



aws INNOVATE

DATA EDITION

23 August, 2022

Derive business insights & detect anomalies using Amazon Athena ML

Hariharan Suresh

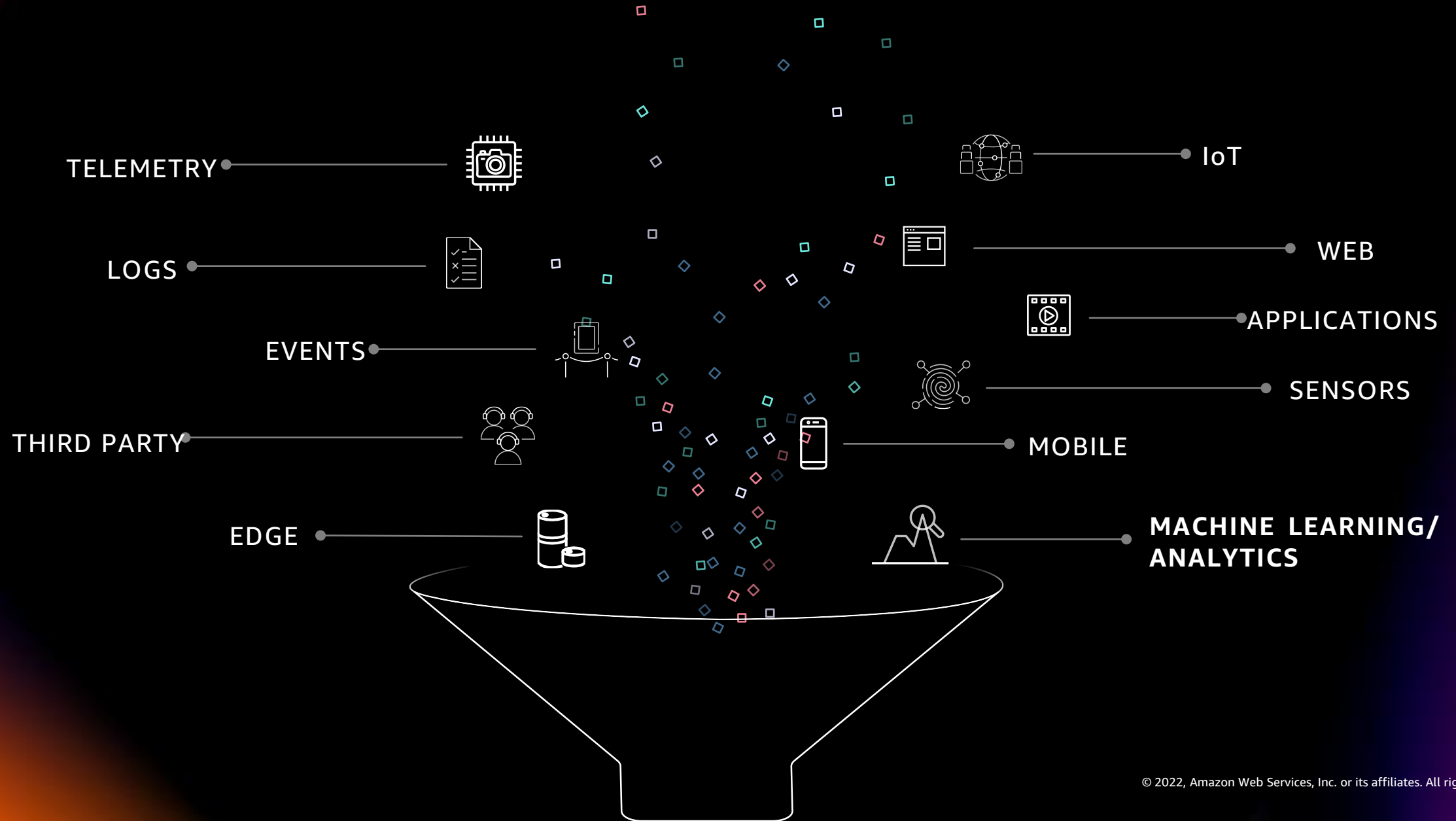
Senior Solutions Architect
Amazon Web Services



Agenda

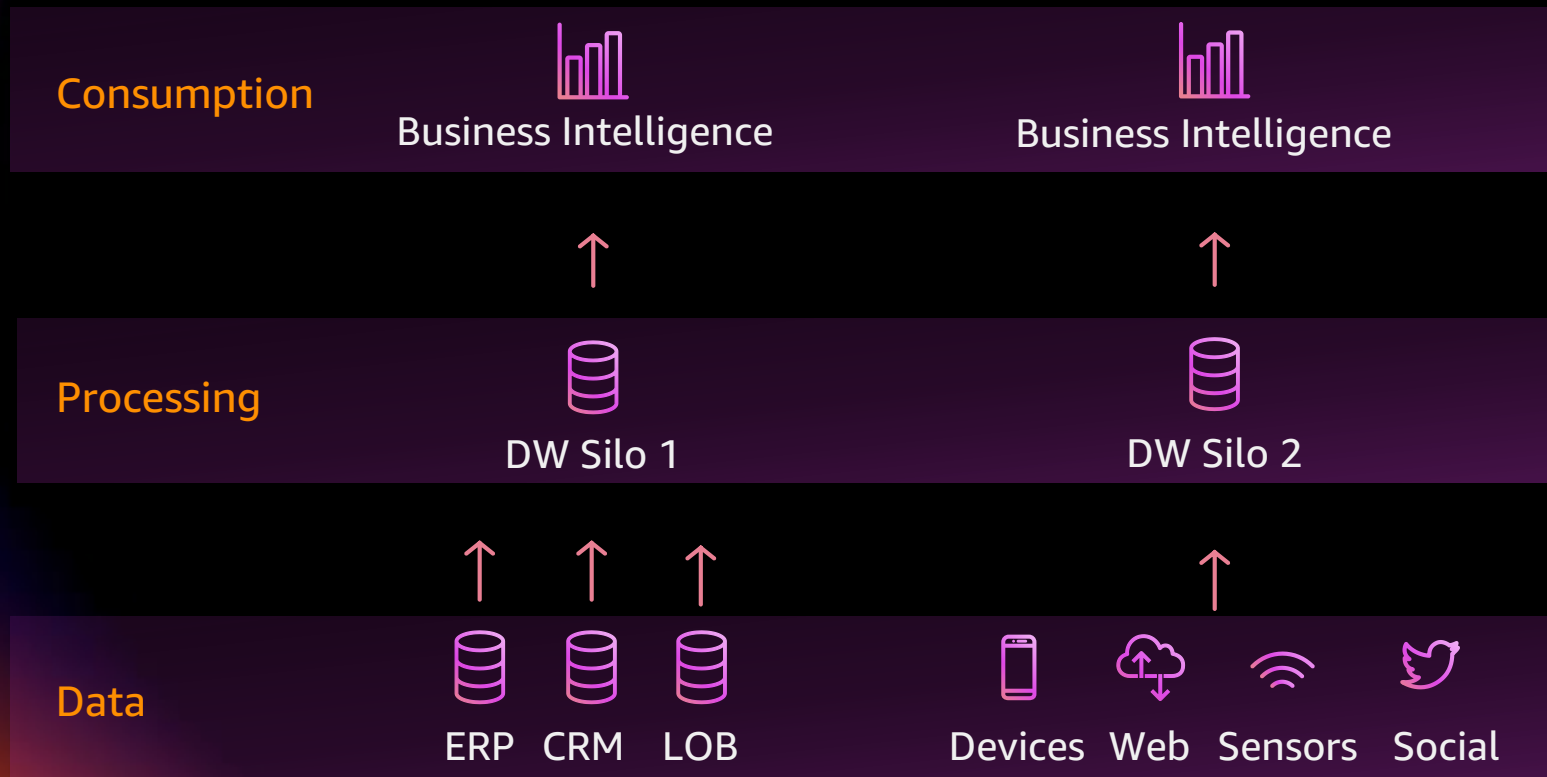
- Challenges to use business data effectively
- How to query massive datasets ?
- Amazon Athena ML
 - Frequently used architecture patterns
 - Demo
 - Benefits

