

AWS State, Local, and Education Learning Days

Salt Lake City, Utah

11:30am – 12:30pm

200
level

Amazon SageMaker
Unified Studio

Transforming Public
Sector Analytics and
AI/ML workloads with
a unified IDE

aws Learning Days
State, Local, and Education

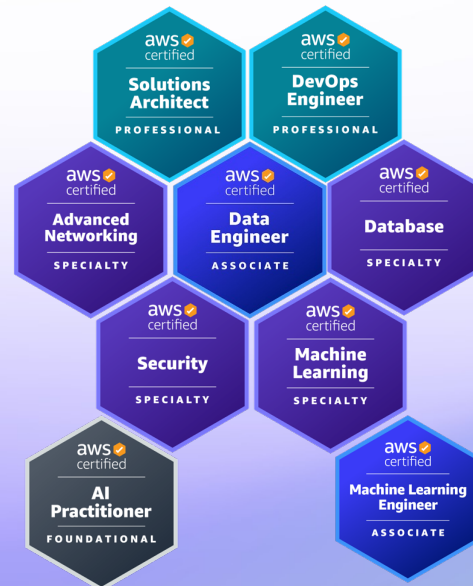


Transforming Public Sector Analytics and AI/ML workloads

with Amazon SageMaker Unified Studio

Aditya Challa (Adi)

Sr. Solutions Architect
Amazon Web Services
aditchal@amazon.com



Integrating data,
analytics, and
AI at scale



Optimizing current
investments for
higher value



What's holding organizations back

Establishing end
to end governance





FREEDOM TO INVENT

Innovate without constraints

Break free from silos and vendor lock-in with best-in-class data foundations from AWS, trusted by millions

19+

years of
innovation



Millions

customers innovating
with data and AI



100PB+

daily data
footprint



The background features a light blue gradient with several circular icons representing different AWS services. The icons are arranged in a scattered pattern around the central text. Some icons are in teal circles, while others are in blue circles. The icons include symbols for data analysis, machine learning, storage, and networking.

How might we make this even
better for our customers?

THE NEXT GENERATION OF



Amazon SageMaker

The center for all your data, analytics, and AI

Unified Studio

SQL analytics

Amazon Redshift
Amazon Athena

Data processing

Amazon EMR
AWS Glue

Model development

Amazon SageMaker AI

Gen AI App development

Amazon Bedrock

Streaming

Amazon MSK
Amazon Kinesis

Business intelligence

Amazon QuickSight

Search analytics

Amazon OpenSearch Service

COMING SOON

AVAILABLE NOW

COMING SOON

Data & AI Governance

Open Lakehouse

Unified Studio

SQL analytics	Data processing	Model development	Gen AI App development	Streaming	Business intelligence	Search analytics
Amazon Redshift Amazon Athena	Amazon EMR AWS Glue	Amazon SageMaker AI	Amazon Bedrock	Amazon MSK Amazon Kinesis	Amazon QuickSight	Amazon OpenSearch Service
				COMING SOON	AVAILABLE NOW	COMING SOON

