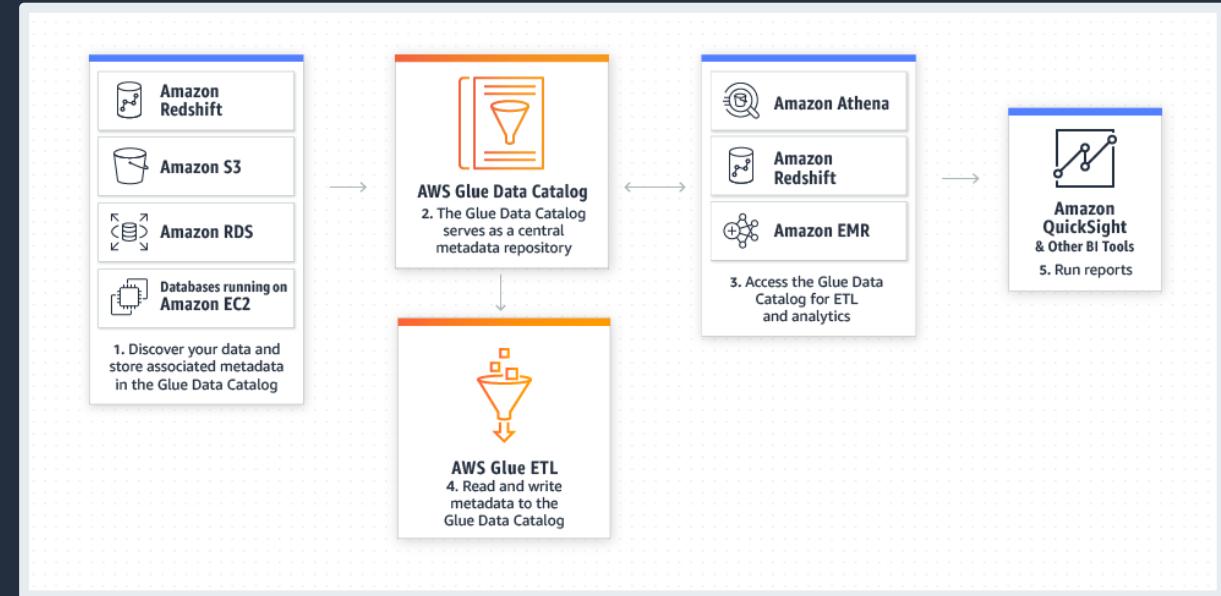
Crawl your data sources, identify data formats and suggest schemas and transformations

Automates the effort in building, maintaining and running ETL jobs

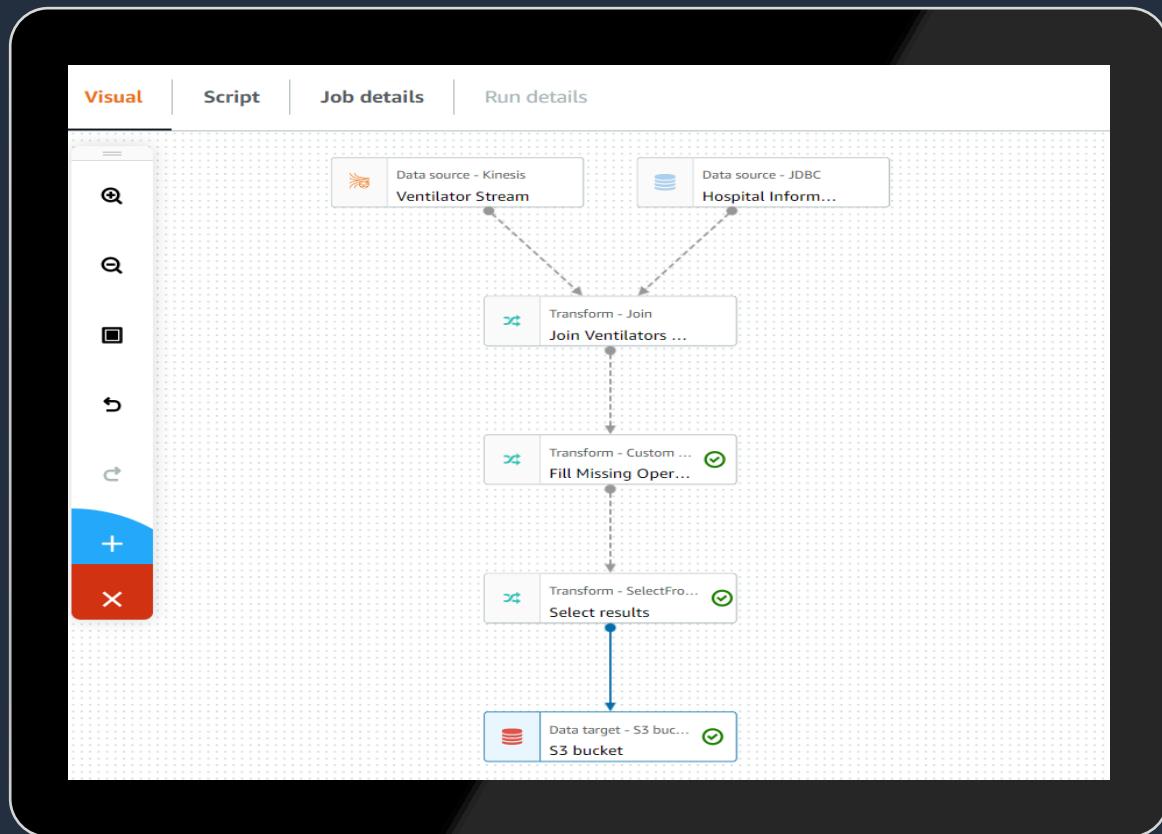
Coming soon—faster job start-up times (under 2 minutes)



© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.



