



UNLOCK DATA & AI POTENTIAL WITH A FULLY ORCHESTRATED HEALTH LAKEHOUSE



Hareesh Veldandi, Percept Health
June 2024



ABOUT HAREESH VELDANDI

19+ Years healthcare industry strategist, and data alchemist



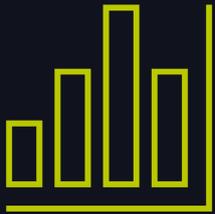
Percept Health

- Founder of Percept Health
- Author: "The Hidden Crisis: Inefficiency, Waste, and Overspending in the American Healthcare System," nominated for MedPage Today's 2022 Readers' Choice Award
- MBA from University of Cambridge
- Advised leading health systems, payers and other orgs: Advocate Health, Kaiser Permanente, WHO etc.
- Experiences at Microsoft, Epic, and Orion Health
- Drives initiatives in data and AI, embracing cloud computing, lakehouses, and generative AI workflows
- Unlocks true potential of liquid healthcare data

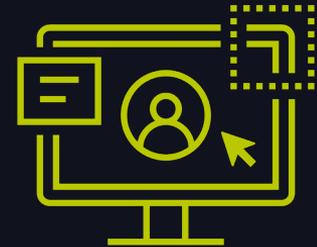
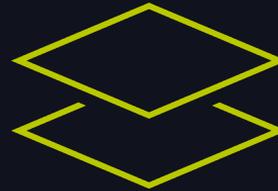
ABOUT PERCEPT HEALTH