Data & AI Governance

Open Lakehouse

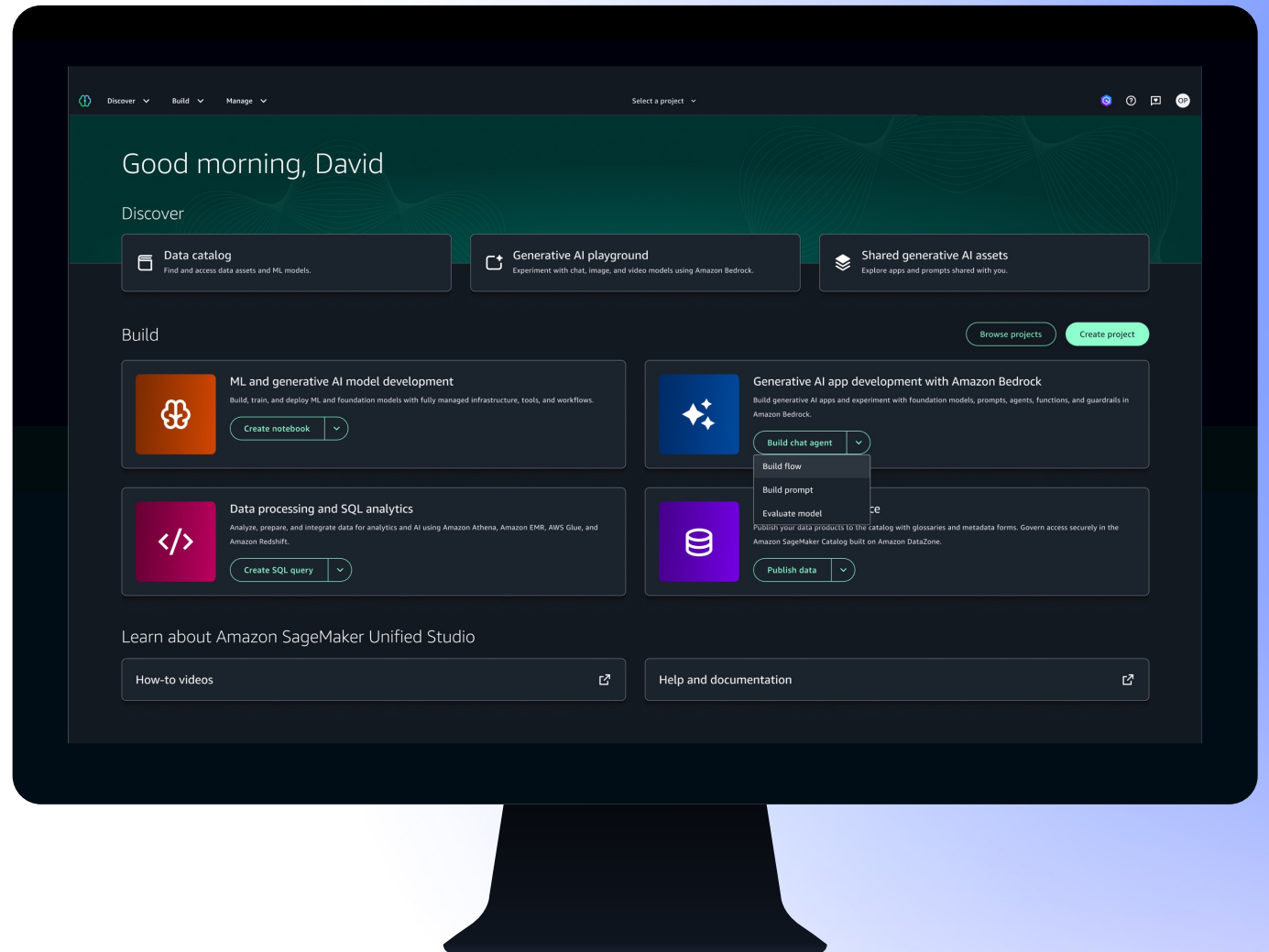
A single data and AI development environment

