

AWS State, Local, and Education Learning Days

Washington, DC



AI/ML For Data and Analytics

Hiren Deliwala (he/him)

SA Manager, SLG-K12
hdaws@amazon.com

Mahinder Chawla (he/him)

Sr. Solutions Architect, SLG
mahindc@amazon.com

Challenges we're hearing



Demand is rising, while resources and capacity to deliver them aren't keeping pace



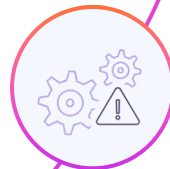
Expectations are rising, with constituents demanding the same digital experiences they get from the private sector



Infrastructure is aging, creating friction across the data lifecycle (capture, storage, management, leveraging)

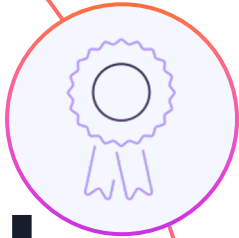


Requirements are becoming more complex, creating barriers to change and blocking adoption of "built for change" solutions



Change is slowing, with risk-averse cultures introducing inertia and thwarting innovation

What internal and external customers are asking for



Better experiences that minimize time to access, provide self-service capabilities, and deliver in joined-up ways

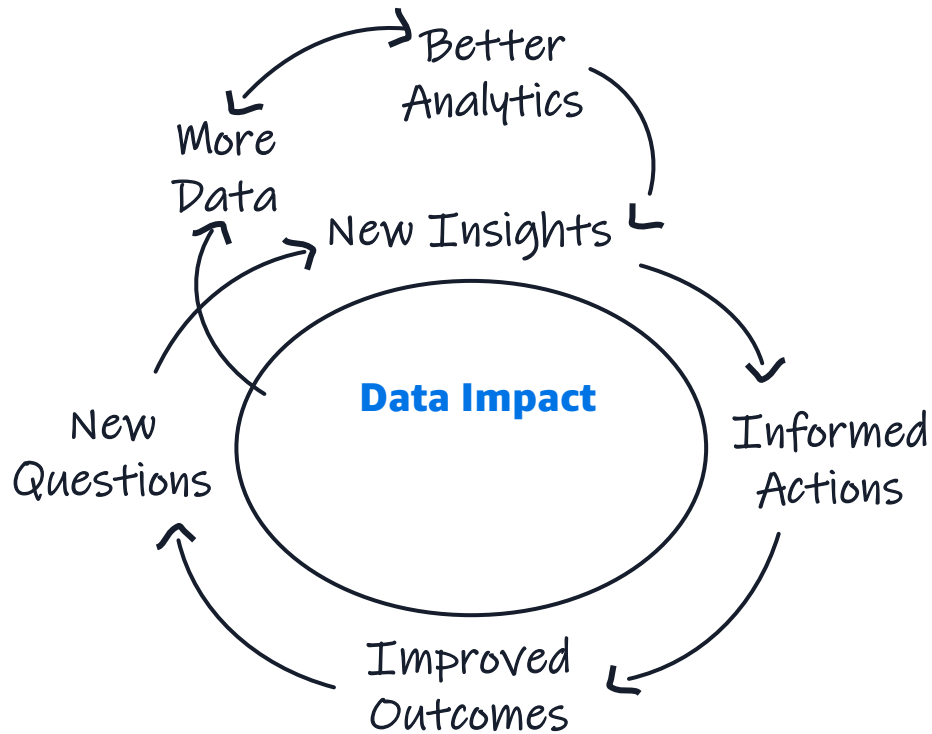


More productive staff who have answers at their fingertips and aren't stuck doing manual work



Empowered teams that focus on outcomes, use make data-informed decisions, and meet legal and policy requirements

The Data Driven Organization



Key Characteristics



Set 'Think Big' goals



Focus on delivering policy priorities with quality



Shared leadership conviction and Business-IT alignment on data ownership



Strong collaboration and agility concerning data products across data producers and consumers



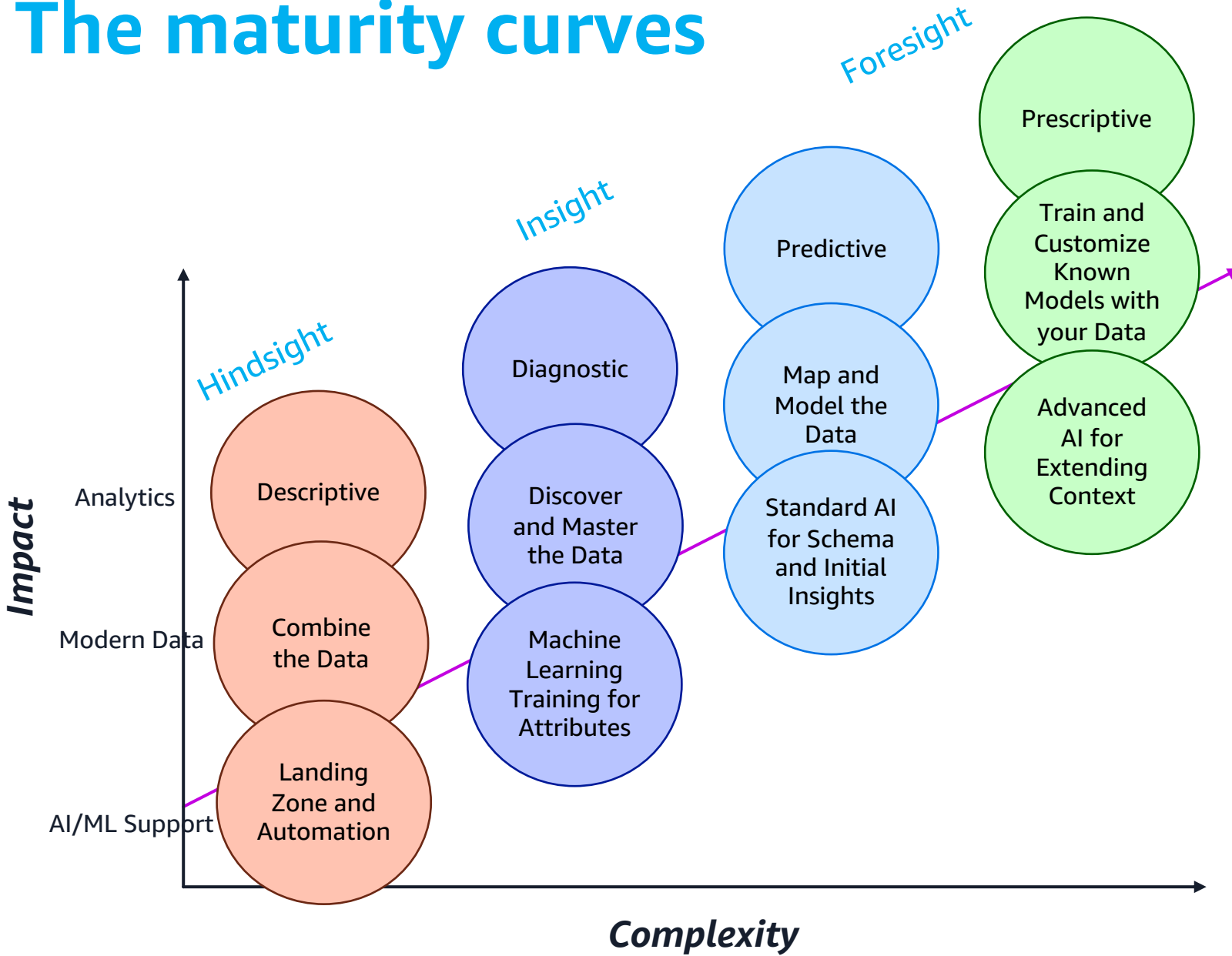
Upskilled and empowered producers and consumers who self-serve



Privacy, security, compliance and federated governance without impeding innovation

“An organization that harnesses data as an **asset drive sustained innovation** and create **actionable insights** to **improve policy making decisions** that reflect outcomes constituents care about, that builds greater **trust**.”

The maturity curves



DESCRIPTIVE

What happened?

DIAGNOSTIC

Why did it happen?

