



# 데이터 수집부터 분석까지, 데이터 파이프라인 구축 함께 시작해요!

Amazon Web Services, Solutions Architect 성재우  
2022.10.12

# 목차

- 데이터 수집, 분석, 파이프라인
- 과거에 했던 일들, 어려웠던 점들
- AWS Modern Data Analytics & Pipeline
- LAB - Demogo Prime Data Pipeline Architecture

# 이번 Webinar 단 1개의 목표

- 우리가 가지고 있는 데이터를 AWS에서 분석하려면 어떻게 파이프라인을 구성해야 하는지?

# 데이터 수집, 분석, 파이프라인

# 데이터 = 전략적 자산

“세계에서 가장 가치 있는  
자원은 더이상 원유가  
아니라 데이터\*”

\*Copyright: The Economist, 2017, David Parkins



# 데이터 수집

 Source

 수집, 저장

 분석

 파이프라인



정형, 반정형,  
비정형 데이터



# 데이터 수집

 Source

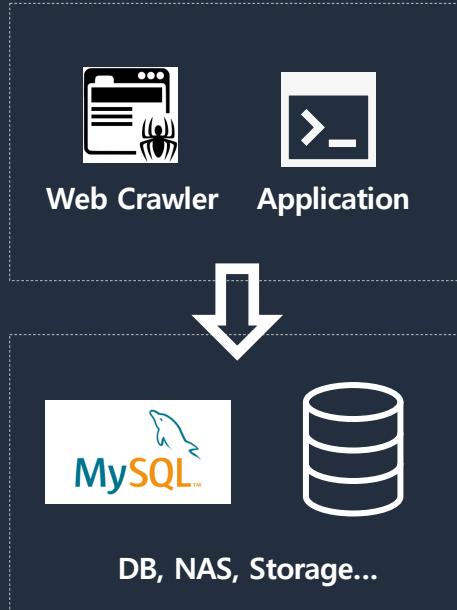
 수집, 저장

 분석

 파이프라인



정형, 반정형,  
비정형 데이터



# 데이터 분석

Source

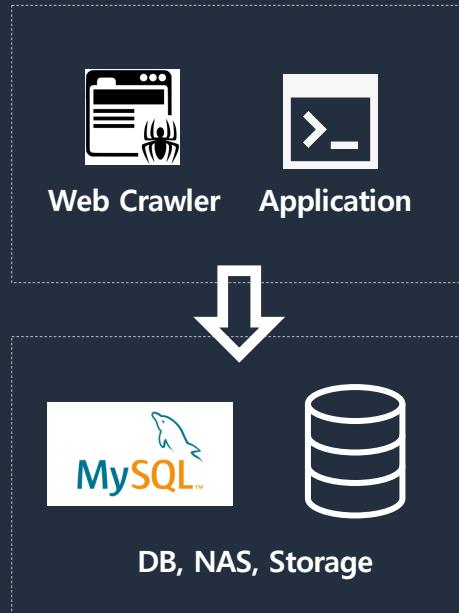
수집, 저장

분석

파이프라인



정형, 반정형,  
비정형 데이터



# 데이터 파이프라인

 Source

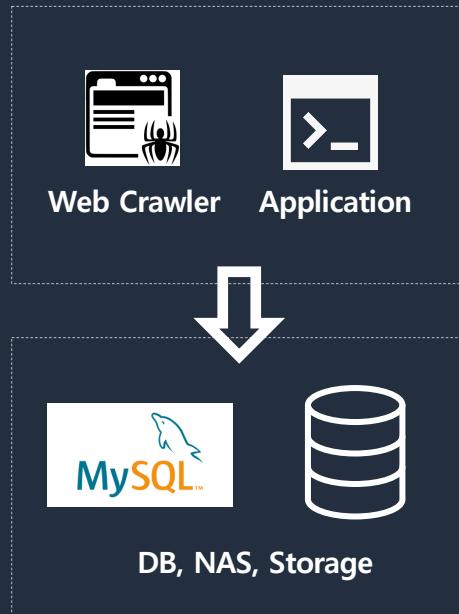
 수집, 저장

 분석

 파이프라인



정형, 반정형,  
비정형 데이터



 Query  
 Report

Automation

# 과거에 했던 일들, 어려웠던 점들

# 데이터 분석 플랫폼 발전 과정



# 기존의 방법으로 어려웠던 점



폭발적으로  
증가하는 데이터



새로운  
데이터 소스



다양해지는  
데이터 유형



많은 곳에서  
데이터를 사용



다양한 애플리케이션에서  
데이터를 사용

# 데이터 Silo 문제

모든 데이터 소스를 한번에 볼 수 있는 단일 뷰가 없는 상황



# 급격한 데이터 볼륨의 증가



상상하는 것 보다  
훨씬 많은 데이터

Data	Data platforms <b>need to</b>
grows $>10x$ every 5 years	live for <b>15</b> years
	scale <b>1,000x</b>

\* IDC, Data Age 2025: The Evolution of Data to Life-Critical Don't Focus on Big Data, Focus on the Data That's Big, April 2017.

## 다양한 데이터 유형



# 그 어느 때보다 다양해진 데이터 유형

# 데이터 분석 기술의 발전



그 어느 때보다  
빠르게 발전하는  
데이터 분석 기술

<-2006-----2010-----2012-----2013--

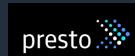
Years ago



Hadoop



Elasticsearch



Presto

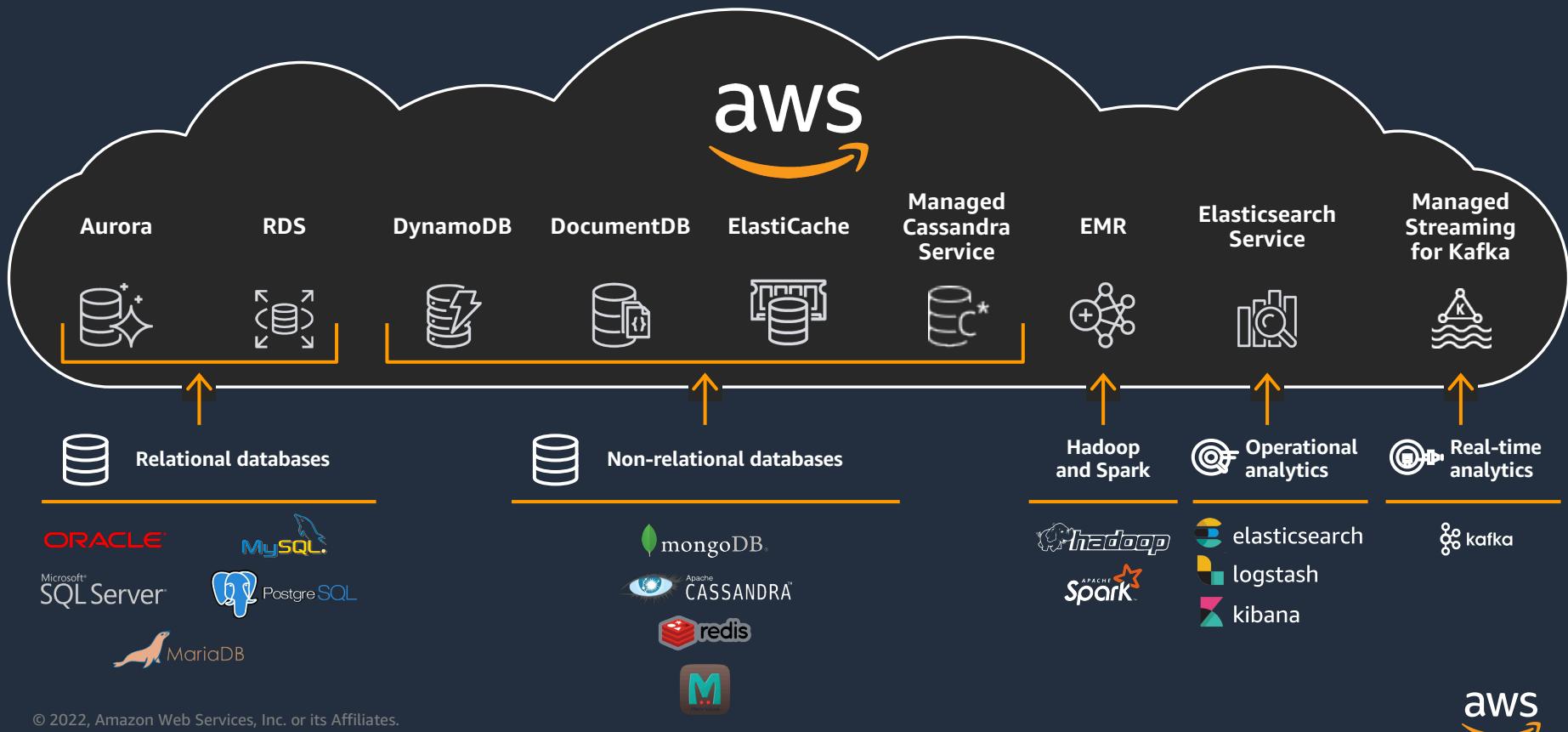


Spark

Didn't exist

# AWS Modern Data Analytics & Pipeline

# Data Trends 그리고 AWS

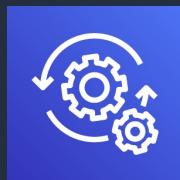


# AWS의 완전 관리형 데이터 분석 포트폴리오



# 완전 관리형 분석 서비스의 특징

직접 서버에 설치해서 운영하지 않아도 되는 완전 관리형 분석 서비스



AWS Managed  
Services

- ✓ 변경 요청
- ✓ 모니터링
- ✓ 패치 관리
- ✓ 보안
- ✓ 백업 서비스



Amazon  
Opensearch Service



Amazon Managed  
Streaming for Kafka



Amazon Redshift



Amazon EMR

# 데이터 레이크 아키텍처로 이동

Amazon S3를 활용한 데이터 레이크 구축



데이터 웨어하우징



분석



기계 학습

DW 아키텍처의 확장 또는 발전

모든 데이터를 모든 형식으로 저장

내구성, 가용성 및 엑사바이트 규모

보안, 규정 준수, 감사 가능

DW에서 예측에 이르기까지 모든 유형의 분석 실행

# AWS가 제공하는 다양한 분석 포트폴리오

## 데이터, 시각화, 참여 & 기계 학습

신규



Data Exchange



QuickSight



Pinpoint



SageMaker



Comprehend



Lex



Polly



Rekognition



Translate

+ 그 외 다양핚

## 분석



Redshift



EMR  
(Spark &  
Hadoop)



AWS Glue  
(Spark &  
Python)