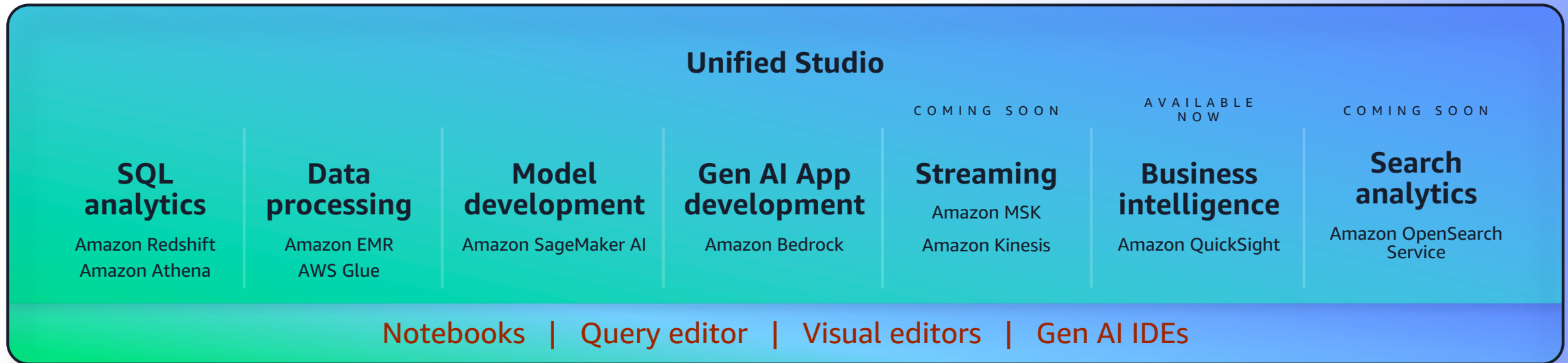
Discover your data and AI assets

Build with a complete toolset for analytics and AI

Collaborate in shared projects

Govern access securely





YOUR DATA AND AI DEVELOPMENT ACCELERATED BY

Amazon Q Developer

the most capable gen AI assistant in software development

Ask questions

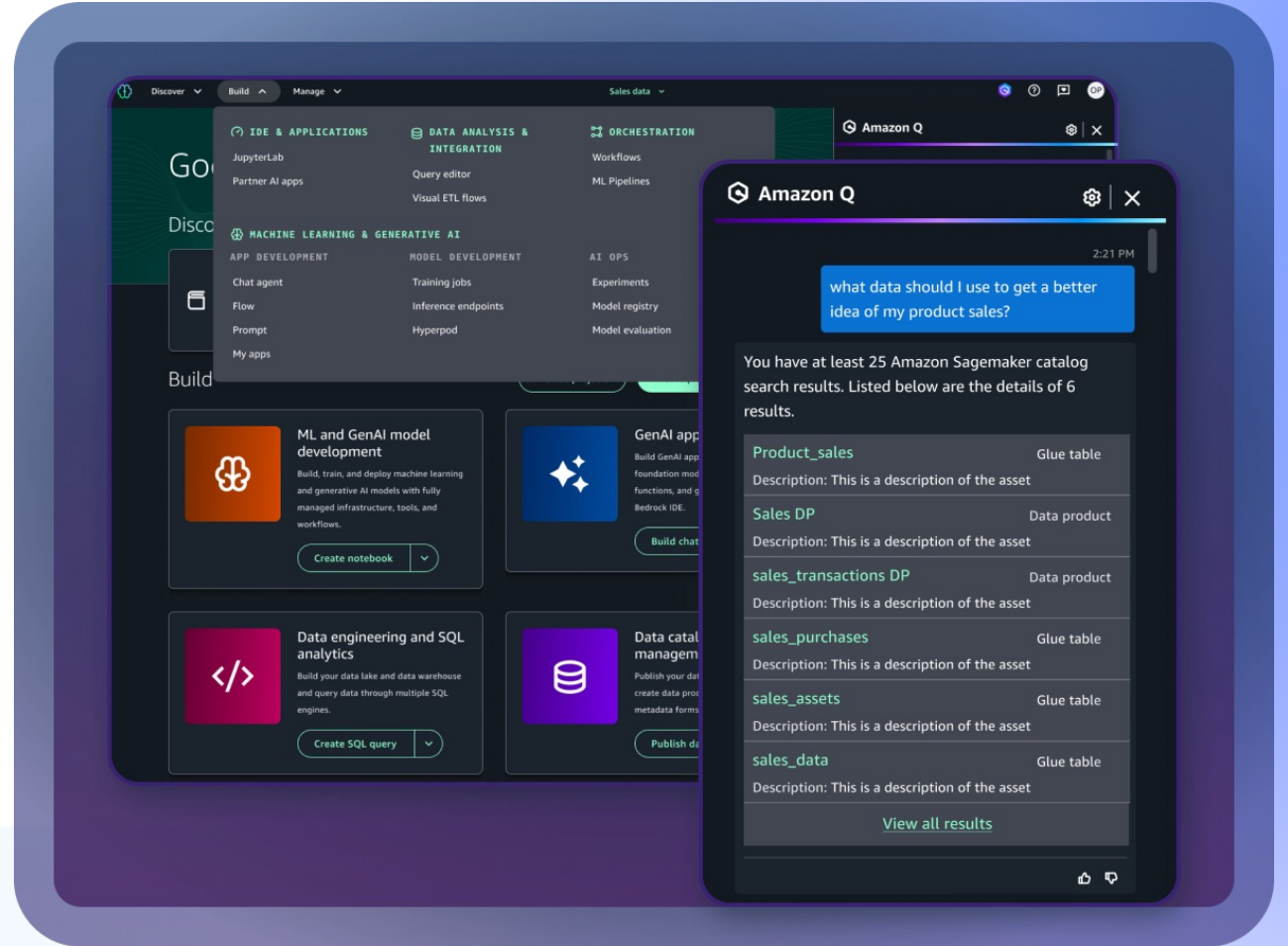
Get answers

Search and discover data

Generate code

Author queries

All through natural language



“ At Carrier, the next generation of Amazon SageMaker is transforming our enterprise data strategy by streamlining how we build and scale data products. SageMaker Unified Studio’s approach to data discovery, processing, and model development has significantly accelerated our lakehouse implementation.”