AWS Glue Studio



ETL developer

Rich **visual interface**

250+ built-in transformations

Profile data to understand data patterns and anomalies

Work on large datasets **at scale**

AWS Glue DataBrew



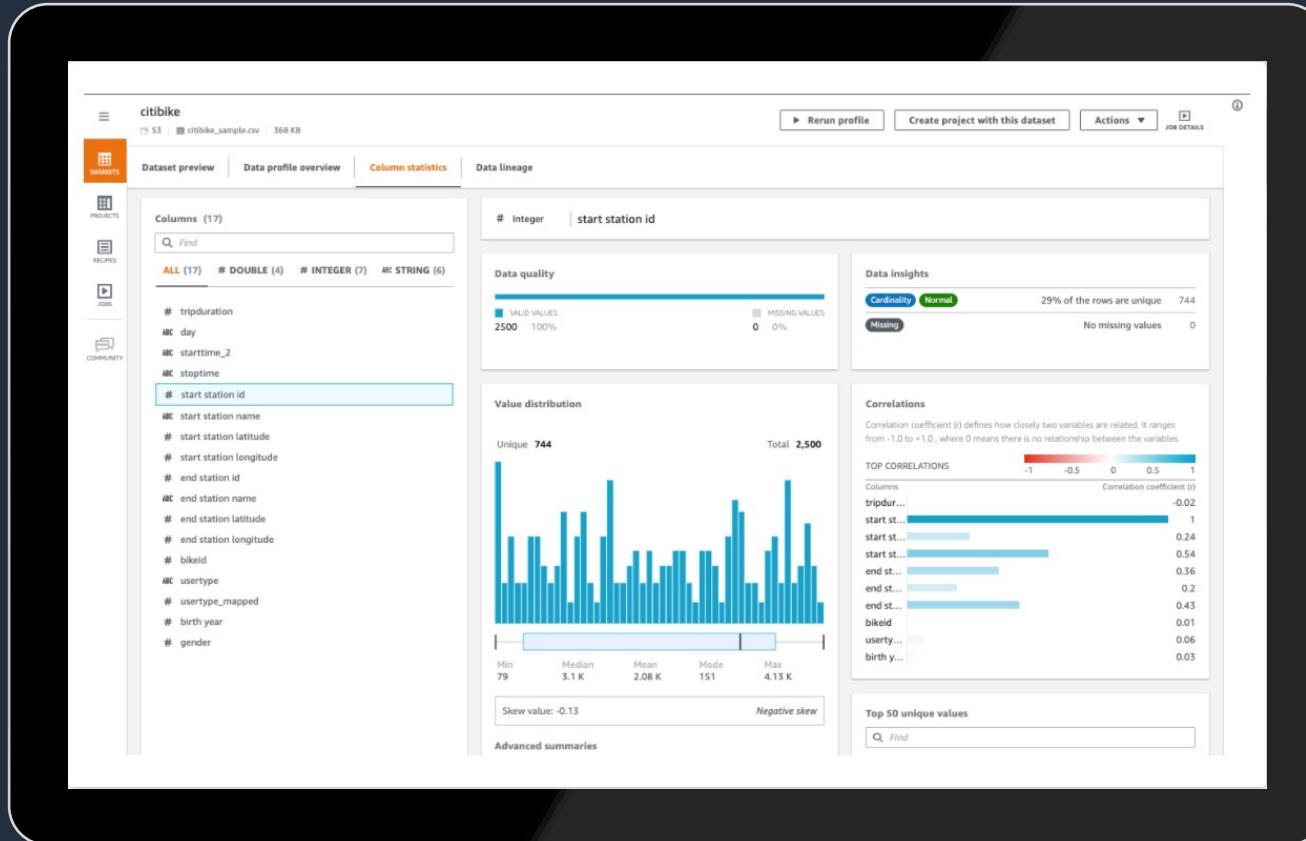
Business Analyst
Data Scientist

Run ETL jobs **without writing code**

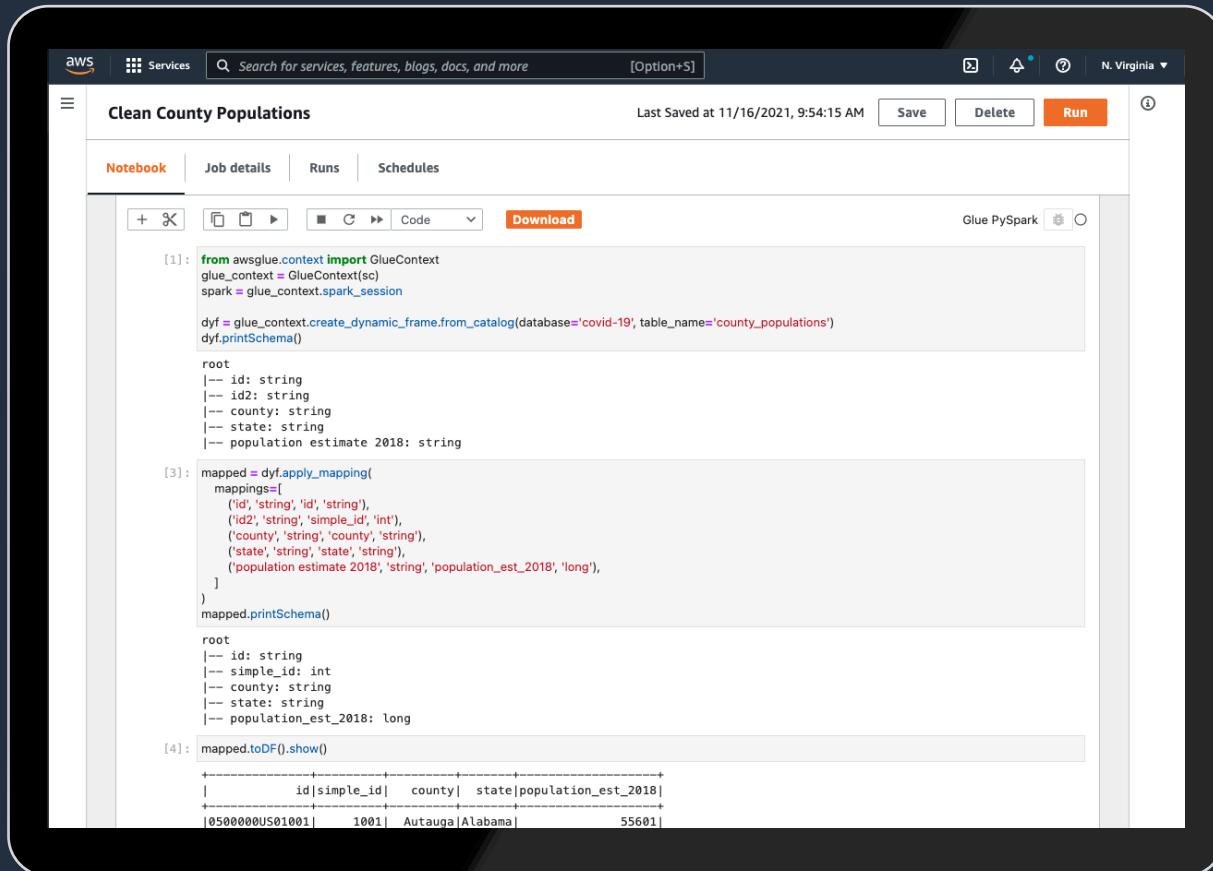
Monitor thousands of jobs through a single pane of glass

Distributed processes

Advanced transforms through code snippets



AWS Glue Notebooks



The screenshot shows the AWS Glue Notebook interface. The notebook titled "Clean County Populations" contains the following PySpark code:

```
from awsglue.context import GlueContext
glue_context = GlueContext(sc)
spark = glue_context.spark_session

df = glue_context.create_dynamic_frame.from_catalog(database="covid-19", table_name="county_populations")
df.printSchema()

root
|-- id: string
|-- id2: string
|-- county: string
|-- state: string
|-- population estimate 2018: string

[3]: mapped = df.apply_mapping(
    mappings=[
        ('id', 'string', 'id', 'string'),
        ('id2', 'string', 'simple_id', 'int'),
        ('county', 'string', 'county', 'string'),
        ('state', 'string', 'state', 'string'),
        ('population estimate 2018', 'string', 'population_est_2018', 'long'),
    ]
)
mapped.printSchema()

root
|-- id: string
|-- simple_id: int
|-- county: string
|-- state: string
|-- population_est_2018: long

[4]: mapped.toDF().show()

+-----+-----+-----+
|      id|simple_id| county| state|population_est_2018|
+-----+-----+-----+
|05000000US01001|     1001| Autauga| Alabama|          55601|
```



Data engineer

Clean and normalize data with a rich visual interface

Choose from 250+ built-in transformations to automate tasks

