aws INNOVATE
DATA EDITION

AUGUST 30,2023

Accelerate insights from real-time streaming data at scale with databases, analytics and ML

Arun Balaji
Principal Prototyping Engineer
AWS India

Gopalakrishnan Subramanian
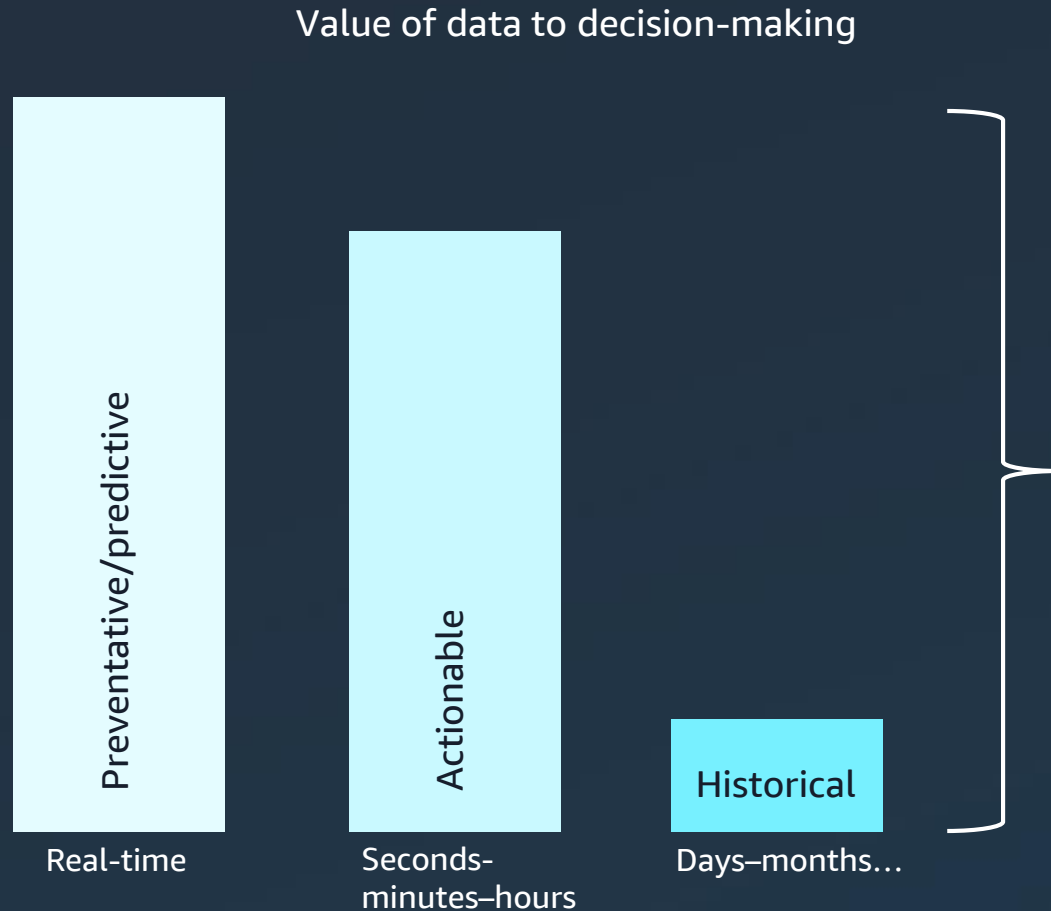Principal Database Specialist Solutions Architect
AWS India

Agenda

- Why invest in real-time data analytics

- Building real-time analytics on AWS

- Use cases and architectures

- Demo: Near-real-time fraud detection using Amazon Redshift Streaming Ingestion with Amazon Kinesis Data Streams and Amazon Redshift ML

- How to get started

Why invest in real-time data analytics ?

The value of data diminishes over time

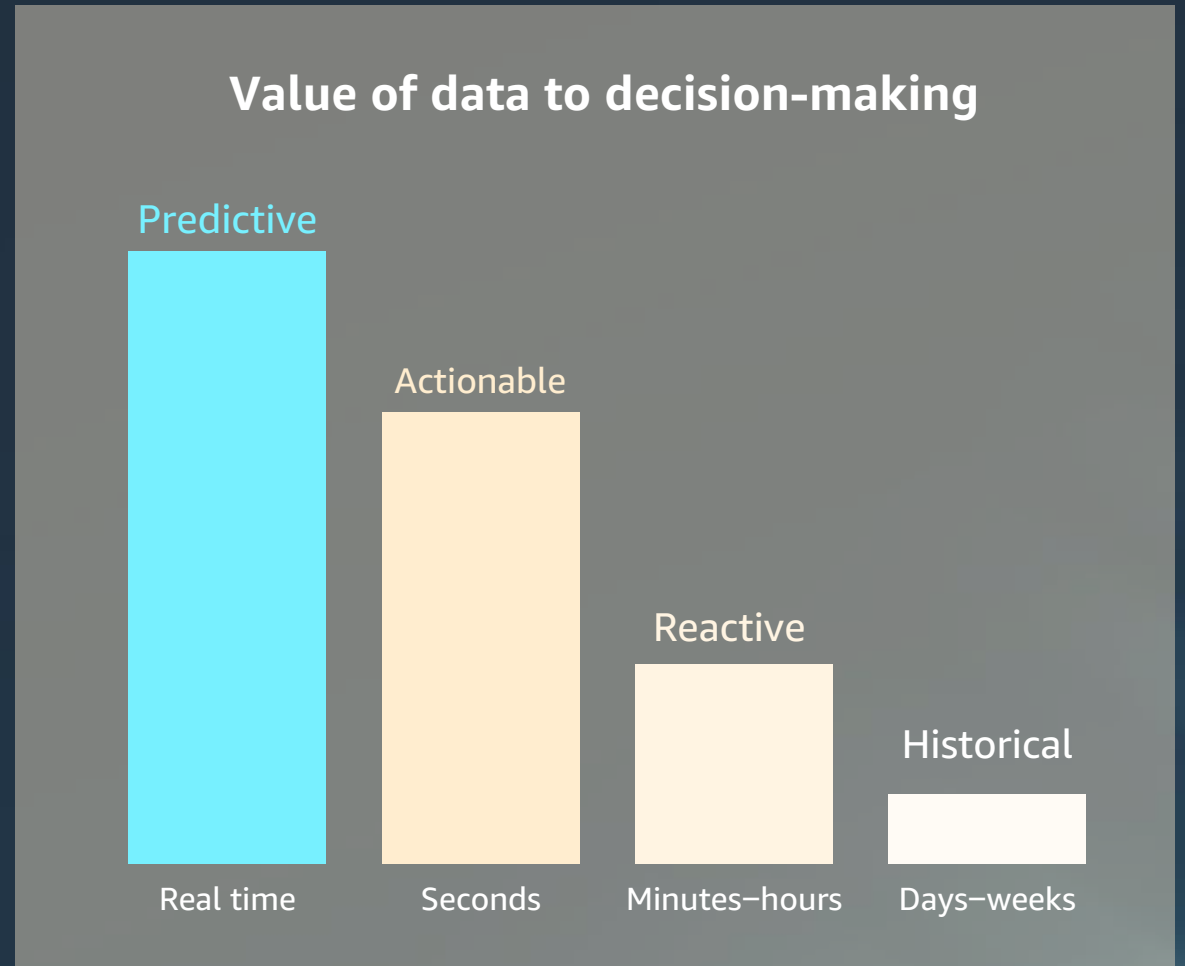Data has a short shelf life of actionability. Real-time data lets you act **as fast as the market dictates.**