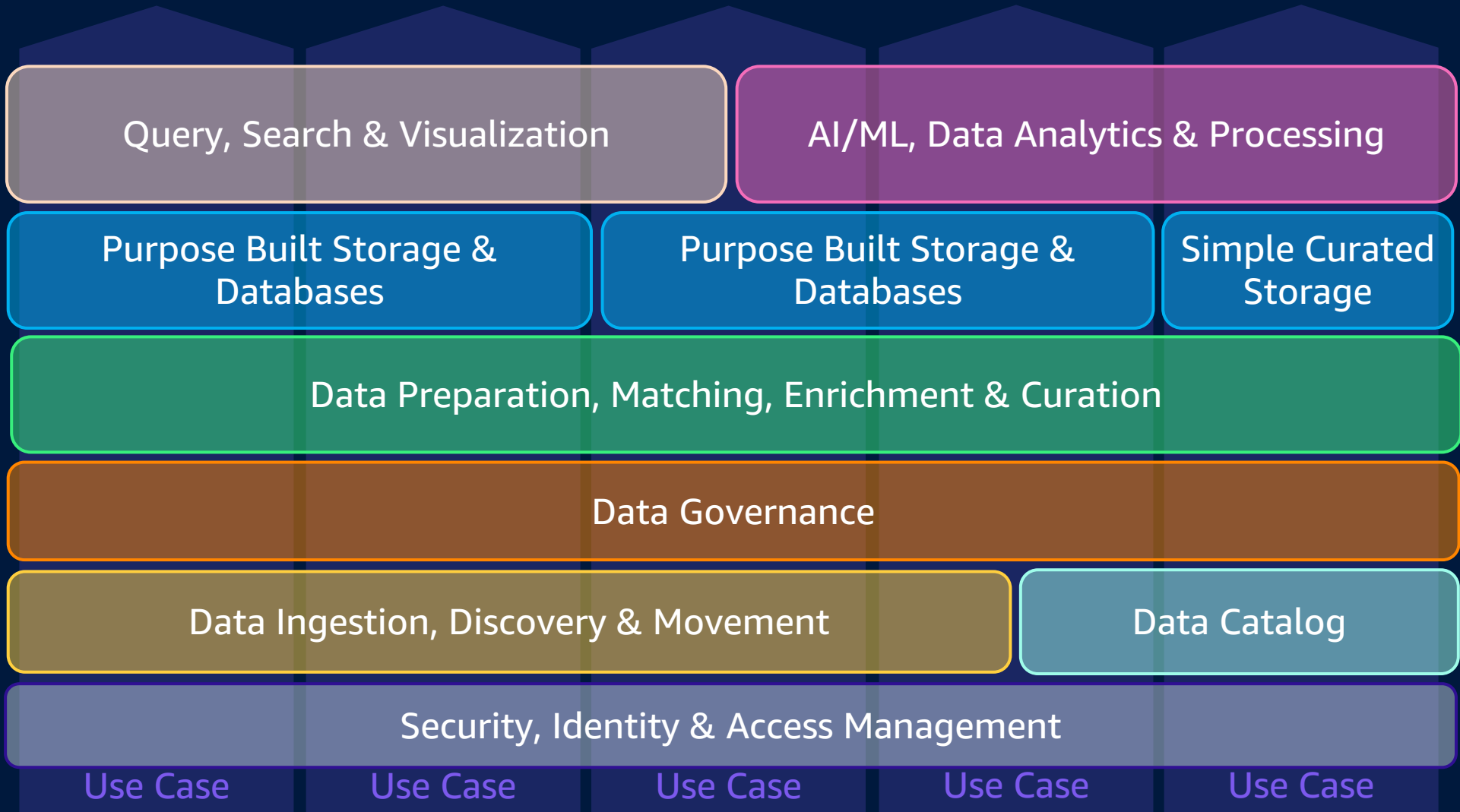
PREDICTIVE

What will happen?

PRESCRIPTIVE

How do we make it happen?

A Modern, Foundational Data Architecture



DevOps /
DataOps

External
Data
Sources &
Products

Build a Modern, Foundational Data Architecture

Built in AI/ML Capabilities

AI/ML enabled/integrating services

Query, Search & Visualization



Athena
Interactive analytics



QuickSight
Visualizations



OpenSearch
Operational Analytics



Redshift
Data warehousing

AI/ML, Data Matching & Processing



SageMaker
ML



Comprehend
NLP



Transcribe
Speech-to-text



Textract
Extract text



Health Lake
Health Data Lake



Comprehend Medical
NLP



Translate
Translation

+ more

Specialized Partner Solutions

Data Preparation, Enrichment & Analysis



Glue
Serverless Data Integration



EMR
Hadoop + Spark



Glue DataBrew
Data Preparation & Transformation



Forecast
Forecasts



Kinesis Data Analytics
Real time

Data Ingestion, Integration & Movement



Glue
Serverless Data Integration



Kinesis Data Streams
Data Streaming



Data Exchange
Data exchange



Clean Rooms



API Gateway
API Management



Kinesis Data Firehose
Streaming ETL



Data Migration Service



AppFlow



Data Pipeline



DataSync

Data Storage & Databases



S3/Glacier



RDS
MySQL, PostgreSQL, MariaDB, Oracle, SQL Server,



Keyspaces (For Apache Cassandra)
Wide column



Neptune
Graph



Aurora
MySQL, PostgreSQL



Redshift
Data warehousing



DocumentDB
Document



Timestream
Time Series

Data Catalog



Glue
Data Catalog



Lake Formation
Data Lakes



DataZone
Governed Analytics

Data Governance

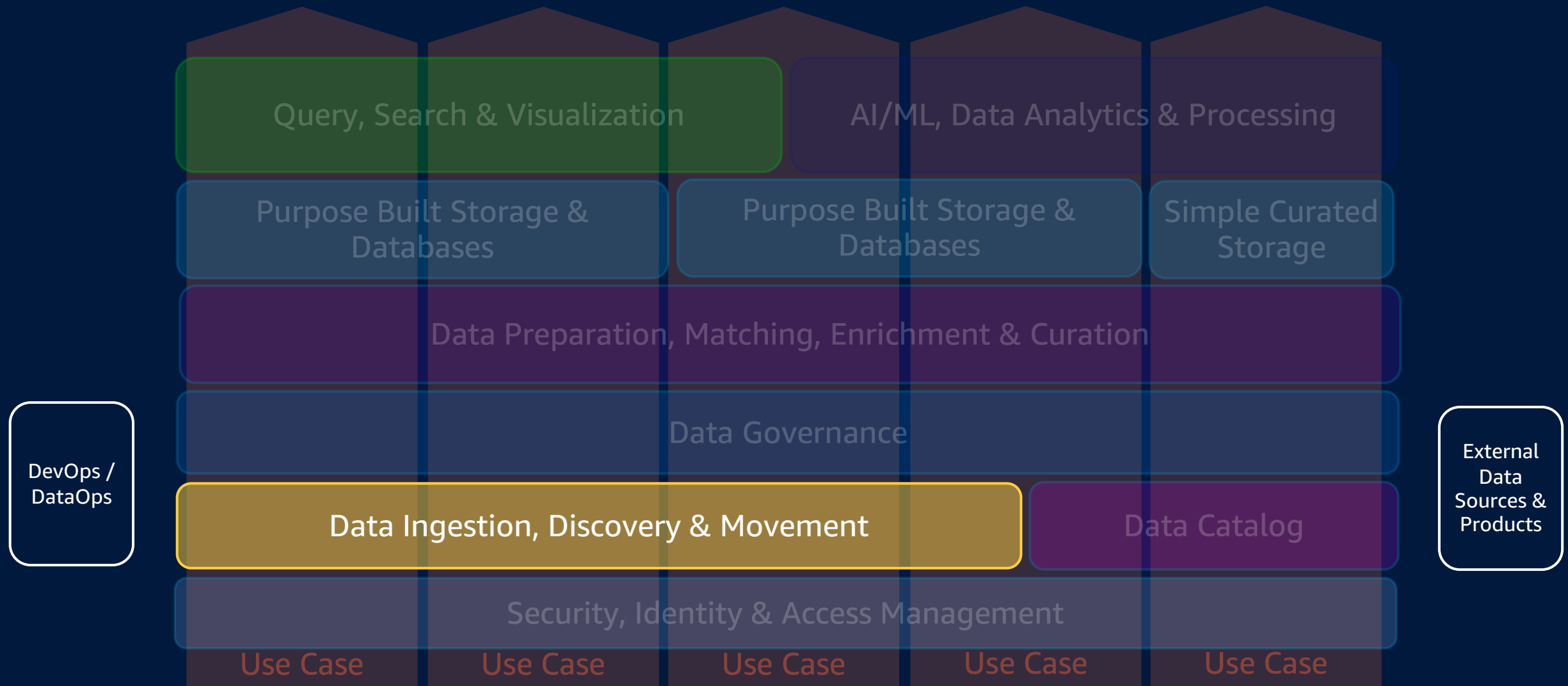


Machine learning

- Data Discovery
- Data Attribution
- Metadata Tags
- Schema Definitions
- Data Catalogues
- Data Protection



A Modern, Foundational Data Architecture



Content analysis and object detection

Extract insights and identify objects of interest from large volumes of 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



Text to Speech Transcription, Classification & Analytics

AN EXAMPLE FROM AWS POST CALL ANALYTICS AND CALL SUMMARIZATION

GenAI Transcript Summary

Summary
The customer asked about the cashback rate of their rewards card, and paid \$50 to the account balance.

AI Customer Sentiment
Positive 😊

AI Agent Sentiment
Neutral 😐

How could the overall experience be improved?
The agent could have not assumed the customer's pronouns, and also asked for their address.

Call Analytics Summary

Issue
I'm calling about my rewards card.

Action Items
No action items detected.

Outcomes
your card balance is \$75.34.