Profile data to understand data patterns and anomalies



Real-time: Amazon Kinesis

- Easily collect, process, and analyze data and video streams in real time



Kinesis Data Streams

Collect streaming data, at scale, for real-time analytics

Kinesis Data Firehose

Prepare and load data streams into data stores and analytics services

Kinesis Data Analytics

Get actionable insights from streaming data in real time

Kinesis Video Streams

Capture, process, and store media streams for playback, analytics and machine learning

Amazon Managed Streaming for Apache Kafka

Fully managed, highly available, and secure Apache Kafka service



**Runs and
manages
Apache Kafka
for you**



**Fully
managed**



**Highly
available**



**Multiple
levels of
security**

Lab Stream 1

▼ Stream 1: Data Movement

▼ Database CDC with DMS

- A. Configure a CDC replication task

Learn More

- B. Replicate data changes

▼ Streaming Ingest with Kinesis

- A. Create DynamoDB table

- B. Configure Kinesis Data Firehose

- C. Load data into DynamoDB

- D. Validate Data in S3

Learn More

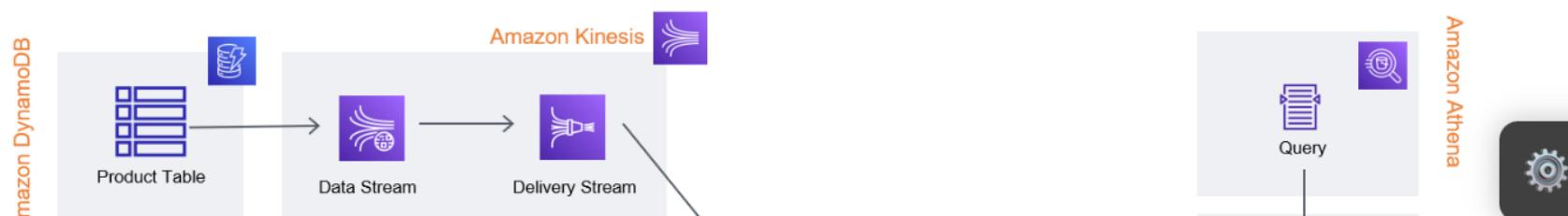
AWS Modern Data Architecture Immersion Day > Stream 1: Data Movement

Stream 1: Data Movement

Seamless data movement is a foundational part of a strong, efficient and effective modern data architecture. A modern data architecture will:

- allow you to quickly and easily ingest data to the data lake ("outside in"), and enable you to start understanding your data and gaining insights without needing to first transform your data into a standard model.
- move data out of the data lake ("inside out") and between purpose built services ("around the perimeter").
- move data at the right speed, whether that in bulk, by capturing changes to the source or in real and near-real time.