Modern apps store massive data



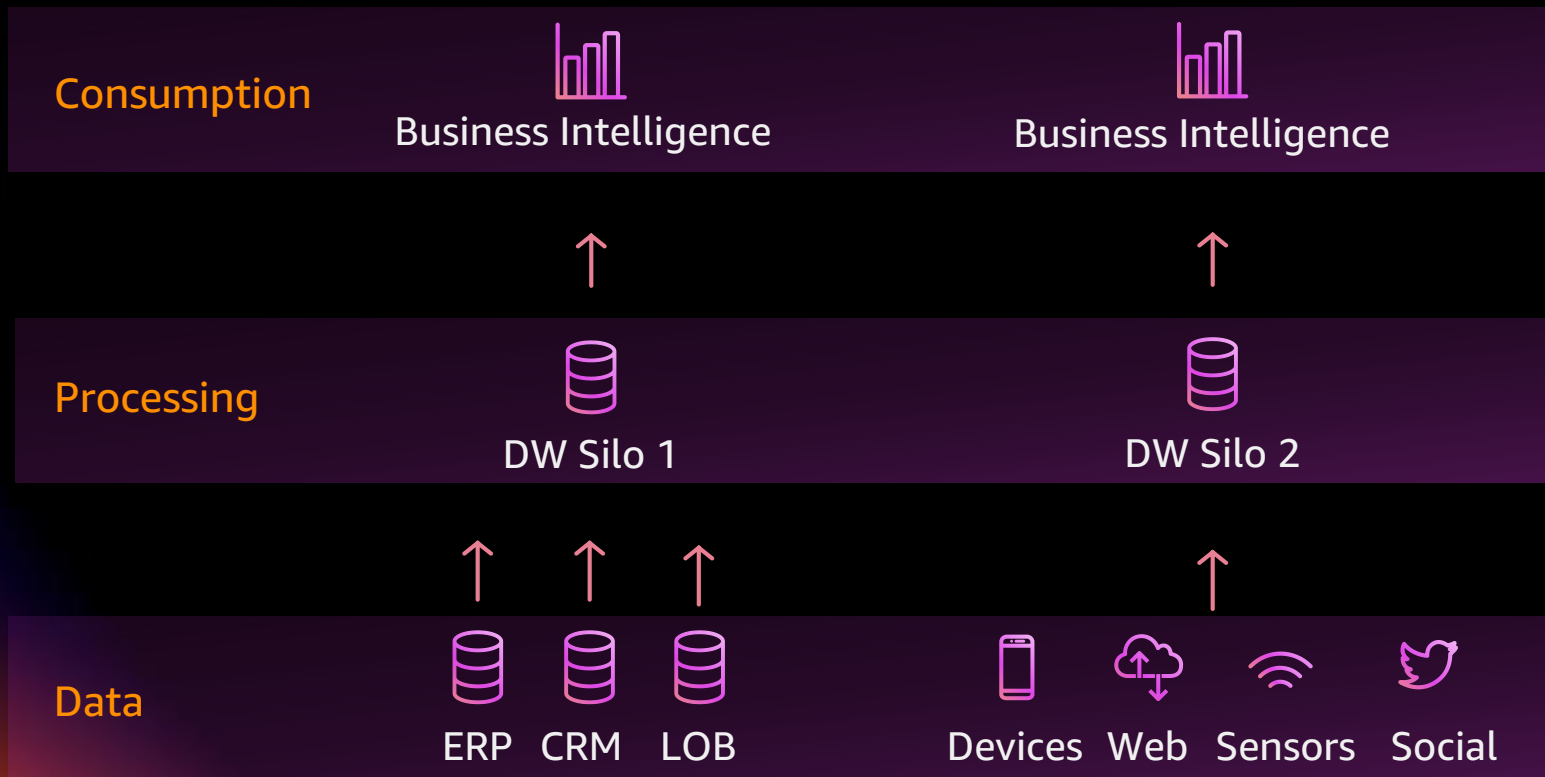
Scalability issues with traditional approach

Traditional Data Footprint



Scalability issues with traditional approach

Traditional Data Footprint



Business Goals



Make better decisions



Improve efficiencies



Respond faster



Uncover opportunities

Interactive querying with Amazon Athena



SERVERLESS

ZERO setup cost

Point to Amazon S3
and start querying



PAY PER QUERY

Pay only for queries
run

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



OPEN AND FLEXIBLE

ANSI SQL
JDBC/ODBC drivers

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



EASY

Serverless: zero
infrastructure, zero
administration

Integrated with
Amazon QuickSight
and other BI tools

Use ML models with Amazon Athena



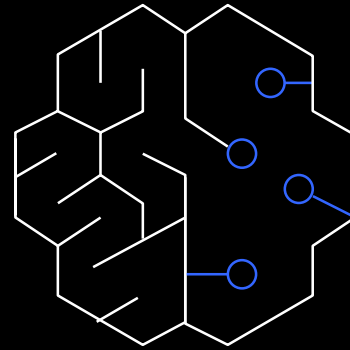
FEDERATED ATHENA QUERY

Select data from any
data source



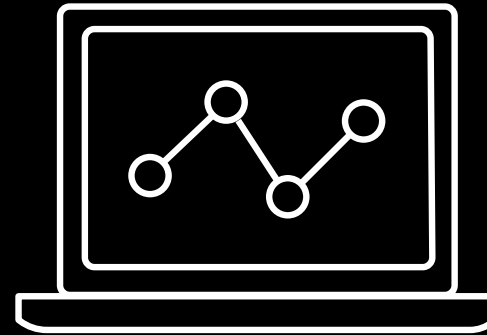
TRANSFORM DATA

Using user defined
functions in
Amazon Athena



TRAIN AND DEPLOY MODEL

On Amazon SageMaker



WRITE UDF

To pre process or post
process data



RUN INFERENCE

On data from any data
source

How to use Amazon Athena ML ?

USING EXTERNAL FUNCTION

```
ml_function_name (variable1 data_type[, variable2 data_type][,...])
```

RETURNS data_type

```
SAGEMAKER 'sagemaker_endpoint'
```

```
SELECT ml_function_name(expression)
```

How to use Amazon Athena ML ?

