

AWS State, Local, and Education Learning Days

Madison, WI

3:15PM to 4:15PM

300
level

AI/ML for data and analytics

Unleash the power of data: Transform public sector insights with AI and ML across your entire enterprise.



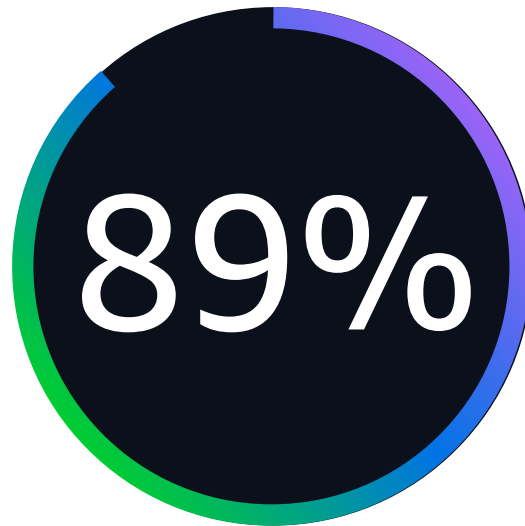
AI/ML For Data and Analytics

Ignatius Narchetty

Solutions Architect, SLG
inarchet@amazon.com

AI is transforming the data landscape

of CDOs are moving forward with Gen AI

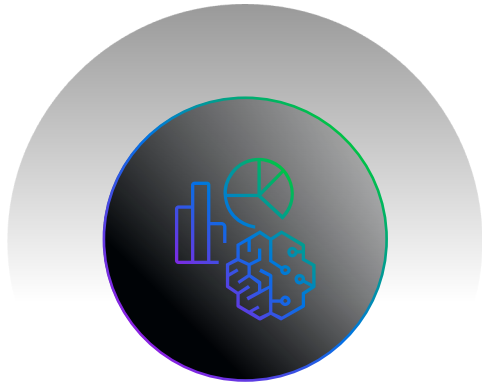


52%

of CDOs view data foundation not being ready for AI

Back to the basics

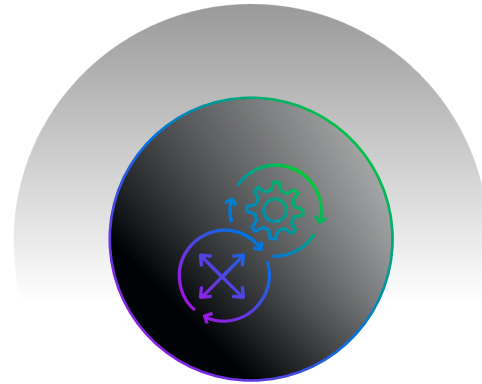
What's holding organizations back



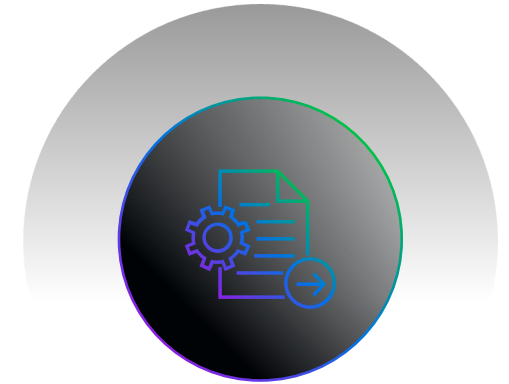
Integrating
analytics and AI



Overcoming
data silos



Governing data
and AI



Scaling with
operational excellence

How can you build a modern data strategy?



Tenets of a Modern Data Foundation

Data as a Product

We believe data deserves quality, accessibility, and usability, driving innovation and value creation

Domain-Driven and User-Centric

We empower business users to take ownership of their data, fostering accountability and agility.

Automate with AI

We use advanced analytics, machine learning and generative AI to automate data processing and uncover hidden patterns.



Continuous Innovation

We foster a culture of continuous learning and innovation to enhance our data capabilities.

Data Drives Decision Making

We believe data should be easily accessible and usable, enabling timely and informed decision-making.