An outcomes partner in the healthcare data journey

Healthcare
Data and AI



Purpose-built Platforms
and Solutions



Percept Health

ABOUT PERCEPT HEALTH

We have experiences of helping millions of lives & thousands of providers

4M+

FFS and Value-based
care lives

50+

Health Systems,
CINs and ACOs

20,000

Providers

\$10B

Total Cost of Care

ABOUT CLIENT: DORAL HEALTH AND WELLNESS

Doral serves patients in NY burroughs, primarily brown and black communities

32

Doral Providers serving
the communities of
New York City

14,000

Home-based visits for
COVID vaccinations
across NYC

10,000

Homebound seniors,
including dual eligibles,
receive care through Doral

2,000+

Care coordinated
patients - receiving RPM
and CM services

45,000+

Patients served in
the last 3 years
through Doral brands

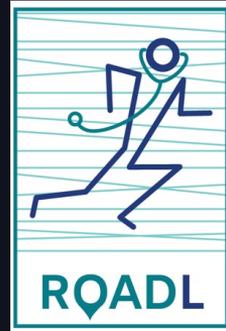
ABOUT CLIENT: DORAL HEALTH AND WELLNESS

Doral is a mini-health system that provides a range of healthcare services



Multi-Specialty &
Primary Care Practice

Caring for Our Community's Health™



Medical Home Visits & Related
Services; Wellness, Virtual and
Remote Care

Healthcare in Real-Time®



Health Integration and
Management Platform

The Health Integration Solution™

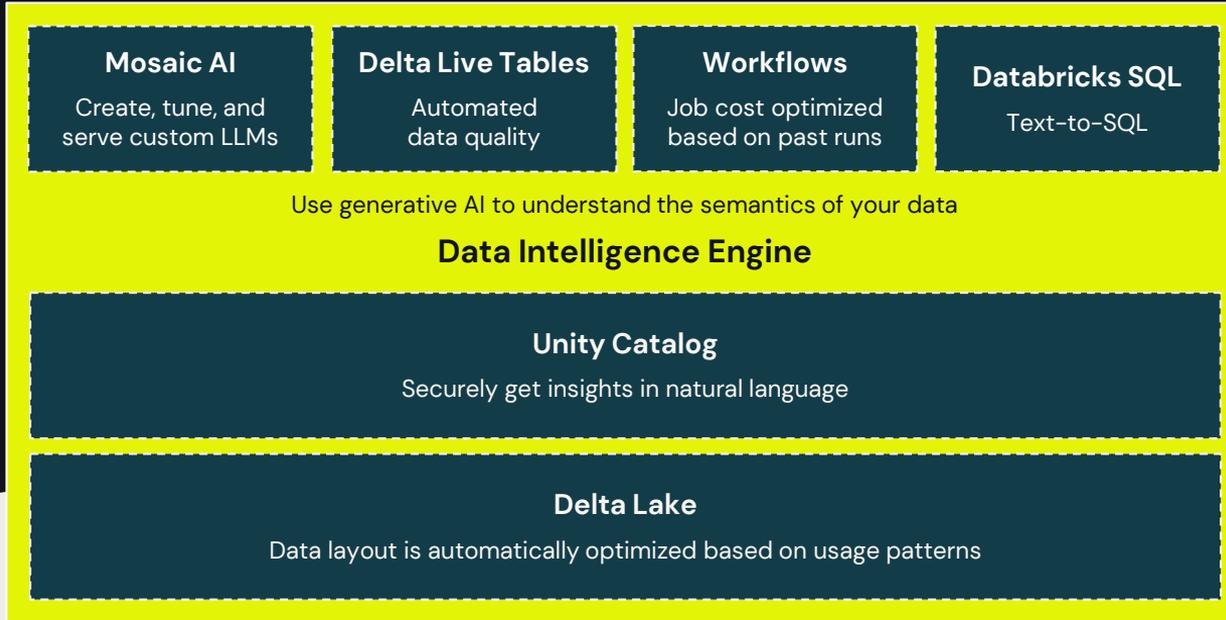
Fully Orchestrated Healthcare Data Platform



Percept Health

DATA INTELLIGENCE PLATFORM

Databricks' vision and industry-leading platform enablers



Use generative AI to understand the semantics of your data

Data Intelligence Engine

Unity Catalog

Securely get insights in natural language

Delta Lake

Data layout is automatically optimized based on usage patterns

Open Data Lake

All Raw Data
(Logs, Texts, Audio, Video, Images)

