



# aws INNOVATE

DATA EDITION

AUGUST 19, 2021

# **Deliver better customer experience with machine learning in real-time**

Aneesh Chandra PN

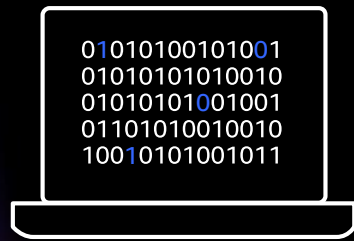
Specialist Solutions Architect, Data & Analytics

Amazon Web Services

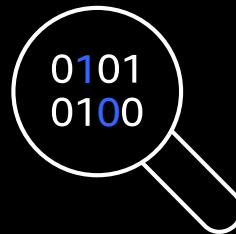


# The opportunity

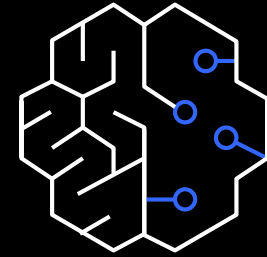
Get more value from your data



DATA



ANALYTICS



MACHINE  
LEARNING

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

Insights from data are perishable & can lose value quickly



Acting on real-time data can help increase customer retention and customer loyalty



Stream processing allows analytical insights to be gathered and acted upon instantly

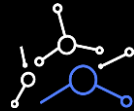
# Common real-time analytics & ML use cases