Athena



Elasticsearch  
Service



Kinesis Data  
Analytics

## 데이터 레이크 인프라 & 관리



S3/Glacier



Lake  
Formation



AWS Glue

## 데이터 이동

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

# AWS 데이터 수집

# AWS Data Lake Pipeline Flow



수집



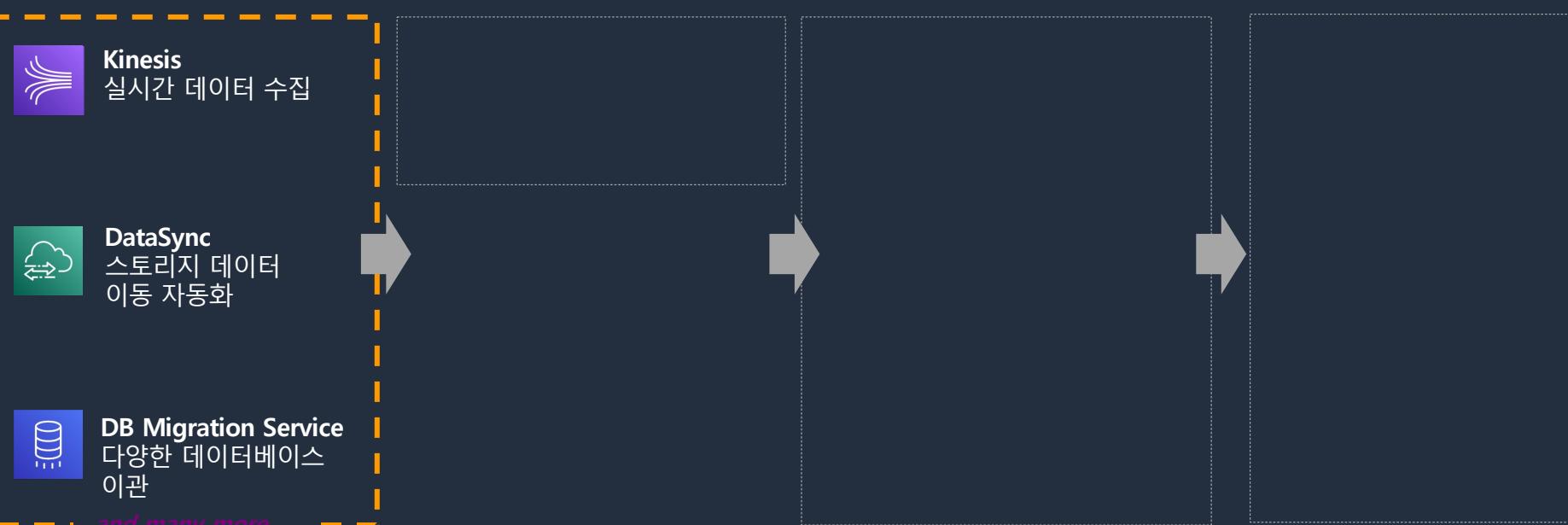
저장



처리 및 분석



소비



# AWS Database Migration Service (DMS)

데이터 마이그레이션 서비스



다양한 DB Source



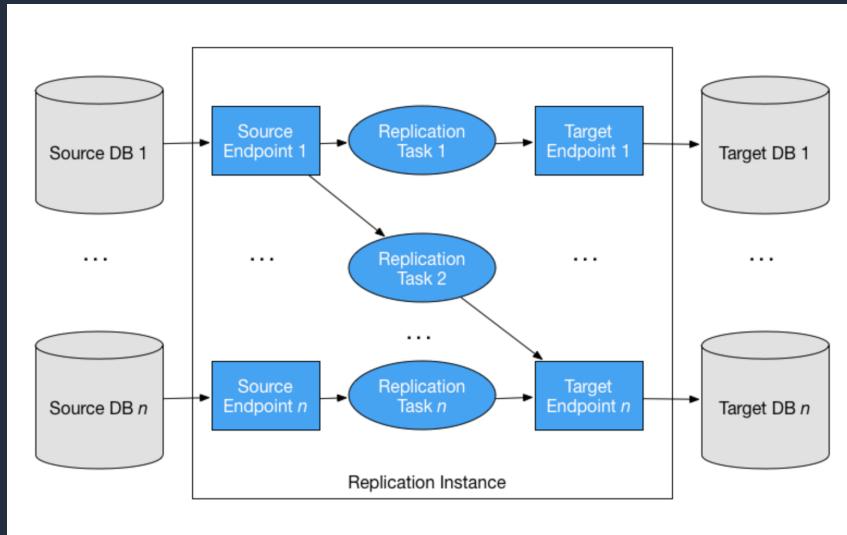
실시간 변경 반영 (CDC)



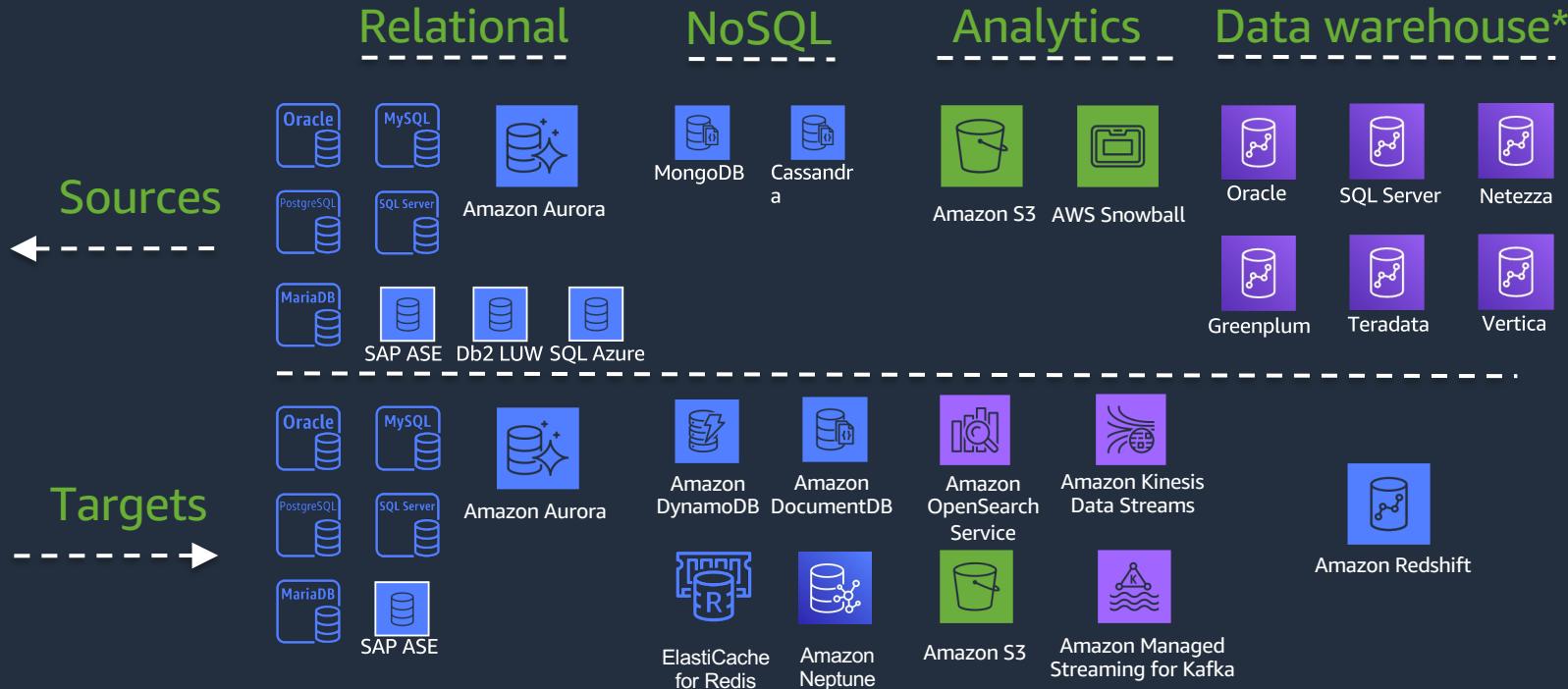
이기종 데이터베이스 마이그레이션

# 데이터 수집 – AWS Database Migration Service(DMS)

- Source Endpoint, Target Endpoint를 지정 후 Replication Task를 생성
- API, SDK, AWS CLI등을 사용하여 Task를 원하는 시점에 실행도 가능