Value of data to decision-making

Preventative/predictive

Real-time

Actionable

Seconds-minutes–hours

Historical

Days–months…

Majority of organizations surveyed believe untimely data inhibited business opportunities[1]

The value of the data for insights reduces drastically **within a day**[2]

1. Learn more
2. Value of data

Failing to **act in real-time** can translate to real losses

Data have a short shelf life of actionability

Value of data to decision-making

Predictive | Actionable | Reactive | Historical

Real time | Seconds | Minutes–hours | Days–weeks

Personalization	Real-time insights	Anomaly detection and fraud prevention	New use cases
Chatbots, live agents, or virtual assistants	Real-time monitoring	IoT and supply chain management	New business models
Loyalty programs	More accurate ML models	Predictive maintenance	New products
Automated customer service	Predictive analytics	Dynamic pricing	New services
Real-time notifications and feedback	Customer 360	Security and threat management	New experiences

Enhance customer experience	**Better and faster decision-making**	**Improve operations**	**Innovate and stay ahead of competitors**

Streaming analytics anatomy

Ingest, process, and analyze high volumes of
high-velocity data from a variety of sources in **real time**
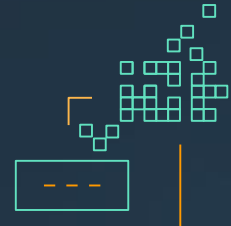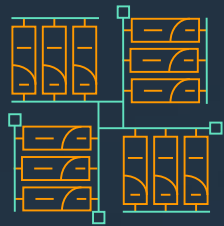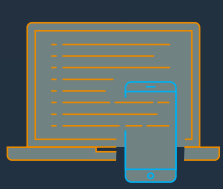
1. Source

Devices and applications that produce real-time data at high velocity

2. Stream ingestion

Data from tens of thousands of data sources can be collected and ingested in real time

3. Stream storage

Data is stored in the order it was received for a set duration of time and can be replayed indefinitely during this time
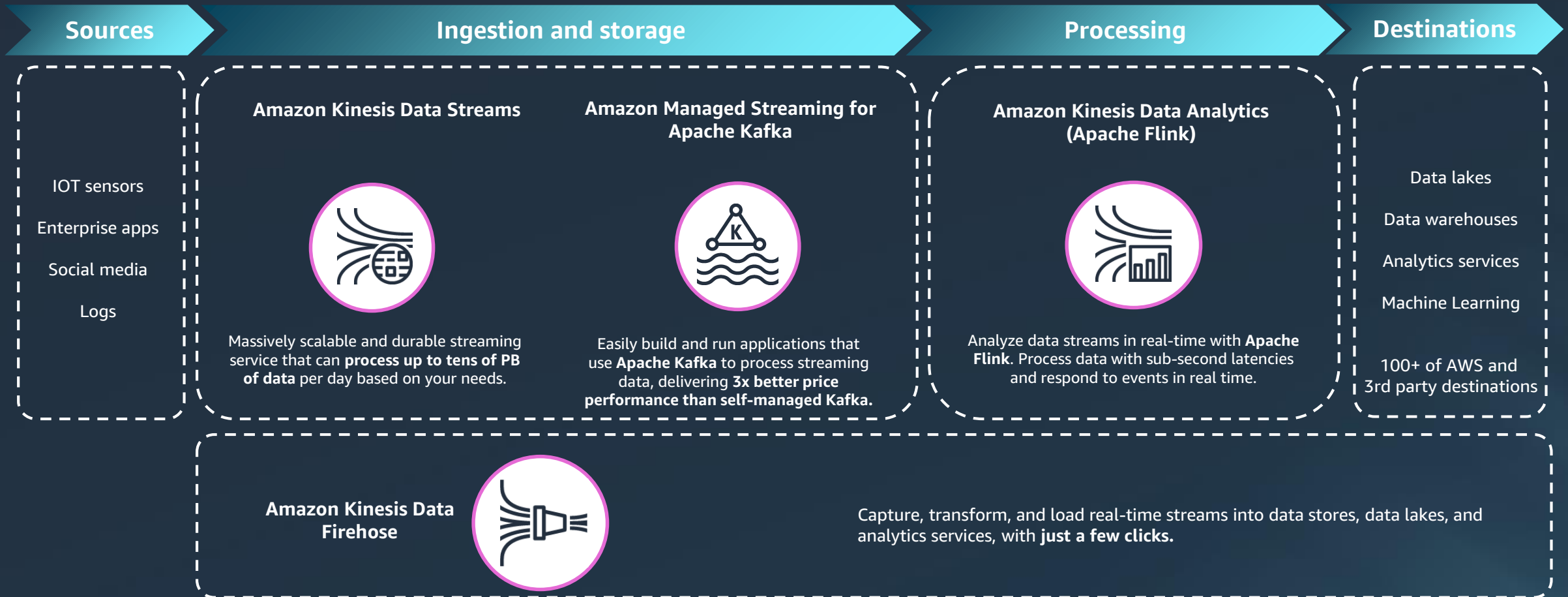
4. Stream processing

Records are read in the order they are produced, enabling real-time analytics or streaming ETL

5. Destination

Data lake/warehouse (most common)

Database, search

Building end-to-end streaming data pipelines is easier with AWS

Sources	Ingestion and storage	Processing	Destinations

Sources

IOT sensors

Enterprise apps

Social media

Logs

Ingestion and storage

Amazon Kinesis Data Streams

Massively scalable and durable streaming service that can **process up to tens of PB of data** per day based on your needs.

Amazon Managed Streaming for Apache Kafka

Easily build and run applications that use **Apache Kafka** to process streaming data, delivering **3x better price performance than self-managed Kafka.**

Processing

Amazon Kinesis Data Analytics (Apache Flink)

Analyze data streams in real-time with **Apache Flink**. Process data with sub-second latencies and respond to events in real time.