USING EXTERNAL FUNCTION

```
m1_function_name (variable1 data_type[, variable2 data_type][,...])  
RETURNS data_type  
SAGEMAKER 'sagemaker_endpoint'  
SELECT m1_function_name(expression)
```

USING EXTERNAL FUNCTION

```
predict_customer_registration(age INTEGER)  
RETURNS DOUBLE  
SAGEMAKER 'xgboost-2019-09-20-04-49-29-303'  
SELECT  predict_customer_registration(age) AS probability_of_enrolling,  
        customer_id  
FROM    "sampledb"."m1_test_dataset"  
WHERE   predict_customer_registration(age) < 0.5;
```

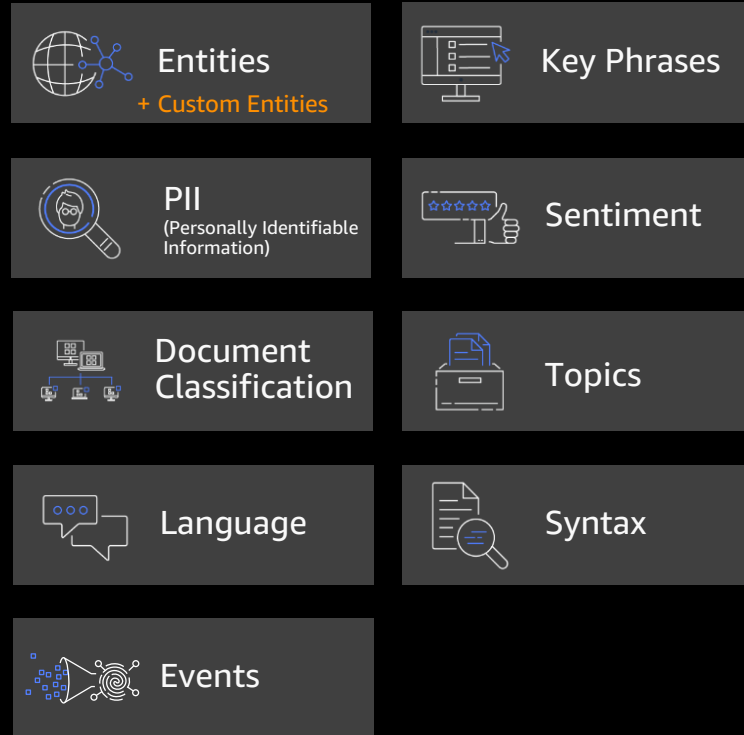
Amazon Athena ML Usage Pattern 1

Derive NLP Insights using Amazon Comprehend

Amazon Comprehend



Automatically
extract insights
from text



Amazon.com, Inc. is located in Seattle, WA
and was founded July 5th, 1994 by Jeff
Bezos. Known to the most customer obsessed
organization, it welcomes thousands of
customers and partners to one of its flagship
events AWS re:Invent every year.