# 데이터 수집 – AWS Database Migration Service(DMS)



\* Supported via AWS SCT data extractors

# AWS DataSync

온라인 데이터 전송 및 마이그레이션



온프레미스에서 데이터 이동



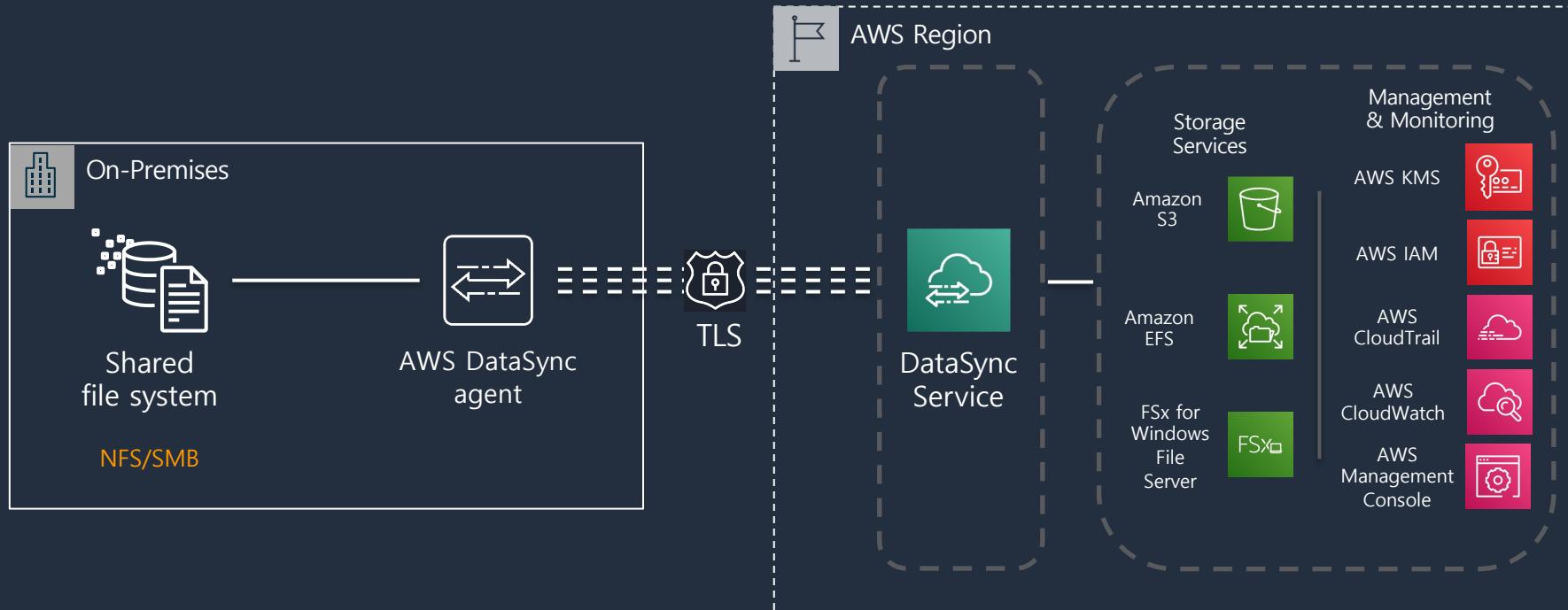
NFS, SMB, HDFS 등 다양한  
프로토콜 지원



다양한 AWS 스토리지 지원  
S3, EFS, FSx, SnowCone

# 데이터 수집 – AWS DataSync

AWS와의 데이터 전송을 단순화, 자동화 및 가속화



# Amazon Kinesis

실 시간으로 데이터 수집, 프로세싱, 분석



Kinesis Data Streams



Kinesis Data Analytics



Kinesis Video Streams



Kinesis Data Firehose

# 데이터 수집 – AWS Kinesis Streams



이상 및 사기 탐지



실시간으로 고객 경험 조정

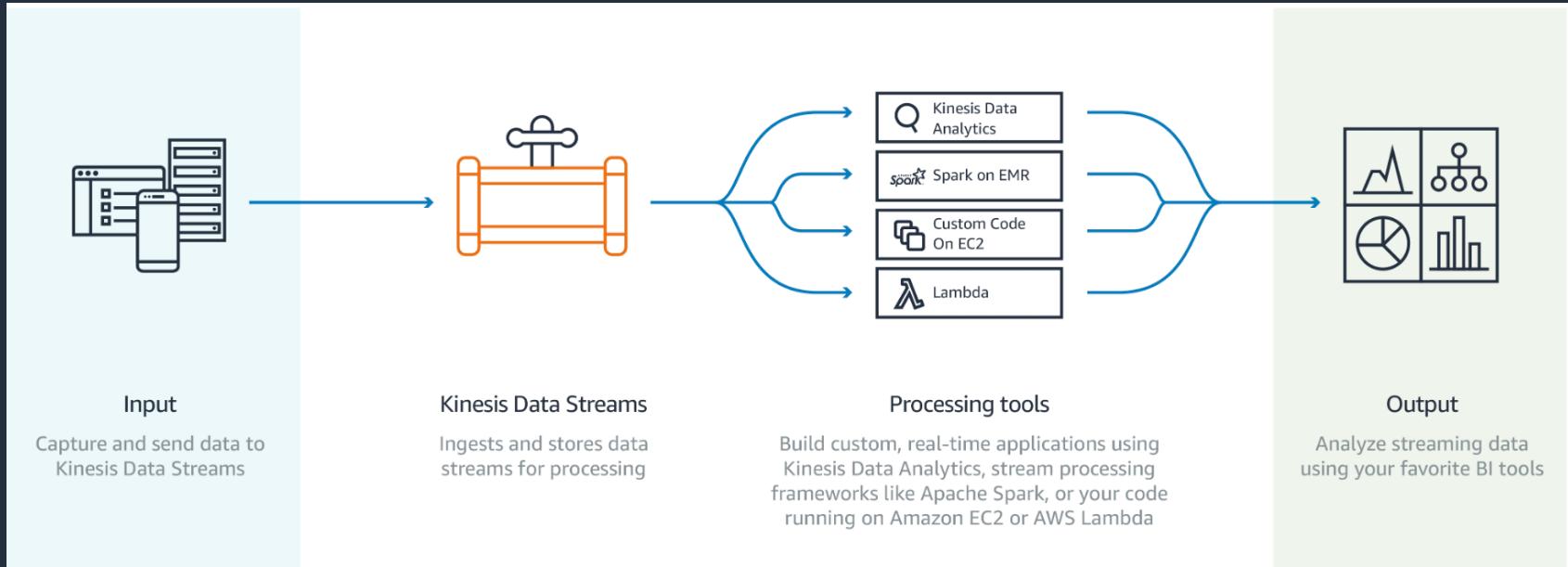


실시간 개인화

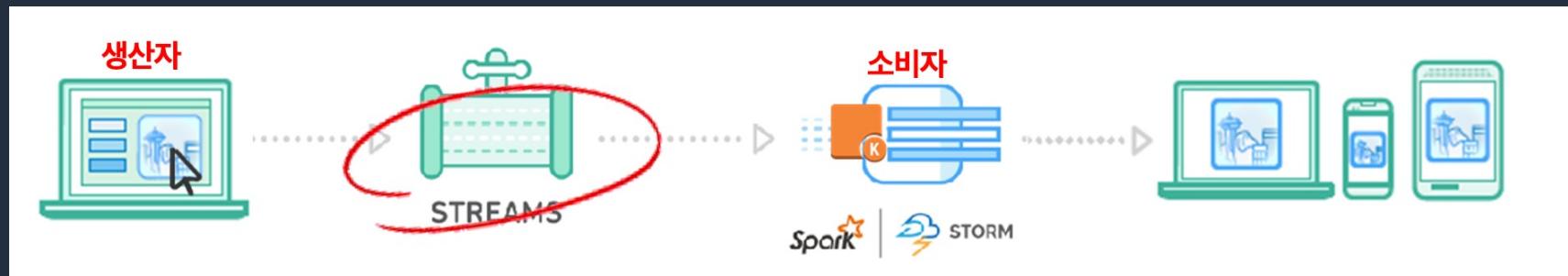


의료 및 응급 서비스 지원

# 데이터 수집 – AWS Kinesis Streams



# 데이터 수집 – AWS Kinesis Streams



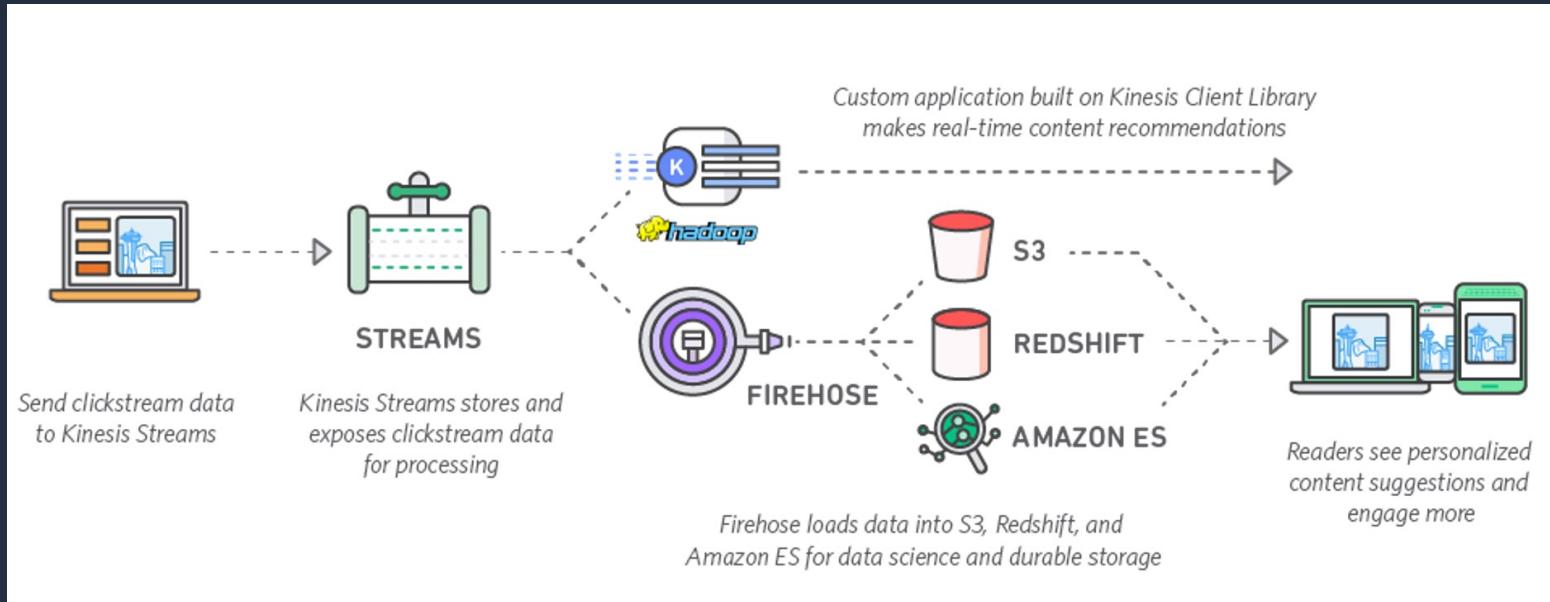
클릭 스트림 데이터 전송

클릭 스트림 데이터를 저장 및  
프로세싱 전송

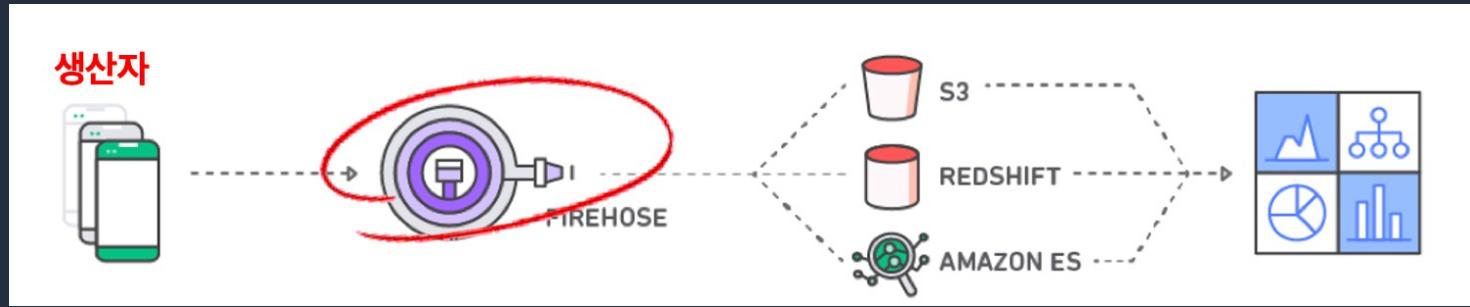
KCL 기반 앱으로  
실시간 개인화 컨텐츠 생성

개인화 된 컨텐츠 제안

# 데이터 수집 – AWS Kinesis Data Firehose



# 데이터 수집 – AWS Kinesis Data Firehose



스트리밍 데이터 캡처 및 전송

FireHose 는 Amazon S3, Redshift 및 Elasticsearch에  
지속적으로 스트리밍 데이터 로드

선호하는 BI 도구를 사용하여  
스트리밍 데이터 분석

# AWS 데이터 저장

# AWS Data Lake Pipeline Flow



수집



저장



처리 및 분석



소비



Kinesis  
실시간 데이터 수집



DataSync  
스토리지 데이터  
이동 자동화



DB Migration Service  
다양한 데이터베이스  
이관

*and many more...*



Amazon S3  
안전하고, 비용  
효율적인 스토리지

*Interoperate with everything*

# Amazon Simple Storage Service(S3)

최고의 확장성, 가용성, 보안 및 성능  
객체 스토리지



99.999999999% 내구성



뛰어난 가용성, 확장성 그리고  
합리적인 비용



보안 및 컴플라이언스

# 데이터 저장 – Amazon Simple Storage Service(S3)

밀리세컨드 내의 지연시간을 가지는 보안성과 높은 확장성, 내구성을 가지는 오브젝트 스토리지  
모든 유형의 데이터 저장소 – 웹사이트, 모바일앱, 기업용 어플리케이션, IoT센서