Transcript

- Agent - 00:03
Hello, thanks for calling Bank. Uh my name is, how can I help you?
- Customer - 00:07
Hi, [Issue]: I'm calling about my rewards card.
- Agent - 00:11
Excellent. Give me a second to look up your customer account for verification purposes. Can you please uh state your first name?
- Customer - 00:20
My first name is [PII]

Home > Call List > Call Details

Call Details Info

Swap Agent/Caller

Call Metadata

- Timestamp: 2023-02-22 12:23:49
- Guid: 102
- Agent: AndrewK
- Call Duration: 03:02
- Entity Recognizer Name: sample-entities.csv
- Language Model: en-US
- Agent Sentiment: Sentiment: 😊 Trend: ↗
- Customer Sentiment: Sentiment: 😊 Trend: ↗

Transcribe Details

- Type: Transcribe Call Analytics
- Job Id: Card2_GUID_102_AGENT_AndrewK_DT_2023-02-22T12-23-49.wav
- File Format: wav
- Sample Rate: 8000
- PII Redaction: Enabled
- Custom Vocabulary: -
- Vocabulary Filter: -
- Average Word Confidence: 97.9%

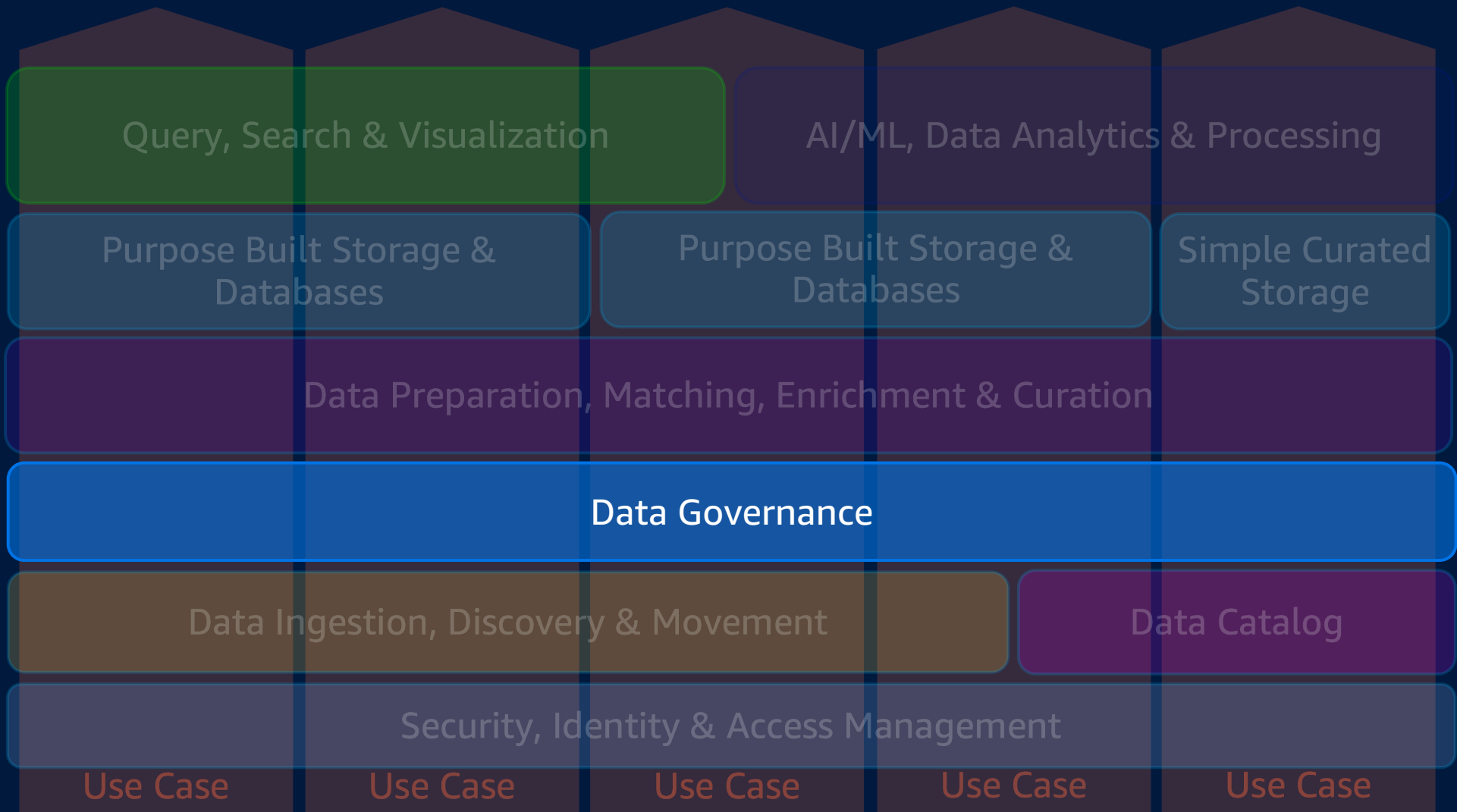
Sentiment

Call Quarter	Customer Score	Agent Score
Q1	0	4
Q2	1	2
Q3	3	3
Q4	4.5	4.5

Speaker Time

Speaker	Percentage
Agent	50.0%
Customer	50.0%
Silence	0.0%

A Modern, Foundational Data Architecture



DevOps /
DataOps

External
Data
Sources &
Products



Data Governance

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

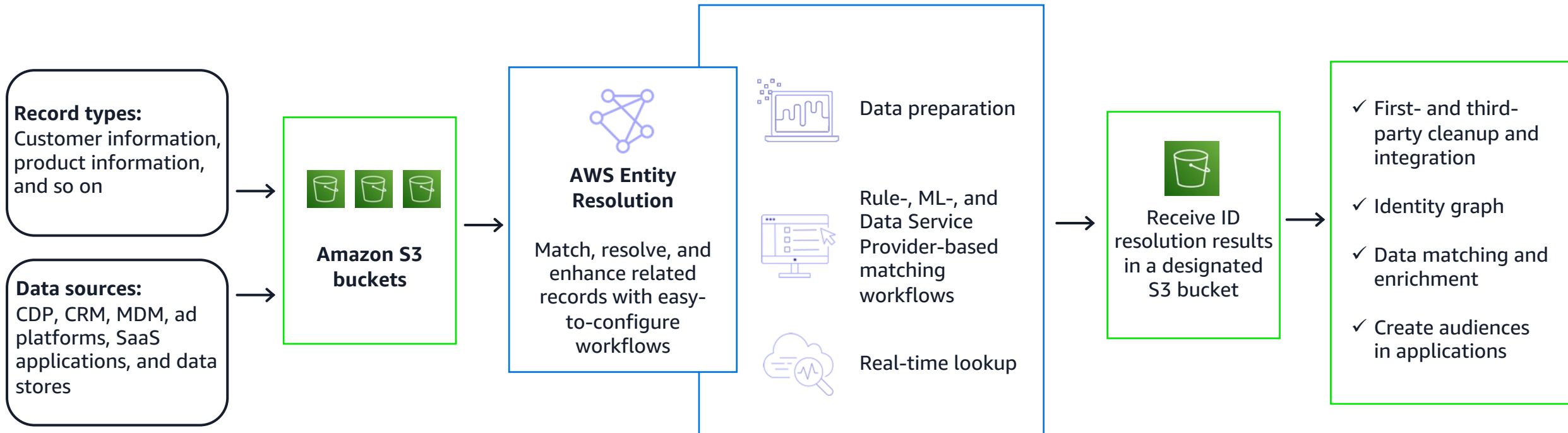
[AWS Glue Data Quality](#) learns patterns on data statistics gathered over time using ML algorithms. It detects anomalies, unusual data patterns and alerts users. It also auto-creates rules to monitor these specific patterns so that you can progressively build data quality rules.

Entity Resolution & Data Matching

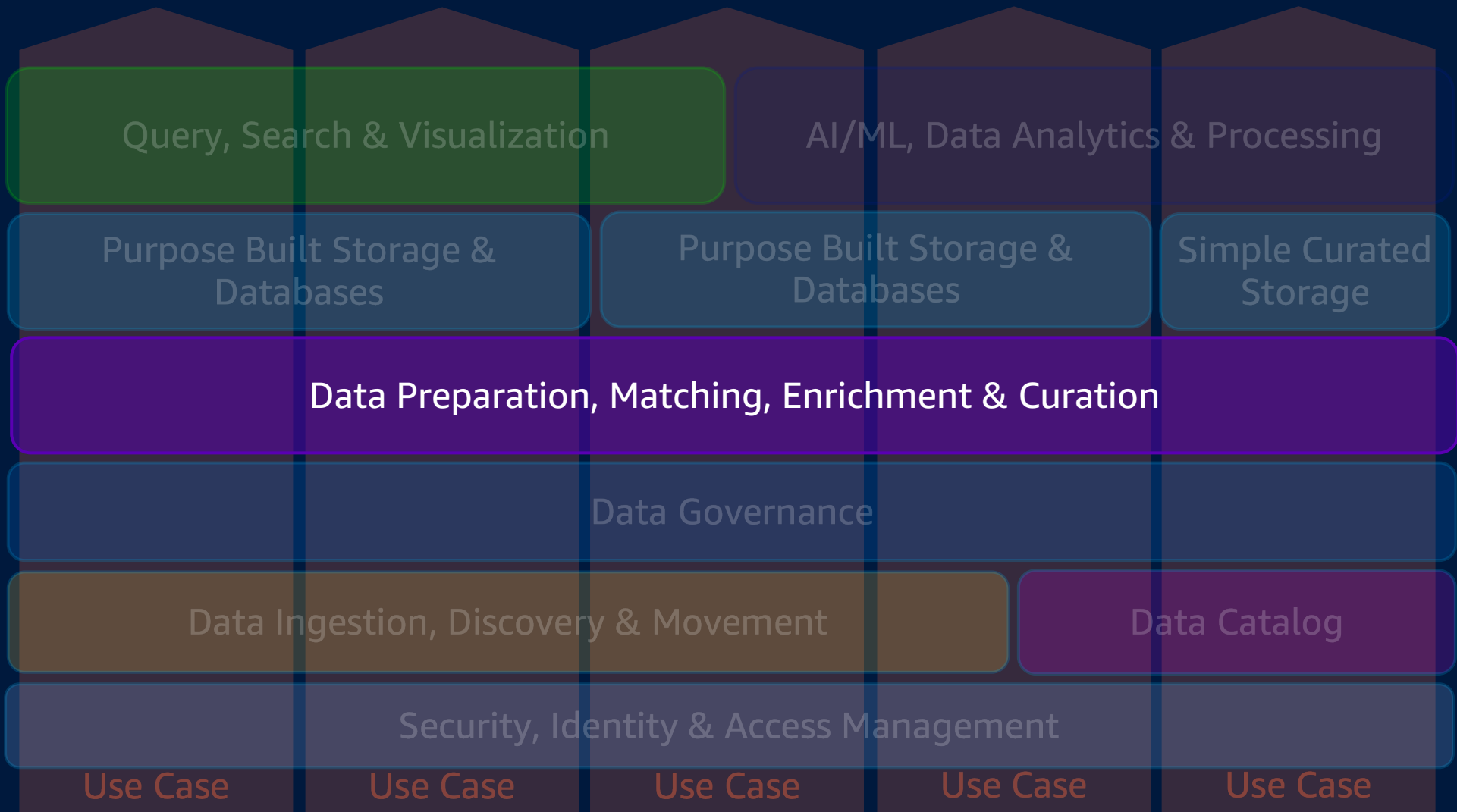
Match, link, and enhance related customer, product, business, or healthcare records stored across multiple applications, channels, and data stores