Amazon.com, Inc. Entity: **ORGANIZATION**

Seattle, WA Entity: **LOCATION**

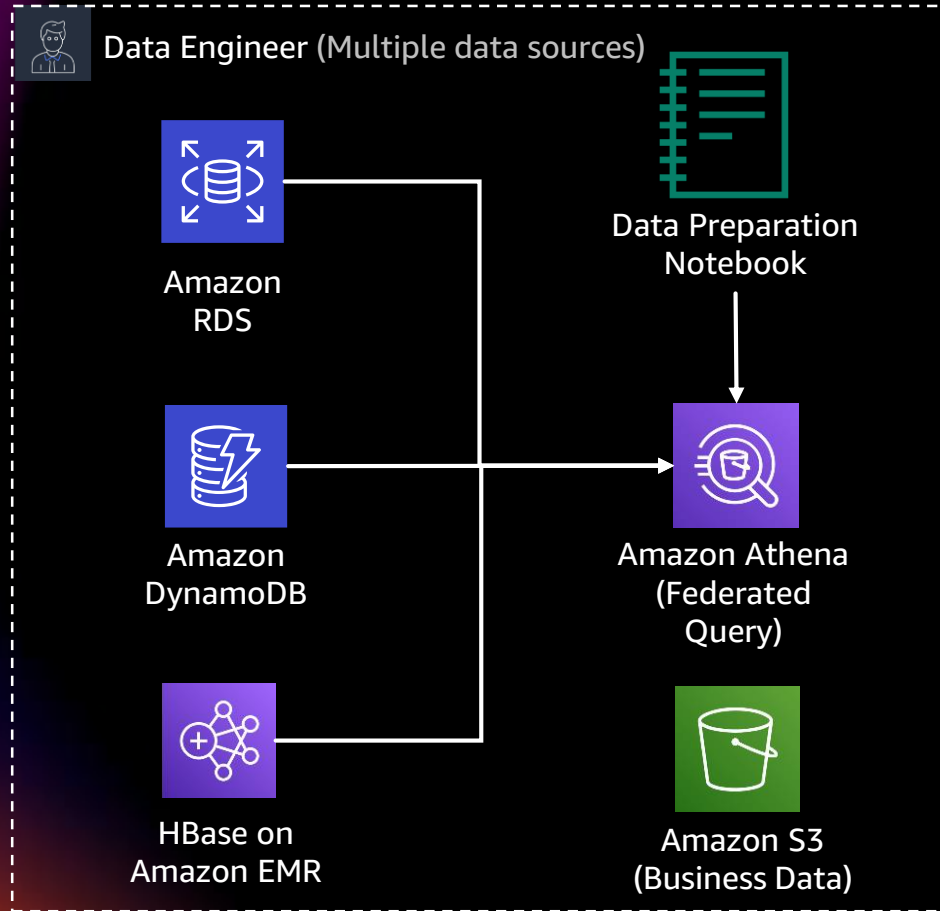
July 5th, 1994 Entity: **DATE**

Jeff Bezos Entity: **PERSON**

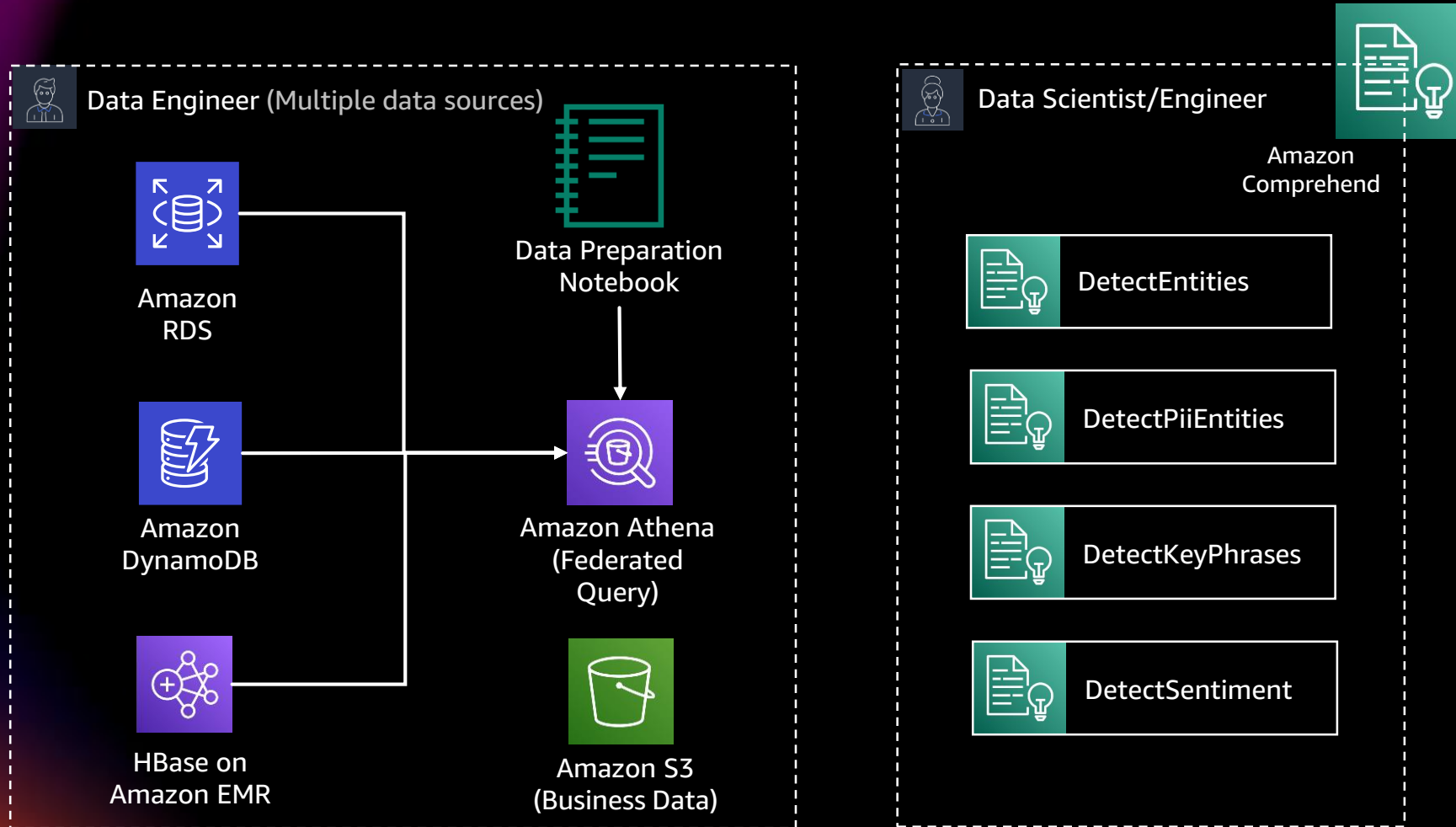
thousands of customers Entity: **QUANTITY**

re:Invent Entity: **EVENT**

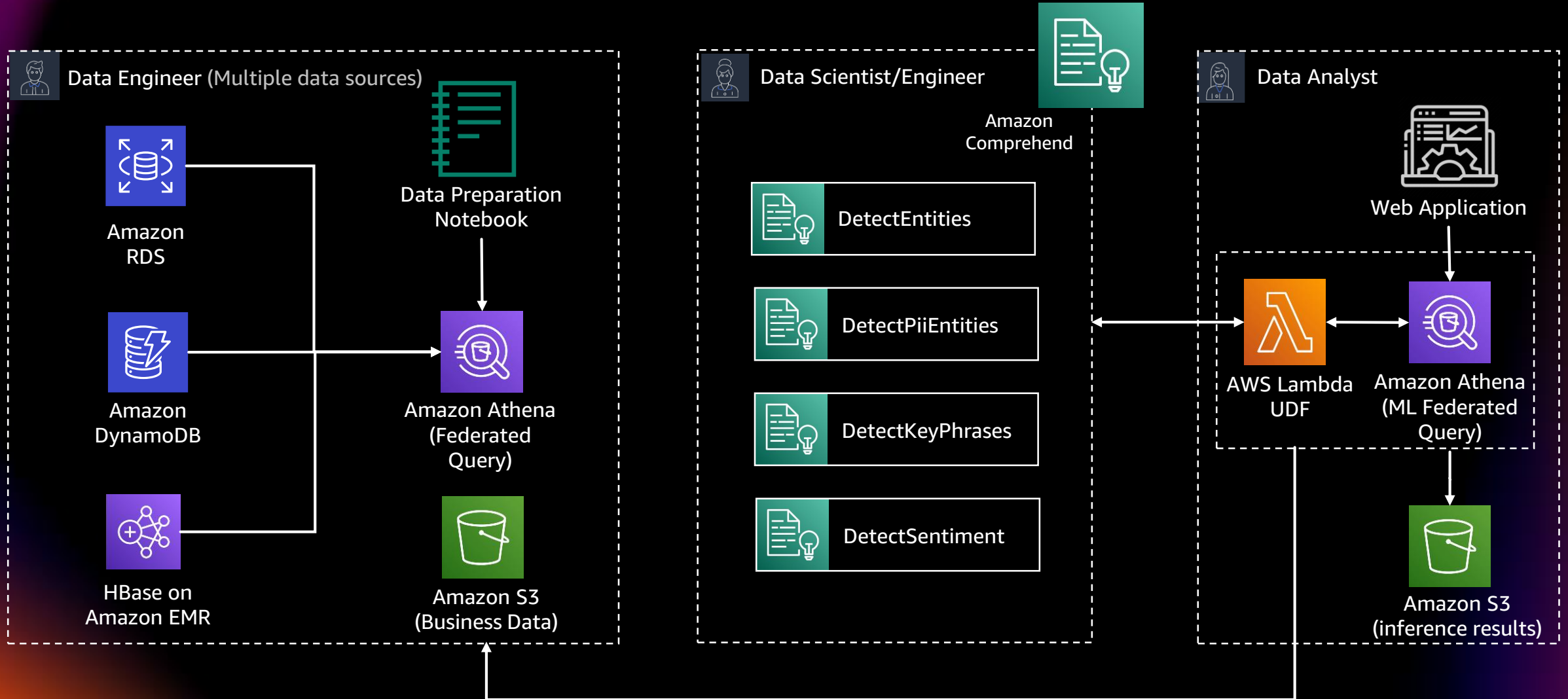
Amazon Athena ML – NLP - Usage Pattern 1



Amazon Athena ML – NLP - Usage Pattern 1



Amazon Athena ML – NLP – Usage Pattern 1



Amazon Athena ML Usage Pattern 1 - Demo

Usage Pattern 1 – Code Samples

#Detect Entities using Amazon Comprehend

USING EXTERNAL FUNCTION

```
detect_entities(text_col VARCHAR, lang VARCHAR)
RETURNS VARCHAR
LAMBDA 'textanalytics-udf'
SELECT detect_entities('content', 'en') as entities;
```

#Detect PII using Amazon Comprehend

USING EXTERNAL FUNCTION

```
detect_pii_entities(text_col VARCHAR, lang VARCHAR)
RETURNS VARCHAR
LAMBDA 'textanalytics-udf'
SELECT detect_pii_entities('Content', 'en') as pii;
```

Amazon Athena ML Usage Pattern 2

Amazon SageMaker JumpStart open source ML model endpoints

Amazon SageMaker JumpStart

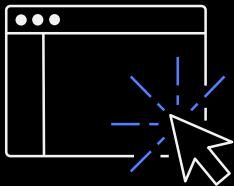


300+ pre-trained, state-of-the-art, open source models from PyTorch Hub, TensorFlow Hub, Hugging Face.