내구성, 가용성과  
확장성



99.99999999%의 내구성  
설계; 데이터는 AWS  
Region 내부에 3곳의  
물리적 공간에 분산  
저장됨; 자동으로 다른  
AWS Region에 복제  
구성 가능

보안과  
컴플라이언스



3가지의 다른 형태의 암호화  
기능을 제공; 리전 간  
전송시에도 데이터 암호화  
전송; CloudTrail로 로그 및  
모니터링하며, ML 기반  
Macie로 민감 데이터를  
찾아내고 보호할 수 있음

즉각적인 쿼리수행



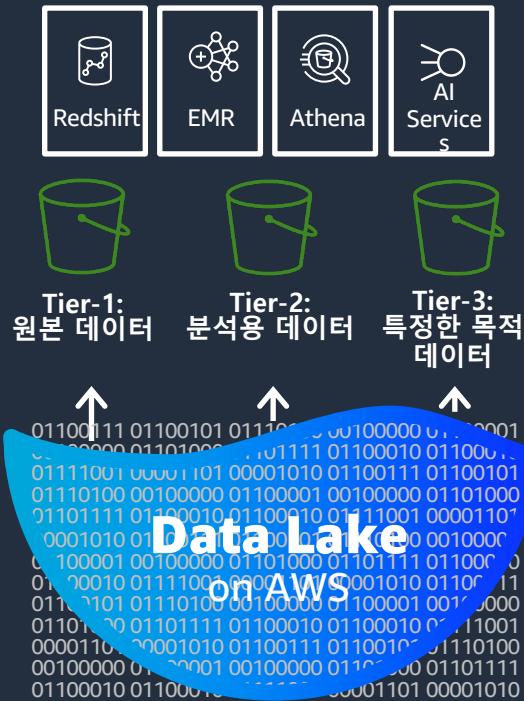
데이터 이전 없이 Data Lake  
에서 분석 및 ML을 실행;  
S3 Select를 사용해서  
데이터의 하위 집합을  
검색하고 분석 퍼포먼스를  
400% 증가

유연한 관리



데이터 사용 트랜드를  
분류, 보고 및 시각화;  
오브젝트에 태그를 붙여  
스토리지 사용과 비용 및  
보안을 확인 가능;  
보관기간 및 Tiering을  
자동화하는 수명주기 관리  
정책 작성

# 데이터 저장 – Amazon S3 설계 방식 (Tiering)



## Tier-1: 원본 데이터

- 원본 데이터의 저장과 보관
- 최소한의 데이터 변환 작업만
- S3의 라이프사이클 기능 활용, S3-IA 또는 Glacier

## Tier-2: 분석용 데이터

- Parquet / ORC 같은 컬럼 방식 포맷의 사용
- 파티션 정책에 따라 분산 및 파티션 유지 관리
- 분석을 위한 최적화

## Tier-3: 특정한 분석 목적 데이터

- 도메인 레벨로 데이터마트 분리
- Use Case에 적합한 구성
- 특정 분석 방식에 적합한 데이터 변경 (ML, AI)

# AWS 데이터 처리

# AWS Data Lake Pipeline Flow



수집



저장



처리 및 분석



소비



# 데이터 카탈로그

## Schema on Read

- 데이터 Schema 확인을 데이터를 읽는 시점에 하는 것
- Insert 할 때 스키마 체크를 하지 않음으로 빈번하게 데이터가 바뀔 수 있는 빅데이터 환경에 최적화된 방법



## Schema on Write

- MySQL, Oracle 등에 해당. 사전에 Table DDL을 설정해 놓고 Insert할 때 Table 정보와 데이터가 다르다면 Error

# AWS Data Lake Pipeline Flow



수집



저장



처리 및 분석



소비



Kinesis  
실시간 데이터 수집



DataSync  
스토리지 데이터 이동 자동화



DB Migration Service  
다양한 데이터베이스 이관

*and many more...*



AWS Glue  
데이터 카탈로그



Amazon S3  
안전하고, 비용 효율적인 스토리지



EMR, Glue  
비정형 데이터 처리



Athena  
ad-hoc 쿼리



Redshift  
데이터 웨어하우스



OpenSearch  
로그 분석 및 검색 엔진



Kinesis Data Firehose  
실시간 스트리밍 분석

*and many more...*

*Interoperate with everything*

# AWS Glue

간편하고 확장 가능한  
서버리스 ETL 서비스



빠른 데이터 통합



확장 가능한 자동화



관리할 서버가 없음

# AWS Glue의 핵심 기능

## SERVERLESS DATA INTEGRATION SERVICE

Scalable Data  
Integration Engine



Built-in data transforms



Execution engine



Monitor

# AWS Glue의 핵심 기능

## SERVERLESS DATA INTEGRATION SERVICE

Scalable Data  
Integration Engine



Built-in data transforms



Execution engine



Monitor

Centralized and Unified  
Data Governance



Glue data catalog



Glue crawlers



Lake formation

# AWS Glue의 핵심 기능

## SERVERLESS DATA INTEGRATION SERVICE

Scalable Data  
Integration Engine



Built-in data transforms



Execution engine



Monitor

Centralized and Unified  
Data Governance



Glue data catalog



Glue crawlers



Lake formation

Connect and  
Ingest Data



Glue connectors



Glue connector marketplace



Variety of interfaces

# AWS Glue의 핵심 기능

## SERVERLESS DATA INTEGRATION SERVICE

Scalable Data  
Integration Engine



Built-in data transforms



Execution engine



Monitor

Centralized and Unified  
Data Governance



Glue data catalog



Glue crawlers



Lake formation

Connect and  
Ingest Data



Glue connectors



Glue connector marketplace



Variety of interfaces

User Productivity  
and Data Ops



Persona specific tools



Productivity tools



Data ops tools

# AWS Glue의 핵심 기능

## SERVERLESS DATA INTEGRATION SERVICE

Scalable Data  
Integration Engine



Built-in data transforms



Execution engine



Monitor

Centralized and Unified  
Data Governance



Glue data catalog



Glue crawlers



Lake formation

Connect and  
Ingest Data



Glue connectors



Glue connector marketplace



Variety of interfaces

User Productivity  
and Data Ops



Persona specific tools



Productivity tools



Data ops tools

# Amazon Athena

표준 SQL을 통한 S3 데이터 분석



즉시 쿼리 가능



쿼리당 비용 지불



오픈 소스 기반으로 표준 SQL 지원

# 데이터 활용 – AWS Athena

- Presto DB Engine 기반
- Serverless : 인프라 관리 불필요, 자동 확장
- S3에서 직접 Query 실행
- 데이터 scan 용량 만큼 과금
- AWS Glue 데이터 카탈로그와 통합
  - Schema on Read



# 데이터 활용 - AWS Athena

Athena    Query Editor    Saved Queries    History    Data sources    Workgroup : primary

Settings    Tutorial    Help    What's new

10+

**Data source** Connect data source  
awsdatacatalog

**Database** ticketdata

Filter tables and views...

**Tables (22)** Create table

- mlb\_data
- name\_data
- nfl\_data
- nfl\_stadium\_data
- parquet\_person
- parquet\_sport\_location
- parquet\_sport\_team
- parquet\_sporting\_event
- parquet\_sporting\_event\_ticket
- person
- player
- seat
- seat\_type
- sport\_division
- sport\_league
- sport\_location
- sport\_team
- sporting\_event
- sporting\_event\_ticket
- sporting\_event\_ticket\_1bb4a008b349ed873527a4c2b9f8ac5f
- ticket\_purchase\_hist
- ticket\_purchase\_hist\_95f83e3d847527d7c4e84a4949d62d2b

**Views (2)** Create view

- sporting\_event\_info
- sporting\_event\_ticket\_info

New query 1 New query 2 New query 3 New query 4 +

```
1 SELECT "mlb_id", "mlb_name", "mlb_pos", "mlb_team", "mlb_team_long", "bats", "throws", "birth_year", "bp_id", "bref_id", "bref_name", "cbs_id", "cbs_name",
2   "cbs_pos", "espn_id", "espn_name", "espn_pos", "fg_id", "fg_name", "lahman_id", "nfbc_id", "nfbc_name", "nfbc_pos", "retro_id", "retro_name", "debut", "yahoo_id",
3   "yahoo_name", "yahoo_pos", "mlb_depth"
4 FROM "Ticketdata"."mlb_data"
5 LIMIT 1000
```

Run query Save as Create (Run time: 1.77 seconds, Data scanned: 395.04 KB)