Destinations

Data lakes

Data warehouses

Analytics services

Machine Learning

100+ of AWS and 3rd party destinations

Amazon Kinesis Data Firehose

Capture, transform, and load real-time streams into data stores, data lakes, and analytics services, with **just a few clicks.**

Realize greater cost savings by running streaming data on AWS

Reduced Infrastructure Costs

✓ **Pay only for what you use**, scale up or down without having to commit to long-term investment or paying for unused resources

✓ **No upfront investment** in hardware and software

Reduced Operational Costs

✓ **Reduce development and maintenance costs** by handing off infrastructure management to AWS

✓ **Free-up IT** so your team can focus on driving innovation

Greater Efficiency and Price Performance

✓ Modernize your technology to realize greater savings

✓ Amazon MSK delivers **3x better price performance than self-managed Kafka**

Customer examples

Accelerate revenue growth with real-time analytics

Poshmark

Aa leading social marketplace for new and second hand style fashion goods with over **80M+ customers and 200M+ available listings**

Goal

Drive **top-line revenue growth** through **enhancing user experience and personalization** and **fraud prevention**

Challenges

The initial approach of using batch processing for personalization and security did not meet expectations for customer experience improvement

Solution

Designed a **real-time personalization and fraud prevention solution** using real-time event capture and real-time data enrichment leveraging AWS streaming data services, machine learning capabilities, and storage solutions
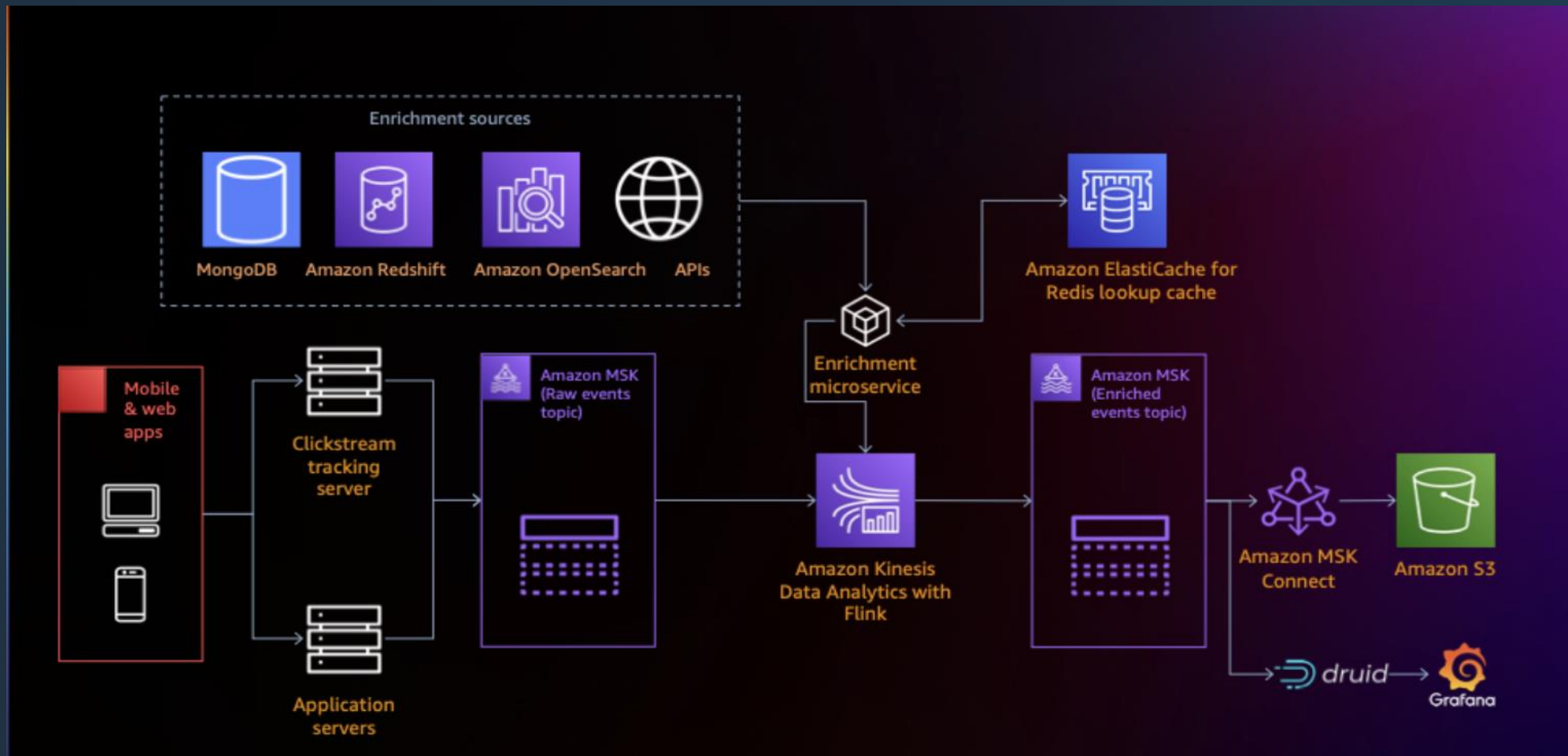
Results

✓ Drove top-line revenue growth, **increasing customer click through rate by 8% (vs batch processing)**

✓ Improved alerting of Account Take Over(ATO)/Fraud from 45% with batch based system to **80% with real-time monitoring**
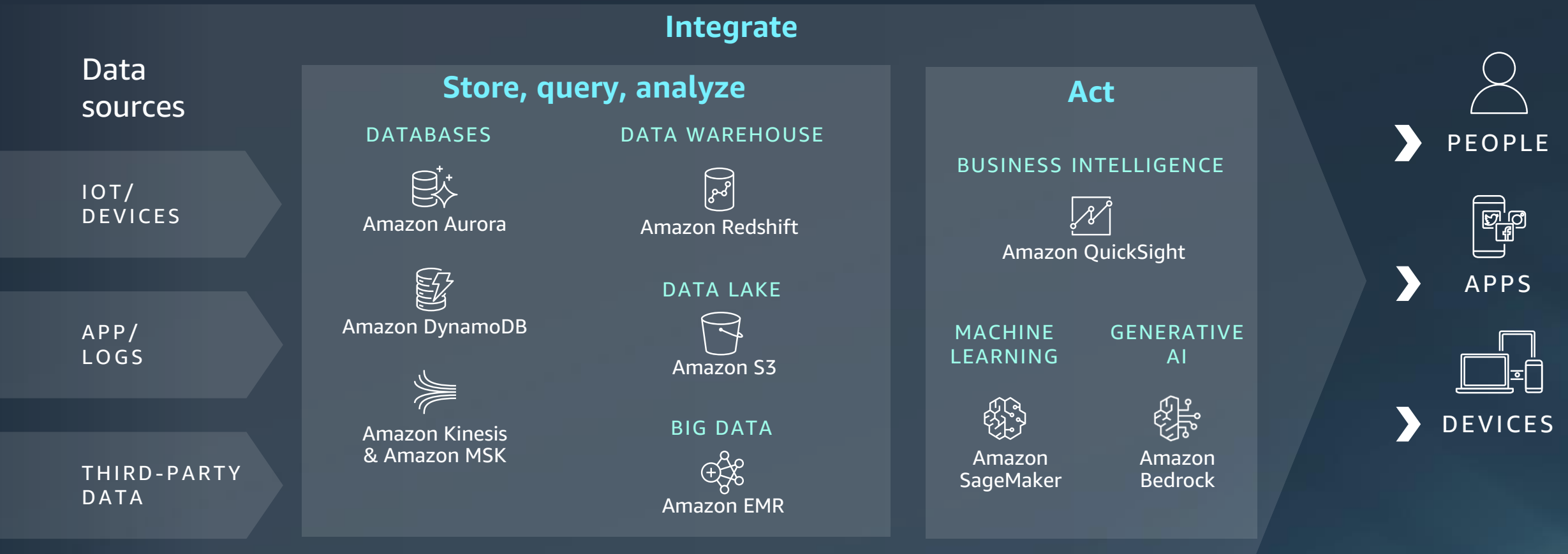
Learn more