18 pre-built solutions
Leverage **out-of-the-box** solutions



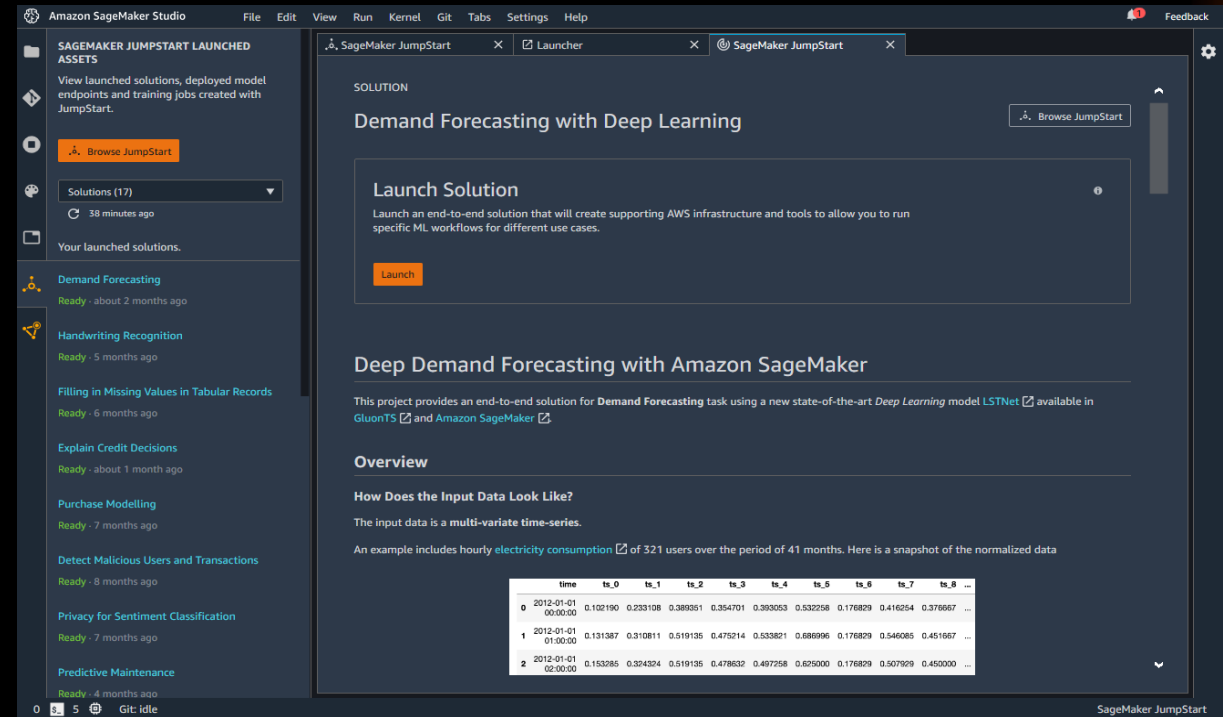
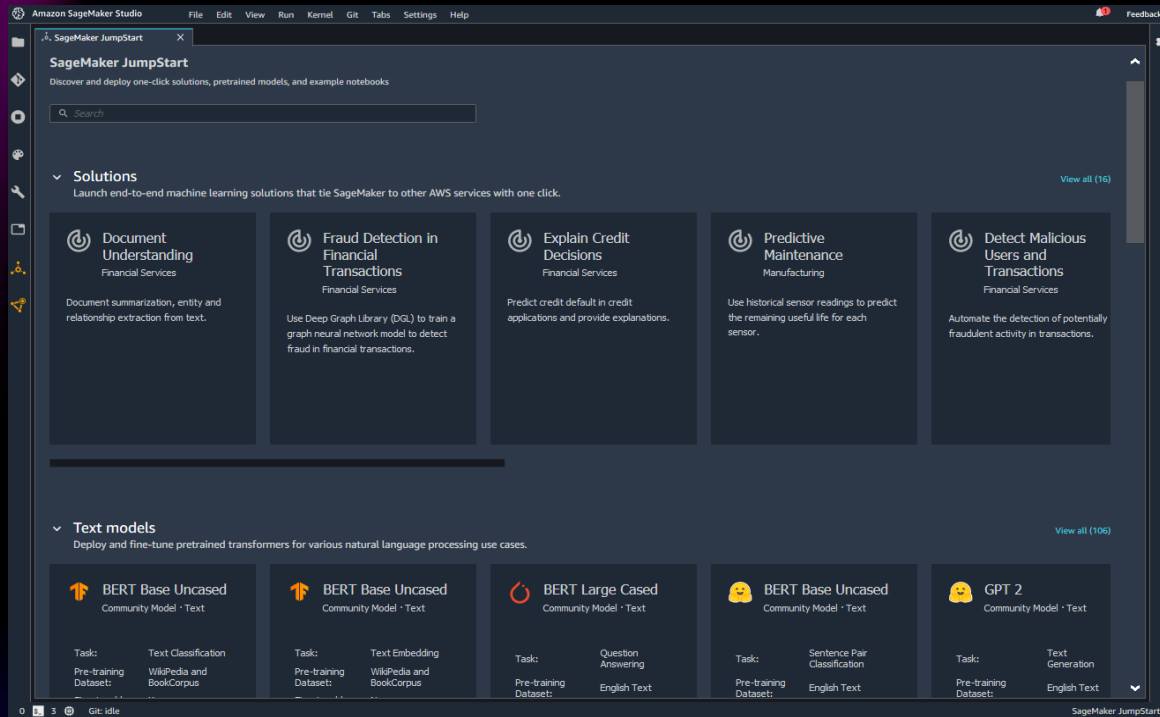
300 open source models
One-click deployable ML models & algorithms



Get started with just a **few clicks**
Customized for business problem

	TASKS	MODELS
TABULAR	Classification, Regression	LightGBM, CatBoost, XGBoost
TEXT	Sentence Classification Text Classification Question Answering Summarization Text Generation, Translation, Named Entity Recognition	BERT, RoBERTa, DistilBERT, Distillbart xsum, GPT2, ELECTRA, & More
VISION	Image Classification Image Embedding Object Detection Semantic Segmentation	ResNet, Inception, MobileNet, SSD, Faster RCNN, YOLO, & More