[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.

ML-powered Matching & Entity Resolution



A Modern, Foundational Data Architecture



Data Preparation, Matching, Enrichment & Curation

ML-based data transformations

ML-generated data insights & data augmentation

- Trends, anomalies, categorization, classification, labeling, sentiment, summarization, etc.

ML services for can be integrated into data processing jobs

- E.g., translation, inference, labeling, etc.

Generative AI powered data integration coding assistance

Generative AI assistance for data integration

The screenshot displays the AWS Glue Studio interface. The top navigation bar includes the AWS logo, 'Services', a search bar, and a keyboard shortcut '[Option+S]'. On the right, it shows the location 'Tokyo' and the user 'Username @ 1234-5678-9012'. The left sidebar contains navigation links for 'AWS Glue', 'Getting started', 'ETL jobs' (with sub-links for Visual ETL, Notebooks, and Job run monitoring), 'Data Catalog tables', 'Data connections', 'Workflows (orchestration)', 'Data Catalog', 'Data Integration and ETL', and 'Legacy pages'. At the bottom of the sidebar are toggle switches for 'Enable compact mode' and 'Enable new navigation'. The main content area is titled 'AWS Glue Studio' and features a 'Create job' section with three options: 'Visual ETL' (Author in a visual interface focused on data flow), 'Notebook' (Author using an interactive code notebook), and 'Script editor' (Author code with a script editor). Below this is an 'Example jobs' section with a 'Create example job' button. The 'Your jobs (2)' section includes a search bar with 'demo' and '2 matches', a refresh button, and 'Actions' and 'Run job' buttons. A table lists the jobs:

<input type="checkbox"/>	Job name	Type	Last modified	AWS Glue version
<input type="checkbox"/>	q-demo-taxi	Glue ETL	4/26/2024, 1:19:07 PM	4.0
<input type="checkbox"/>	q-demo	Glue ETL	4/25/2024, 3:41:38 PM	4.0

The bottom of the interface shows 'CloudShell' and 'Feedback' links on the left, and copyright information '© 2024, Amazon Web Services, Inc. or its affiliates.' along with 'Privacy', 'Terms', and 'Cookie preferences' on the right.



A Modern, Foundational Data Architecture



DevOps /
DataOps

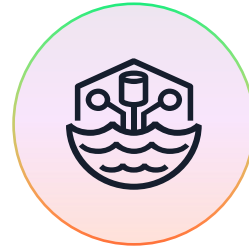
External
Data
Sources &
Products

AWS brings ML closer to data



Databases

+



Data warehouses
+ data lakes

+



Business
intelligence tools

ML Purpose-built databases

- Databases with in-built AI/ML support like vector stores (Amazon Aurora PostgreSQL-Compatible & RDS for PostgreSQL, Amazon Neptune ML, Amazon MemoryDB for Redis, Amazon Document DB, and Amazon OpenSearch)
- Integrate ML into queries using run-time inference inside SQL as done in Amazon Redshift ML and Athena ML

Inference at query time – Link Prediction

Without inference

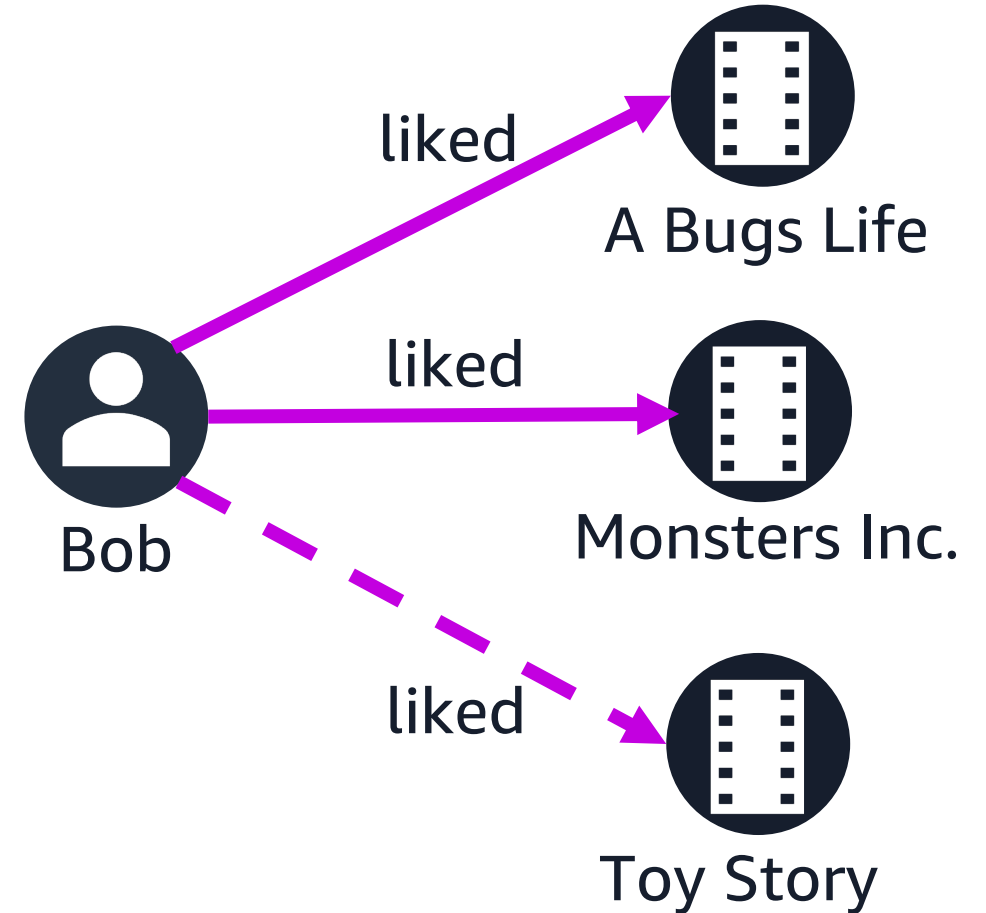
```
g.V().has('name', 'Bob').  
out('liked').  
hasLabel('movie').values('title')
```

- ⇒ A Bugs Life
- ⇒ Monsters Inc.