In the next series of exercises, we'll use various services to seamlessly ingest data to the data lake.



Break



Agenda - AWS Modern Data Architecture

Time	Topic
13:00 – 13:20	LAB: Streaming Data Ingestion
13:20 – 13:35	LAB: Fined Grained Access Control - Column
13:35 – 14:35	Purpose Built Analytics Services
14:35 – 14:50	Break
14:50 – 15:20	LAB: Query with Amazon Redshift
15:20 – 15:45	LAB: Creating Federated Query to Redshift through Athena
15:45 – 16:00	LAB: Building Data Visualization with QuickSight
16:00 – 16:20	Recap / Learning Resources

Continue our lab! Please go to:
go.jansat.co/md-lab



Wi-fi: Guest
Password: BrokenWires@@2019



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

Lab Stream 2

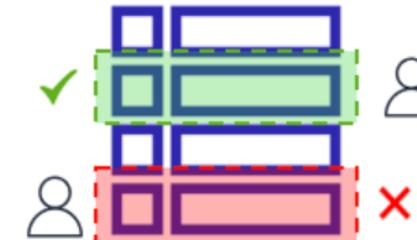
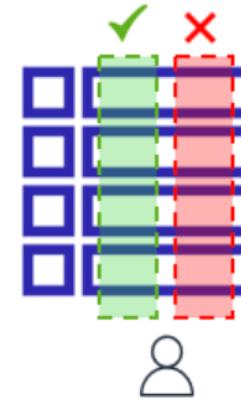
▼ Stream 2: Data Lake Security

▼ Column Level Access Control

- A. Configure Users and Permissions
- B. Verify Permissions

Stream 2: Data Lake Security

The data lake forms a core component of the modern data architecture. In this series of workshops, we'll continue to explore data lake permissions and the catalog.





Analytics services

Patiphan Pinkeaw

Solutions Architect (Boy)
Amazon Web Services



**Big Data
Processing**



**Data
Warehousing**



**Real-time
Analytics**



**Operational
Analytics**



**Interactive
Query**



**Serverless
Data processing**

Data warehousing: Amazon Redshift

- First and most popular cloud data warehouse

Data lake & AWS integration



Analyze exabytes of data across data warehouse, data lakes, and operational database

Query data across various analytics services

Best performance, most scalable



3x faster with RA3*

10x faster with AQUA*
*vs other cloud DWs

Adds unlimited compute capacity on-demand to meet unlimited concurrent access

Most secure & compliant



AWS-grade security (eg. VPC, encryption with KMS, CloudTrail)

All major certifications such as SOC, PCI, DSS, ISO, FedRAMP, HIPPA

Lowest cost



Cost-optimized workloads by paying compute and storage separately

1/10th cost of Traditional DW at \$1000/TB/year

Up to 75% less than other cloud data warehouses & predictable costs

- Easily Run Spark, Hadoop, Hive, Presto, HBase, and more big data apps on AWS

Latest versions



Updated with latest open source frameworks within 30 days

Low cost



50–80% reduction in costs with EC2 Spot and Reserved Instances
Per-second billing for flexibility

Use S3 storage



Process data in S3 securely with high performance using the EMRFS connector

Easy



Fully managed no cluster setup, node provisioning, cluster tuning

Operational Analytics

Fully managed, scalable, secure, Elasticsearch service

**Open source
Elasticsearch APIs,
Kibana, and Logstash**



Open-source Elasticsearch APIs
Managed Kibana
Integration with Logstash

Fully managed



Deploy Elasticsearch clusters
in minutes: simplified hardware
provisioning, software
installation/patching, failure
recovery, backups, and
monitoring

**Scalable, secure,
and compliant**



Scale clusters up/down via a
single API call or a few clicks
Secured network isolation
with VPC, encrypt data
at-rest and in-transit
Compliant: HIPPA, PCI DSS,
and ISO

**Pay only for
what you use**



Cost-optimized workloads
No upfront fee or
usage requirement
Critical features built-in:
encryption, VPC support,
24x7 monitoring

Amazon Athena

- Serverless, interactive query service

Query instantly



Zero setup cost

Point to S3 and start querying

Pay per query



Pay only for queries run

Save 30–90% on per-query costs through compression

Use S3 storage



ANSI SQL

JDBC/ODBC drivers

Multiple formats, compression types, and complex joins and data types

Easy



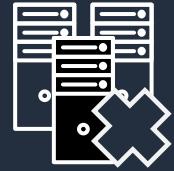
Serverless: zero infrastructure, zero administration