Real-time analytics solution architecture

Streaming analytics with AI/ML

End-to-end data foundation

Integrate

Data sources

IOT/ DEVICES

APP/ LOGS

THIRD-PARTY DATA

Store, query, analyze

DATABASES
Amazon Aurora

Amazon DynamoDB

Amazon Kinesis & Amazon MSK

DATA WAREHOUSE
Amazon Redshift

DATA LAKE
Amazon S3

BIG DATA
Amazon EMR

Act

BUSINESS INTELLIGENCE
Amazon QuickSight

MACHINE LEARNING
Amazon SageMaker

GENERATIVE AI
Amazon Bedrock

PEOPLE

APPS

DEVICES

Catalog and govern | AWS Lake Formation and Amazon DataZone

The AWS AI/ML Stack

AI SERVICES

Generative AI

SCALING GENERATIVE AI
Amazon Bedrock `ANNOUNCED`

AMAZON BUILT FOUNDATIONAL MODEL
Amazon Titan `ANNOUNCED`

Specialized AI Services

BUSINESS PROCESSES	**SEARCH**	**CONVERSATION**	**CODE + DEVOPS**	**INDUSTRIAL**	**HEALTH**
Amazon Personalize	Amazon Kendra	Amazon Lex	Amazon CodeGuru	Amazon Monitron	Amazon HealthLake
Amazon Forecast		Contact Lens	Amazon CodeWhisperer	Amazon Lookout for Equipment	Amazon Comprehend Medical
Amazon Fraud Detector		Voice ID	Amazon DevOps Guru	Amazon Lookout for Vision	Amazon Transcribe Medical
Amazon Lookout for Metrics		Amazon Transcribe Call Analytics			Amazon Omics

Core

TEXT		**SPEECH**		**VISION**		
Amazon Translate	Amazon Comprehend	Amazon Polly	Amazon Transcribe	Amazon Textract Panorama	Amazon Rekognition	AWS

AMAZON SAGEMAKER

JUMPSTART `NEW`
Use pre-built models in SageMaker

CANVAS
No-code ML for business analysts

STUDIO LAB
Learn ML

GROUND TRUTH
Label data

STUDIO IDE

Prepare data	Geospatial ML	Build with notebooks	Train models	Deploy in production
Store features			Tune parameters	Manage and monitor

---------------------- **CI/CD | GOVERNANCE | RESPONSIBLE ML** ----------------------

ML FRAMEWORKS & INFRASTRUCTURE

PyTorch, TensorFlow   Amazon EC2   CPUs   GPUs   `NEW` AWS Inferentia   `NEW` AWS Trainium   Habana Gaudi   FPGA

AWS brings ML closer to data

Databases + Data warehouses + data lakes + Business intelligence tools

AMAZON
AURORA ML

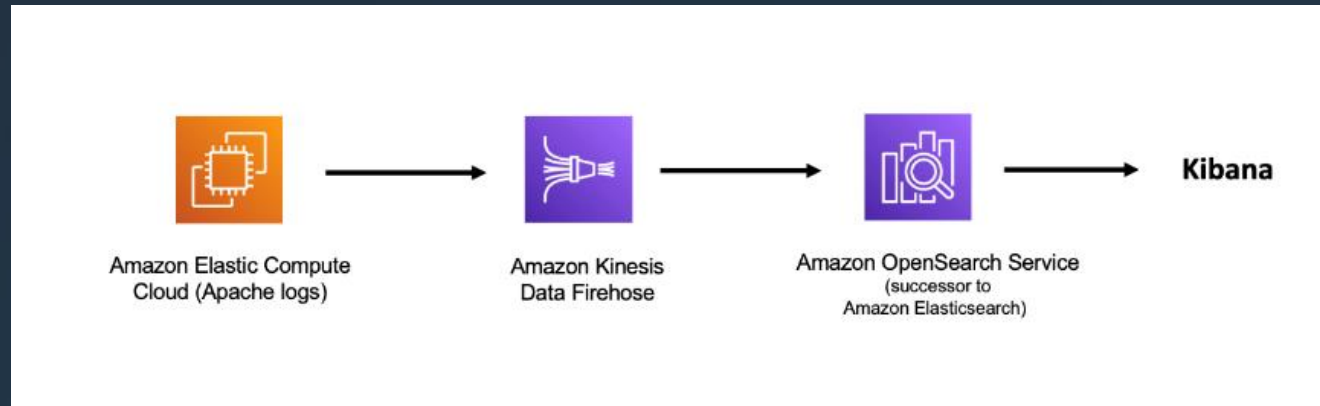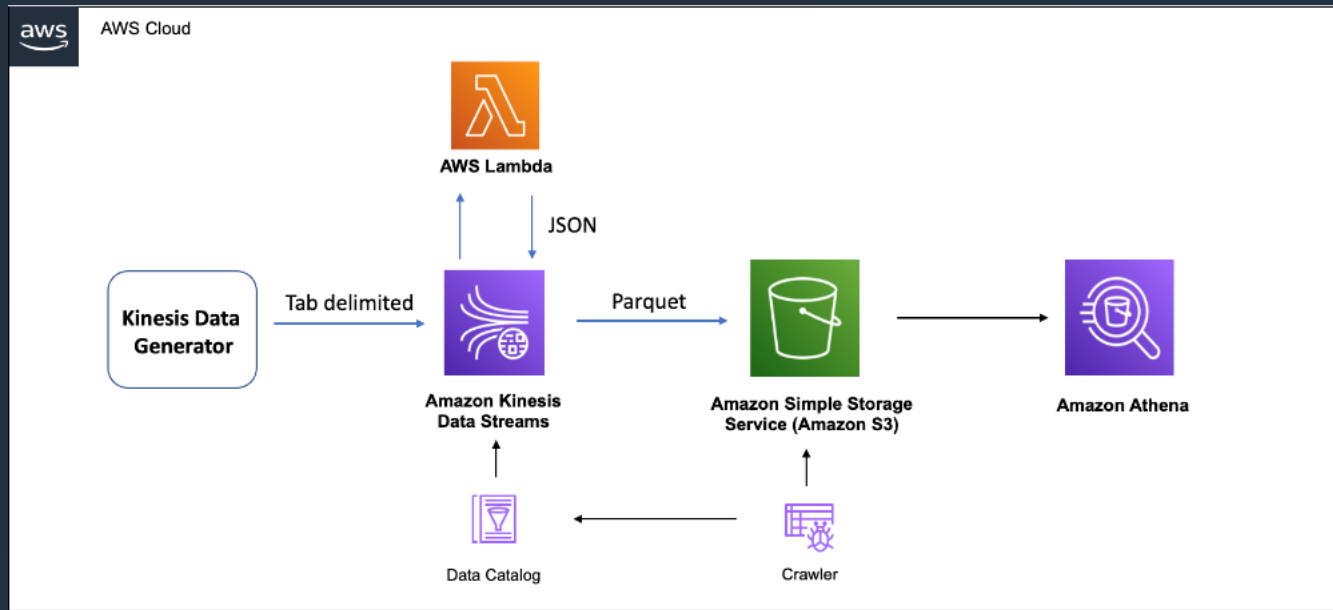AMAZON
NEPTUNE ML