Proactive Data Governance

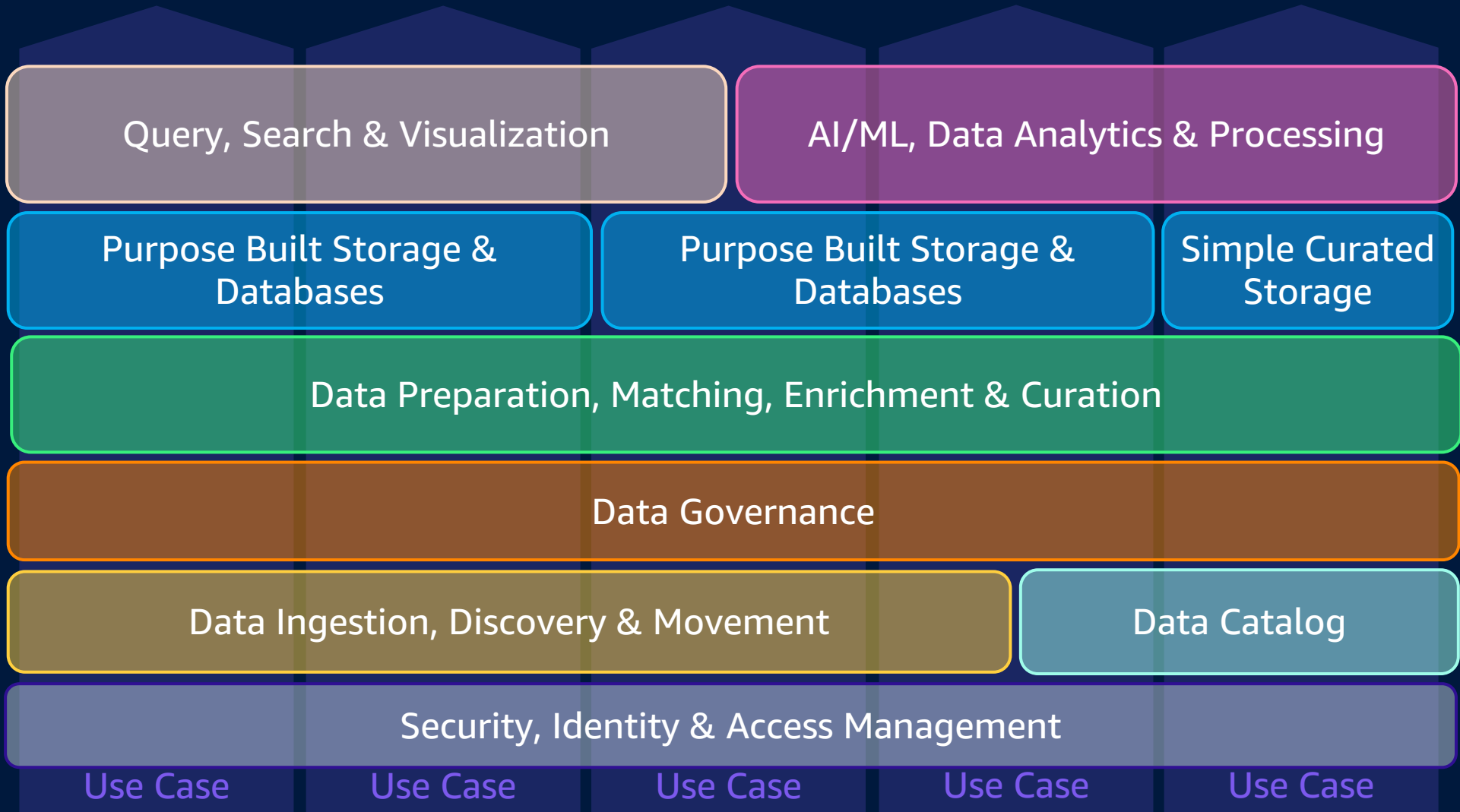
We ensure data quality, security, and compliance without compromising agility and responsiveness

Open and Unified Data Platform

We provide a cohesive, enterprise-wide data platform that integrates various data sources and tools.