Integrated with QuickSight

Serverless analytics

Deliver on-demand analytics on the data lake

Analytics



Serverless
Zero infrastructure
Zero administration



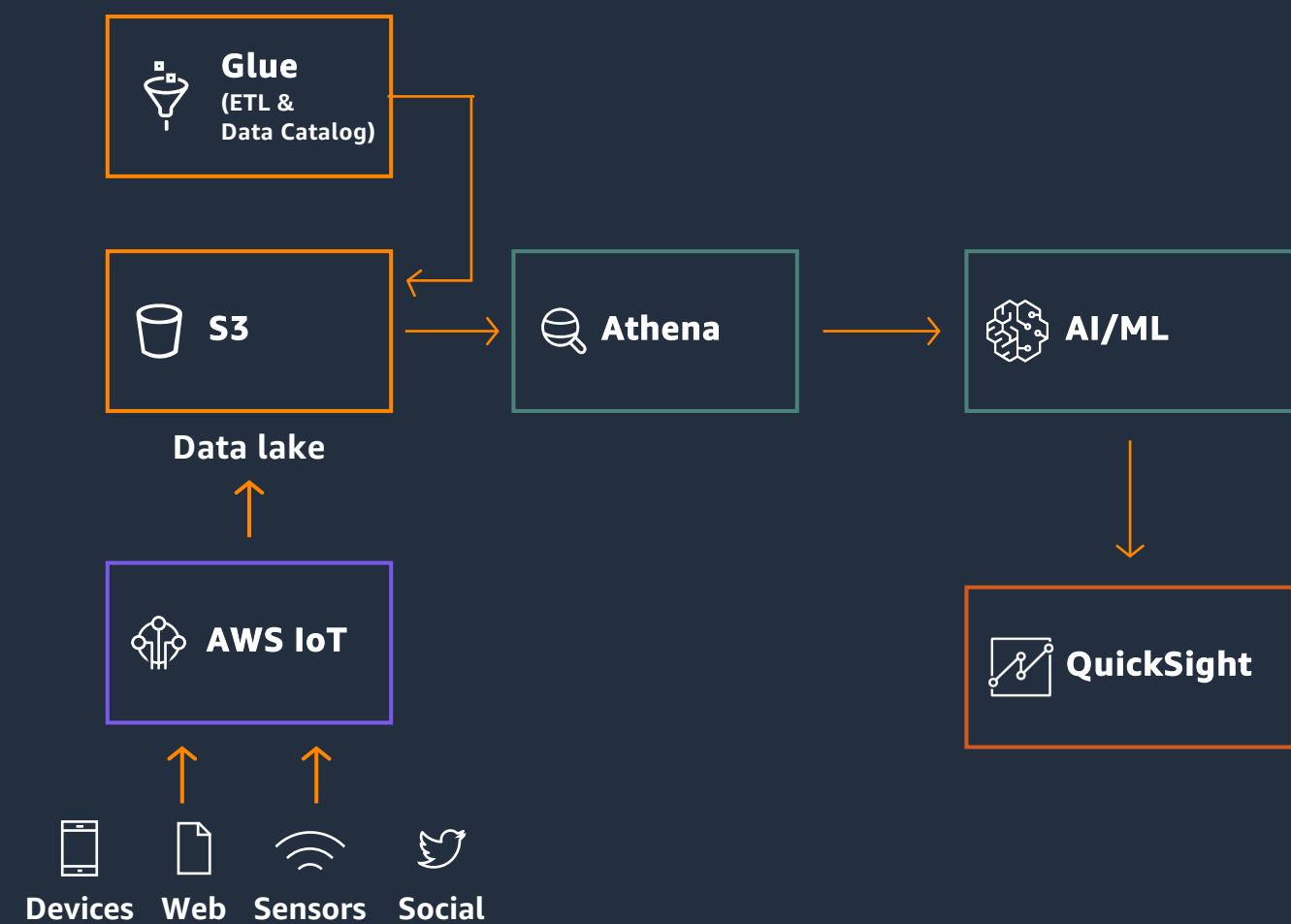
Never pay for
idle resources



Automatically scales
resources with usage



Availability and fault
tolerance built in





Data, visualization, engagement, & machine learning services

Data, visualization, engagement, & machine
learning



Data



Dashboards



Digital User Engagement



Predictive Analytics

Data lakes for machine learning

Data, visualization,
engagement, & ML

- Easier to discover relevant data
- More data makes more accurate and complete models
- More data sources provide more context and nuance
- More compute resources available when needed
- More specialized compute resources when needed
- Granular control over what kinds of data is seen
- Costs reduced by separating storage from compute



AWS Data Exchange

Easily find and subscribe to 3rd-party data in the cloud

Data, visualization,
engagement, & ML

Quickly find diverse
data in one place



>1,000 data products

>80 data providers
including include Dow
Jones, Change Healthcare,
Foursquare, Dun &
Bradstreet, Thomson
Reuters, Pitney Bowes,
Lexis Nexis, and Deloitte

Easily analyze data

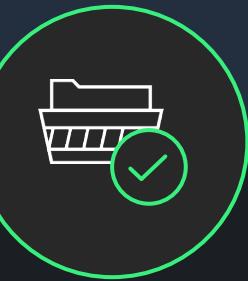


Download or copy data to S3

Combine, analyze, and model
with existing data

Analyze data with EMR,
Redshift, Athena, and AWS
Glue

Efficiently access
3rd party data



Simplifies access to data: No
need to receive physical
media, manage FTP
credentials, or integrate with
different APIs

Minimize legal reviews and
negotiations



Amazon QuickSight

First BI service built for the cloud with pay-per-session pricing & ML insights

Data, visualization,
engagement, & ML

Elastic Scaling



Auto-scale 10 to 10K+
users in minutes

Pay-as-you-go

Serverless



Create dashboards in
minutes

Deploy globally
without provisioning a
single server

Deeply integrated with AWS services



Secure, Private access to
AWS data

Integrated S3 data lake
permissions through AWS IAM

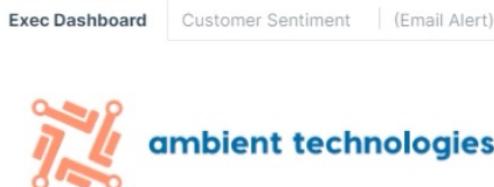
API Support



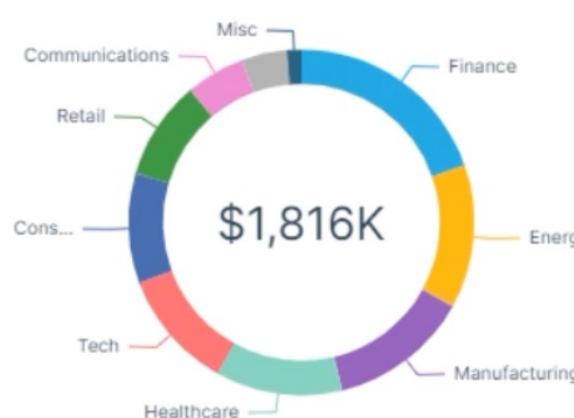
Programmatically onboard
users and manage content