JUSTIN MCDOWELL
DIRECTOR, DATA PLATFORM & DATA ENGINEERING
CARRIER

Carrier

True business impact comes
from getting a **full view of your data**

Unified Studio

SQL analytics

Amazon Redshift
Amazon Athena

Data processing

Amazon EMR
AWS Glue

Model development

Amazon SageMaker AI

Gen AI App development

Amazon Bedrock

Streaming

COMING SOON
Amazon MSK
Amazon Kinesis

Business intelligence

COMING SOON
Amazon QuickSight

Search analytics

COMING SOON
Amazon OpenSearch Service

Data & AI Governance

Open Lakehouse



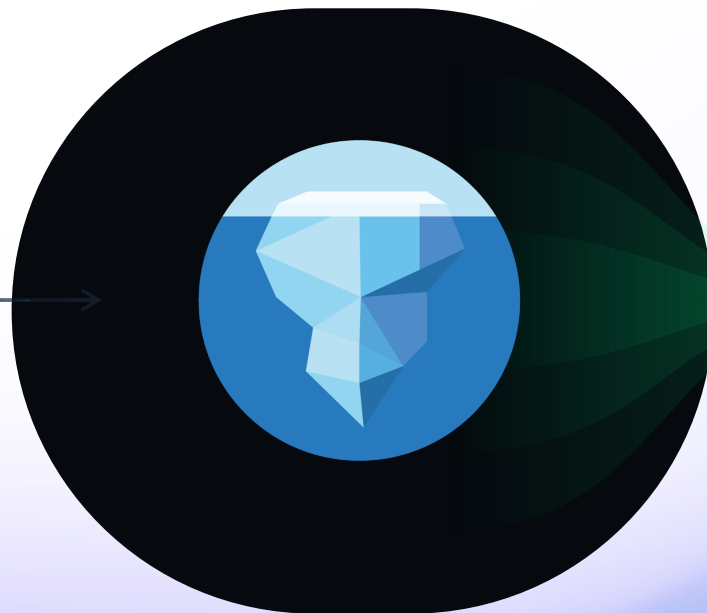
The best place
to run data lakes

Open access to most data sources

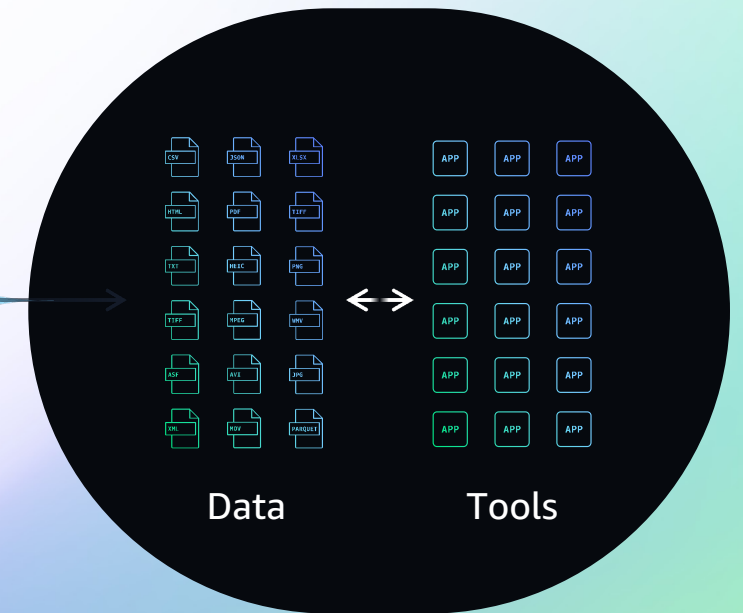
OPEN TABLE FORMATS



EMERGING STANDARD



INTEROPERABILITY





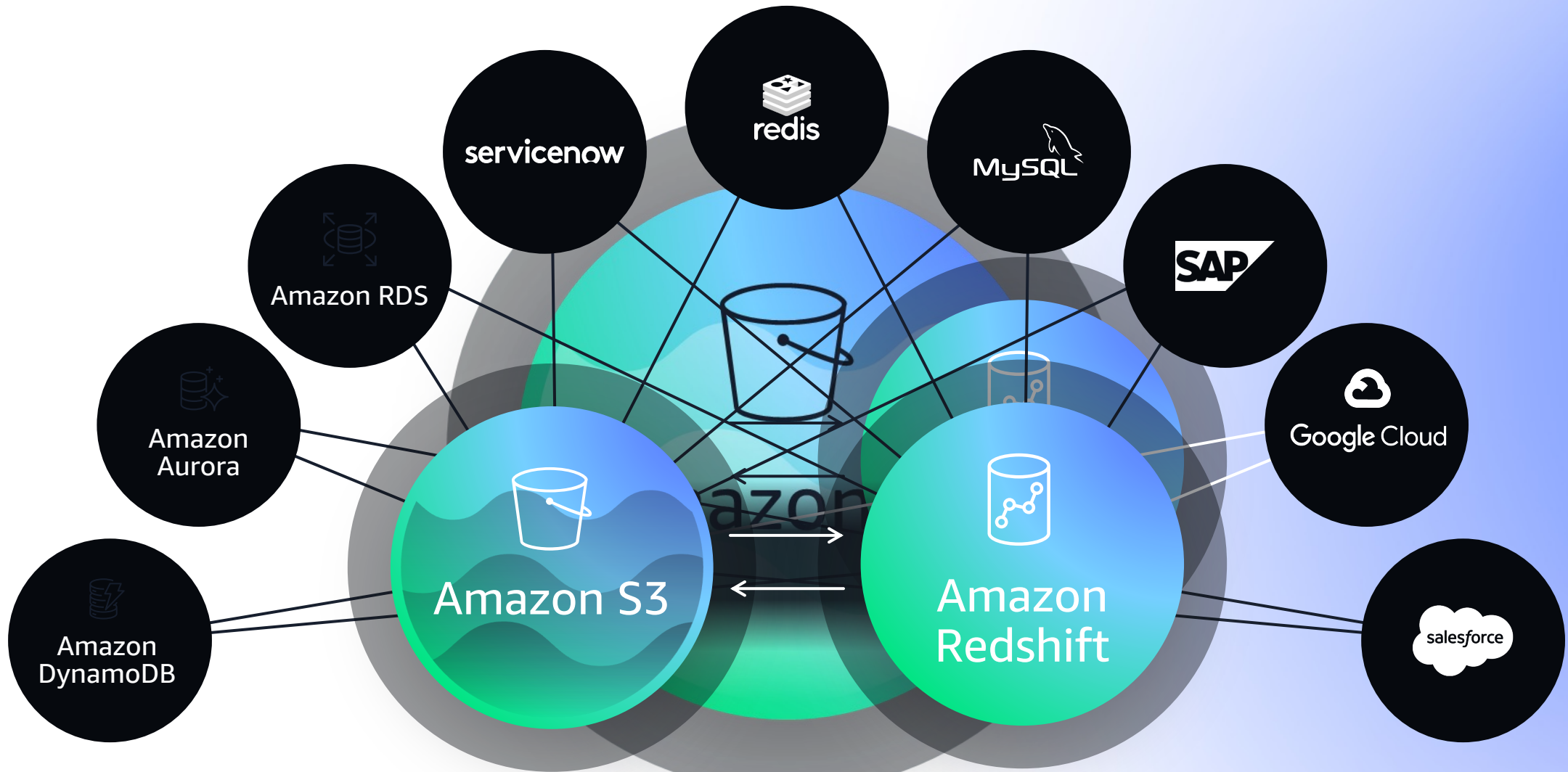
S3 Tables



```
{  
  "Effect": "Allow",  
  "Action": [  
    "s3tables:GetTableData",  
    "s3tables:PutTableData"  
  ]  
}
```

Fully managed Apache Iceberg tables in S3





The best place to run data lakes
 The best place to run data lakes
 petabyte-scale data warehouse

Unified Studio

Data & AI Governance

Read and write from your preferred Apache Iceberg compatible compute engines

APACHE ICEBERG OPEN API

Amazon S3
data lakes



Amazon Redshift
data warehouses



Fine-grained access control



Zero-ETL integrations

Aurora

DynamoDB

RDS

Streaming data – MSK, Kinesis

OpenSearch vector data

ServiceNow

Salesforce

Zoho CRM

Instagram Ads

SAP

Salesforce Pardot

Facebook Ads

Zendesk

Federated querying + 100s of AWS Glue connectors





“The new Amazon SageMaker Lakehouse support for zero-ETL integrations from applications will be a game-changer for us. With zero-ETL, we anticipate being able to synchronize over 3,000 Iceberg tables containing customer, digital, clinical trial, and market access data effortlessly, significantly simplifying our data integration workflow. We anticipate that this new feature will reduce our data integration time by 70% and cut related costs by 50%.”