Anomaly and fraud detection



Tailoring customer experience in real-time



Empowering IoT Analytics



Nourishing Marketing campaigns

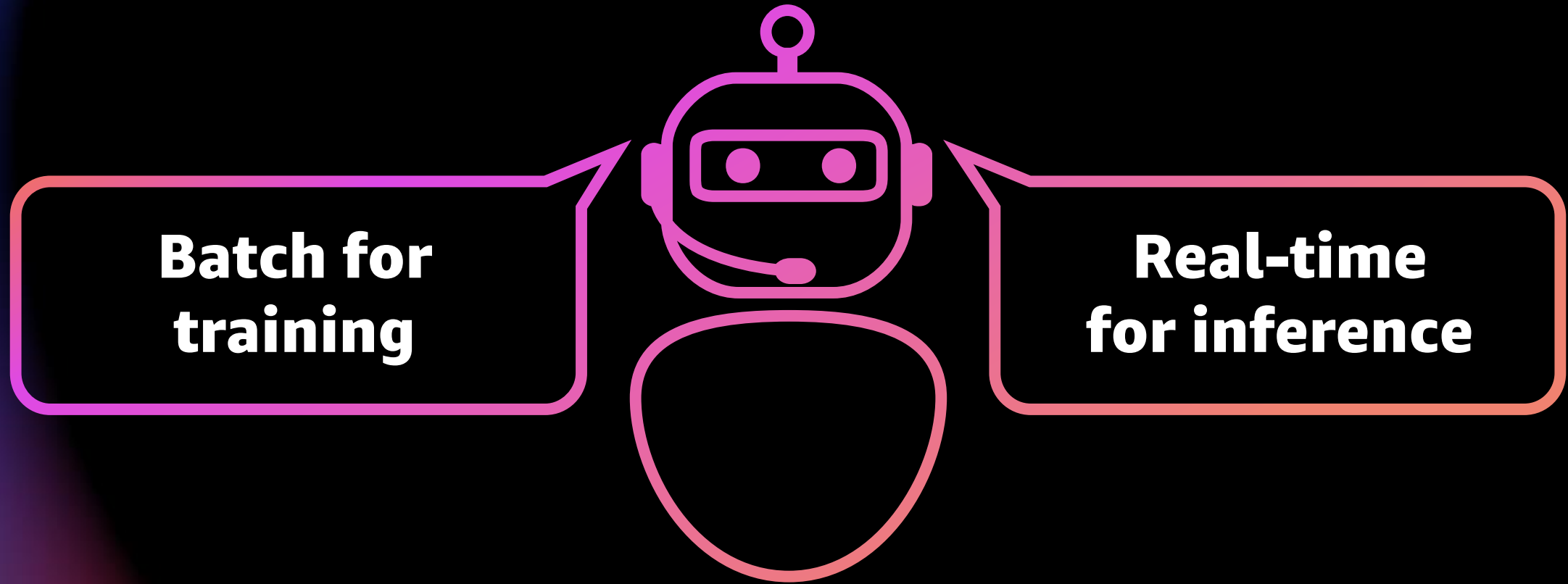


Real-time personalization



Supporting healthcare and emergency services

# Features are the foundation of **high-quality models**



# Challenges with real-time analytics & ML



Difficult to setup & manage



Tricky to scale



Feature drift, duplication



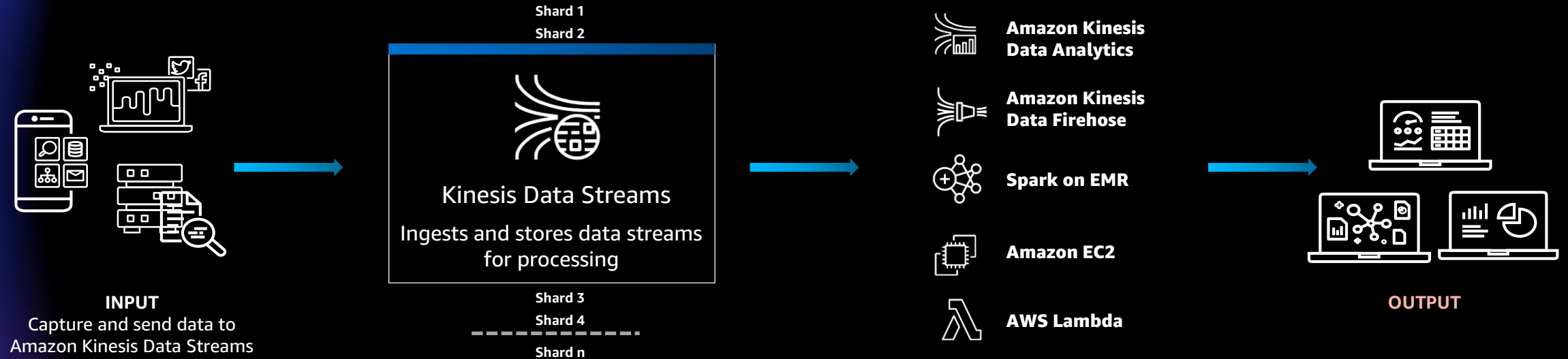
Slow model development/deployment

# Building real-time ML on AWS



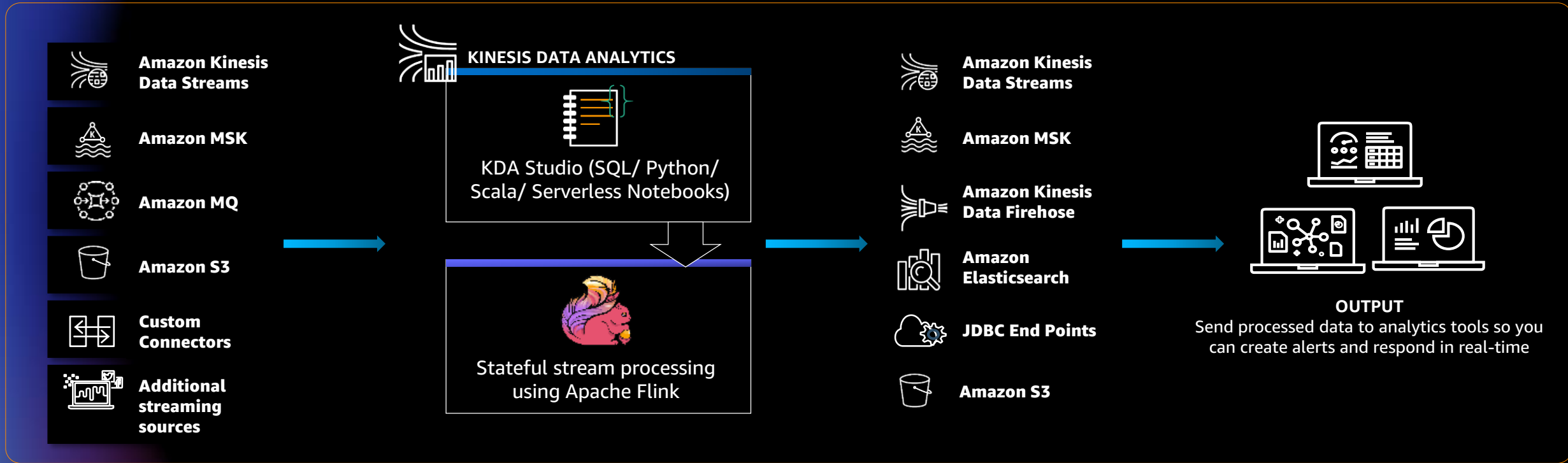


# Amazon Kinesis Data Streams



- Easy administration and low cost
- Real-time, elastic performance
- Secure, durable storage
- Available to multiple real-time analytics applications
- Average latency of 200ms with one standard consumer
- Enhanced Fan Out offers typical average latency of 70 ms