Easily embed in your apps

WW Revenue ▾ Type a question about your data



Sales by Industry



Quarter

Quarter	Strategic	SMB	Enterprise
Q2' 2023	\$51,000	\$49,292	\$33,731
Q1' 2023	\$33,551	\$63,530	\$21,814
Q4' 2022	\$94,031	\$100,846	\$41,016
Q3' 2022	\$43,915	\$79,942	\$20,757
Q2' 2022	\$42,418	\$69,477	\$23,475
Q1' 2022	\$26,578	\$46,031	\$19,987
Kevin 120AWS (Guest)	\$89,844	\$40,208	

MoM Sales

\$2,609 ↑

MoM Customer Orders

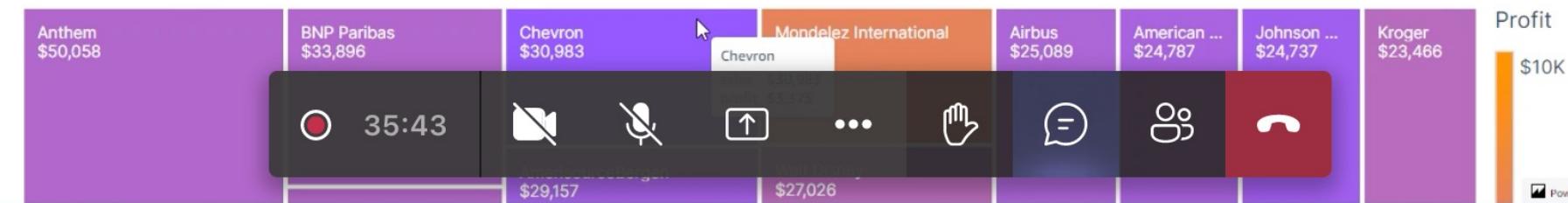
-26 ↓

Forecast

Total sales is forecasted to be \$48,657 for Aug 2024



Top 25 Customers



Anomalies Detected

The top anomaly detected on Mar 2023 was:

- Sales for **Mondelez International** at \$14,052, which was **higher** than the expected \$505

Region

All

Customer Segmentation



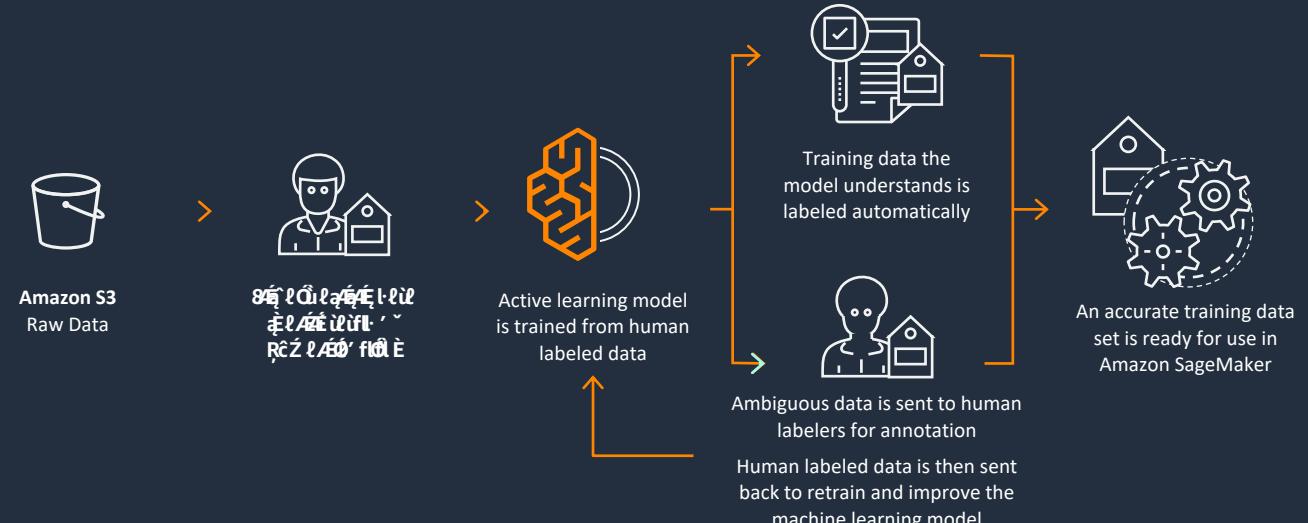
SHOWING TOP 48 IN LONGITUDE, LATITUDE AND TOP 3 IN REGION



Predictive insights with AWS ML & AI services

Data, visualization,
engagement, & ML

- **AI services** that enable developers to plug-in pre-built AI functionality into their apps
- **ML platform services** that make it easy for any developer to get started and get deep with ML
- **ML frameworks and interfaces** for machine learning practitioners



Lab Stream 4

▼ Stream 4: Act On Your Data

- ▶ Sentiment Analysis With SageMaker

▼ Dashboards with QuickSight

- A. Signup for QuickSight
- B. Create a Data Source for the Data Lake
- C. Create a Data Source for the Aurora database
- D. Create a Data Source for the Redshift Cluster (Optional)
- E. Visualize the Data