A Modern, Foundational Data Architecture

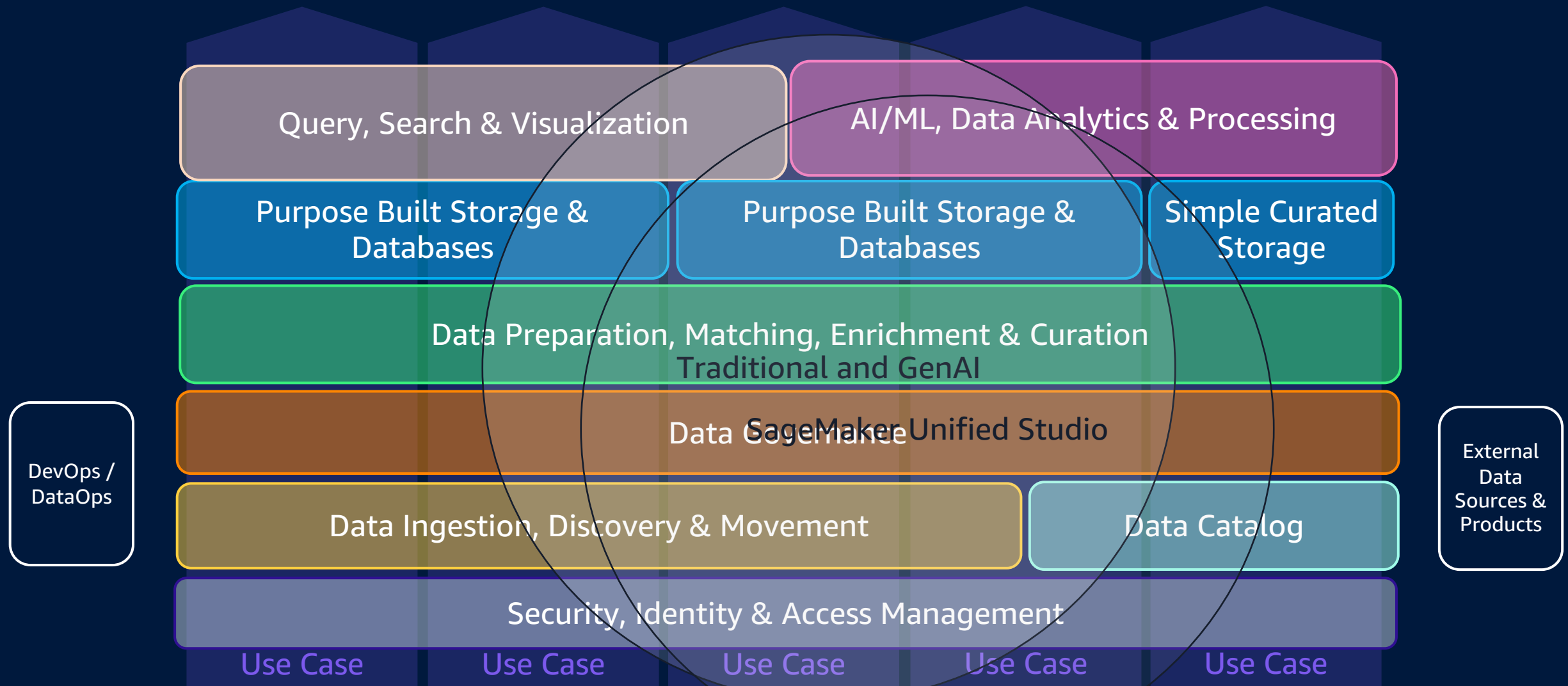


DevOps /
DataOps

External
Data
Sources &
Products



A Modern, Foundational Data Architecture



THE NEXT GENERATION OF



Amazon SageMaker

The center for all your data, analytics, and AI

Amazon SageMaker

Unified Studio

COMING SOON

COMING SOON

SQL Analytics

Amazon Redshift

Data Processing

Amazon EMR
AWS Glue
Amazon Athena

Model Development

Amazon SageMaker AI

GenAI App Development

Amazon Bedrock

Streaming

Amazon MSK
Amazon Kinesis

Business Intelligence

Amazon QuickSight

Search Analytics

Amazon OpenSearch Service

Notebooks | Query editor | Visual editors | Gen AI IDEs

Experiment, train, and deploy AI models with Amazon SageMaker AI

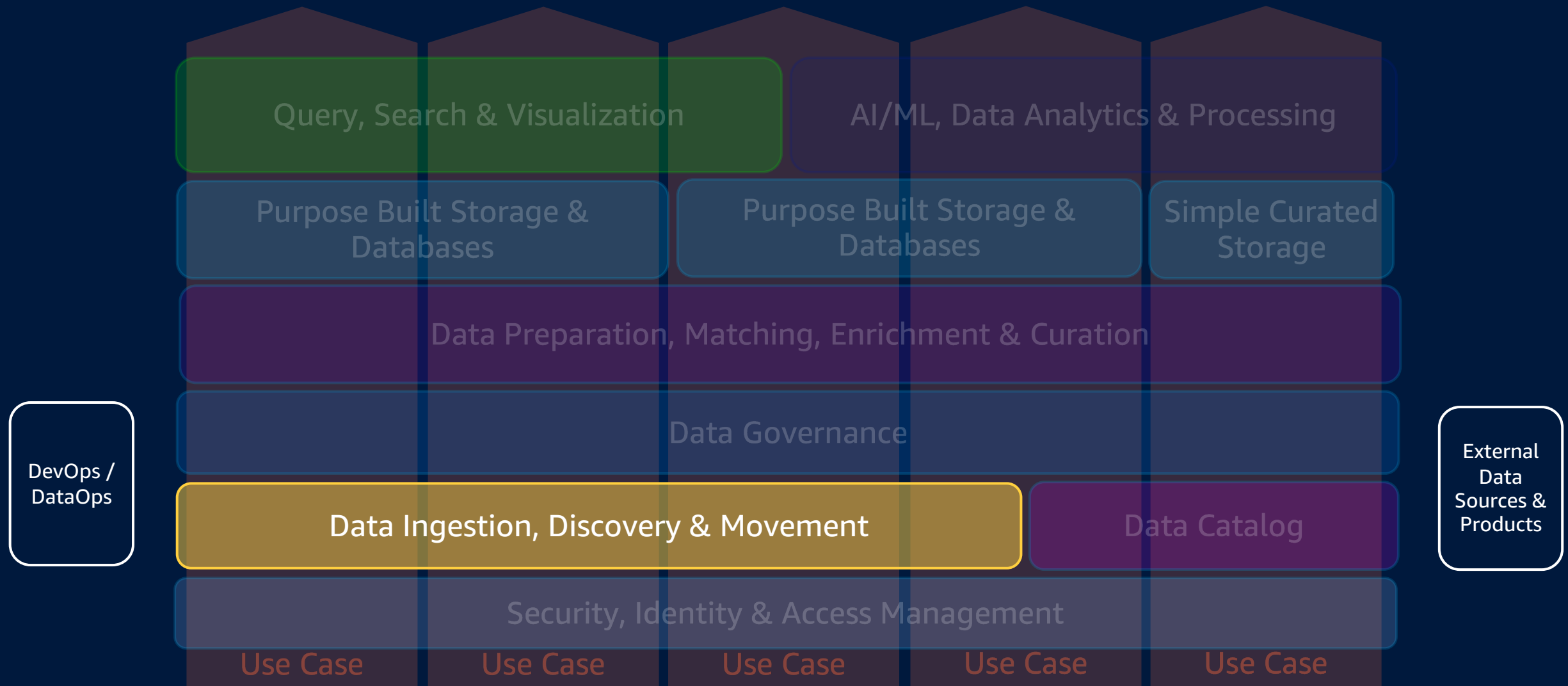
Build custom Gen AI apps with Amazon Bedrock

Prepare and integrate data with Amazon EMR

Run SQL queries with Amazon Redshift



A Modern, Foundational Data Architecture



Data Discovery, Ingestion and Movement

Insights from unstructured content

ML-powered services to extract information from unstructured data

Analyze text with natural language processing (NLP) to identify topics, extract entities, understand sentiment, and classify documents with **Amazon Textract, Amazon Rekognition, and Amazon Comprehend**

Translate content at scale with **Amazon Translate**

Content analysis and object detection

Identify objects of interest from images and videos with **Amazon Rekognition**

Detect personal protective equipment (PPE) to improve worker safety

Analyze vehicle traffic and pedestrian and bicycle safety

Detect objects of interest in video and reduce human effort required to review footage

Transcription, classification and analysis

Automatically convert speech to text with **AWS Transcribe**

Extract key business insights from customer calls, video files, clinical conversations.

Analyze media content

Subtitle your on-demand and broadcast content to increase accessibility and improve customer experience