With inference

```
g.with("Neptune#ml.endpoint", "ENDPOINT").  
V().has('name', 'Bob').  
out('liked').with("Neptune#ml.prediction").  
hasLabel('movie').values('title')
```

- ⇒ Toy Story



Train & use models in Amazon Redshift ML

TRAIN



```
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



Amazon
SageMaker

PREDICT



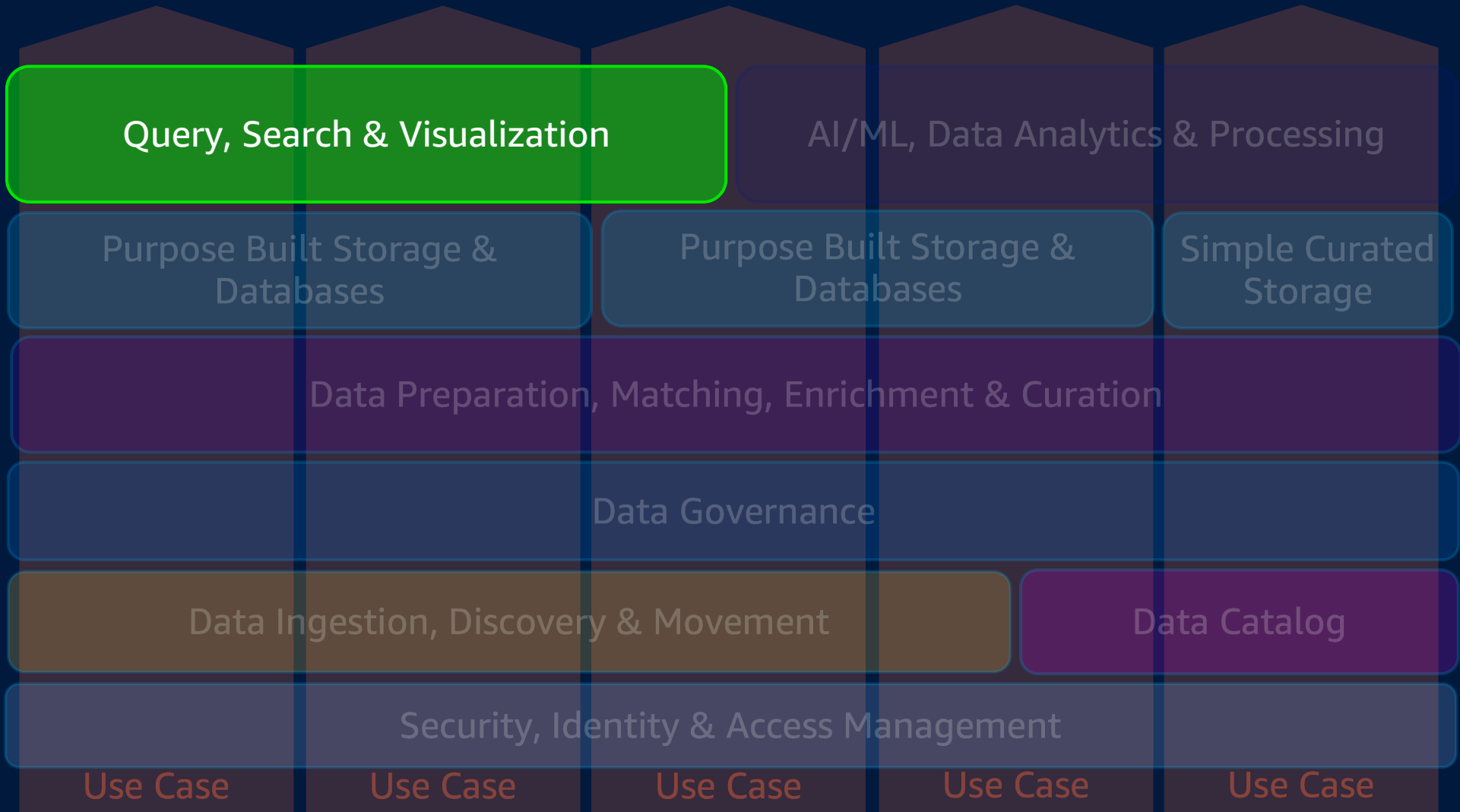
```
SELECT n.id, n.firstName, n.lastName,
predict_customer_churn(n.age, c.zip ...)
AS activity_prediction
FROM new_customers n
WHERE n.marital_status = 'single'
```



Amazon
Redshift

Example: Train a model on data from before 2020-01-01, then use the prediction function on the testing set. The following query displays the predictions of whether customers who signed up after 2020-01-01 will go through churn or not.

A Modern, Foundational Data Architecture



DevOps /
DataOps

External
Data
Sources &
Products

AI/ML in Query, Search & Visualization

- Natural Language insights from GenAI database query
 - text-to-sql
 - natural language search
 - Visualization generation
- Content & data summarization & classification
 - AI-assisted data story telling

AI-powered dashboard authoring experience

A NEW DASHBOARD BUILDING EXPERIENCE POWERED BY GENERATIVE BI

Build visuals

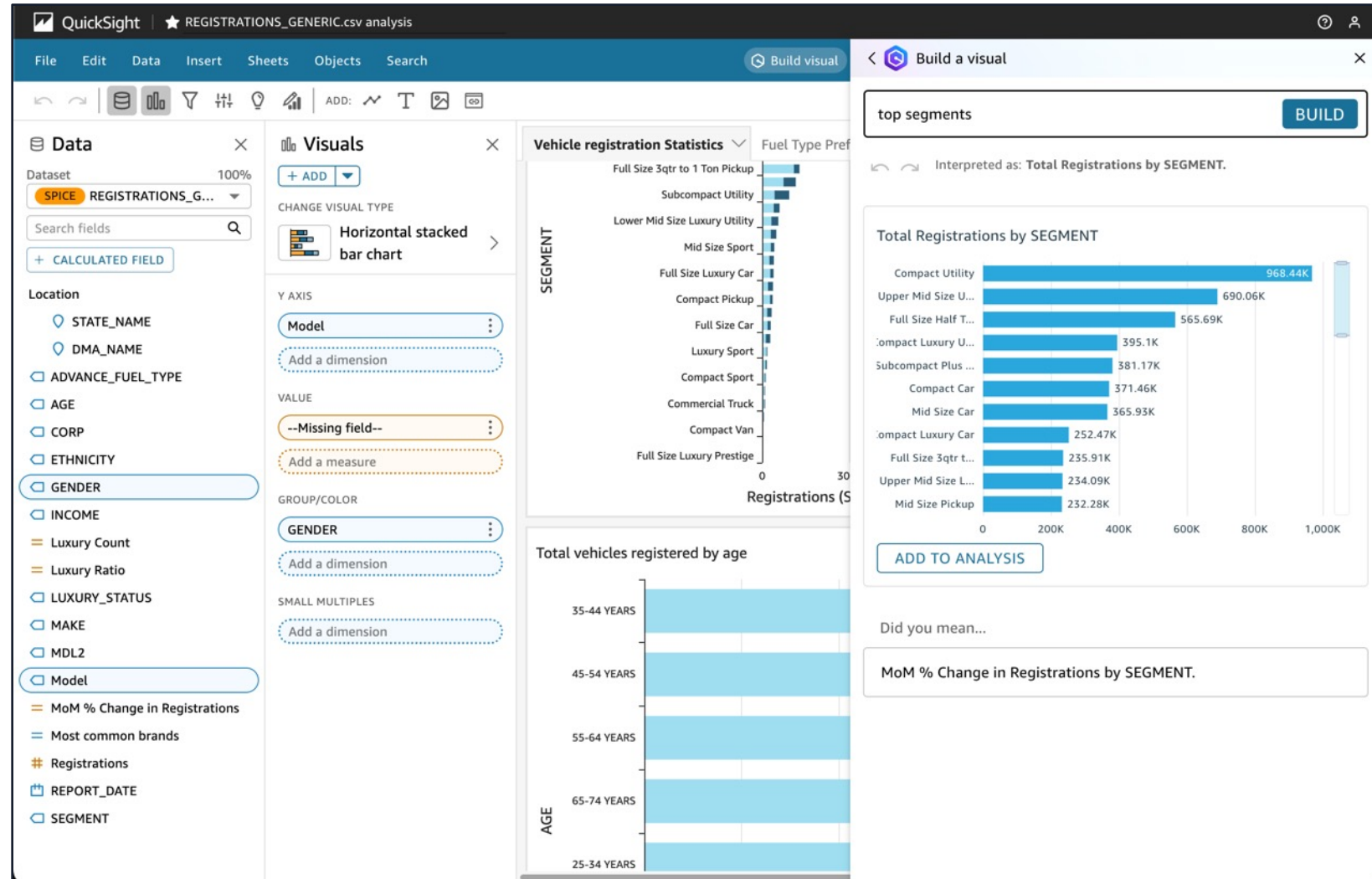
Use natural language to quickly build visuals for dashboards and reports

Create calculations

Build calculations using natural language without looking up or learning specific syntax

Refine visuals

Quickly update visuals by describing desired formats using natural language



AI-assisted storytelling

IMPACTFUL DATA STORYTELLING TO DRIVE ACTIONS

Interpret data for others

Help others derive meaning from data and reach conclusions to drive decisions

Generate stories using AI

Produce cohesive, powerful, and insightful narratives by analyzing only a few words of data

Create refined content

Control AI verbosity, customize narrative text, and apply stunning visual themes to bring content to life