AMAZON
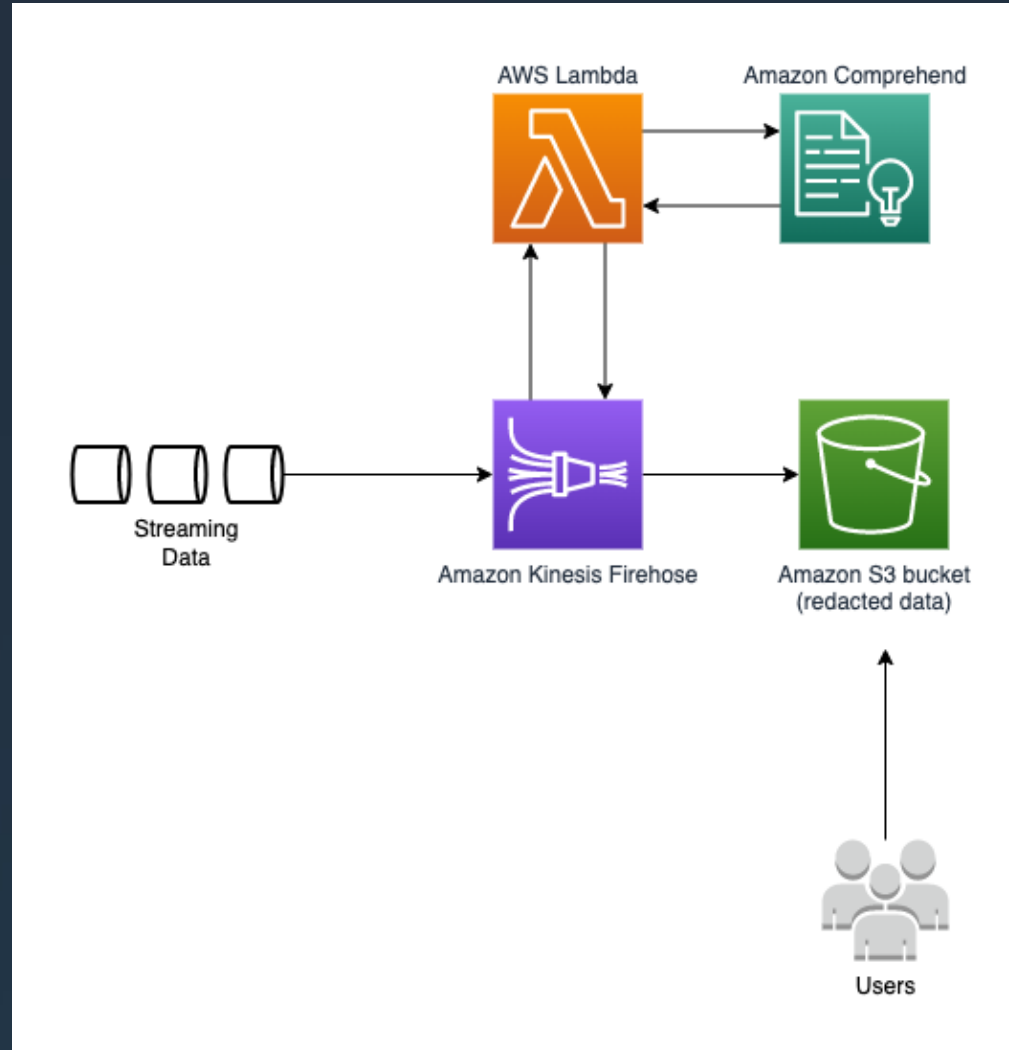REDSHIFT ML

AMAZON
ATHENA ML

AMAZON
QUICKSIGHT ML

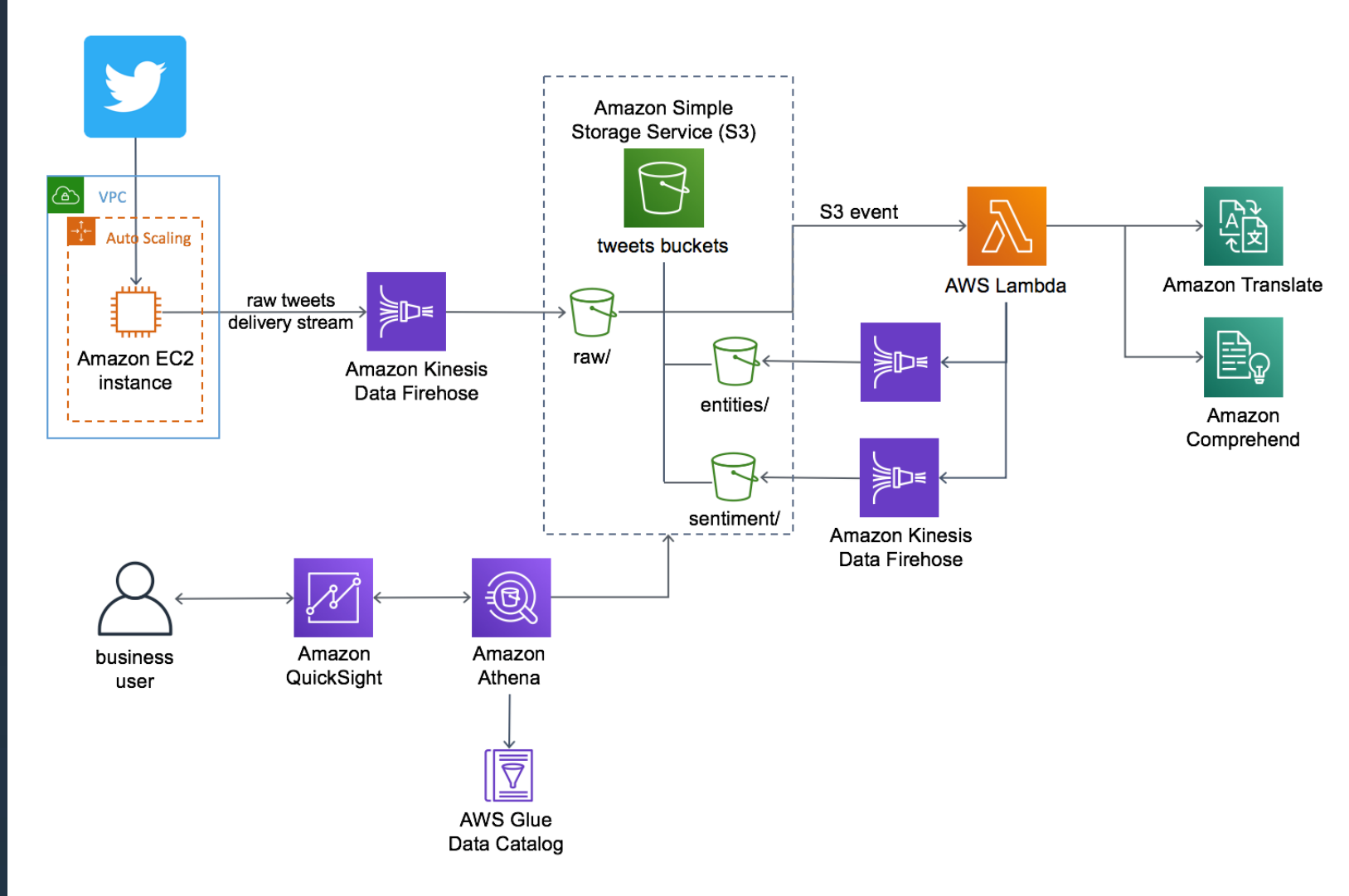# Bringing it together: Use cases and architectures