Name	Modified
document_insights_t...	29m ago
document_insights.jp...	39s ago
getting_started.ipynb	29d ago
README.md	last mo.
Untitled.ipynb	14d ago

```

[ ]: !python -m pip install q amazon-textract-response-parser --upgrade
!python -m pip install q amazon-textract-caller --upgrade
!python -m pip install q amazon-textract-prettyprinter==0.0.16
!python -m pip install q amazon-textract-textractor --upgrade

[ ]: #Generate Code Here

[ ]: # Example usage
try:
    bucket_name = 'smus-unstructureddata-bucket-1'
    object_key = 'documents/invoice.png'

    image_bytes = read_image_from_s3(bucket_name, object_key)
    print(f"Successfully read image, size: {len(image_bytes)} bytes")

    # You can now use the image_bytes for further processing
    # For example, you could convert it to a PIL Image:
    from PIL import Image
    import io
    from IPython.display import display
    image = Image.open(io.BytesIO(image_bytes))
    display(image)

except Exception as e:
    print(f"Failed to read image: {e}")

[ ]: import boto3
import botocore
import sagemaker
import pandas as pd
from IPython.display import Image, display, JSON
from textractcaller.t_call import call_textract, TextractFeatures, call_textract_expense
from textractprettyprinter.t_pretty_print import convert_table_to_list
from trp import Document
import os
textract = boto3.client('textract')
expense_resp = textract.analyze_expense(Document={'Bytes': image_bytes})

[ ]: summary_entities_values = []
summary_fields = {}
expense_item = {}

for expense_doc in expense_resp["ExpenseDocuments"]:
    for field in expense_doc["SummaryFields"]:
        kvs = {}
        if "LabelDetection" in field:
            if "ValueDetection" in field:
                kvs[field["LabelDetection"]["Text"]] = field["ValueDetection"]["Text"]
            else:
                kvs[field["Text"]] = field["ValueDetection"]["Text"]
        summary_entities_values.append(kvs.copy())
        kvs = None

    for line_item_group in expense_doc["LineItemGroups"]:
        for line_items in line_item_group["LineItems"]:
            for field in line_items["LineItemExpenseFields"]:
                print(field)

```

Configure space Learn more

Amazon Q Free Tier I am Amazon Q, your generative AI assistant. Ask me anything about Amazon SageMaker Unified Studio and other AWS services. Or, choose a sample question below to start a conversation.

- What is SageMaker Unified Studio?
- Find me data on marketing
- How do I connect to a Redshift cluster?
- How can I build a Gen AI playground in Bedrock?

To learn more, see Amazon Q developer. Please help us improve by leaving feedback.

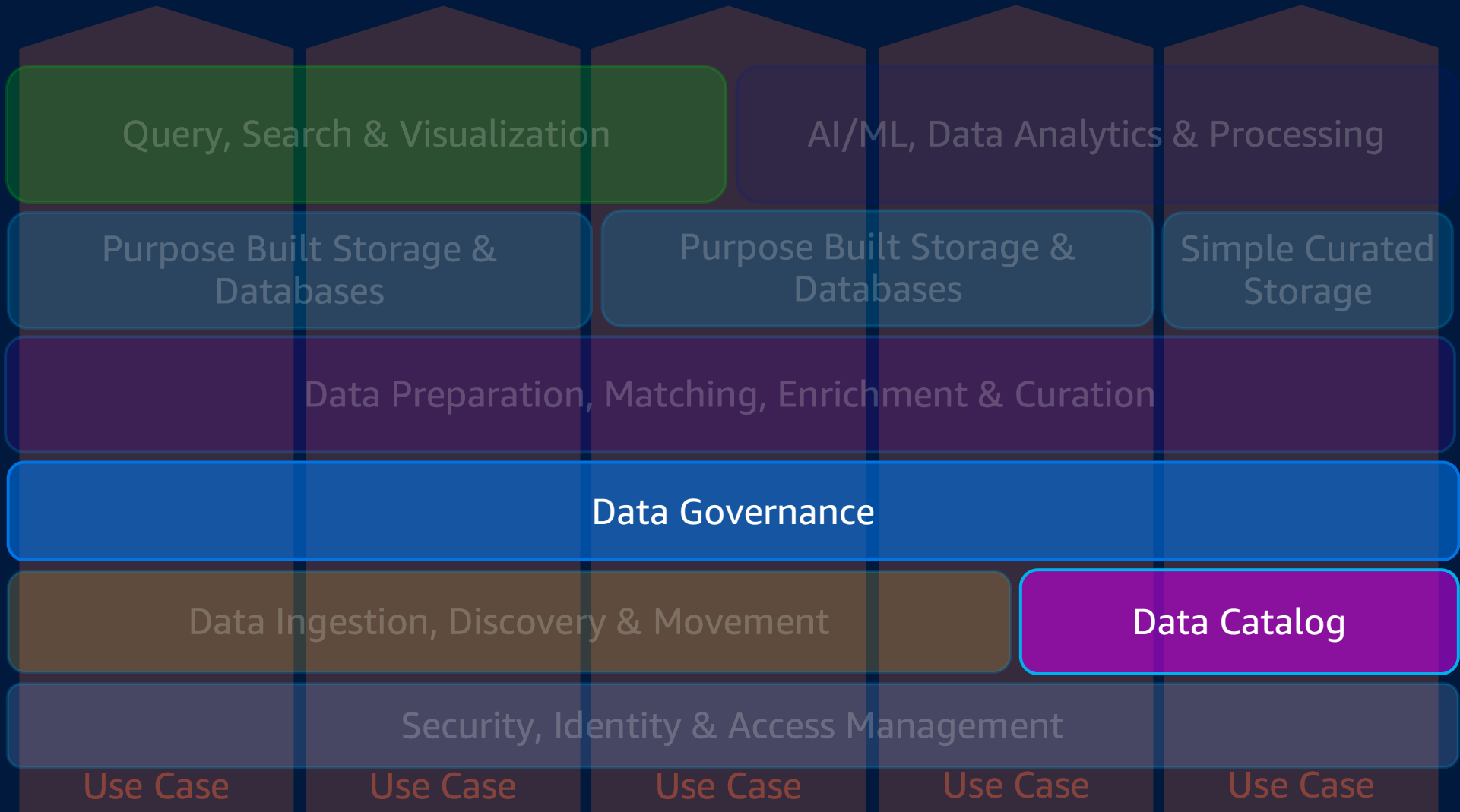
Additional Info Amazon Q may retain chats to provide and maintain the service. Amazon Q stores and processes your data in the N. Virginia (us-east-1) Region, even if the Amazon SageMaker Unified Studio domain is in a different AWS Region.

Ask me anything about SageMaker Unified Studio. Max 3500 characters. Amazon Q Developer uses generative AI. You may need to verify responses. See the AWS Responsible AI Policy.

Discussion

How might the ability to extract insights from unstructured data change your organization's data strategy? What new data sources might become valuable that were previously too difficult to process?