Format query Clear

Use Ctrl + Enter to run query, Ctrl + Space to autocomplete

## Results

	mlb_id	mlb_name	mlb_pos	mlb_team	mlb_team_long	bats	throws	birth_year	bp_id	bref_id	bref_name	cbs_id	cbs_name
1	+5.065600000000000e+05	Alexi Amarista	3B	SD	San Diego Padres	L	R	1989	+5.588900000000000e+04	amarial01	Alexi Amarista	1735053	Alexi A
2	+4.582100000000000e+05	Alexi Casilla	2B	TB	Tampa Bay Rays	S	R	1984	+4.579900000000000e+04	casilal01	Alexi Casilla	1103724	Alexi C
3	+4.683960000000000e+05	Alexi Ogando	P	ATL	Atlanta Braves	R	R	1983	+4.991000000000000e+04	ogandal01	Alexi Ogando	1174266	Alexi O
4	+4.696860000000000e+05	Alfredo Aceves	P	NYY	New York Yankees	R	R	1982	+4.692800000000000e+04	aceveal01	Alfredo Aceves	1638980	Alfredo
5	+4.516280000000000e+05	Alfredo Figaro	P	TEX	Texas Rangers	R	R	1984	+5.245300000000000e+04	figaral01	Alfredo Figaro	1654334	Alfredo
6	+5.540540000000000e+05	Alfredo Gonzalez	C	CWS	Chicago White Sox	R	R	1992	+6.920100000000000e+04			2210083	Alfredo
7	+5.012450000000000e+05	Alfredo Marte	LF	BAL	Baltimore Orioles	R	R	1989	+5.175500000000000e+04	marteal01	Alfredo Marte	1956711	Alfredo
8	+4.305800000000000e+05	Alfredo Simon	P	CIN	Cincinnati Reds	R	R	1981	+4.558100000000000e+04	simonal01	Alfredo Simon	448968	Alfredo
9	+4.553780000000000e+05	Ali Solis	C	LAD	Los Angeles Dodgers	R	R	1987	+5.236200000000000e+04	solisal01	Ali Solis	1946524	Ali Solis
10	+4.888520000000000e+05	Allan Dykstra	1B	TB	Tampa Bay Rays	L	R	1987	+5.781300000000000e+04	dykstal01	Allan Dykstra	1960245	Allan D
11	+5.018000000000000e+05	Allen Craig	1B	BOS	Boston Red Sox	R	R	1984	+5.104300000000000e+04	craigal01	Allen Craig	1661498	Allen C
12	+5.439030000000000e+05	Allen Webster	P	PIT	Pittsburgh Pirates	R	R	1990	+5.878700000000000e+04	webstal01	Allen Webster	1799467	Allen W
13	+6.072370000000000e+05	Amir Garrett	P	CIN	Cincinnati Reds	R	L	1992	+7.094600000000000e+04			2053474	Amir G
14	+4.448430000000000e+05	Andre Ethier	RF	LAD	Los Angeles Dodgers	L	L	1982	+4.596200000000000e+04	etheian01	Andre Ethier	490390	Andre E
15	+5.165890000000000e+05	Andre Rienzo	P	MIA	Miami Marlins	R	R	1988	+5.663800000000000e+04	rienzan01	Andre Rienzo	2027388	Andre R
16	+5.927430000000000e+05	Andrelton Simmons	SS	LAA	Los Angeles Angels	R	R	1989	+6.711900000000000e+04	simmoan01	Andrelton Simmons	1918584	Andrelton S
17	+4.332170000000000e+05	Andres Blanco	2B	PHI	Philadelphia Phillies	S	R	1984	+3.283500000000000e+04	blancan01	Andres Blanco	392116	Andres B

# AWS 데이터 활용

# AWS Data Lake Pipeline Flow



수집



저장



처리 및 분석



활용



Kinesis  
실시간 데이터 수집



DataSync  
스토리지 데이터 이동 자동화



DB Migration Service  
다양한 데이터베이스 이관

*and many more...*



AWS Glue  
데이터 카탈로그와 ETL



Amazon S3  
안전하고, 비용 효율적인 스토리지



EMR  
비정형 데이터 처리



Redshift  
데이터 웨어하우스



Athena  
ad-hoc 쿼리



OpenSearch  
로그 분석 및 검색 엔진



Kinesis  
실시간 스트리밍 분석

*and many more...*

*Interoperate with everything*



QuickSight  
비즈니스 인텔리전스



Data Exchange  
데이터 구독



ML & AI  
머신러닝 개발

*and many more...*

# Amazon QuickSight

인사이트 도출을 위한 클라우드  
기반 데이터 시각화 도구



오토스케일링 및 서비스



내부/외부 사용자 접근



AWS 서비스와 결합



머신러닝을 통한 인사이트 도출

# 데이터 활용 – AWS QuickSight



## Cloud native = No 서버 = 오토스케일

관리, 유지 관리, 배포할 서버나 소프트웨어가 없이 수십  
사용자부터 시작해서 수천 사용자까지 확장



## AWS와 완전 통합

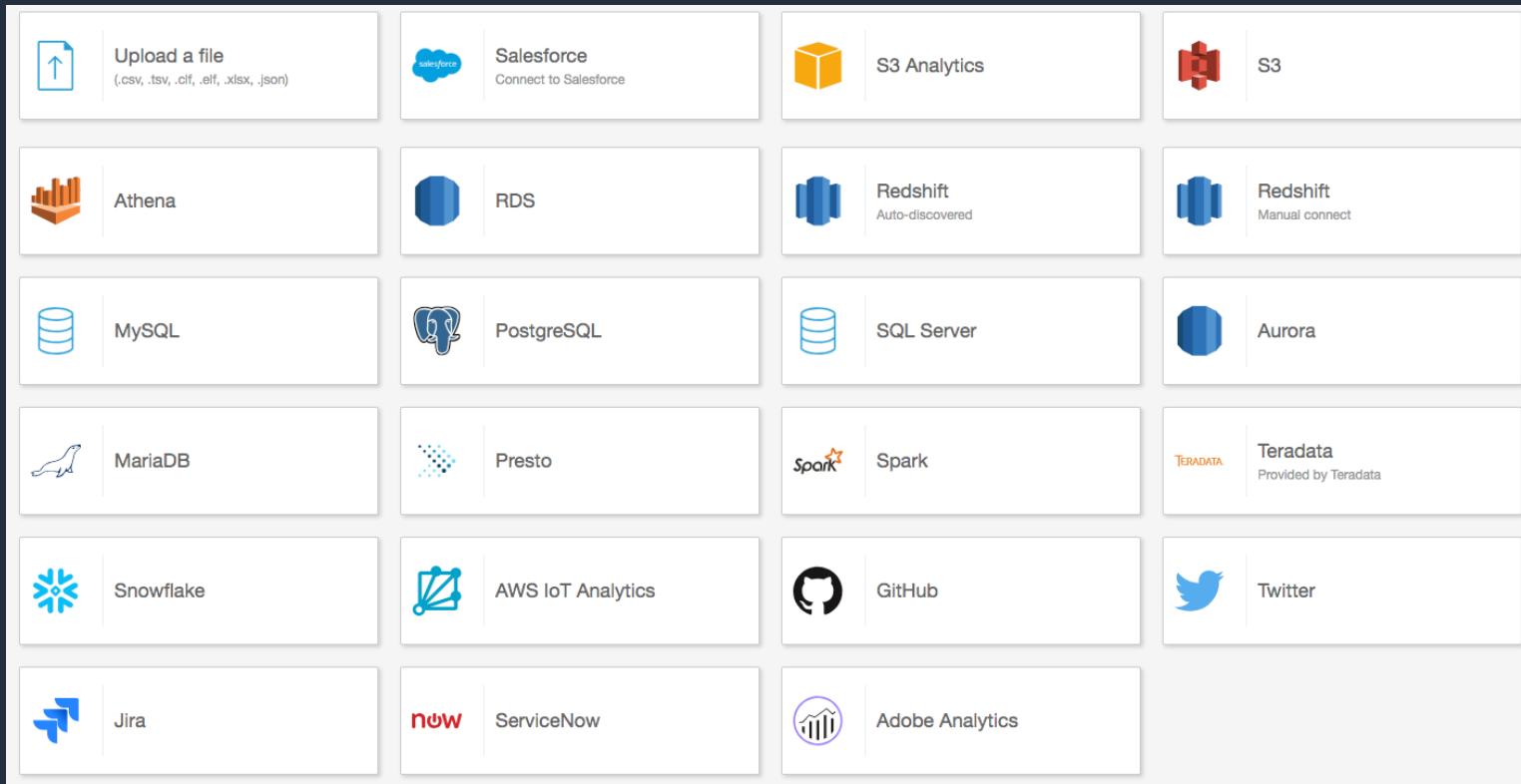
AWS에서 End-to-end 분석 구축. 안전한 프라이빗 VPC 액세스,  
세분화된 액세스 제어, ML 통합



## 모두를 위한 Insight

모든 사용자에게 액세스 권한을 제공하고 사용량에 대해서만  
지불. 선 결제 없음. 비활성 사용자에 대한 비용 없음

# 데이터 활용 – AWS QuickSight



# 데이터 활용 – AWS QuickSight

Total Sales (All)

\$43B

Total Net Sales(All)

\$22B

TOTAL SALES(YoY)

Percent Difference  
-28% ↓

2021  
\$817M

Average MAU

7,782K명

Total Sales Per Year



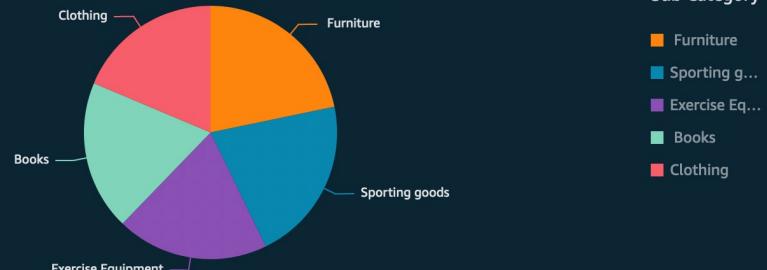
Total Sales by categories



Total Sales per Channel



Total Sales by sub categories



# AWS 데이터 파이프라인

# Amazon MWAA

AMAZON MANAGED WORKFLOW  
OF APACHE AIRFLOW



관리형 오케스트레이션 서비스



IAM, KMS를 통한 향상된 보안



오픈 소스 기반으로 Airflow 호환

# 데이터 활용 - AWS MWAA



Setup



Scaling



Security



Upgrades



Maintenance

- 신속한 Airflow 배포
- 오픈 소스 Airflow와 통일
- Airflow UI와 동일
- Console, CLI, API, or CloudFormation
- 원활한 Worker Scaling, 구성이 필요하지 않음
- Celery Executor 사용
- ECS Fargate
- IAM과 통합된 AuthN 및 AuthZ
- VPC 또는 퍼블릭 보안 엔드 포인트가 있는 Airflow UI
- Workers 는 AWS 서비스에 대한 간소화된 액세스를 위해 IAM 실행 역할 수행
- 업그레이드를 위한 유지 관리 기간
- 장애 발생 시 스냅샷 및 롤백
- CloudWatch 를 사용한 통합 모니터링
- Multi-AZ
- 장애 발생 시 자동으로 재시작

# DAG

```
from airflow import DAG

default_args = {
    'owner': 'airflow',
    'depends_on_past': False,
    'email': ['ricsue@amazon.com'],
    'email_on_failure': False,
    'email_on_retry': False
}

DAG_ID = 'daily_dw_ingest'

dag = DAG(
    dag_id=DAG_ID,
    default_args=default_args,
    description='First Apache Airflow DAG',
    schedule_interval=None,
    start_date=days_ago(2),
    tags=['devcon', 'demo'],
)
```

dag\_id=daily\_dw\_ingest

## Task

```
move_file = BashOperator(  
    task_id='move_current_file',  
    bash_command="cd {work_dir} && mv  
{source_file} {destination_file}"  
    dag=dag  
)
```



task\_id=move\_current\_file

## Task



```
move_file = BashOperator(  
    task_id='move_current_file',  
    bash_command="cd {work_dir} && mv  
{source_file} {destination_file}"  
    dag=dag  
)
```



## Operators

BashOperator  
PythonOperator  
DummyOperator  
...