Use case 1: Simple use case of real time log analytics with Amazon Kinesis Firehose
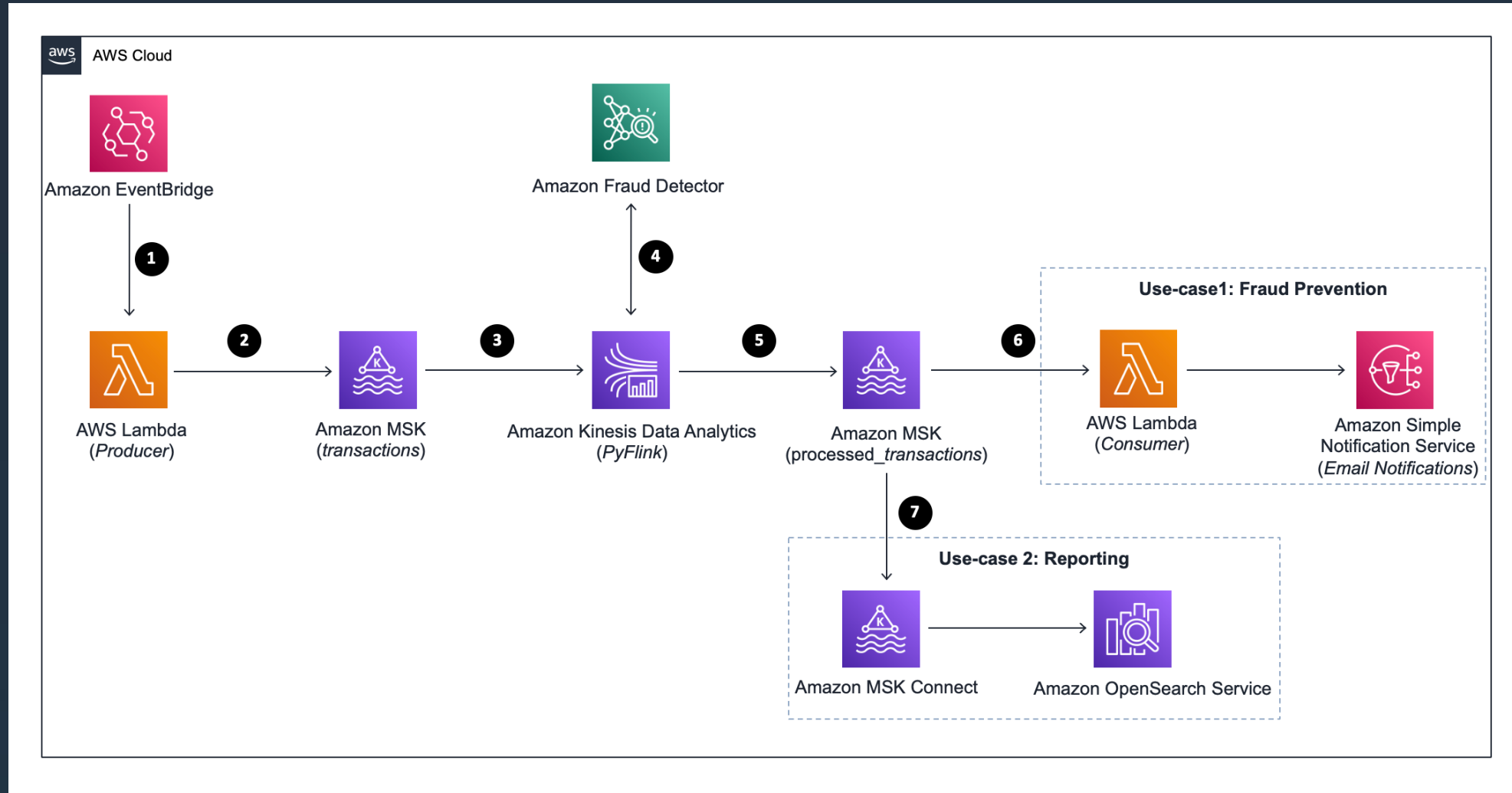
Use case 2: Redact sensitive data from streaming data in near-real time using Amazon Comprehend and Amazon Kinesis Data Firehose

Use case 3: AI-driven social media dashboard

Use case 4: Build and visualize a real-time fraud prevention system using Amazon Fraud Detector

Score user activity in real-time with Amazon Fraud Detector

Identify potential fraudulent activity with a fully managed service that uses ml models and fraud detection expertise from AWS and amazon.com