A Modern, Foundational Data Architecture



DevOps /
DataOps

External
Data
Sources &
Products



Data Catalog

Business Data Catalog

Gen AI powered Business Catalog Generation

Generate context and data aware business metadata to compliment the technical metadata

Semantic Search

Efficiently explore and understand large datasets

Utilizes natural language processing (NLP) to understand the context and meaning behind search queries

Enables users to find relevant data across various data sources using intuitive, human-like language

Govern

Govern data access with Fine grained access controls and publish/subscribe workflows

Protect and secure data and models by providing access to only the data that they need.

Increase trust and transparency with lineage, data quality and access monitoring

Unified Studio

Data & AI Governance

Built into
Unified
Studio



Data



Models



Gen AI



Compute

SageMaker Catalog

Built on
Amazon
DataZone

Data
quality

Data
classification

Data & ML
lineage

Permissions and
subscriptions workflows

Amazon Q
generated context


Guardrails


Responsible AI

Auditable –
Cost logging
and monitoring

Good morning, Root

Discover

 **Catalog**
Find and access data assets and ML models.

 **Generative AI playground**
Experiment with chat, image, and video models with Amazon Bedrock.


 **Shared generative AI assets**
Explore apps and prompts shared with you.

Build


Your projects

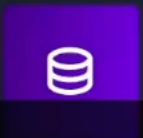
[Browse all projects](#) [Create project](#)

Project name	Status	Description	Created at (UTC-05:00)
smus-project-2	Active	-	August 03, 2025, 10:01
smus-project-1	Active	-	July 16, 2025, 19:53

 **ML and generative AI model development**
Build, train, and deploy ML and foundation models with fully managed infrastructure, tools, and workflows.
[Create notebook](#)

 **Generative AI app development with Amazon Bedrock**
Build generative AI apps and experiment with foundation models, prompts, agents, functions, and guardrails in Amazon Bedrock.
[Build chat agent](#)

 **Data processing and SQL analytics**
Analyze, prepare, and integrate data for analytics and AI using Amazon Athena, Amazon EMR, AWS Glue, and Amazon Redshift.
[Create SQL query](#)

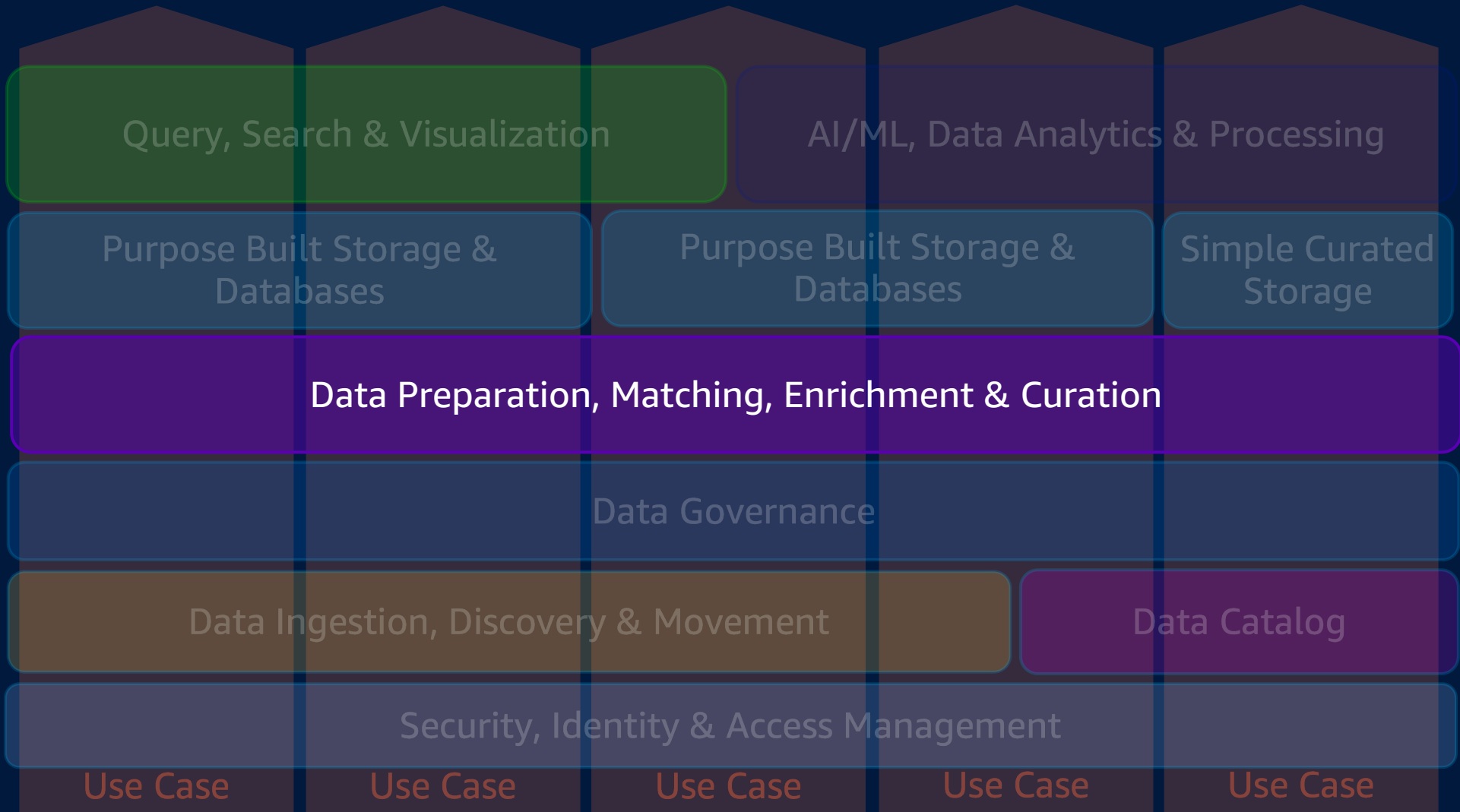
 **Data and AI governance**
Publish your data products to the catalog with glossaries and metadata forms. Govern access securely in the Amazon SageMaker Catalog built on Amazon DataZone.
[Publish data](#)

Discussion

What is the importance of a business data catalog?