Share up-to-date governed data

Quickly update and disseminate data at any time

The screenshot displays the Amazon QuickSight interface. The main dashboard is titled "Interests Analysis" and features a radar chart titled "Revenue by primary interest and source". The chart has 14 axes representing different interest categories: Unspecified, Travel, Technology, Sports, Outdoor, Home and, Health, Food and Dining, Finance, Fashion, Entertainment, Education, Wellness, Automotive, and Arts and Culture. The "Unspecified", "Fashion", and "Health" categories show the highest revenue. To the right of the chart, there is a text block that reads: "As shown in the interests analysis graph, fashion, health, and beauty/wellness generated the highest total lead revenues of \$46,113, \$31,041, and \$29,286, respectively. This suggests interests like fashion and wellness have a significant impact on performance and that marketing campaigns should be tailored to target these high-value interest segments." In the foreground, a "Build story" overlay is visible, which is a beta feature. It prompts the user to "Describe your data story in simple language and add the visuals you want to include." Below this prompt, there is a text input field containing the text: "Build a story about marketing campaign performance over time. Describe top campaigns and account managers. How can we improve overall campaign success?". At the bottom of the overlay, there are several options for adding visuals: "Total impression...", "Monthly impress...", "Most effective i...", and "Comparing reve...". There is also an "ADD VISUALS" button and a "BUILD" button.

AI answers to questions of data on demand

DEEP INSIGHTS AT YOUR FINGERTIPS

Executive summaries of dashboards

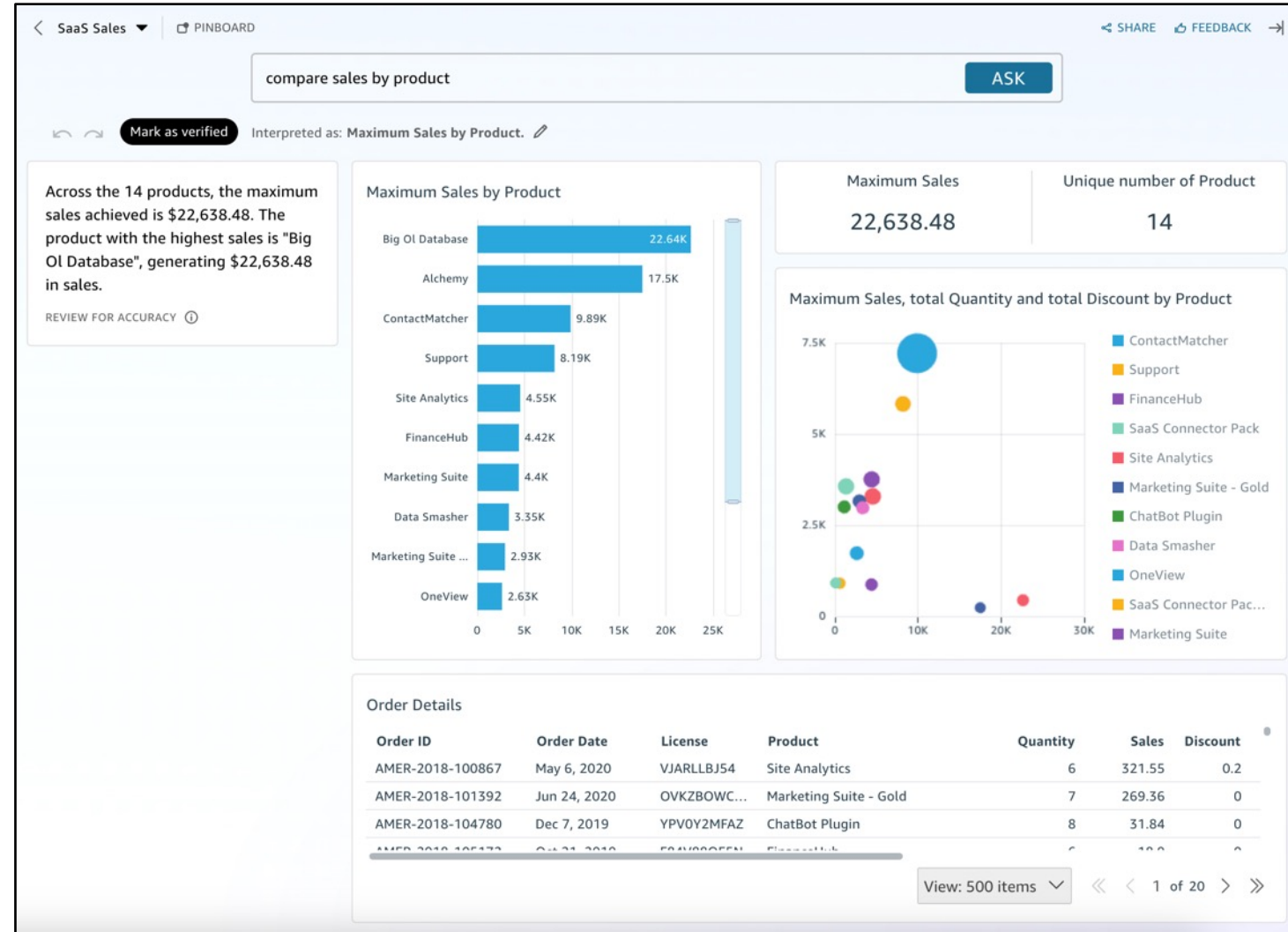
Instant summaries of key dashboard insights in natural language explaining top movers, outliers, and more

Powerful Q&A for nonexperts