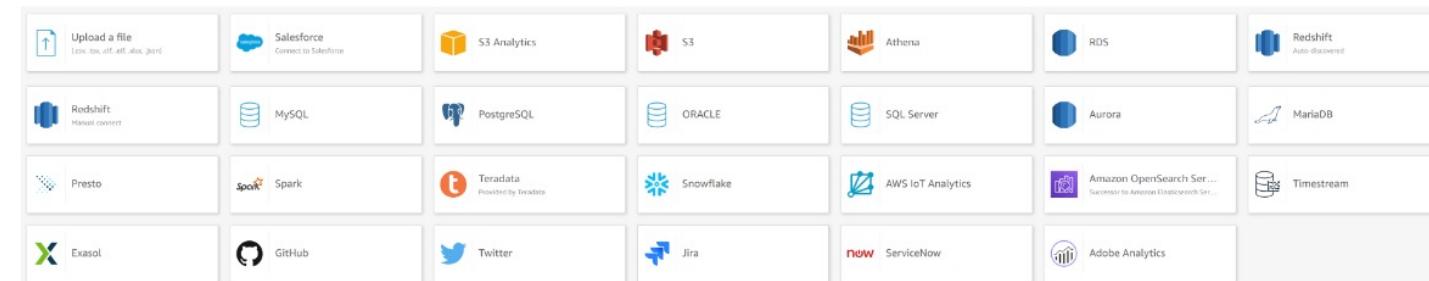
Learn More

- ▶ Log Analytics with OpenSearch

[AWS Modern Data Architecture Immersion Day](#) > [Stream 4: Act On Your Data](#) >
[Dashboards with QuickSight](#)

Dashboards with QuickSight

Amazon QuickSight is cloud native, serverless BI solution, which integrates with a wide range of services to produce compelling dashboards and visualizations. Amazon QuickSight allows everyone in your organization to understand your data by asking questions in natural language, exploring through interactive dashboards, or automatically looking for patterns and outliers powered by machine learning.



Lab Stream 5

AWS Modern Data Architecture Immersion Day > Stream 5: Data Sharing

Stream 5: Data Sharing

Data Sharing is an important part of scaling a modern data platform for Large enterprises. In these labs, we'll look at some ways you can explore data sharing on AWS within a modern data platform.

▼ Stream 5: Data Sharing

▼ Cross Account Sharing with Lake Formation

A. Prepare your Data Product

B. Share with other accounts

C. Consume data from another account

[Learn More](#)

► References

[Previous](#)

[Next](#)





Service Screener

Automated Well Architected Reviews

Nonthapat Kaewamporn (Tontan)

Associate Solutions Architect
Amazon Web Services (Thailand)

What is Service Screener?

Service Screener is a tool that runs automated checks on AWS environments and provides recommendations based on the AWS Well Architected Framework.

Demo



Checks as of today!

Services



Frameworks



Why does it exist?

- Time Consuming
 - Looking through the questionnaire (2-4 hours / pillars)
 - Perform fixes (2 weeks – 8 weeks)
- Lack of skills to perform fixes
- Unsure of the impacts on the changes

New Features in Q1!

- Host report in S3
- Excel Extraction Report
- JSON Output
- Serverless Services (Lambda)



Service Screener

Screenshot of the AWS GitHub repository for Service Screener v2. The repository is public and has 32 commits. The 'About' section describes it as a tool for automating checks on AWS environments based on best practices.

Code | Issues | Pull requests | Actions | Projects | Security | Insights

main · 1 branch · 0 tags

About

A tool that allows AWS customers to automate checks on their environment and services based on AWS best practices and provide recommendations on how to improve.

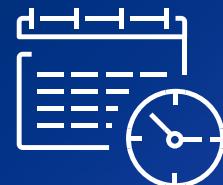
- Readme
- Apache-2.0 license
- Code of conduct
- Security policy
- 3 stars
- 4 watching
- 2 forks
- Report repository



go.jansat.co/ssv2

AWS Builders Online Series

Thai sessions



13 กรกฎาคม 2566

ลงทะเบียนเลย!
go.aws/3CmLOd8



recap:
**AWS Modern Data
Architecture**



Summary and resources

AWS Whitepapers and Guides

Expand your knowledge of the cloud with AWS technical content authored by AWS and the AWS community, including technical whitepapers, technical guides, reference material, and reference architecture diagrams

- Overview of Amazon Web Services
- Introduction to AWS Security
- Understanding Your Application Readiness when Migrating to AWS
- Web Application Hosting in the AWS Cloud: Best Practices
- Overview of Deployment Options on AWS

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

Best Practices for Security, Identity, & Compliance

AWS Architecture Center

The screenshot shows the AWS Architecture Center homepage. At the top, there's a navigation bar with links for Contact Us, Support, English, My Account, Sign In to the Console, Products, Solutions, Pricing, Documentation, Learn, Partner Network, AWS Marketplace, Customer Enablement, Events, Explore More, and a search icon. Below the navigation is a secondary navigation bar with links for AWS Architecture Center, Technology Categories, Video Series, AWS Well-Architected, Libraries, and More Resources.