# Amazon Kinesis Data Analytics



- Interact with streaming data in real time using SQL, Python, Scala and Java or integrated Apache Flink applications
- Deploy KDA studio adhoc analysis as a durable state application with in KDA for Apache Flink
- Build fully managed and elastic stream processing applications

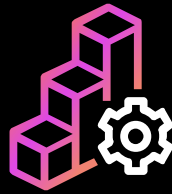
# Amazon SageMaker Feature Store: Securely store, discover, and share features for machine learning



Online  
and offline



Millisecond  
latency



Consistent  
features

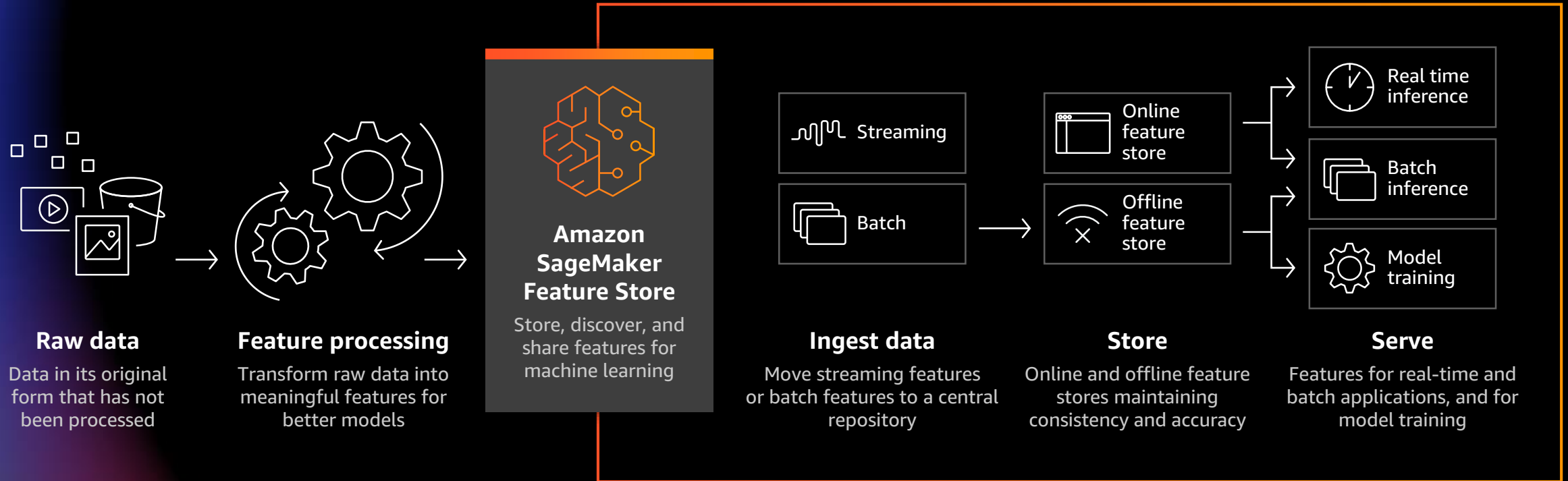


Visually  
search



Sharing and  
collaboration

# Amazon SageMaker Feature Store

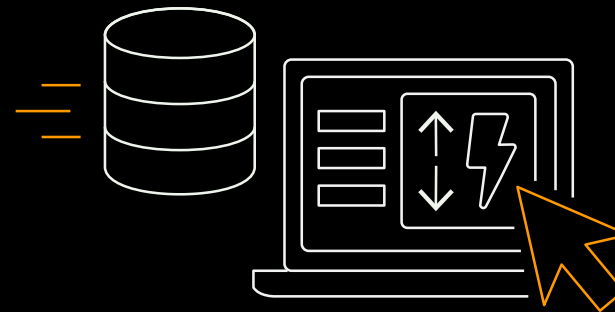


# Support for separate feature stores



## Online feature store

- Primarily used for real time predictions
- Use cases such as real-time fraud detection
- Latest copy of feature data
- High throughput writes
- Low millisecond latency reads



## Offline feature store

- Primarily used for batch predictions and model training
- Historical record of feature data
- High throughput writes
- <15 minutes read after write consistency