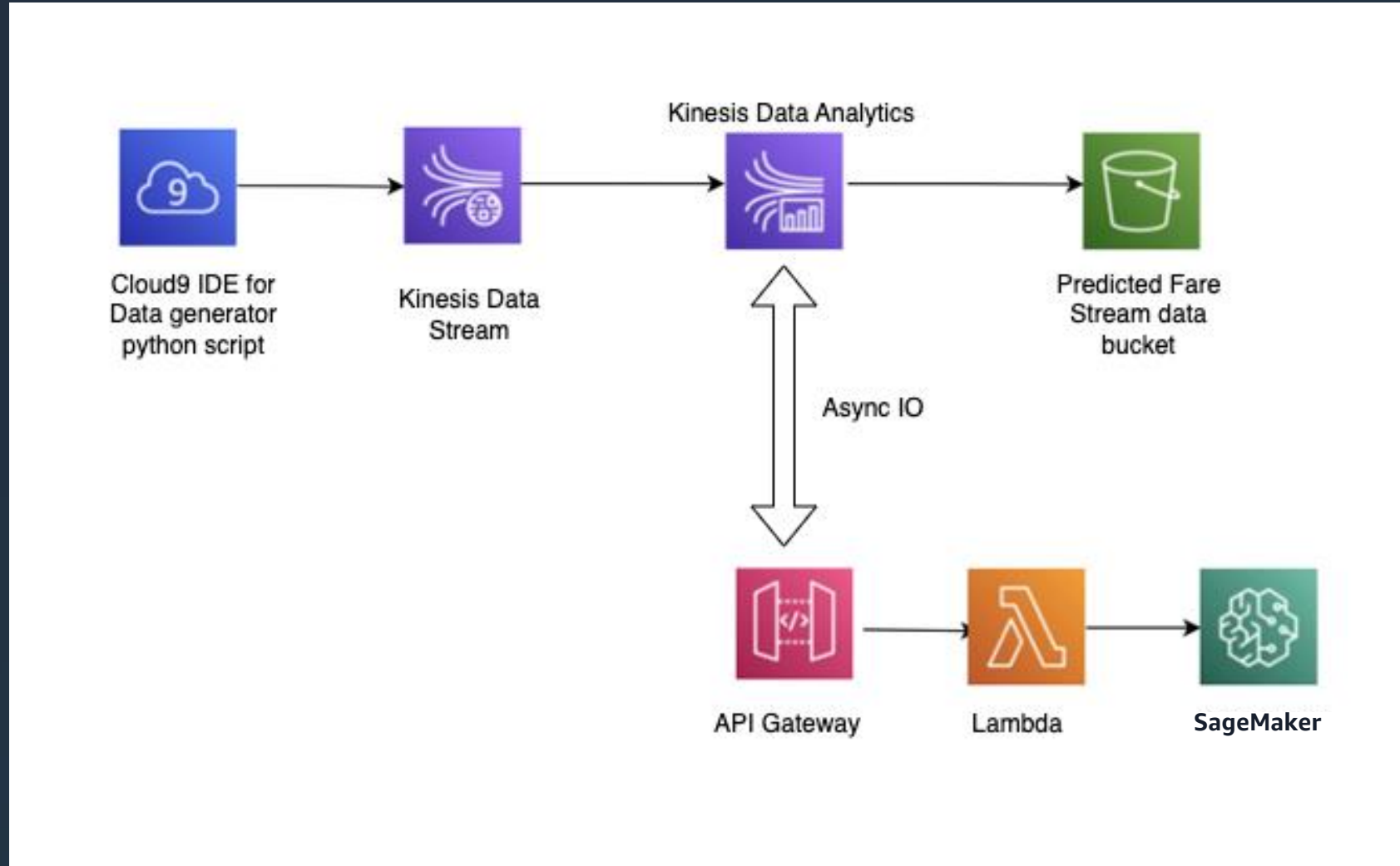User performs an online activity → App or website captures details about the activity, session, and user → Details are sent to Amazon Fraud Detector activity fraud model API → Outcomes are sent back in real-time to allow automation of decisions

Use case 5: Real-time analytics and ML inference – Predict fares

Challenges with building time-series solutions

RELATIONAL DATABASES
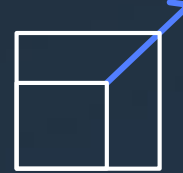
Inefficient at processing time-series data

Data management issues with rigid schema

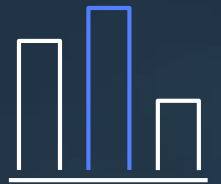Limited integrations for ML, analytics, and data collection

EXISTING TIME-SERIES SOLUTIONS

Difficult to scale for large volumes of data

Minimal data lifecycle management

Real-time and historical data are decoupled

Addressing time-series data challenges with Amazon Timestream

Fast, scalable, and serverless time-series database

Serverless and easy to use

No servers to manage or instances to provision; software patches, indexes, and database optimizations are handled automatically

Performance at scale

Capable of ingesting trillions of events daily; the adaptive SQL query engine provides rapid point-in-time queries with its in-memory store, and fast analytical queries through its magnetic store
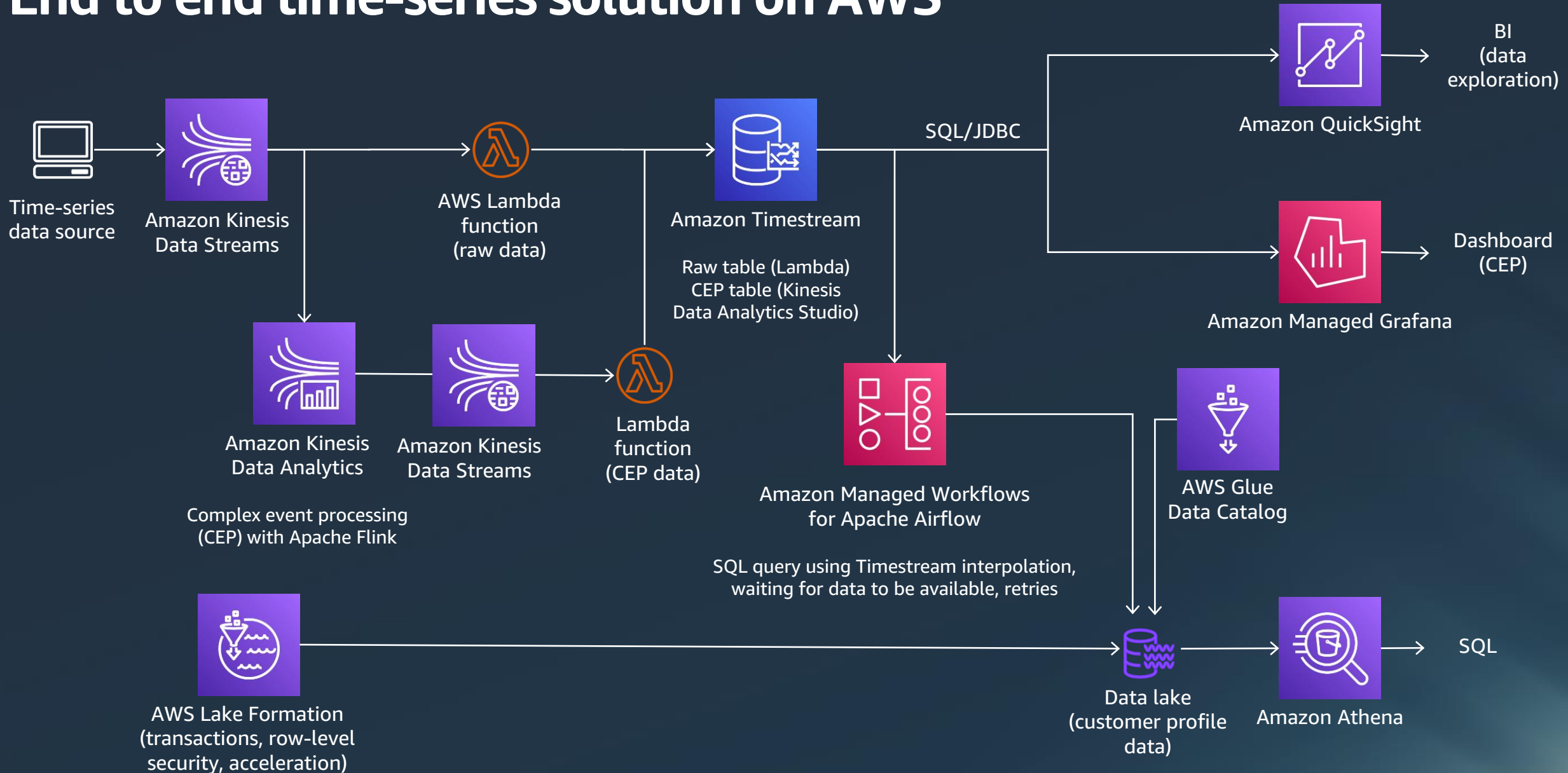
Purpose built for time-series data