Organizations can effectively manage, understand, and utilize their data assets.

A Modern, Foundational Data Architecture



DevOps /
DataOps

External
Data
Sources &
Products



Preparation, Matching, Enrichment, Curation

Data Protection

ML-powered sensitive data identification and redaction

[Amazon Macie](#) is a data security service that uses machine learning (ML) and pattern matching to discover and help protect your sensitive data.

Data Quality

Use ML to detect anomalies and hard-to-detect data quality issues

[AWS Glue Data Quality](#) learns patterns on data statistics gathered over time using ML algorithms. It detects anomalies, unusual data patterns and alerts users. It also auto-creates rules to monitor these specific patterns so that you can progressively build data quality rules.

Entity Resolution & Data Matching

Match, link, and enhance related customer, product, business, or healthcare records stored across multiple applications, channels, and data stores

[AWS Entity Resolution](#) and [AWS LakeFormation FindMatches](#) enable you to identify matching records in your dataset even when the records do not have a common unique identifier and no fields match exactly.

Home > Projects >

Overview

Project overview

Data

Compute

Members

Project catalog

- Assets
- Subscription requ
- Data sources
- Metadata entities

IDE & APPLICATIONS

- JupyterLab
- Spaces
- Partner AI Apps

DATA ANALYSIS & INTEGRATION

- Query Editor
- Visual ETL flows
- Data processing jobs

ORCHESTRATION

- Workflows
- ML Pipelines

MACHINE LEARNING & GENERATIVE AI

APP DEVELOPMENT	MODEL DEVELOPMENT	AI OPS
Chat agent	Jumpstart models	MLflow
Flow	Training jobs	Model registry
Prompt	Inference endpoints	Model evaluations
My apps	HyperPod	

Actions

New

Repository and branch: datazone-bf4eakefve5k87-dev · main

Type
—
JupyterLab
JupyterLab
JupyterLab
JupyterLab

Readme

Welcome

SageMaker Unified Studio is an integrated development environment that brings together familiar tools from AWS analytics and AI/ML services for data processing, SQL analytics, ML model development, and generative AI application development to help teams collaborate and bring data products to market faster...

[View more](#)

Project details | Project profile parameters | JDBC connection details

Project details

Project name smus-project-1	Project ID bf4eakefve5k87	Last modified July 16, 2025, 19:59 (UTC-05:00)	Amazon S3 location s3://amazon-sagemaker-339713032974-us-east-1-612b0d545750/dzd_coyergqsvwepbb/bf4eakefve5k87/dev
Status Active	Domain ID dzd_coyergqsvwepbb	Creation date July 16, 2025, 19:53 (UTC-05:00)	Project role ARN arn:aws:iam::339713032974:role/datazone_usr_role_bf4eakefve5k87_awsjezwum9p5lz
Created by dzrootdomainuser	Project profile name smus-project-profile	Domain unit name dz-domain	Update status Up to date
Description —			

dzd_coyergqsvwepbb.sagemaker.us-east-1.on.aws/projects/bf4eakefve5k87/visual-editor

Discussion

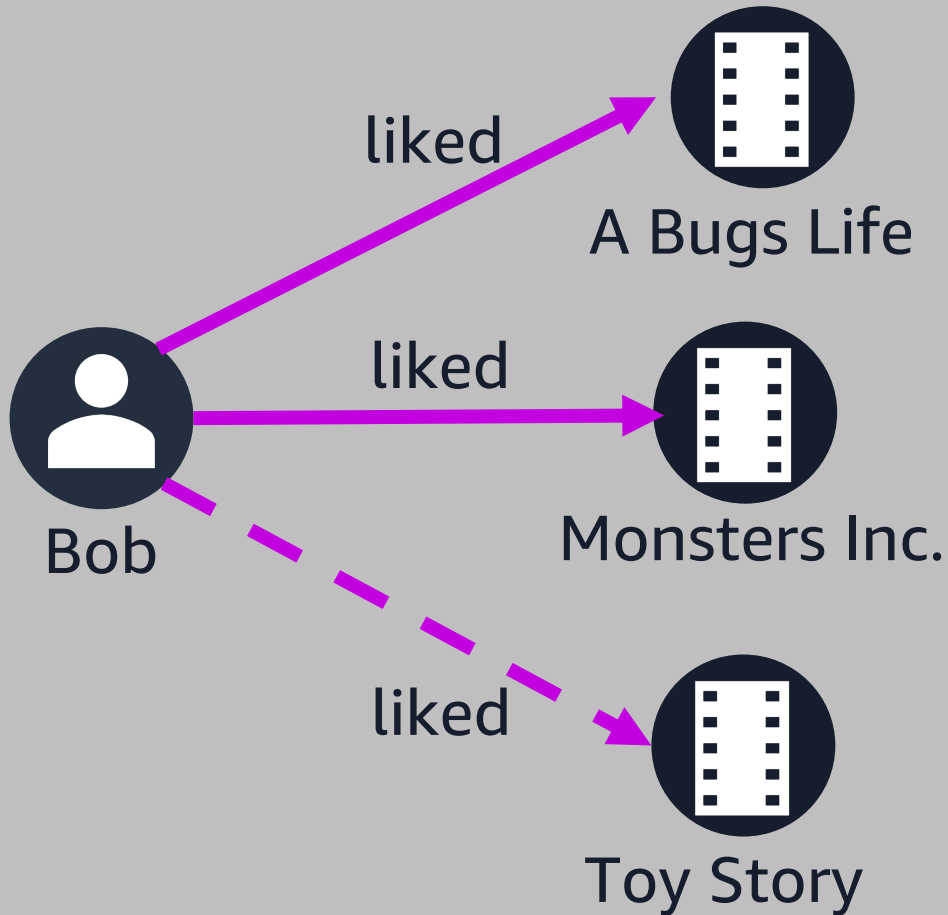
What is the difference between ETL and ELT?