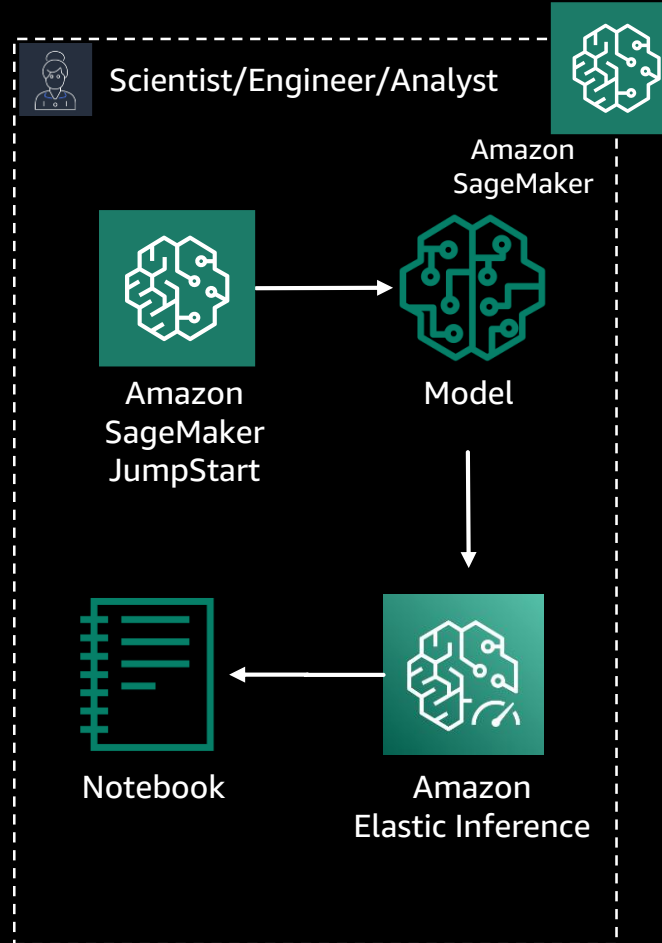
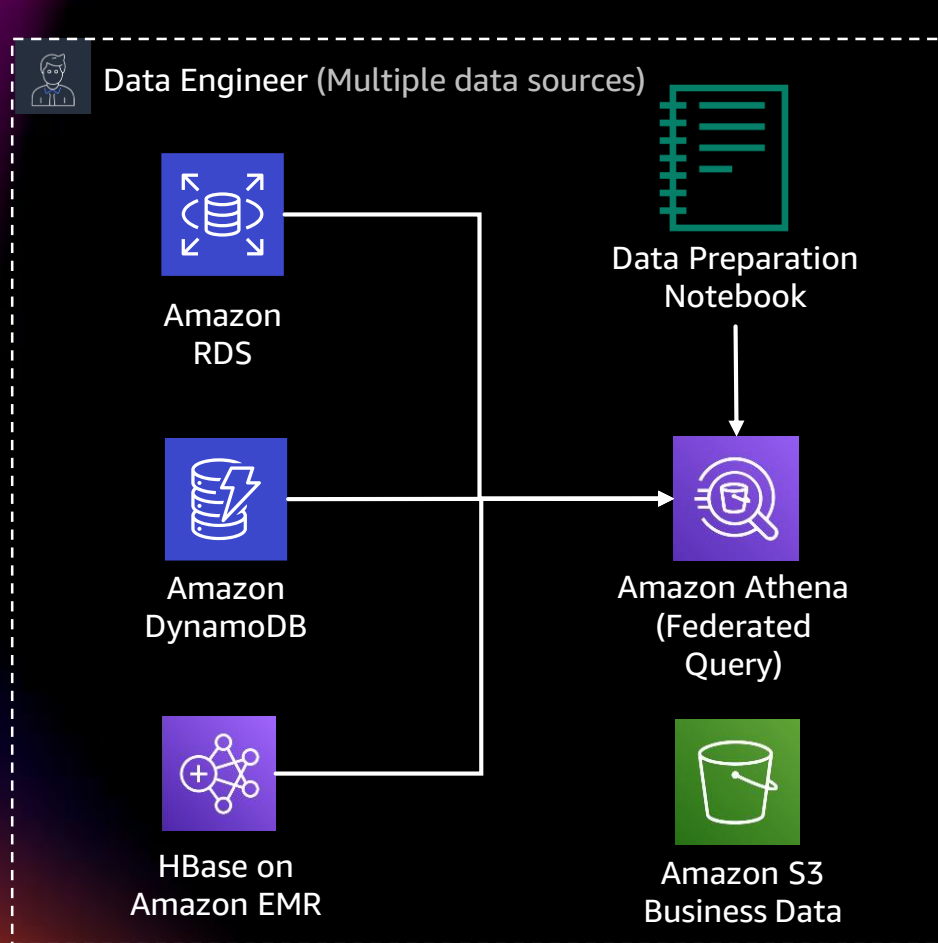
Easily launch pre-built ML solutions



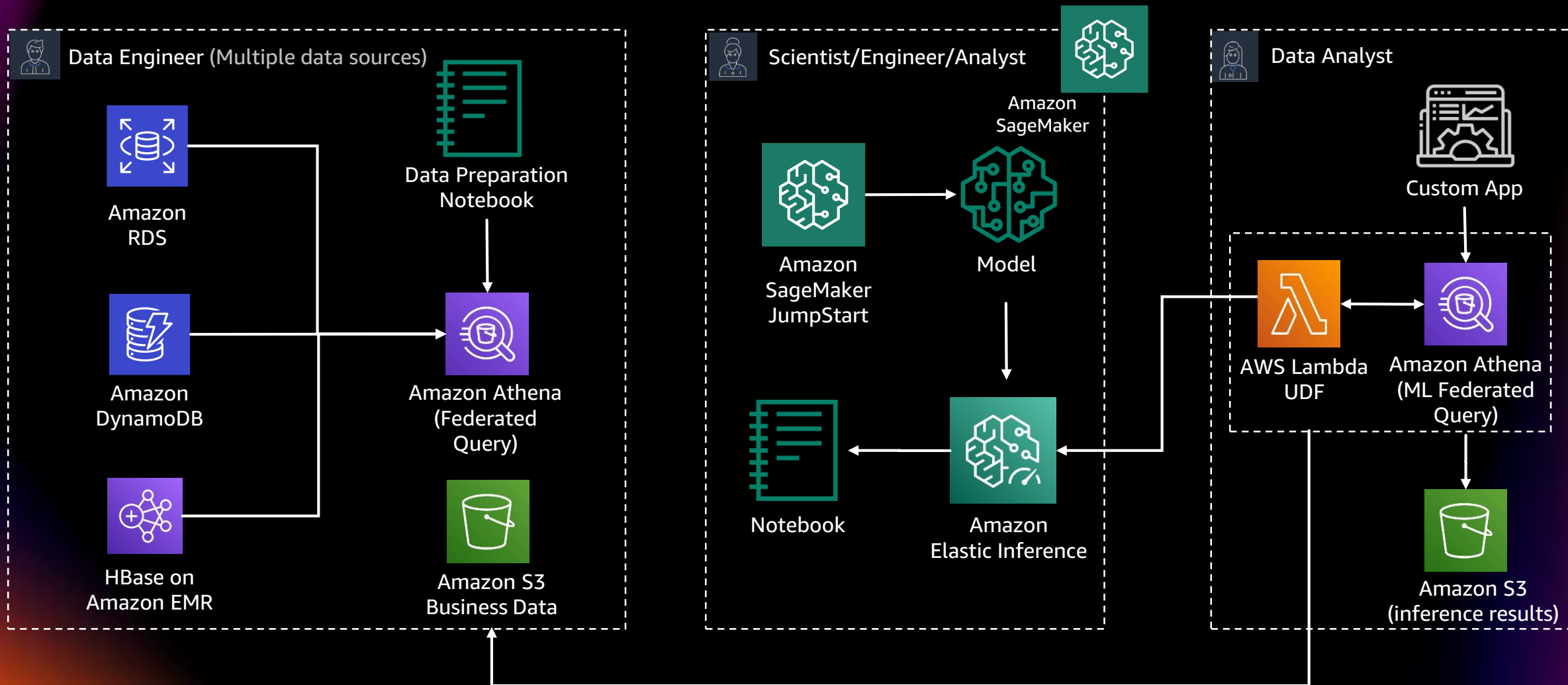
Browse and search Amazon Sagemaker JumpStart content to explore solutions, models, example notebooks, blogs, and video tutorials

Launch solutions through AWS CloudFormation with single click, Open pre-populated notebooks for solutions to solve the business problems end to end