Built-in analytics using standard SQL with added interpolation and smoothing functions to identify trends, patterns, and anomalies
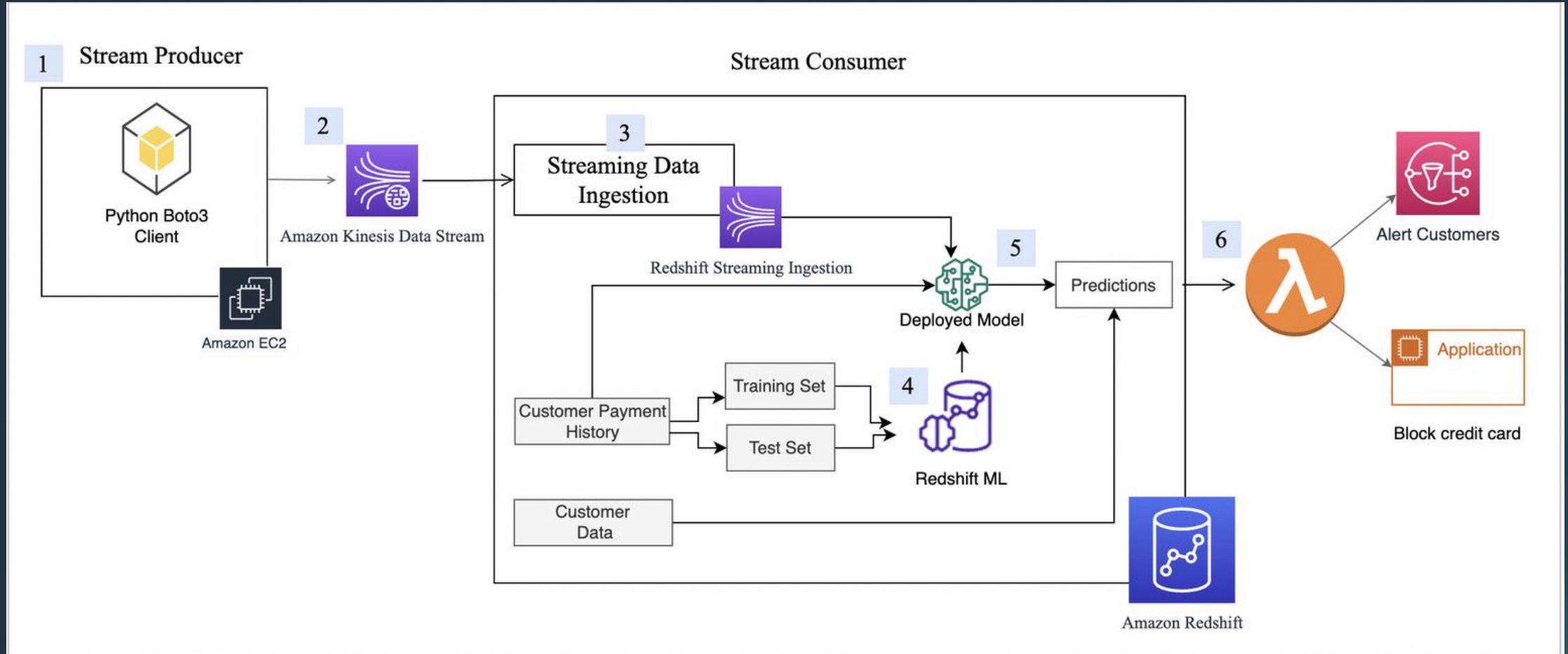
Secure from the ground up

All data is encrypted in flight and at rest using AWS Key Management Service (AWS KMS) with customer-managed keys (CMKs)

End to end time-series solution on AWS

Time-series data source → Amazon Kinesis Data Streams → AWS Lambda function (raw data) → Amazon Timestream

Amazon Timestream
Raw table (Lambda)
CEP table (Kinesis Data Analytics Studio)

Amazon Kinesis Data Analytics → Amazon Kinesis Data Streams → Lambda function (CEP data)

Complex event processing (CEP) with Apache Flink

SQL/JDBC → Amazon QuickSight → BI (data exploration)

Amazon Managed Grafana → Dashboard (CEP)

Amazon Managed Workflows for Apache Airflow

SQL query using Timestream interpolation, waiting for data to be available, retries

AWS Glue Data Catalog

AWS Lake Formation (transactions, row-level security, acceleration)

Data lake (customer profile data) → Amazon Athena → SQL

Demo : Near-real-time fraud detection using Amazon Redshift Streaming Ingestion with Amazon Kinesis Data Streams and Amazon Redshift ML

Demo

Recap

- Why is real time analytics important

- What is streaming ingestion with AWS

- How to integrate AWS AI/ML with streaming data

- Where to apply: Use cases and architectures

- Demo: How to build near real time fraud detection

Getting started

Try it yourself

- Near real time fraud detection - [here](#)

- Streaming ingestion to feature store - [here](#)

- Real time ML inference workshop - [here](#)

- Customer story - [here](#)

- AWS solution example - [here](#)

Visit the AWS Data Resource Hub

A modern data strategy can help you manage, act on, and react to your data so you can make better decisions, respond faster, and uncover new opportunities. Dive deeper with these resources today.

- Maximizing business value with data: 6 essential data-driven use cases for leaders
- The ultimate guide to developing an end-to-end data strategy
- 5 key elements of a data-driven culture
- A builder's guide to AWS analytics services
- … and more!

https://tinyurl.com/data-hub-aws

Visit resource hub

AWS Training and Certification

Access 600+ free digital courses with AWS Skill Builder

Focus on the cloud skills and services that are most relevant to you across 30+ AWS solutions, including digital self-paced learning plans and ramp-up guides

- Build your future in the AWS Cloud at your own pace
- Advance your skills and knowledge with learning plans
- Validate your cloud expertise with AWS Certification

LEARN YOUR WAY SKILLBUILDER.AWS

Thank you for attending AWS Innovate – Data Edition

We hope you found it interesting! A kind reminder to **complete the survey.**
Let us know what you thought of today's event and how we can improve the event experience for you in the future.

aws-apj-marketing@amazon.com

twitter.com/AWSCloud

facebook.com/AmazonWebServices

youtube.com/user/AmazonWebServices

slideshare.net/AmazonWebServices

twitch.tv/aws

Thank you!