YANNICK MISTELI
HEAD OF ENGINEERING,
GLOBAL PRODUCT STRATEGY AT ROCHE



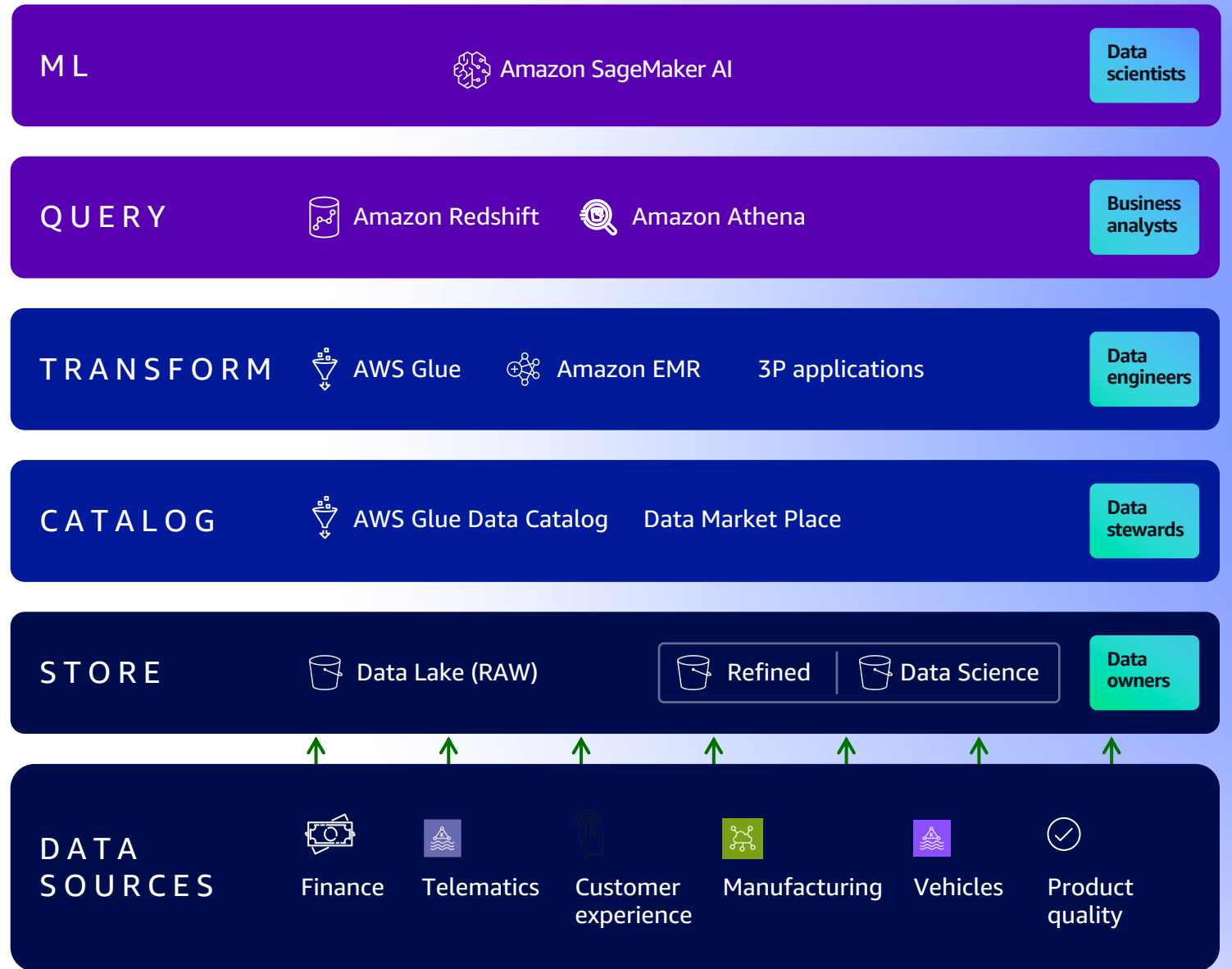


VALUE

Scale without limits, but stay
within your budget



Our customers have invested in varied data sources and tools



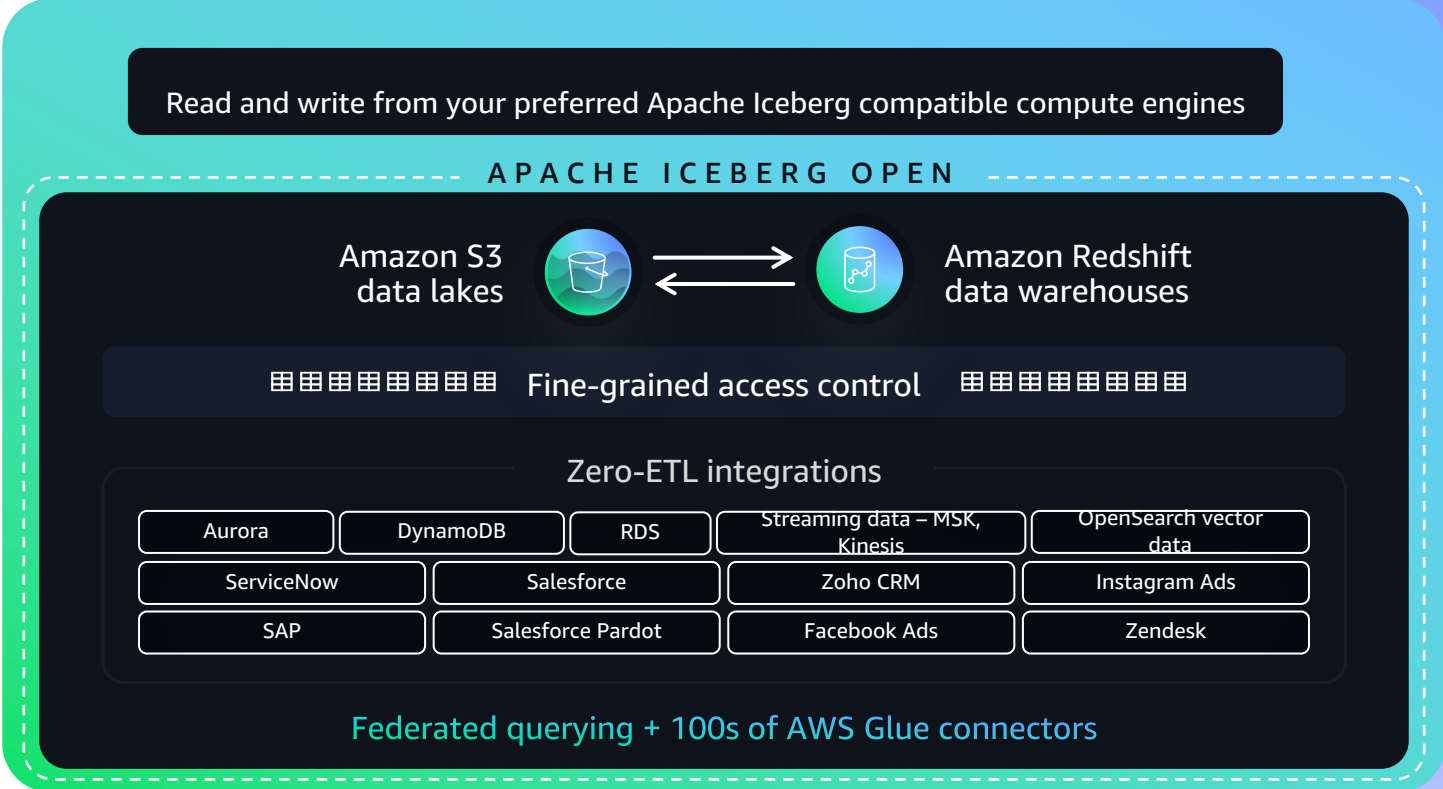
Future-proofing your investment in data

SageMaker Lakehouse

No matter where your data lives
No matter which tools you prefer

Automatically bring data to
the point of analytics and AI with
zero-ETL capabilities.

Query in place from your
data store with federated queries





TRUSTED DATA
FOR TRUSTED AI

**AWS provides built-in
governance across the data
and AI lifecycle to securely
transform raw data into AI-
ready assets**



Data sources

Structured

Unstructured

Data prep
and enrichment

Cleaning | Enriching | Generating | Business
context | Transformation | Aggregation







Data sources

Data prep
and enrichment







Metadata

- Find
- Share
- Understand
- Access
- Act
- Monitor

-  Gen AI / RAG
-  Classifications
-  Forecasting
-  SQL/Spark analytics

Analytics and AI

-  Recommendations
-  Business decision intelligence
-  Graph analytics
-  Anomaly detection



Unified Studio

SQL analytics

Amazon Redshift
Amazon Athena

Data processing

Amazon EMR
AWS Glue

Model development

Amazon SageMaker AI

Gen AI App development

Amazon Bedrock

Streaming

COMING SOON
Amazon MSK
Amazon Kinesis

Business intelligence

COMING SOON
Amazon QuickSight

Search analytics

COMING SOON
Amazon OpenSearch Service

Data & AI Governance

Open Lakehouse

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



“ To address siloed data sets spread across our automotive operations, we are implementing Amazon SageMaker to **unify and govern data** across our connected car, sales, manufacturing, and supply chain units. This approach allows us to search, discover, and share data effortlessly, laying the groundwork to pre-empt quality issues, increase customer satisfaction, and accelerate development of generative AI applications.”