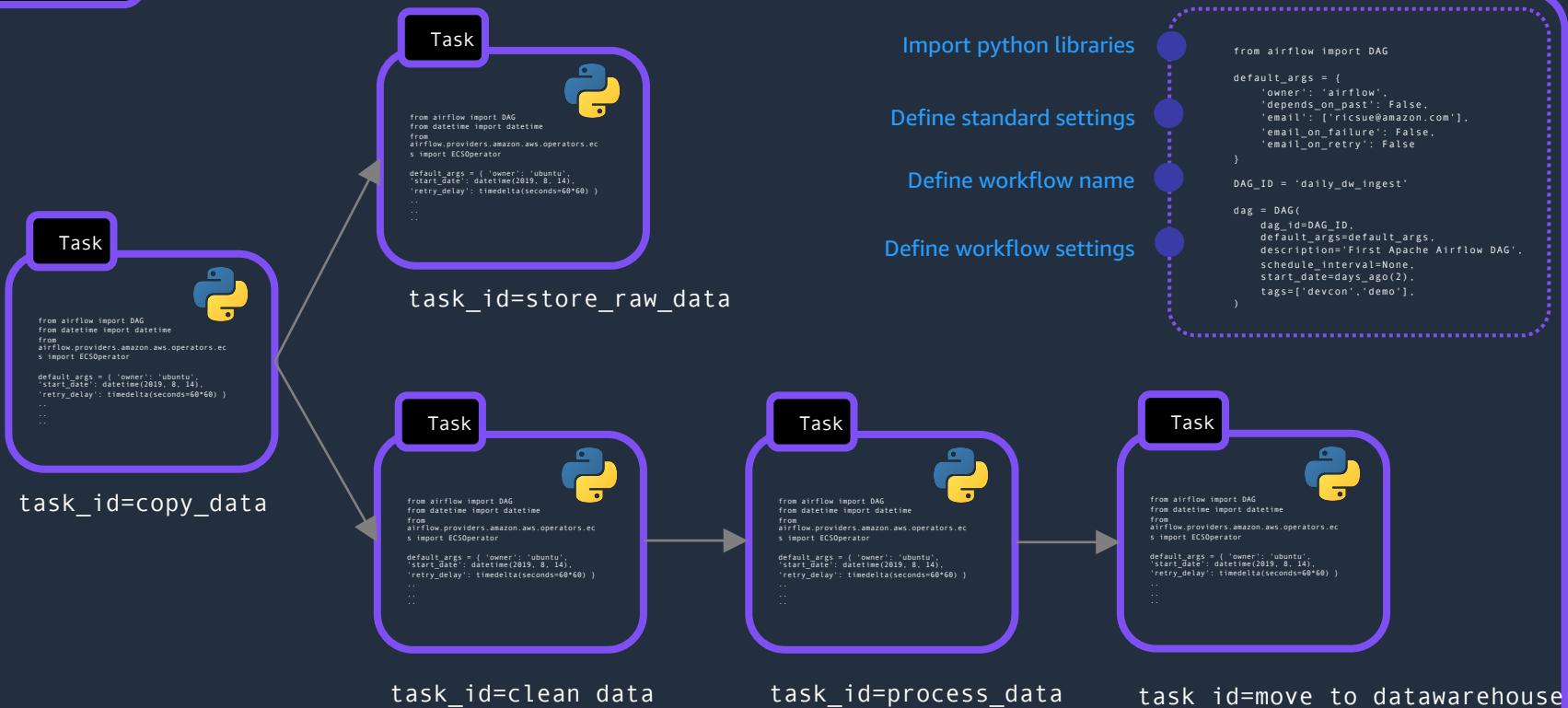
import

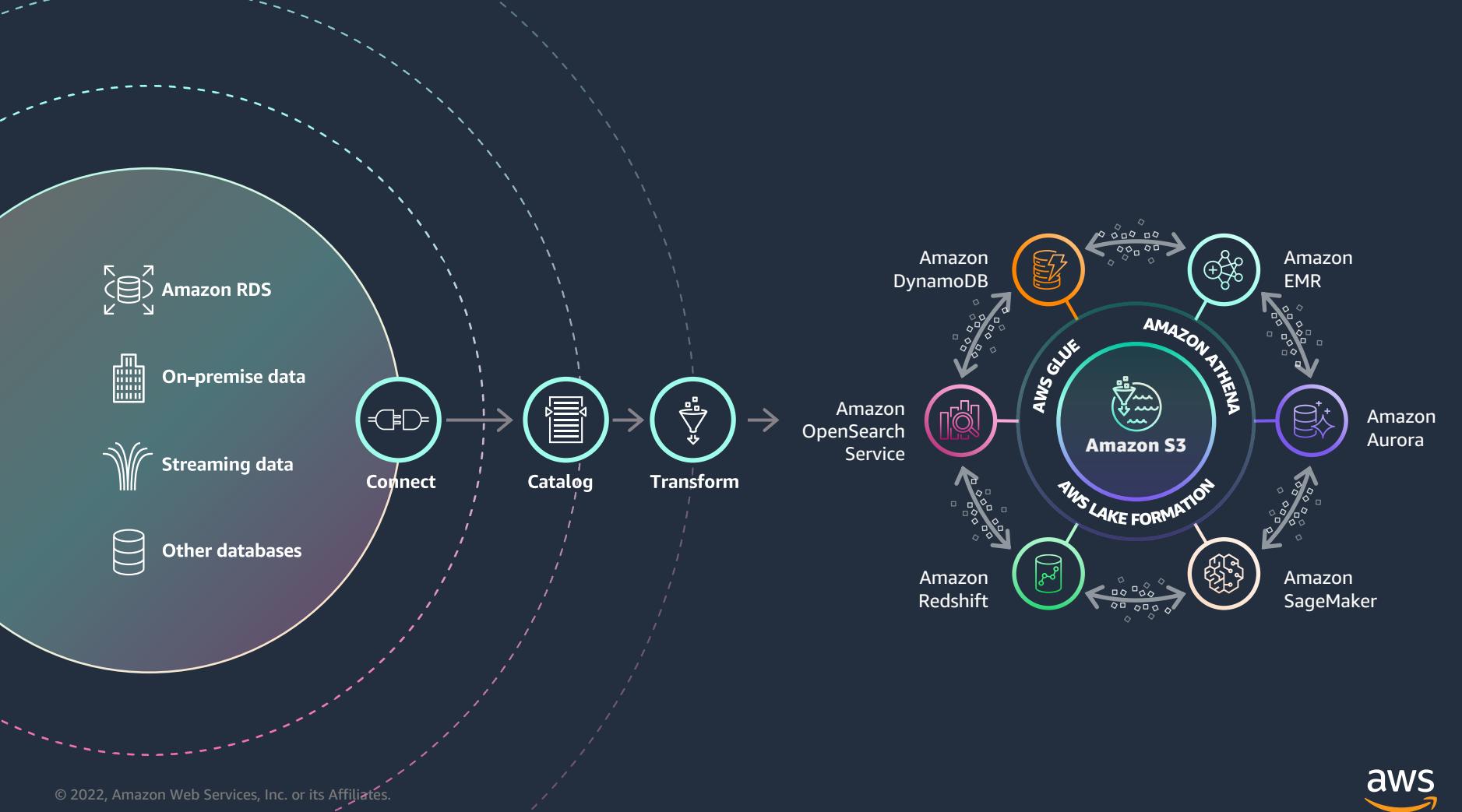
# Operators



- Airbyte
- Alibaba
- Amazon
- Apache Beam
- Apache Cassandra
- Apache Drill
- Apache Druid
- Apache HDFS
- Apache Hive
- Apache Kylin
- Apache Livy
- Apache Pig
- Apache Pinot
- Apache Spark
- Apache Swoop
- Asana
- Celery
- IBM Cloudant
- Kubernetes
- Databricks
- Datadog
- Dingding
- Discord
- Docker
- Elasticsearch
- Exasol
- Facebook
- File Transfer Protocol (FTP)
- Github
- Google
- gRPC
- Hashicorp
- Hypertext Transfer Protocol (HTTP)
- Influx DB
- Internet Message Access Protocol (IMAP)
- Java Database Connectivity (JDBC)
- Jenkins
- Jira
- Microsoft Azure
- Microsoft PowerShell Remoting Protocol (PSRP)
- Microsoft SQL Server (MSSQL)
- Windows Remote Management (WinRM)
- MongoDB
- MySQL
- Neo4J
- ODBC
- OpenFaaS
- Opsgenie
- Oracle
- Pagerduty
- Papermill
- Plexus
- PostgreSQL
- Presto
- Qubole
- Redis
- Salesforce
- Samba
- Segment
- Sendgrid
- SFTP
- Singularity
- Slack
- Snowflake
- SQLite
- SSH
- Tableau
- Telegram
- Trino
- Vertica
- Yandex
- Zendesk

# DAG





# Demogo Prime Lab Architecture

# Five Original Source Types



---

MySQL(RDB)  
정형데이터

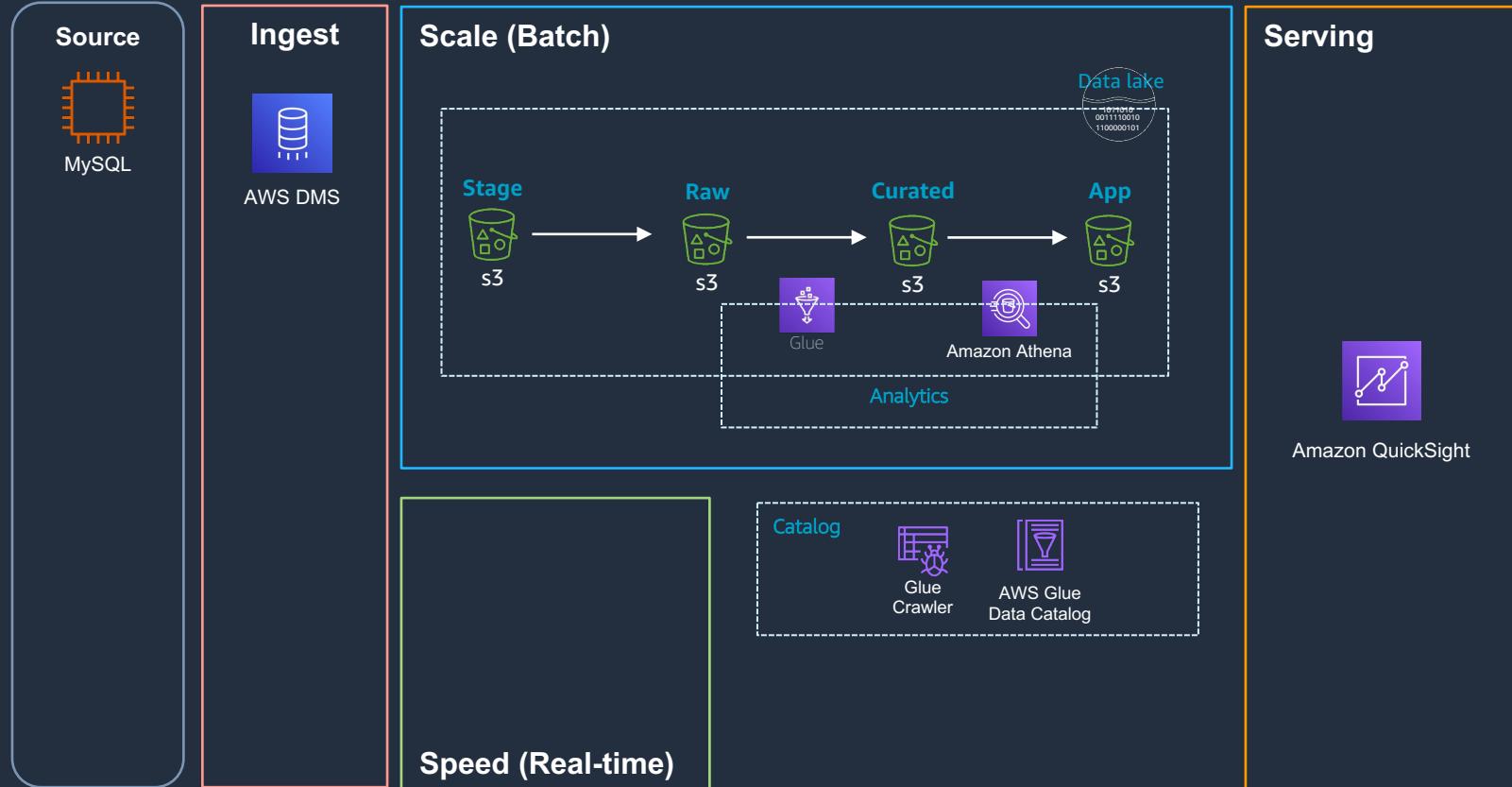
MongoDB(NoSQL  
반정형 데이터)