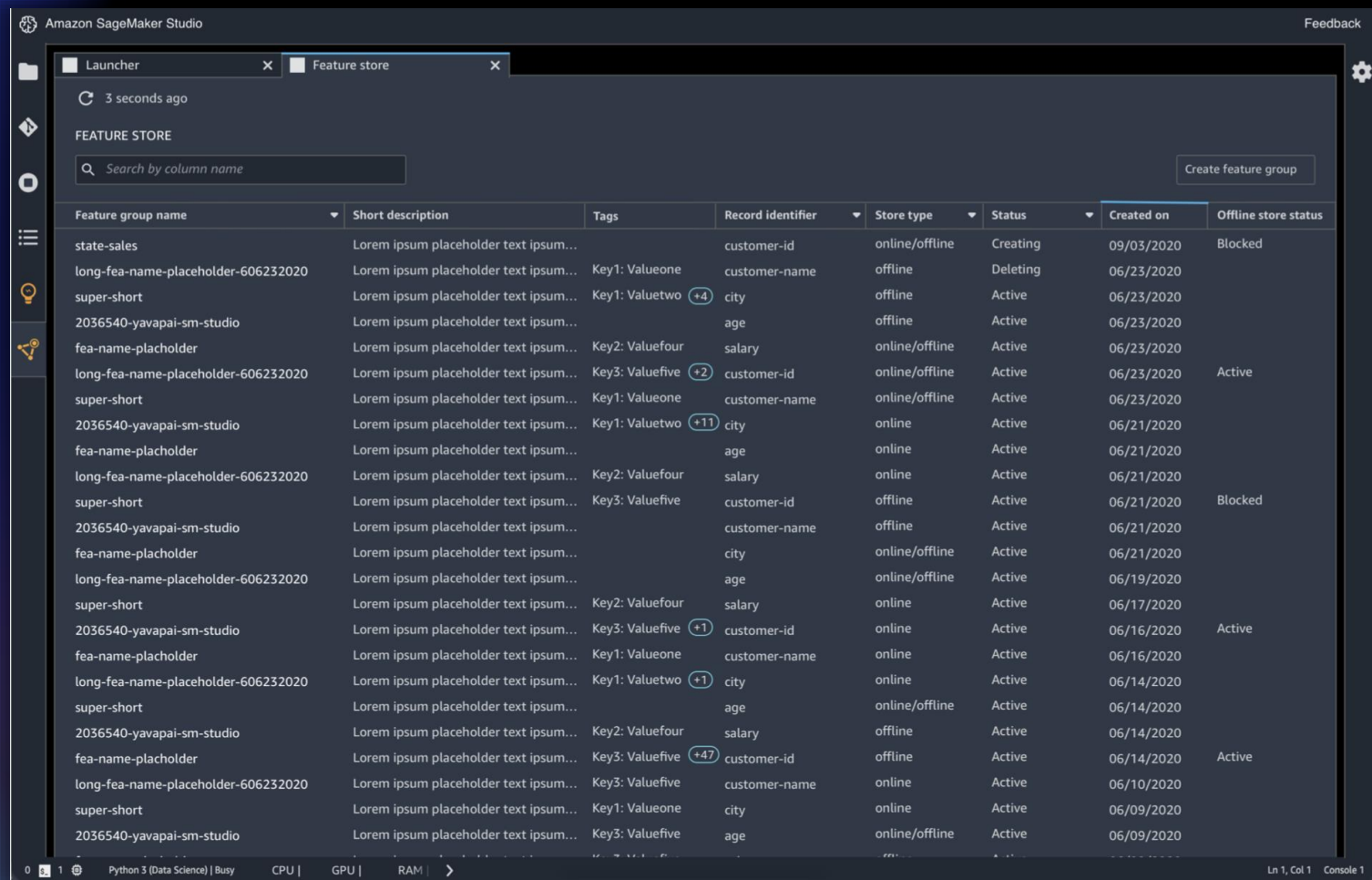
# Manage features using Feature Groups

The screenshot displays the 'Create feature group' page in the Amazon SageMaker Studio console. The left sidebar shows the 'Feature store' section. The main content area is titled 'FEATURE GROUP DETAILS' and includes the following sections:

- Feature group details:** Fields for 'Feature group name' (State Sales (USA)) and 'Description' (Report about sales for CEO).
- Storage configuration:** Options to 'Enable online storage' and 'Enable offline storage'. It includes fields for 'Online store encryption key' and 'Offline store encryption key', as well as 'Offline S3 bucket location' and 'Offline S3 bucket role ARN'.
- Data catalog:** An option to 'Auto create AWS Glue table' and fields for 'Data catalog name', 'Database name', and 'Table name'.