Suggested questions and “what’s in my data” show what can be asked

Multivisual answers with narrative insight summaries explain answer context

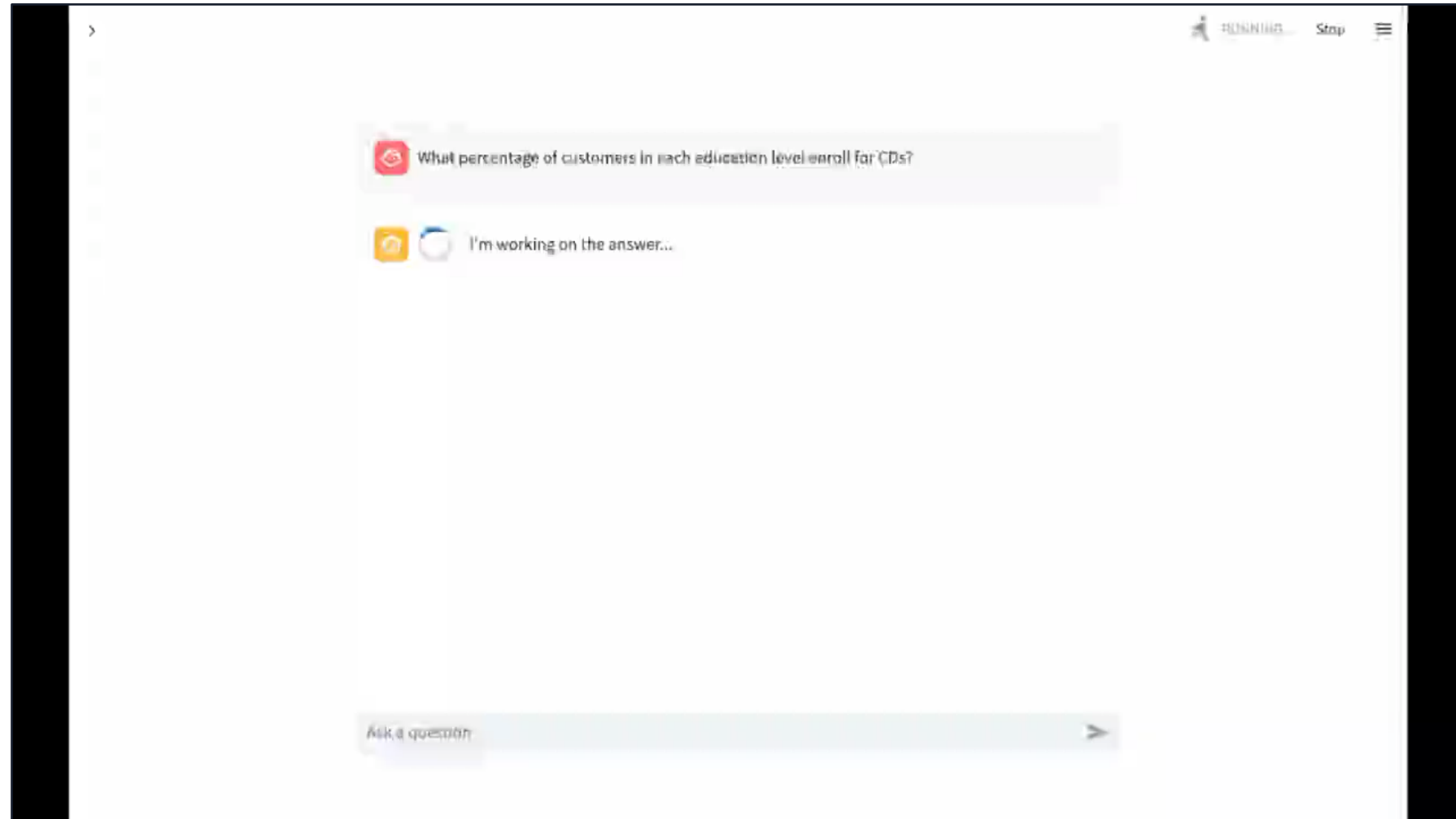
Support for vague questions and “did you mean” alternatives enable iterative fact-finding



Demo - GenAI Text to SQL

This market research assistant utilizes a modern data architecture and generative AI techniques to empower an organization's market research efforts.

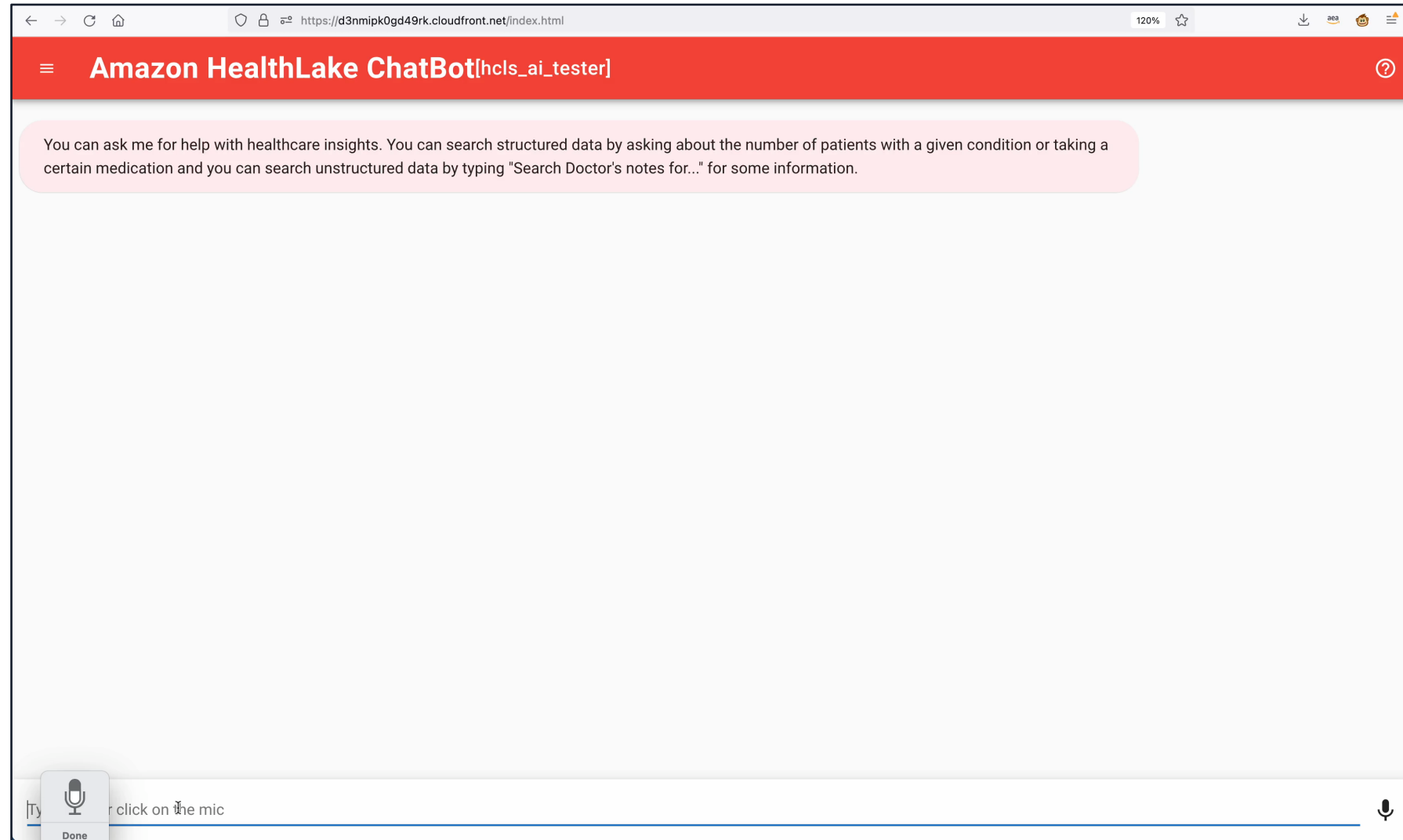
By integrating various data sources and leveraging natural language capabilities, the assistant allows non-technical users to easily query and analyze market research data using conversational natural language.



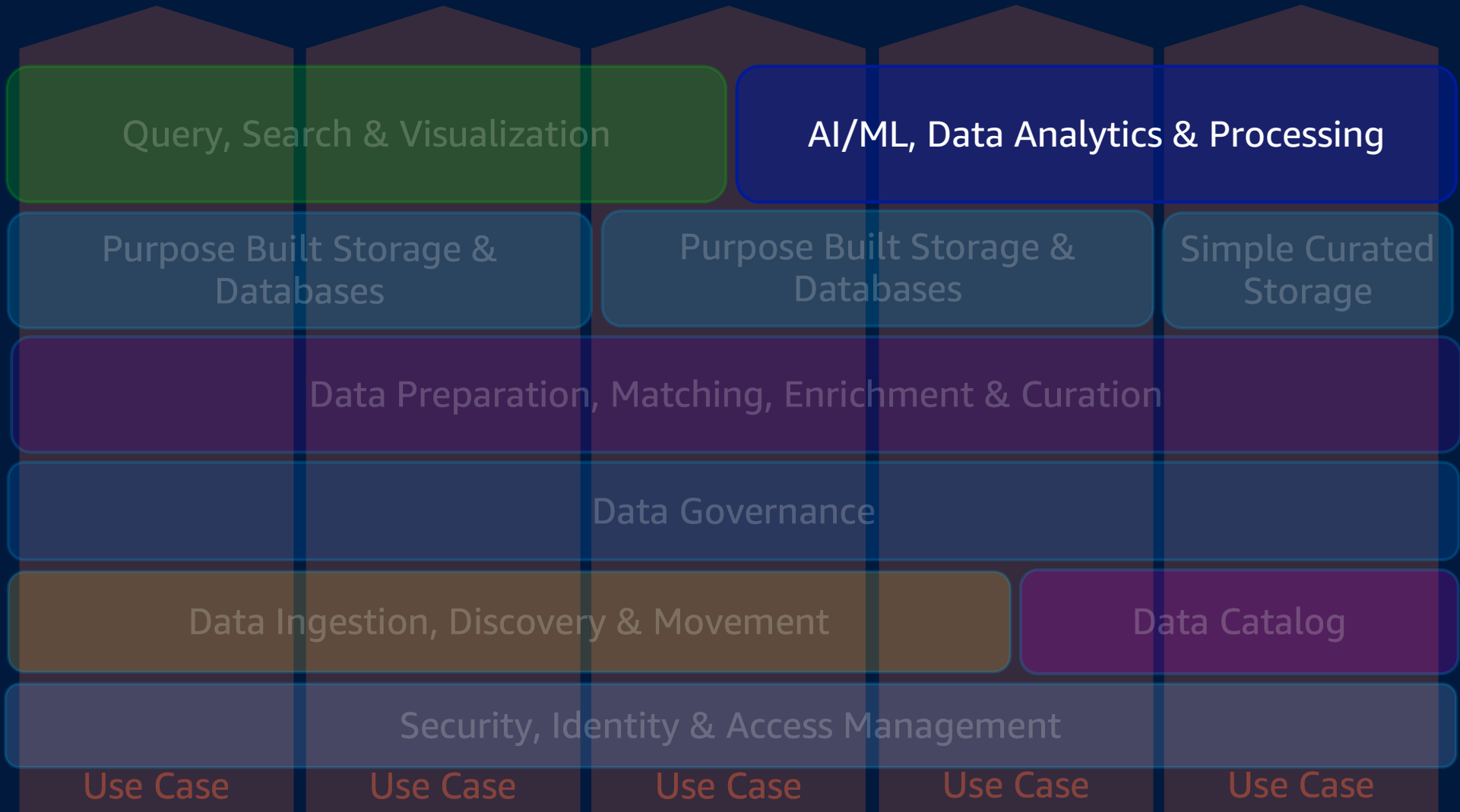
Demo – conversationally interact with EHR data

By combining HealthLake with a large language model, healthcare providers can interact conversationally with their data, gaining insights and making decisions faster than ever before.

One way to use a large language model with HealthLake is through a chatbot interface. Users can ask questions about both their structured data (e.g., Electronic Health Records (EHR)) or their unstructured data (e.g., doctor's notes).



A Modern, Foundational Data Architecture



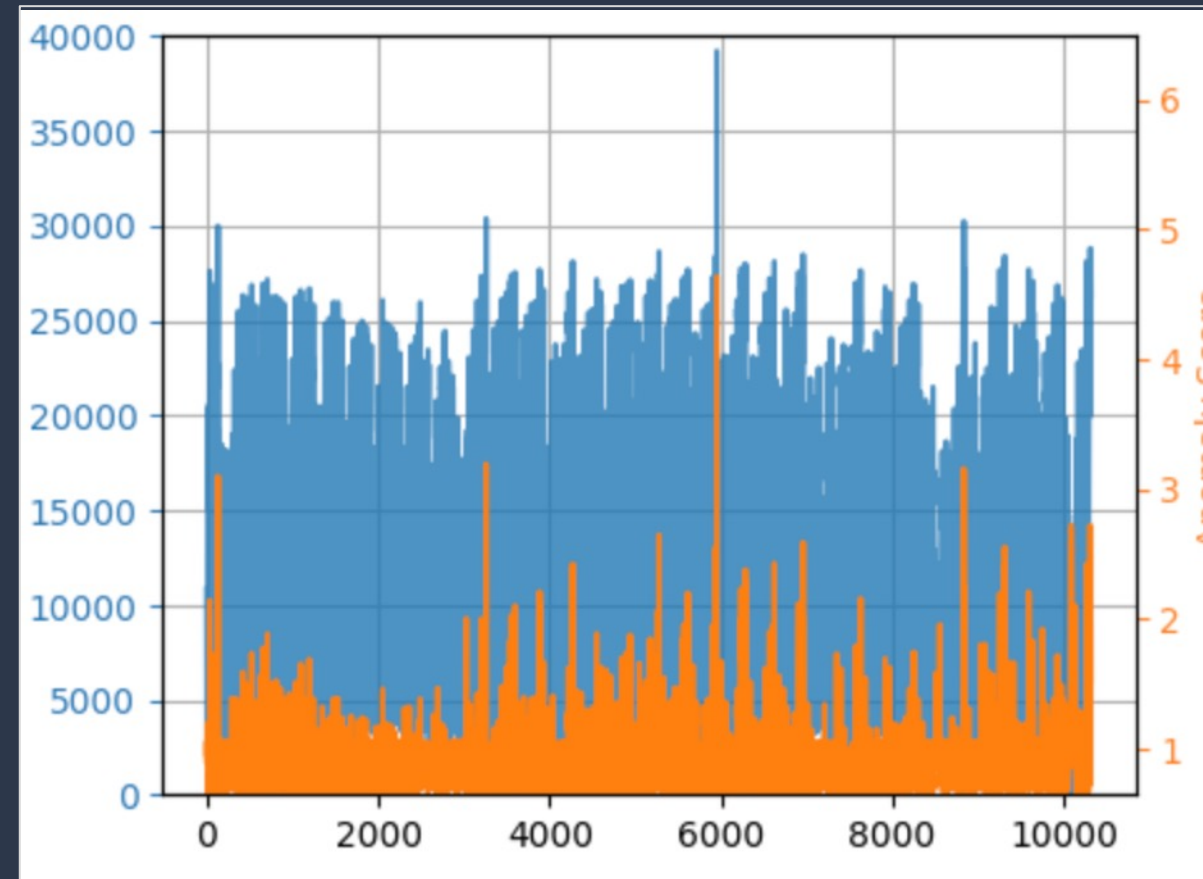
DevOps /
DataOps

External
Data
Sources &
Products

Fraud detection and prevention

Detect and prevent fraud, waste, and abuse

- Enhance accuracy and speed to help detect and prevent waste fraud and abuse
- Managed service approach with prebuilt ML models for fraud detection
- Supervised and unsupervised models for developing highly targeted models to utilize customer data as part of fraud prevention efforts



Predictions and forecasts from IoT and sensor data

Leverage data from smart cities and facilities

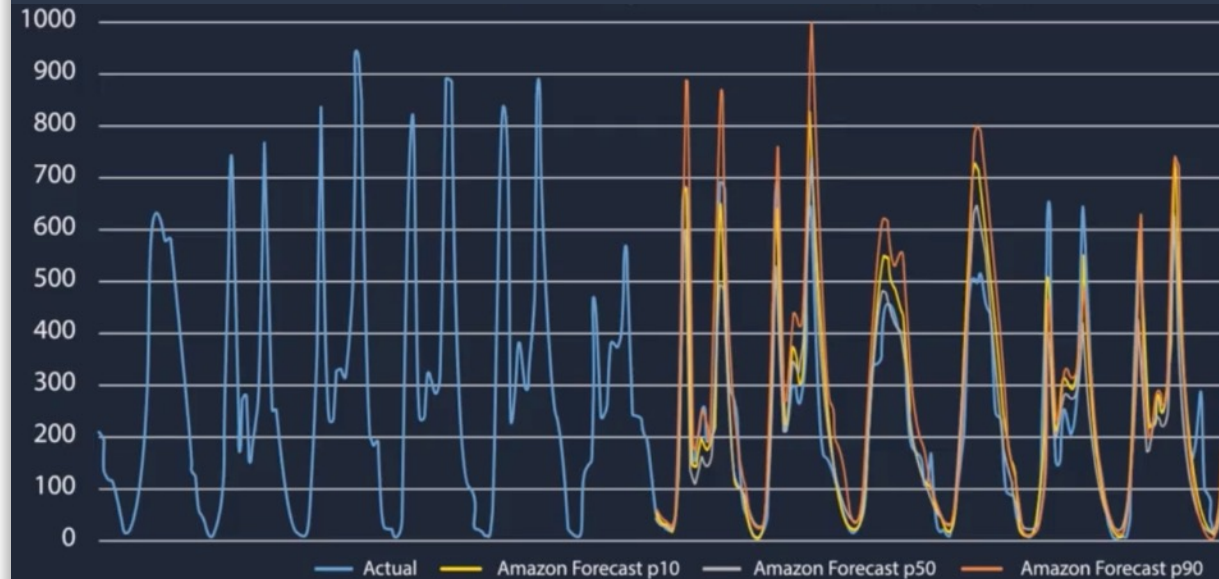
- Smart cities
- Predictive maintenance
- Facility management



Machine learning to forecast trends and support decisions