Window  
FileServer  
비정형 데이터

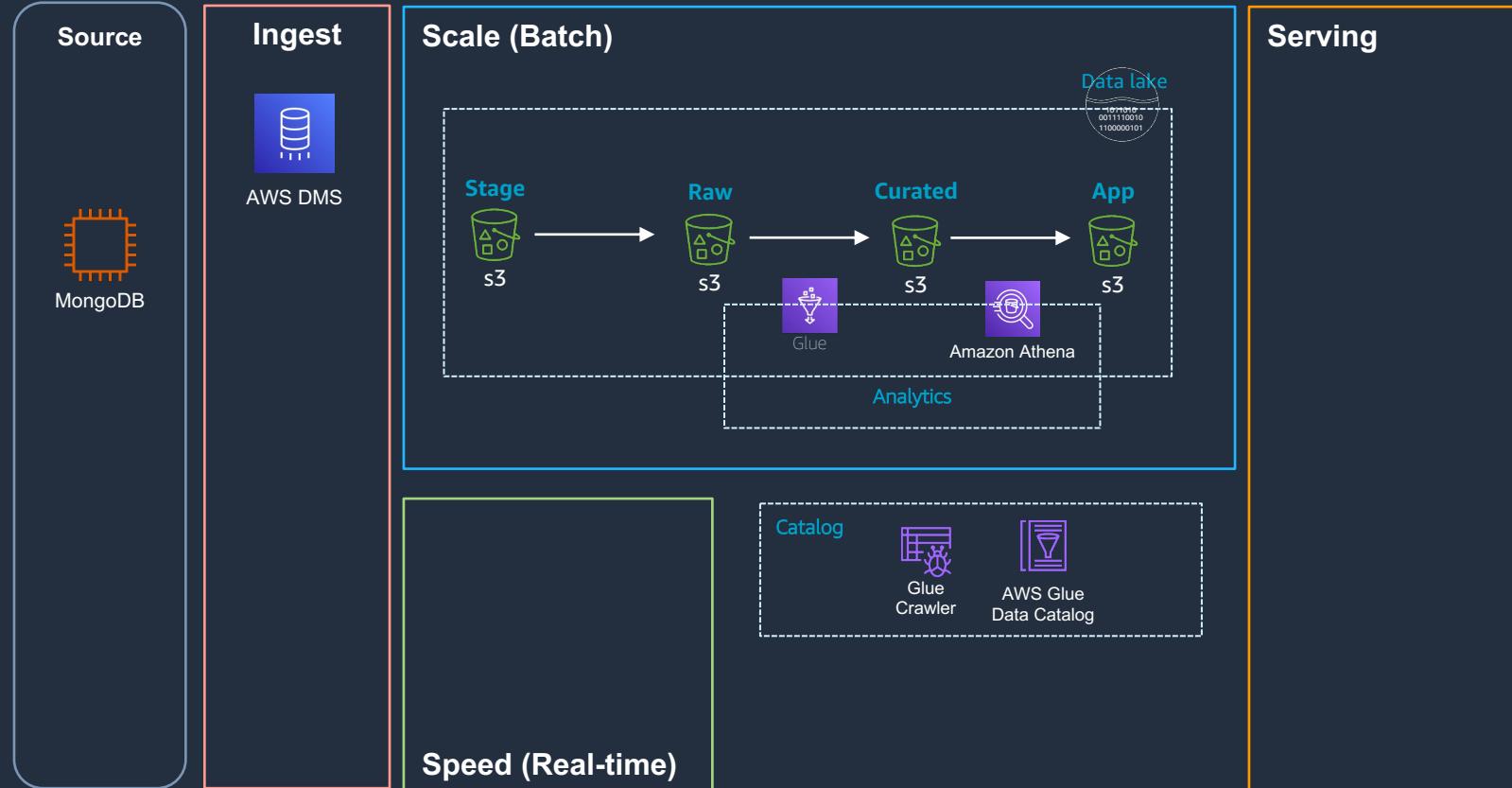
Event(Streaming)  
반정형 데이터

External Data  
반정형 데이터

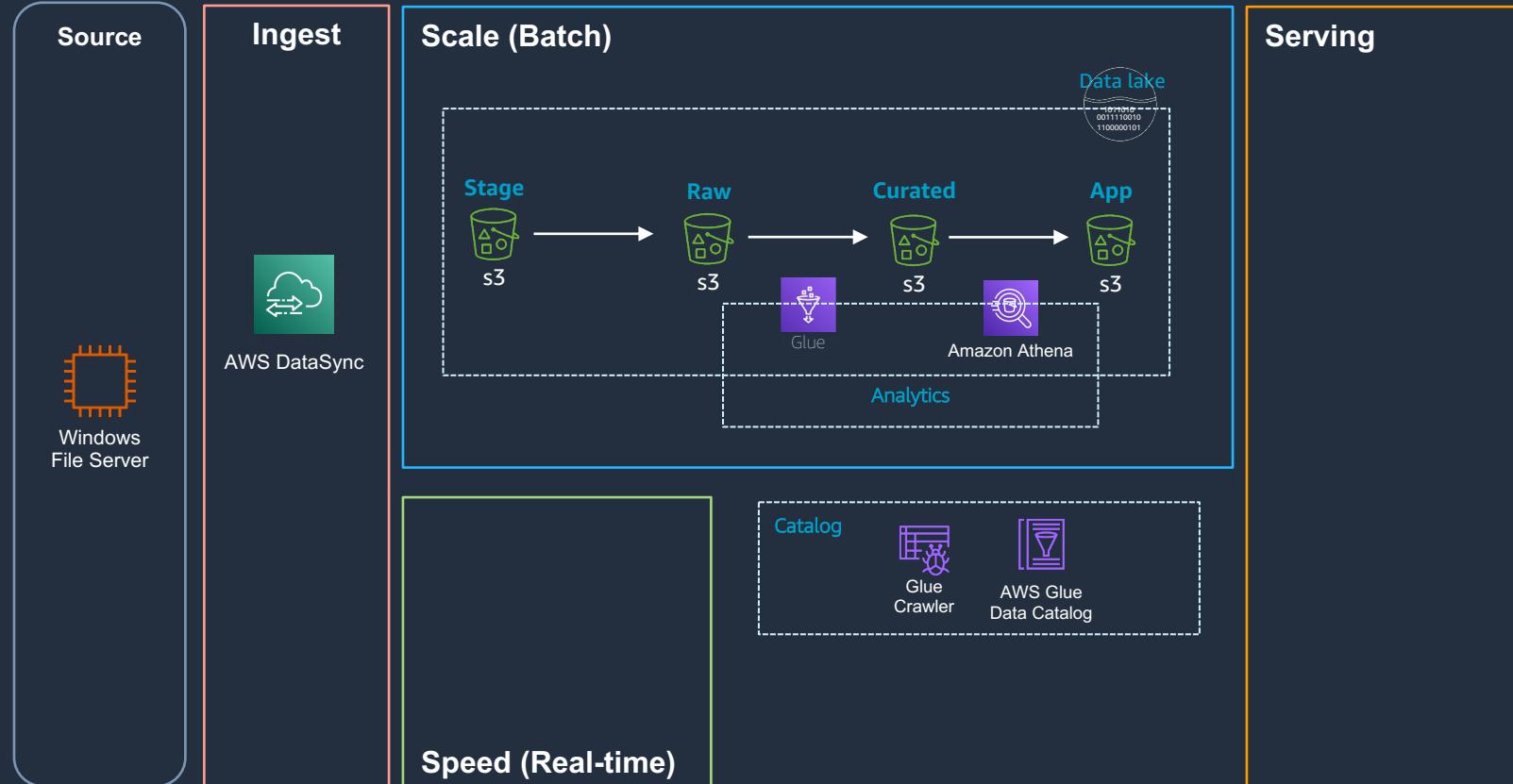
# Building Data Pipeline with multiple sources