A Modern, Foundational Data Architecture



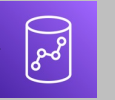
ML Purpose Built Storage & Databases

Link Prediction in graph databases (Amazon Neptune ML)



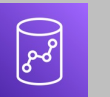
Redshift ML (Train and use models)

```
CREATE MODEL customer_churn
FROM (SELECT c.age, c.zip,
c.monthly_spend, c.monthly_cases,
c.active AS label
FROM customer_info_table c)
TARGET label
FUNCTION predict_customer_churn
```



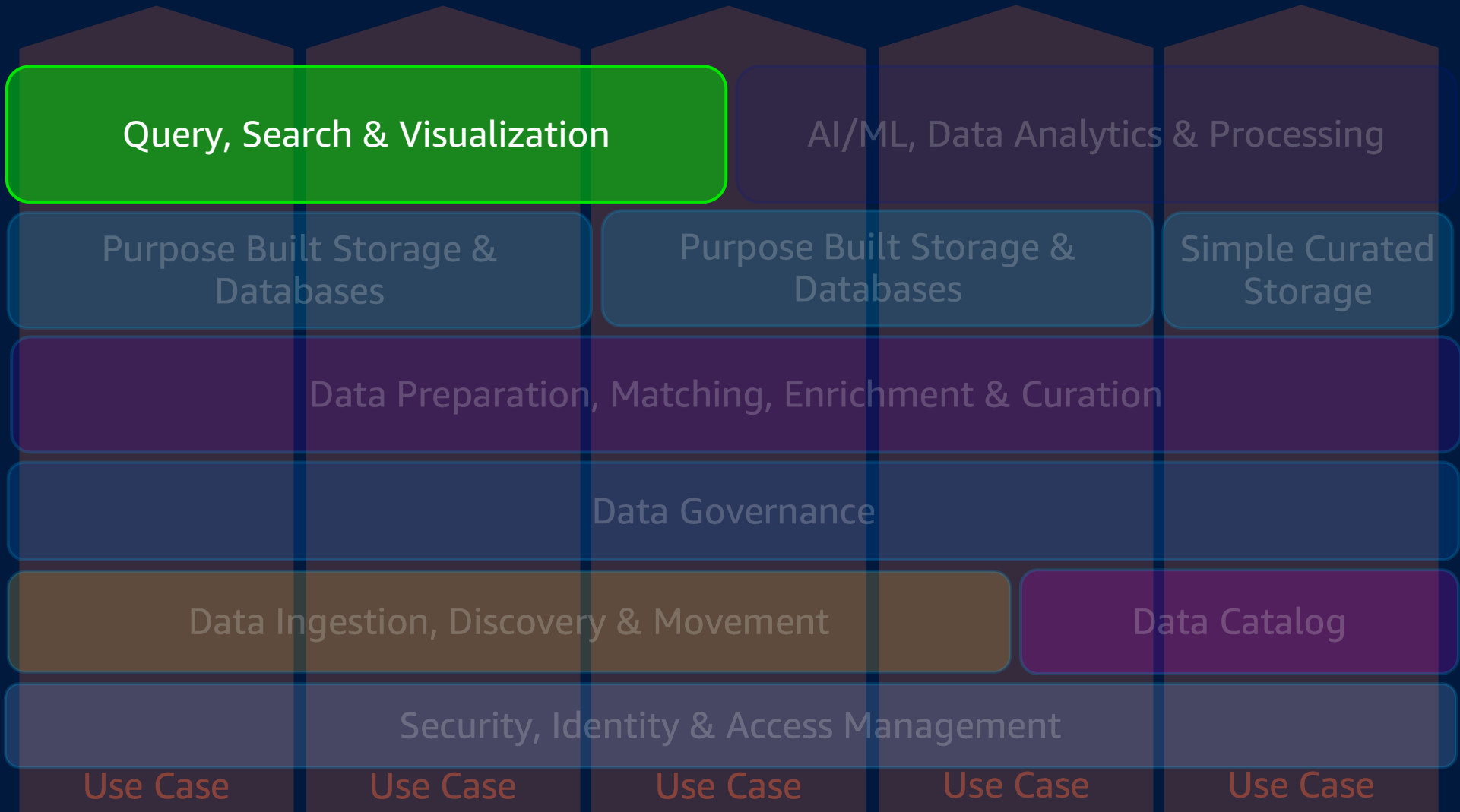
Amazon Redshift

```
SELECT n.id, n.firstName,
n.lastName,
predict_customer_churn(n.age, c.zip
...)
AS activity_prediction
FROM new_customers n
WHERE n.marital_status = 'single'
```