KAMAL DISTELL

VP OF DATA, ANALYTICS, PLATFORMS,
AND DATA SCIENCE, TMNA

TOYOTA MOTOR NORTH AMERICA



Common Public Sector Use Cases

Data Science Professor

I want to create reproducible classroom AI labs for students using notebooks and curated datasets.



Data Analyst

I want to load data from multiple departments into a Lakehouse and analyze with GenAI generated dashboards.



Bioinformatics Researcher

I want to publish curated and annotated genomics datasets for others to discover and apply AI.



Data Steward

I want to build a data mesh by publishing curated datasets with GenAI-generated descriptions and tags.



Clinical Data Scientist

I want to unify EPIC health data into a lakehouse and train predictive models in JupyterLab notebooks.



Data Strategist

I want to integrate DMV service records and use GenAI to summarize trends and improve staffing decisions.



SageMaker Unified Studio for ML

Data

Data Scientists can **easily discover datasets** in the SM Catalog with admin governance

Collaboration

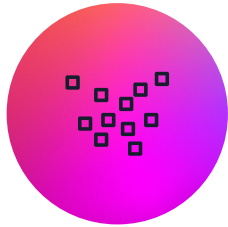
Projects support **team collaboration** as well as resource isolation

Unified

One platform for end-to-end **data + ML tooling for different personas**



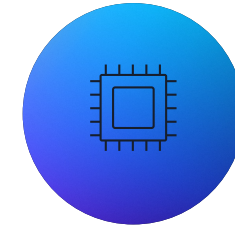
Easily govern your entire data and ML stack



One managed platform for different data and ML users, linked with your Identity Provider



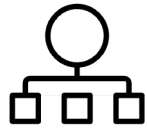
Domain, project, and user level governance across the best-in-class services



No licenses - only pay for the services you use

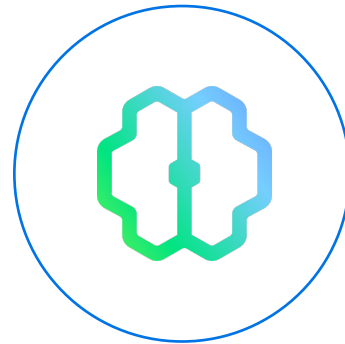
SageMaker Unified Studio for ML

SageMaker Unified Studio



ML platform owners:

1. **Easy infrastructure set up** through pre-defined IaC blueprints
2. **Streamlined project creation** and permission management
3. Enable users **to easily publish, discover, and subscribe to data assets**



ML Engineers, Data Scientists:

1. **Simplified Data Access** and Exploration
2. **Flexibility of compute** at the notebook cell level
3. **One platform** for datasets, FMs, fine-tuning and GenAI deployment



ML Use Cases

Fine-Tune an LLM

Data Processing for Data Scientists

Automated Evaluation, Deployment,
and Monitoring

ML Platform Administration

Build GenAI Applications

SageMaker Studio

Attached Fsx for Lustre, managed MLflow, managed oss VSCode* and JupyterLab

End to End Utilization

Run IDEs, Training or Inference on Hyperpod *(IDEs on roadmap)*

Task Governance and Metrics

Increase compute resource utilization and view utilization

Hyperpod

Orchestration

EKS or Slurm

Clusters

Resilient Environment: automated health checks, auto-repair



Bring structured and unstructured data together for GenAI

GEN AI APPLICATIONS



Amazon Bedrock Knowledge Bases



Amazon SageMaker



Your data is your competitive advantage. It should stay that way.



Security



Resilience



Governance

Well Architected guidance

Recovery

Replication

Audits

Data classification

Incidence response

Multi-az configurations

Encryption at rest and in transit

Object versioning

Row level and column level security

IAM policies

143 security certifications

No sharing customer data trained for LLMs with model providers

Backup and recovery

Data protection and resilience



AWS is there with you, no matter where you are in your journey

Receive peer-level executive guidance



AWS Data Strategy

Jumpstart your data and AI flywheel



AWS Data-Driven Everything

Modernize your data foundation



AWS re:Imagine Data

Guide generative AI use cases



AWS Generative AI Innovation Center

Assess your organizational maturity



AWS Data Strategy Diagnostic

Transform your business



AWS Partner

Additional resources



[An integrated experience](#)
with Amazon SageMaker
Unified Studio



[Foundational Blocks](#)
of Amazon SageMaker
Unified Studio



[Generative AI](#)
using Amazon SageMaker
Unified Studio



Thank you!

Aditya Challa

Sr. Solutions Architect
AWS

aditchal@amazon.com



Please complete the survey
for this session



Track : Artificial Intelligence and Machine Learning
Session : Transforming public sector analytics and AI/ML workloads with SageMaker Unified Studio