- Store features in collections called Feature Groups
- Configure feature groups for online and/or offline storage
- Create data catalog for Feature Groups
- Manage comprehensive metadata using Feature Group tags

# Search and discover features using Feature Store



The screenshot shows the Amazon SageMaker Studio interface with the 'Feature store' tab selected. A search bar at the top left contains the text 'Search by column name'. A 'Create feature group' button is located at the top right. Below the search bar is a table with the following columns: Feature group name, Short description, Tags, Record identifier, Store type, Status, Created on, and Offline store status. The table contains 20 rows of feature groups, each with a unique name, a placeholder description, tags, a record identifier, a store type, a status, a creation date, and an offline store status.

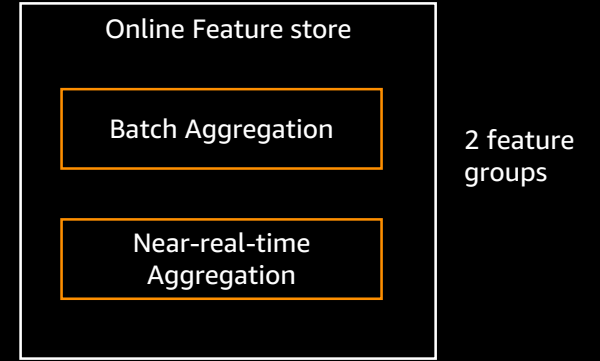
Feature group name	Short description	Tags	Record identifier	Store type	Status	Created on	Offline store status
state-sales	Lorem ipsum placeholder text ipsum...		customer-id	online/offline	Creating	09/03/2020	Blocked
long-fea-name-placeholder-606232020	Lorem ipsum placeholder text ipsum...	Key1: Valueone	customer-name	offline	Deleting	06/23/2020	
super-short	Lorem ipsum placeholder text ipsum...	Key1: Valuetwo +4	city	offline	Active	06/23/2020	
2036540-yavapai-sm-studio	Lorem ipsum placeholder text ipsum...		age	offline	Active	06/23/2020	
fea-name-placholder	Lorem ipsum placeholder text ipsum...	Key2: Valuefour	salary	online/offline	Active	06/23/2020	
long-fea-name-placeholder-606232020	Lorem ipsum placeholder text ipsum...	Key3: Valuefive +2	customer-id	online/offline	Active	06/23/2020	Active
super-short	Lorem ipsum placeholder text ipsum...	Key1: Valueone	customer-name	online/offline	Active	06/23/2020	
2036540-yavapai-sm-studio	Lorem ipsum placeholder text ipsum...	Key1: Valuetwo +11	city	online	Active	06/21/2020	
fea-name-placholder	Lorem ipsum placeholder text ipsum...		age	online	Active	06/21/2020	
long-fea-name-placeholder-606232020	Lorem ipsum placeholder text ipsum...	Key2: Valuefour	salary	online	Active	06/21/2020	
super-short	Lorem ipsum placeholder text ipsum...	Key3: Valuefive	customer-id	offline	Active	06/21/2020	Blocked
2036540-yavapai-sm-studio	Lorem ipsum placeholder text ipsum...		customer-name	offline	Active	06/21/2020	
fea-name-placholder	Lorem ipsum placeholder text ipsum...		city	online/offline	Active	06/21/2020	
long-fea-name-placeholder-606232020	Lorem ipsum placeholder text ipsum...		age	online/offline	Active	06/19/2020	
super-short	Lorem ipsum placeholder text ipsum...	Key2: Valuefour	salary	online	Active	06/17/2020	
2036540-yavapai-sm-studio	Lorem ipsum placeholder text ipsum...	Key3: Valuefive +1	customer-id	online	Active	06/16/2020	Active
fea-name-placholder	Lorem ipsum placeholder text ipsum...	Key1: Valueone	customer-name	online	Active	06/16/2020	
long-fea-name-placeholder-606232020	Lorem ipsum placeholder text ipsum...	Key1: Valuetwo +1	city	online	Active	06/14/2020	
super-short	Lorem ipsum placeholder text ipsum...		age	online/offline	Active	06/14/2020	
2036540-yavapai-sm-studio	Lorem ipsum placeholder text ipsum...	Key2: Valuefour	salary	offline	Active	06/14/2020	
fea-name-placholder	Lorem ipsum placeholder text ipsum...	Key3: Valuefive +47	customer-id	offline	Active	06/14/2020	Active
long-fea-name-placeholder-606232020	Lorem ipsum placeholder text ipsum...	Key3: Valuefive	customer-name	online	Active	06/10/2020	
super-short	Lorem ipsum placeholder text ipsum...	Key1: Valueone	city	online	Active	06/09/2020	
2036540-yavapai-sm-studio	Lorem ipsum placeholder text ipsum...	Key3: Valuefive	age	online/offline	Active	06/09/2020	