Amazon Redshift

A Modern, Foundational Data Architecture



DevOps /
DataOps

External
Data
Sources &
Products



Query, Search and Visualize with QuickSight

Author dashboards

Build Visuals using natural language for dashboards and reports

Create calculations using natural language without learning syntax

Refine visuals by describing desired formats using natural language

AI-assisted storytelling

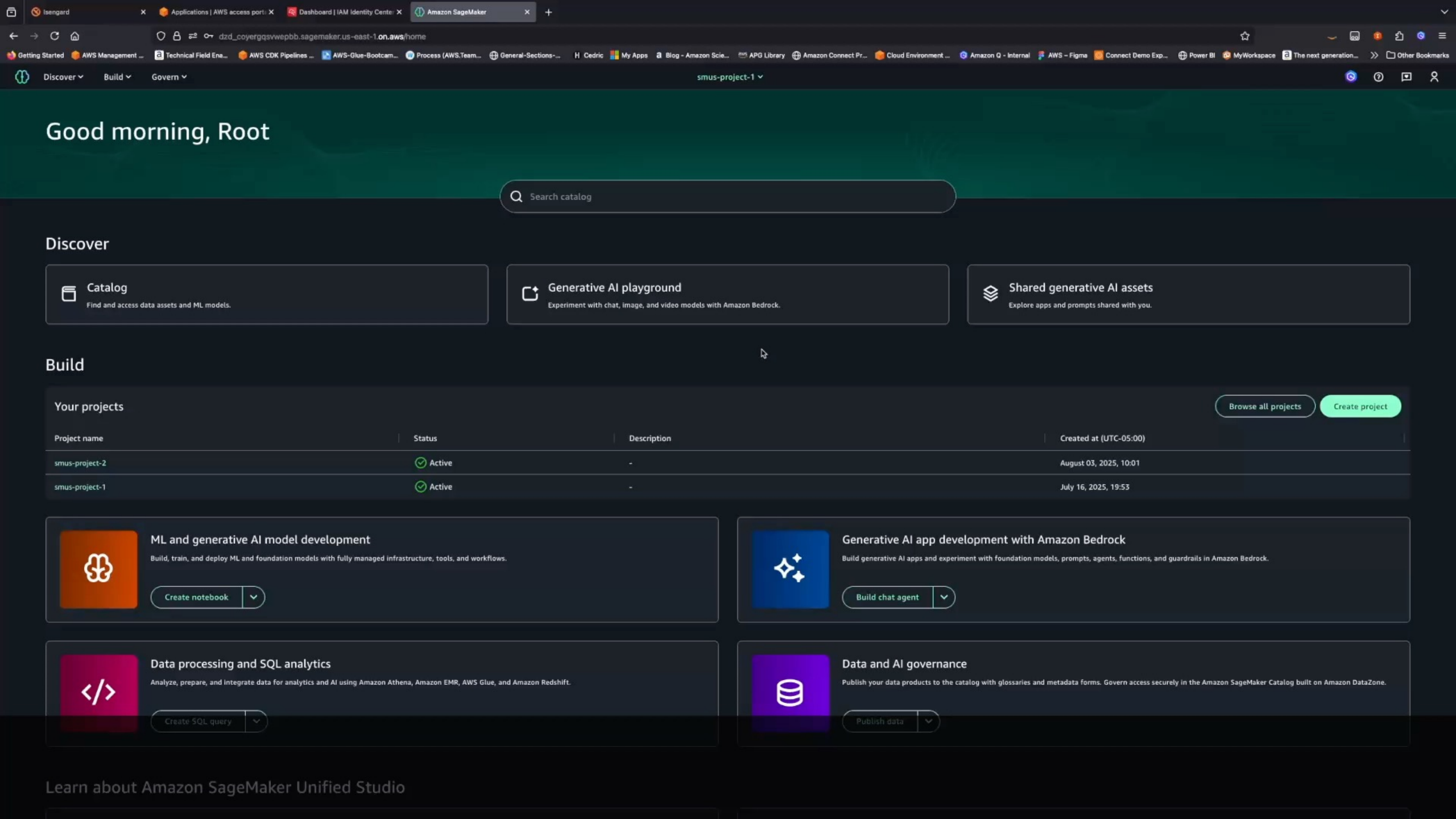
Help others derive meaning from data to reach conclusions

Produce cohesive, powerful and insightful narratives

Control verbosity and customize narrative text and apply visual themes

What-if analysis with Scenarios

Find answers to complex problems through guided step-by-step for an in-depth analysis



Good morning, Root

Search catalog

Discover

Catalog
Find and access data assets and ML models.

Generative AI playground
Experiment with chat, image, and video models with Amazon Bedrock.

Shared generative AI assets
Explore apps and prompts shared with you.

Build

Your projects Browse all projects [Create project](#)

Project name	Status	Description	Created at (UTC-05:00)
smus-project-2	Active	-	August 03, 2025, 10:01
smus-project-1	Active	-	July 16, 2025, 19:53

ML and generative AI model development
Build, train, and deploy ML and foundation models with fully managed infrastructure, tools, and workflows.
[Create notebook](#)

Generative AI app development with Amazon Bedrock
Build generative AI apps and experiment with foundation models, prompts, agents, functions, and guardrails in Amazon Bedrock.
[Build chat agent](#)

Data processing and SQL analytics
Analyze, prepare, and integrate data for analytics and AI using Amazon Athena, Amazon EMR, AWS Glue, and Amazon Redshift.
[Create SQL query](#)

Data and AI governance
Publish your data products to the catalog with glossaries and metadata forms. Govern access securely in the Amazon SageMaker Catalog built on Amazon DataZone.
[Publish data](#)

[Learn about Amazon SageMaker Unified Studio](#)

Discussion

What BI tools have you used?

Demo – Putting it all together

ENHANCE STUDENT RETENTION WITH AMAZON SAGEMAKER UNIFIED STUDIO

This demo showcases how to tackle a critical education sector challenge - ***Student Retention*** - by building a comprehensive data-driven solution using AWS services. It gives participants hands-on experience stepping into the roles of a cross-functional team at Octank University, a fictitious institution facing increasing dropout rates. Through a series of guided exercises, you'll learn how to leverage Amazon SageMaker, data analytics, machine learning, and generative AI to identify at-risk students and create personalized intervention strategies.





SageMaker Unified Studio for Student Engagement

This interactive demo shows how to use SageMaker to identify at-risk students and create personalized engagement plans.

Next

Made with  Storylane

Key Takeaways

- **AI is transforming the data landscape** - 89% of CDOs are moving forward with Generative AI, though 52% view their data foundation as not ready for AI implementation. AI and GenAI technologies can lower these barriers by simplifying data preparation and integration.
- **Amazon SageMaker Unified Studio** - provides an integrated environment where both traditional ML and GenAI capabilities work together across the data stack, combining SQL analytics, data processing, model development, and GenAI application building to reduce implementation barriers.
- **Organizations can overcome key data challenges** - by strategically applying AI and GenAI across their data architecture - Addressing data silos, simplifying analytics integration, enhancing governance, and enabling operational excellence with less technical overhead.



Thank you!

Ignatius Narchetty

Solutions Architect, EDU
inarchet@amazon.com

Please complete the survey
for this session



Track: Data & analytics track
Session: AI/ML For Data and Analytics