Best Practices for Security, Identity, & Compliance

Featured Content

Identity & Access Management
Self-guided learning materials to help you understand identity security.
• Documentation: [Security Best Practices in IAM](#)

Detection
Information about monitoring services that can help you detect and eliminate suspicious activity.
• Documentation: [AWS Security Hub User Guide](#)

Infrastructure Protection
Holistic guidelines and trainings to help you prevent attacks and protect your business.
• Whitepaper: [AWS Best Practices for DDoS](#)

A day in the life of a security professional

Protecting secrets, keys, and data:

Revitalize your security with the AWS Security Reference Architecture

Filter by: Clear all filters ▾ Content Type: Patterns, Reference Architecture Diagrams, AWS Solutions, Guidance, Technical Guides, Whitepapers

Did this page help you? Yes, No, Feedback

Search the AWS Architecture Center

1-6 (318) PATTERN UPDATED, GUIDE NEW, PATTERN UPDATED



© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.



go.jansat.co/architect-center

AWS Resources in one page!



**AWS
Whitepapers
and Guides**
go.aws/3TvXGjk



**AWS Well-
Architected**
go.aws/3CG0nZ4



**AWS Quick
Starts**
go.aws/3MEETQI



**AWS Solutions
Library**
go.aws/3yNqETY

AWS Thailand Local Blog



AWS

ติดต่อเรา การสนับสนุน ▾ บัญชีของฉัน ▾ ลงซื้อเข้าใช้ สร้างบัญชี AWS

ผลิตภัณฑ์ โซลูชัน ราคา เอกสารประกอบ เรียนรู้ เครือข่ายคู่ค้า AWS Marketplace การพัฒนาลูกค้า กิจกรรม สำหรับคุณ > Q

หน้าหลักของบล็อก รุ่น ▾

AWS Thai Blog

**แนะนำขั้นตอนการ Claim เบอร์โทรศัพท์บน Amazon Connect**
by Vittawat Nooket | on 16 MAY 2023 | in [Amazon Connect](#), [Contact Lens For Amazon Connect](#) | [Permalink](#) | [Share](#)

หลังจากที่ได้มีการสร้าง Amazon Connect เส็จเรียบร้อยแล้ว คุณสามารถขอเบอร์โทรศัพท์ไว้ใช้สำหรับระบบ Contact center ได้และยังสามารถตั้งเป็นระบบ production เพื่อให้ลูกค้าของคุณสามารถติดต่อผ่านช่องทางต่างๆได้ หรือสามารถใช้เป็นเบอร์สำหรับการทดสอบร่วมระบบทำงานเป็นปกติ คุณไม่ต้องกังวลเรื่องการซื้อต่อคืนผู้ให้บริการโทรศัพท์หรือจ่ายเงินค่าสายที่ต้องใช้งาน. Amazon Connect จะช่วยให้การจัดการการซื้อต่อคืนผู้ให้บริการโทรศัพท์หรือจ่ายเงินค่าสายที่คุณจำเป็นต้องใช้ได้เอง โดยอุปกรณ์สามารถเชื่อมต่อผ่าน AWS support และระบุความต้องการไปในครั้งเดียว

**เริ่มต้นใช้งาน Amazon EKS ให้ตรงตาม best practice ด้วย EKS Blueprints**
by Wiriyang Pipatsakulroj | on 16 MAY 2023 | in [Amazon Elastic Kubernetes Service](#), [Containers](#) | [Permalink](#) | [Share](#)

อัปเดตข่าวประจำสัปดาห์, AWS Best Practices, AWS Solutions
และอื่น ๆ ฉบับภาษาไทย!



go.jansat.co/AWSThaiBlog



AWS User Group Community (Thailand)



AWS User Group (Thailand)

Public group · 11.9K members

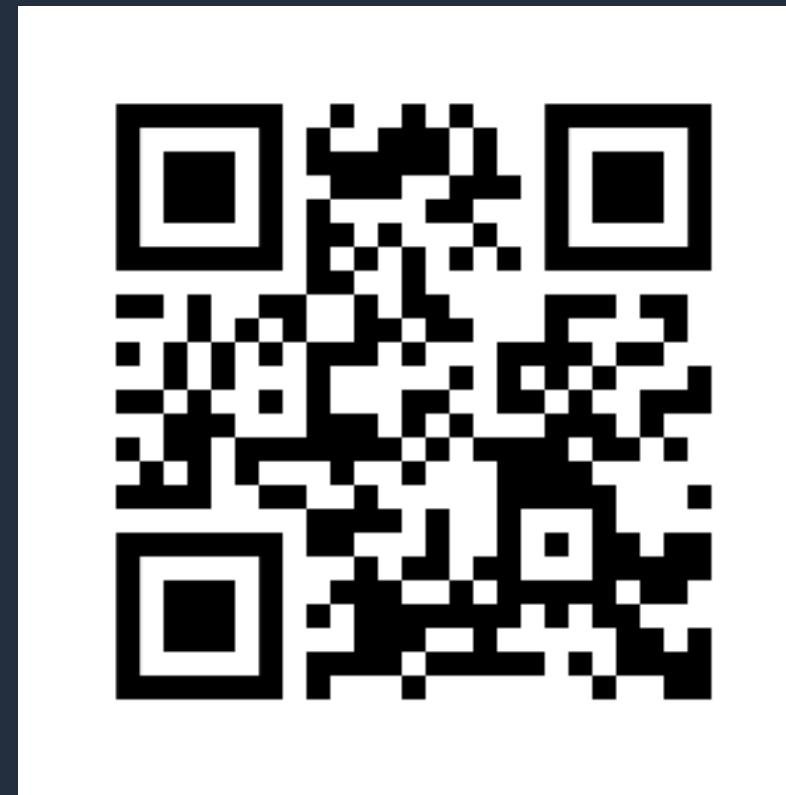
Join Group

About Discussion Featured Topics Events Media



go.jansat.co/AWSUserGroup

Please take a moment to fill out the survey below.



go.jansat.co/md-sv

Kahoot



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

Download today's presentation



go.jansat.co/md-dl

Contact:
Poonsiri Wongwiseskij

poonsiri@amazon.com



Thank you!