- Search features individually or by groups visually with SageMaker Studio
- Discover features by name, description, tags, and other metadata
- Understand how features are grouped relevant to ML applications

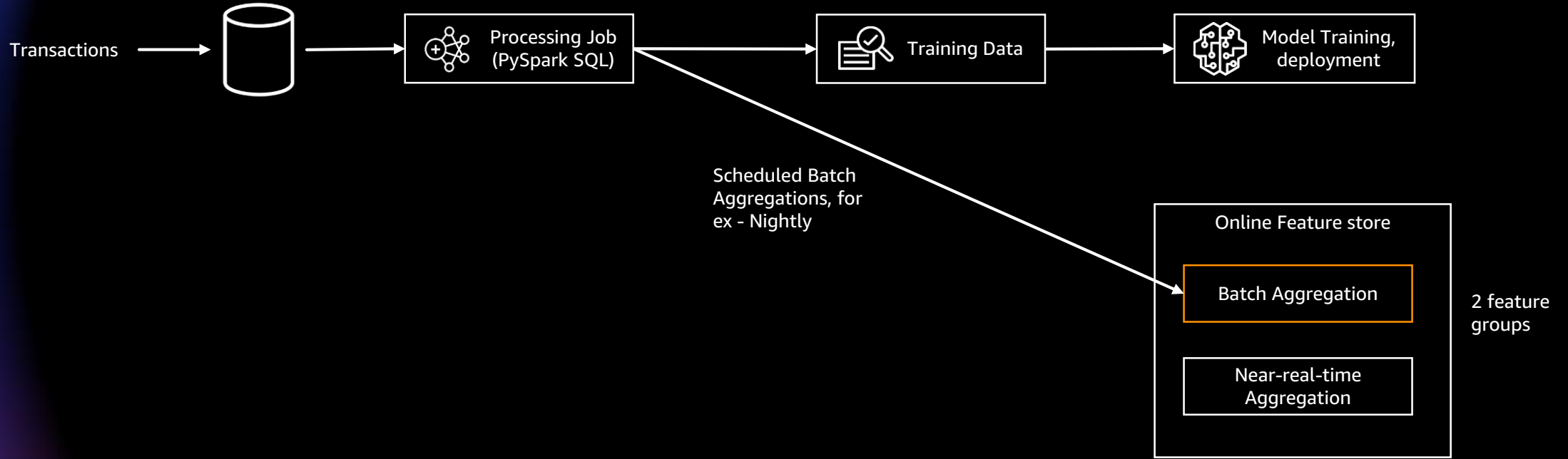
# Streaming aggregation architecture



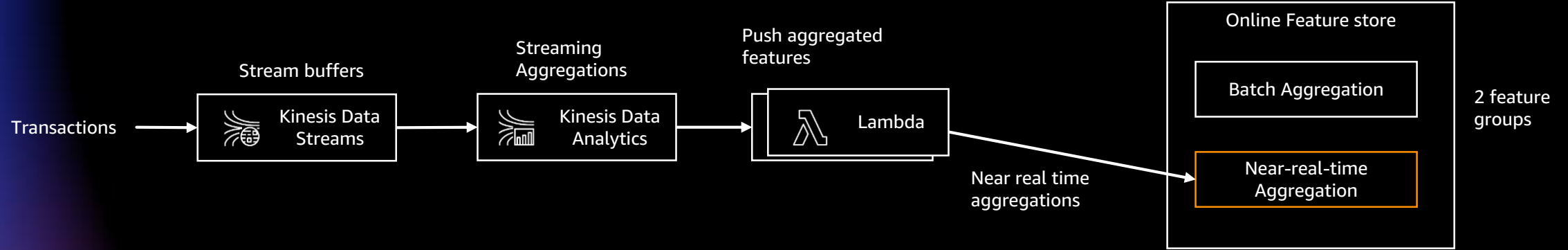
# Streaming aggregation architecture



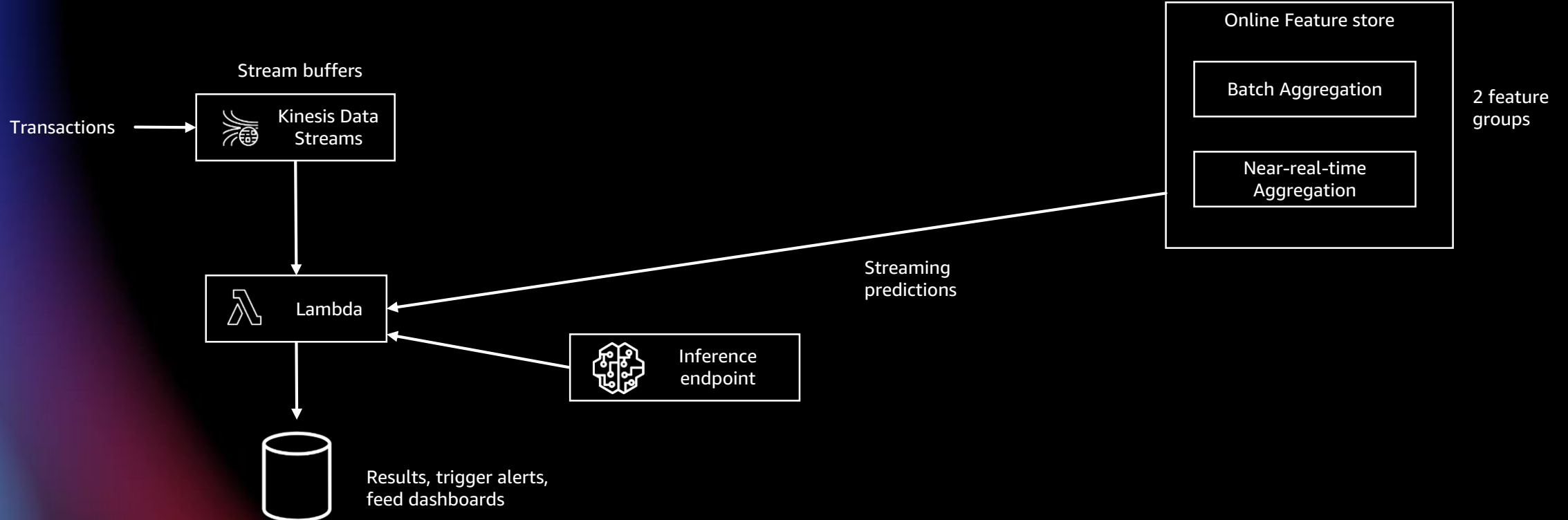
# Streaming aggregation architecture



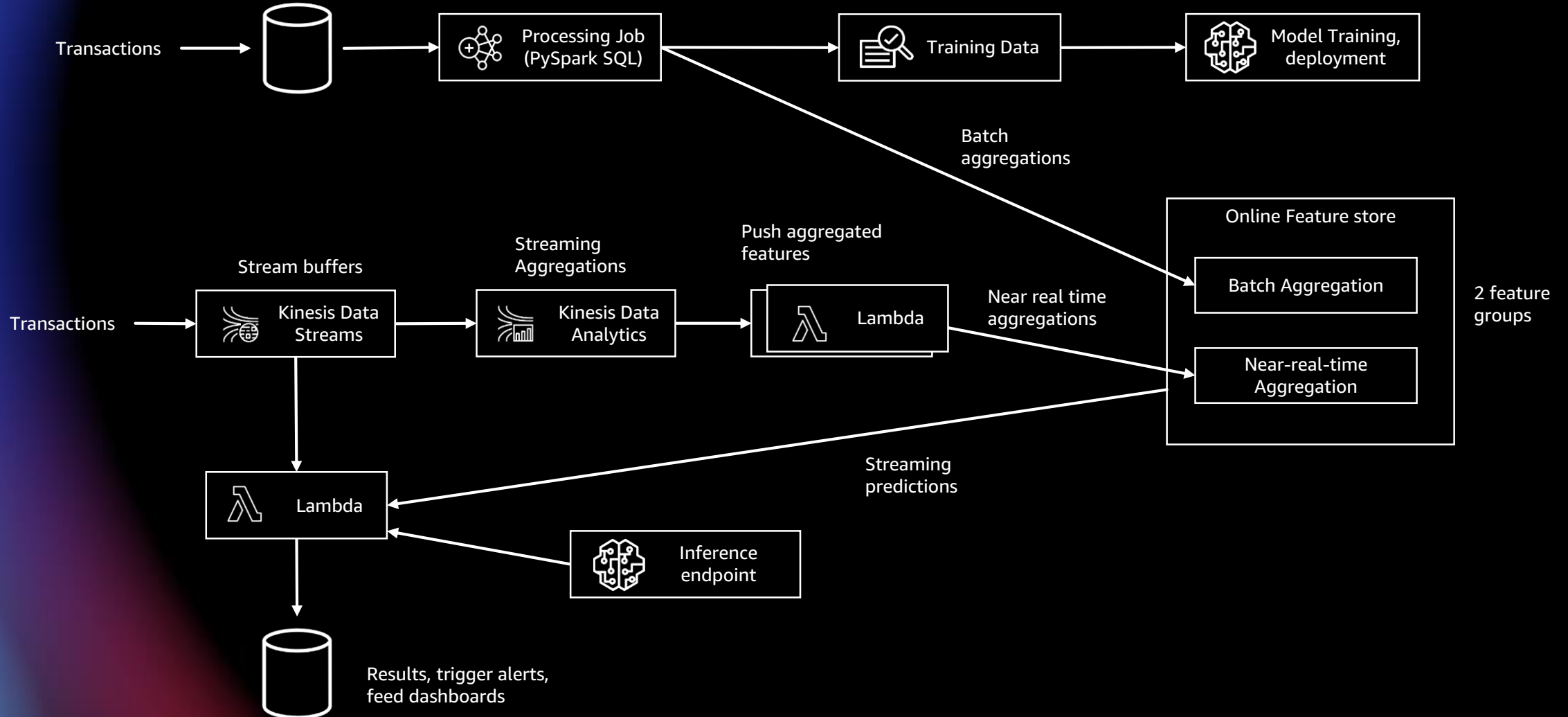
# Streaming aggregation architecture



# Streaming aggregation architecture

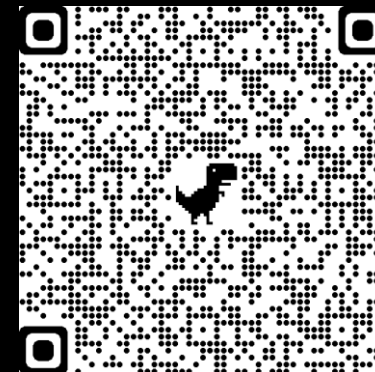


# Streaming aggregation architecture



# Demo

Credit Card Fraud detection



**Blog post:**

Using streaming ingestion with Amazon SageMaker Feature Store to make ML-backed decisions in near-real time

# Visit the AWS Data Resource Hub

Dive deeper with these resources, get inspired and learn how you can use data to make better decisions and innovate faster.

- Building a winning data strategy
- The new leadership mindset for data & analytics