Amazon Athena ML – Usage Pattern 2



Amazon Athena ML – Usage Pattern 2



Amazon Athena ML Usage Pattern 2 - Demo

Usage Pattern 2 – Code Samples

#Detect Entities using Amazon SageMaker JumpStart & UDF – AWS Lambda

USING EXTERNAL FUNCTION

```
detect_entities_jumpstart_ner(content VARCHAR)
```

```
RETURNS VARCHAR
```

```
LAMBDA 'textanalytics-smjs-ner-udf'
```

```
SELECT review_headline_en,
```

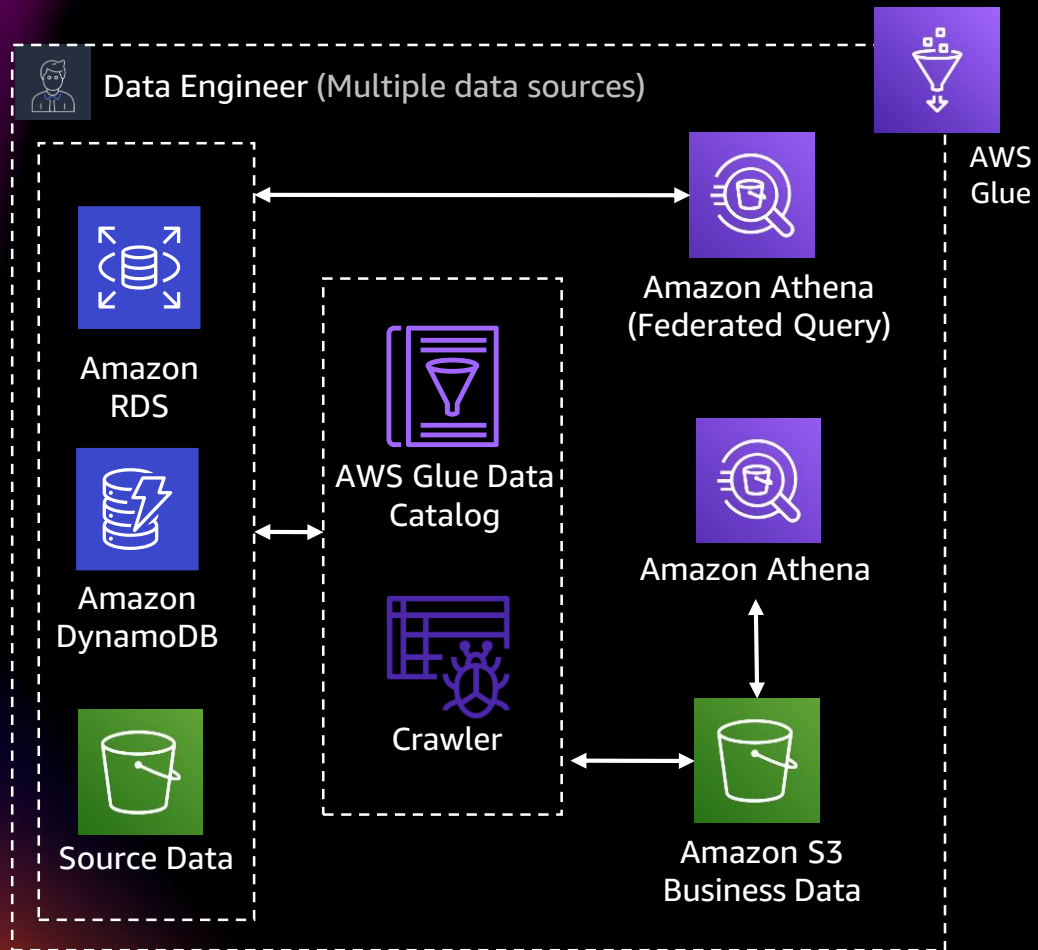
```
    detect_entities_jumpstart_ner(review_headline_en) AS entities
```

```
FROM sampled_b.amazon_reviews;
```