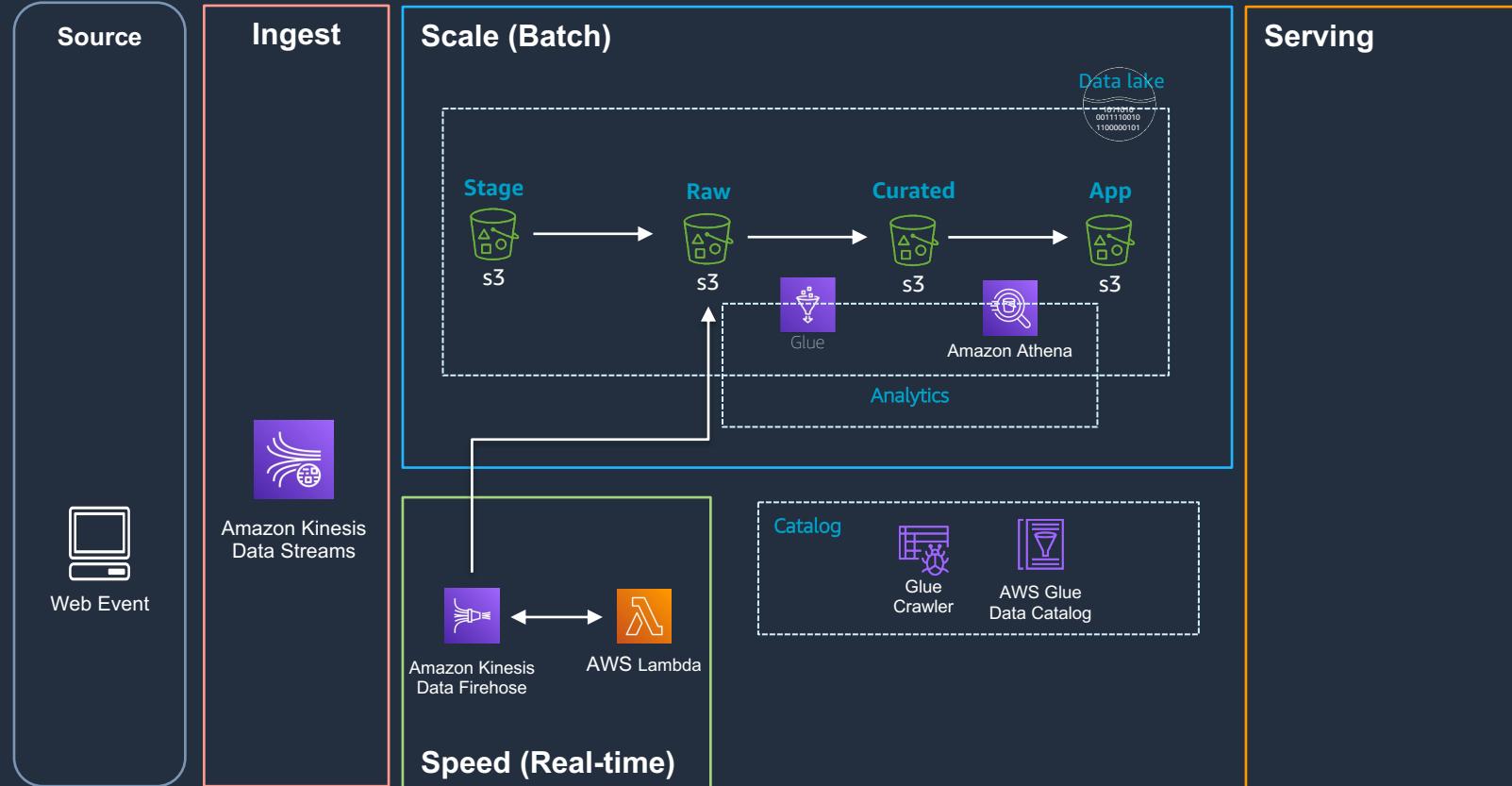
# Building Data Pipeline with multiple sources



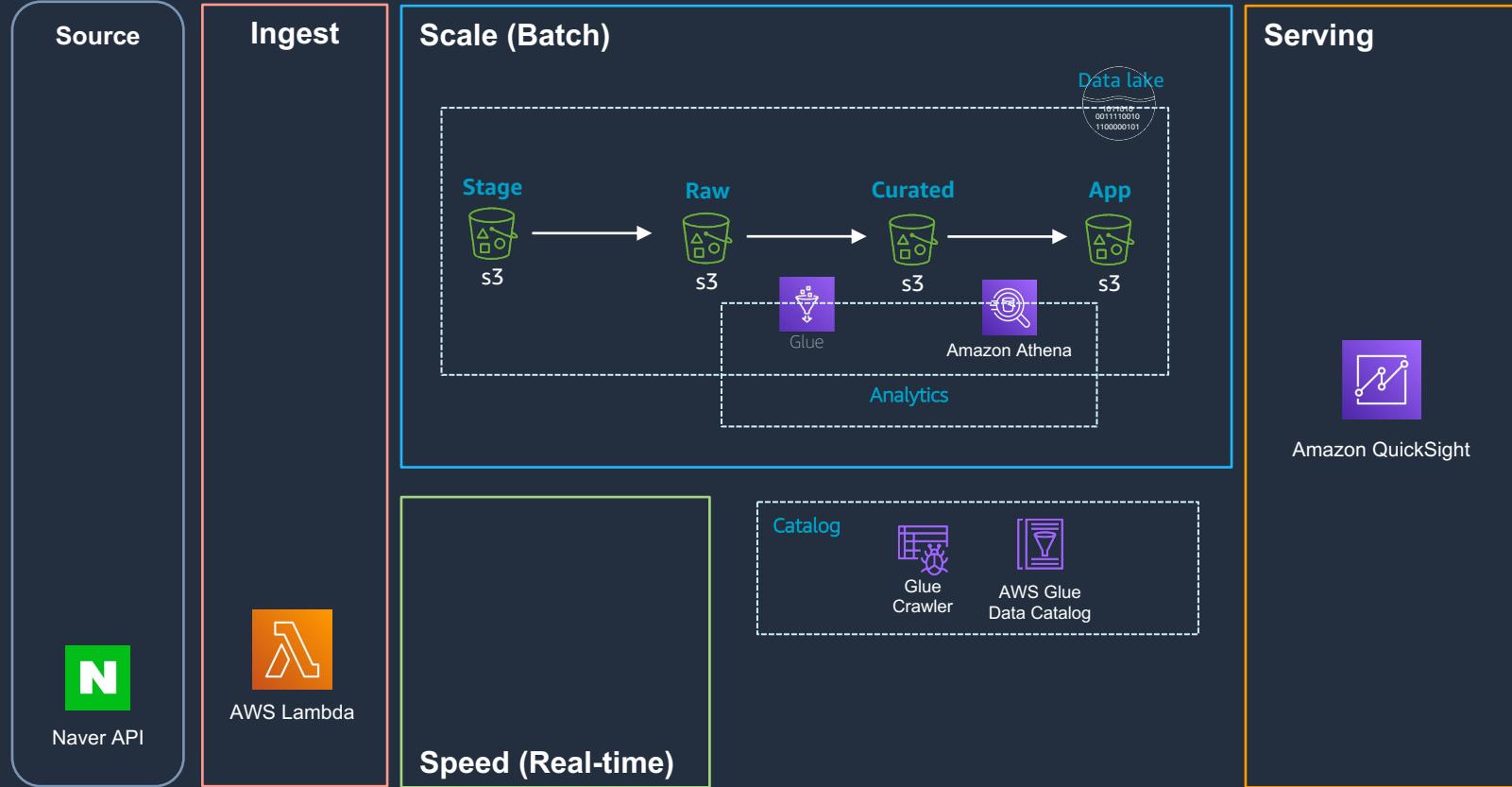
# Building Data Pipeline with multiple sources



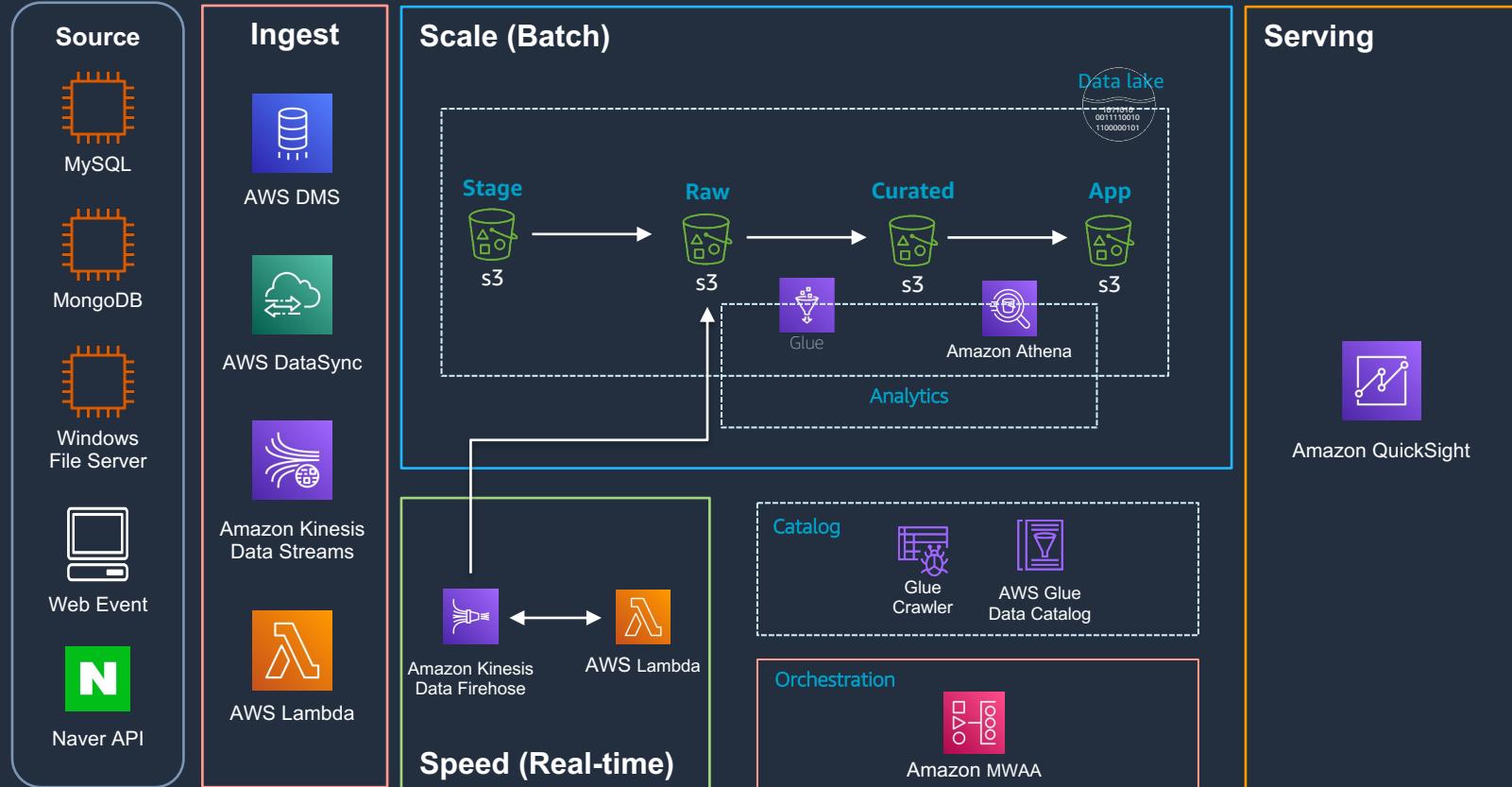
# Building Data Pipeline with multiple sources



# Building Data Pipeline with multiple sources



# Building Data Pipeline with multiple sources



# Thank you!

Solutions Architect, 성재우

© 2020, Amazon Web Services, Inc. or its Affiliates.