- Harness data to reinvent your organization
- Put your data to work with a modern analytics approach
- Breaking free from on-premises database constraints
- Cloud storage adoption: From cost optimization to agility & innovation
- A strategic playbook for data, analytics, and machine learning
- ... and more!



<https://tinyurl.com/aws-data-resource>

Visit resource hub






# AWS Training and Certification

## Empower your teams with comprehensive training

By building skills with AWS Training and Certification, businesses and individuals can see the bigger picture understanding the reasoning behind every data point. As training progresses and teams become data-fluent, previously hidden insights come into view.



Build data skills to  
**unlock any insight**

### Leverage free digital training

Learn how to harness the world's most valuable resource: data. Access digital and virtual instructor-led courses on data analytics and databases built by the experts at AWS and start your learning journey to become data-driven.

[Take a digital course »](#)



### Get certified

Earn industry-recognized credibility and set tangible goals for success with industry-recognized certifications, like *AWS Certified Data Analytics – Specialty*.

[Learn more »](#)



### Ramp-up your skills

Deep dive into new topics and focus on knowledge gaps at your own pace with the *AWS Ramp-Up Guide: Database* and *AWS Ramp-Up Guide: Data Analytics*. With a wide range of whitepapers, blog posts, videos, webinars and peer resources available for data professionals to leverage for independent learning.

[Download ramp-up guides »](#)

# 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](mailto:aws-apj-marketing@amazon.com)



[twitter.com/AWSCloud](https://twitter.com/AWSCloud)



[facebook.com/AmazonWebServices](https://facebook.com/AmazonWebServices)



[youtube.com/user/AmazonWebServices](https://youtube.com/user/AmazonWebServices)



[slideshare.net/AmazonWebServices](https://slideshare.net/AmazonWebServices)



[twitch.tv/aws](https://twitch.tv/aws)

# Thank you!