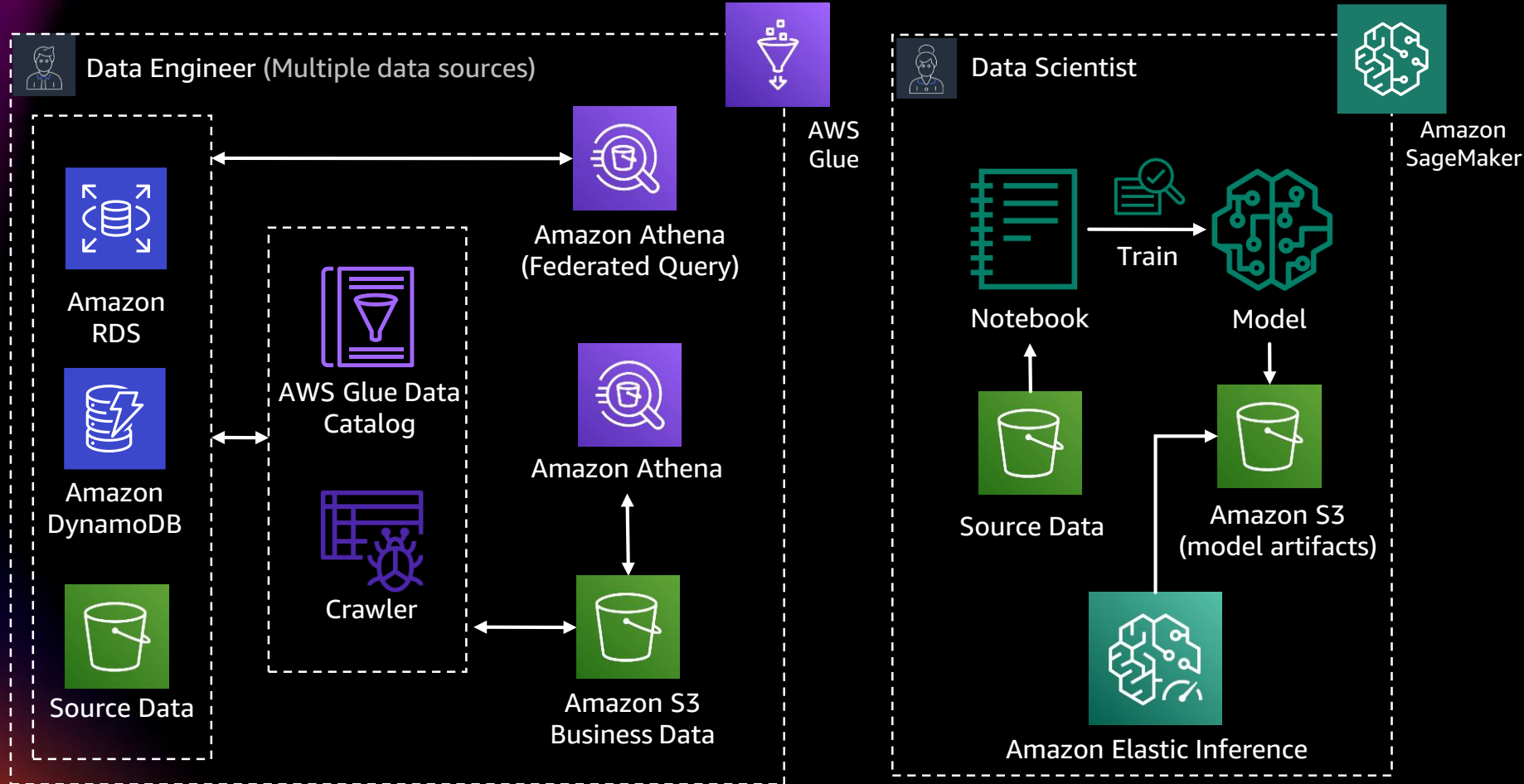

Amazon Athena ML Usage Pattern 3

Detect Anomalies using Amazon SageMaker built ML models

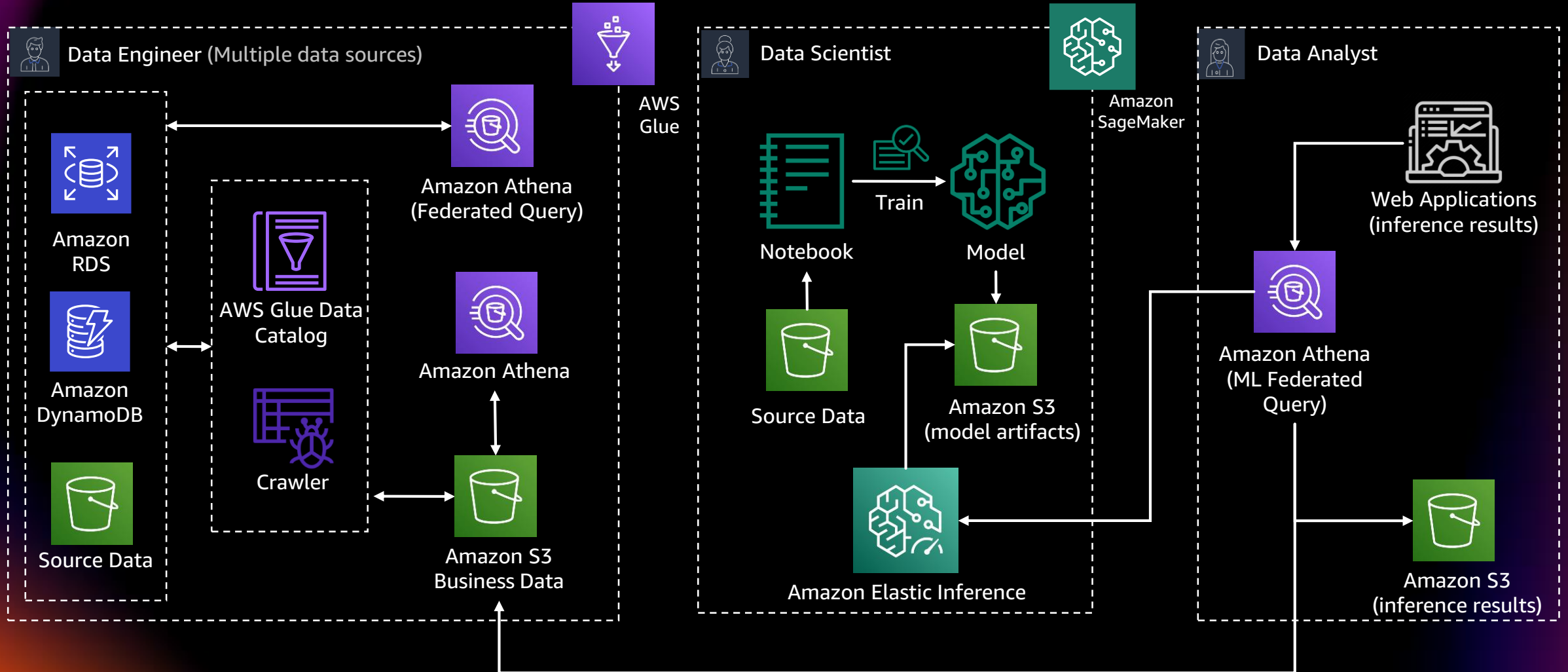
Amazon Athena ML – Usage Pattern 3



Amazon Athena ML – Usage Pattern 3



Amazon Athena ML – Usage Pattern 3



Amazon Athena ML Usage Pattern 3 - Demo

Usage Pattern 3 – Code Samples

#Compute Anomaly Score using Amazon SageMaker Random Cut Forest Model

USING EXTERNAL FUNCTION

detect_anomaly(b INT)

RETURNS DOUBLE

SAGEMAKER 'randomcutforest-2022-06-23-13-44-34-342'

SELECT o_orderdate, COUNT(*) AS number,

detect_anomaly(CAST(COUNT(*) AS INT)) as anomaly_score

FROM "lambda:mysql".sales.orders

GROUP BY o_orderdate

ORDER BY detect_anomaly(CAST(COUNT(*) as INT)) DESC LIMIT 100;

Amazon Athena ML - Benefits



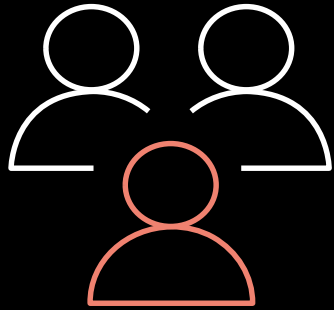
No ML Expertise
Needed



Easily power workloads
with ML APIs



Pay As You Use



Flexible to Use
By Multiple Personas



Serverless
Architecture

Key takeaways

- Challenges to use business data effectively
- Efficient querying of massive datasets
- Amazon Athena ML architecture patterns
 - Derive NLP insights using Amazon Comprehend
 - Amazon SageMaker JumpStart open source ML model endpoints
 - Detect anomalies using Amazon SageMaker built ML models
 - Benefits

Getting started

- [Analyze Text](#) with SQL Functions using Amazon Comprehend
- Experiment with open source models [using Amazon SageMaker JumpStart](#) with ease
- Use [Anomaly Detection ML Inference](#) with Amazon Athena ML
- [Query ML Models](#) with Amazon Athena
- Learn integrating [AWS Lambda with Amazon Athena for implementing UDFs](#)
- Check out this [blog](#) for enabling visualizations with Amazon Athena outputs

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.

- Harness data to reinvent your organization
- In unpredictable times, a data strategy is key
- Make data a strategic asset
- Rewiring your culture to be data-driven
- Put your data to work with a modern analytics approach
- ... and more!



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

[Visit resource hub](#)

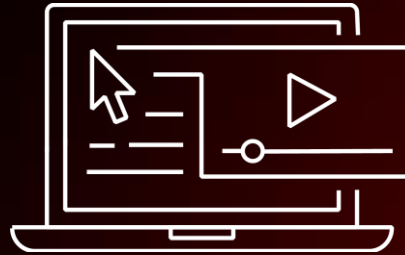
AWS Training and Certification for Data and Analytics



AWS Data & Analytics FREE Training Resources

Discover how to harness data, one of the world's most valuable resources, and innovate at scale.

<https://bit.ly/3Ntlhy7>



AWS Data Analytics Learning Plan

This learning plan expose you to the fastest way to get answers from all your data to all your users. It can also help prepare you for the AWS Certified Data Analytics - Specialty certification exam.

<https://bit.ly/3wBVjD1>



AWS Certified Data Analytics - Specialty

Earning AWS Certified Data Analytics – Specialty validates expertise in using AWS data lakes and analytics services.

<https://go.aws/3lwF0RR>

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!