Accurate, time series forecast with machine learning

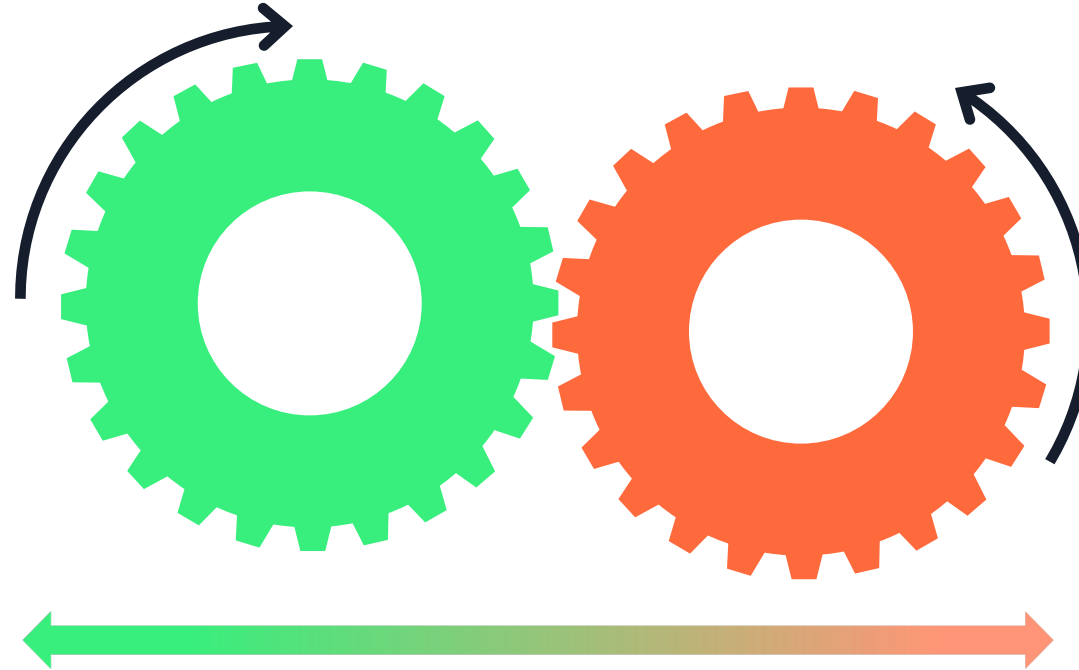
- Predicting service demand or program activities
- Allocating resources to optimize impact and outcomes for citizens
- Financial planning and revenue / cost forecasts



No Code Predictive Analytics Value Proposition

Accelerate data science teams

Do more with your current team by using low-code machine learning tools in order to get to the desired outcomes faster.



Enable business teams

Give business teams the ability to do ML without any code, scaling the number of people who can create ML powered insights, forecasts, and predictions

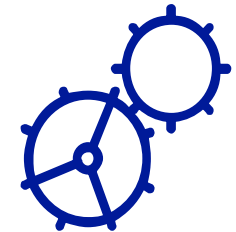
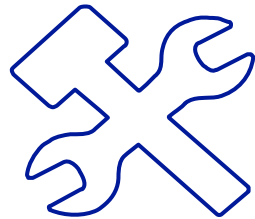
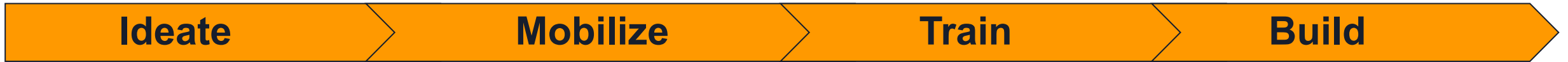
Collaborate together

Leverage the integrated capabilities of Amazon QuickSight and Amazon SageMaker Canvas making it easy for business users to use machine learning or for data scientists to make changes on the models business users build and creating one place for all the analytics and machine learning in a team or organization.

What can we do to maximize the opportunities presented by AI/ML and GenAI?

- Develop and begin implementing data modernization strategy & modern data architecture
- Encourage learning & managed experimentation
- Participate in public sector efforts to develop best practices
- Adopt an iterative approach: Think Big, Start Small, Scale Fast

AWS is here to help



Executive Visioning

Executive Briefing
Center (EBC)

Digital Innovation (DI)

Data-Driven Everything
(D2E)

Digital & Data
Transformation (D2T)

Immersion Day(s)

ML Embark

Game Days

Analytics Acceleration POC

ML Solutions Lab

PACE (POC/MVP)

ProServe or Partner SOW



Thank you!

Hiren Deliwala (he/him)

SA Manager, SLG-K12
hdaws@amazon.com

Mahinder Chawla (he/him)

Sr. Solutions Architect, SLG
mahindc@amazon.com

Please complete the survey
for this session



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