DATA
INTELLIGENCE
PLATFORM IS THE
FOUNDATION TO
MANY USE CASES

DORAL'S DATA INTELLIGENCE PLATFORM

Built and operated by Percept Health

External Systems

ACOs / CINs / Payers

HIE

HHA Platform

RPM / CCM

Medical Devices

Internal Systems

EMR 1

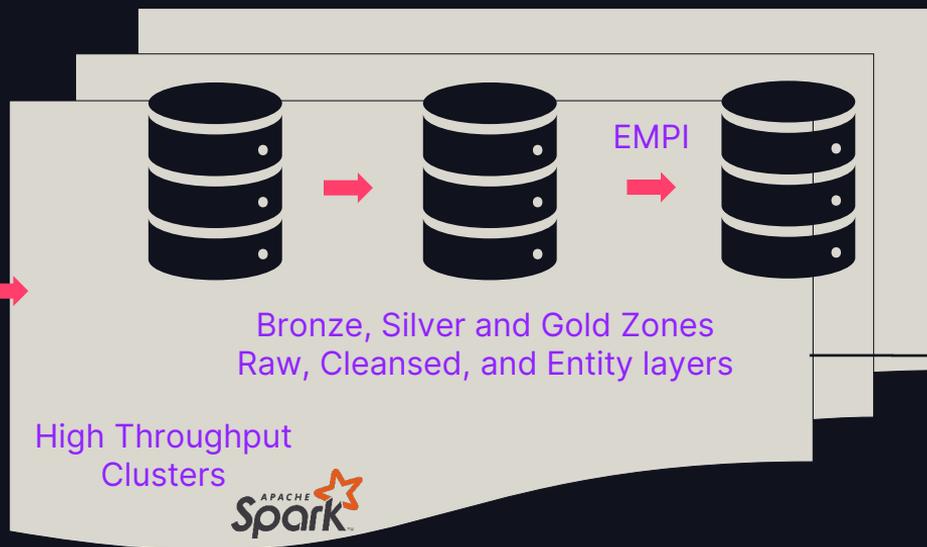
EMR 2

HR System

Active Directory

CRM

APIs and Data Extracts



Doral leadership and care management users



Serving Layer



END-TO-END ORCHESTRATION

Workflows / Jobs artifacts inside Databricks

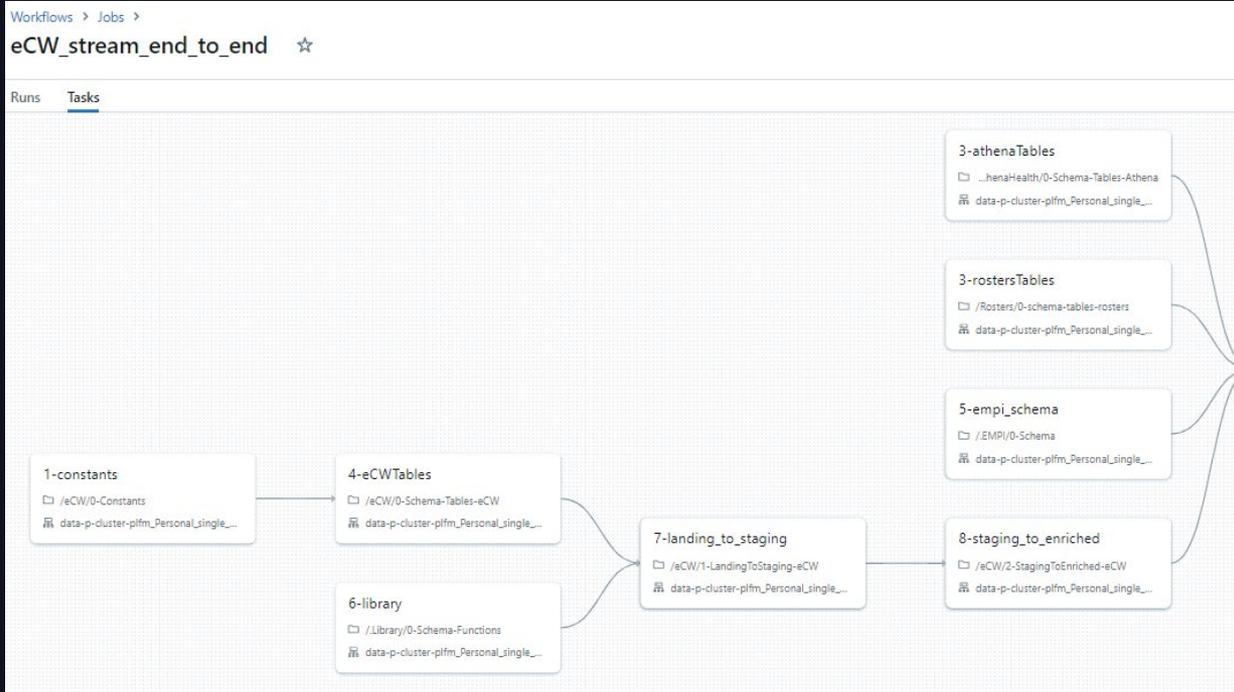
The screenshot displays the Databricks Workflows page. The left sidebar contains navigation options: New, Workspace, Recents, Catalog, Workflows (selected), Compute, SQL, SQL Editor, Queries, Dashboards, Alerts, Query History, SQL Warehouses, Data Engineering, Job Runs, Data Ingestion, Delta Live Tables, Machine Learning, Playground, Experiments, and Features. The main content area is titled 'Workflows' and includes tabs for Jobs, Job runs, and Delta Live Tables. A search bar for 'Filter jobs' and a checkbox for 'Only jobs owned by me' are present. A 'Create job' button is in the top right. Below is a table of jobs with columns for Name, Created by, Trigger, and Recent runs.

Name	Created by	Trigger	Recent runs
☆ activedirectory_stream_to_enriched	devops	Scheduled	🟢🟢🟢🟢🟢
☆ athena_stream_end_to_end	devops	Scheduled	🟢🟢🟢🟢🟢
☆ eCW_stream_end_to_end	devops	Scheduled	🟢🟢🟢🟢🟢
☆ empeon_stream_to_enriched	devops	Scheduled	🟢🟢🟢🟢🟢
☆ gaps_stream_to_end_to_end	devops	Scheduled	🟢🟢🟢🟢🟢
☆ healthix_stream_to_end_to_end	devops	Scheduled	🟢🟢🟢🟢🟢
☆ healthviewx_stream_to_end_to_end	devops	Paused - Scheduled	— — — — —
☆ hhaexchange_end_to_end	devops	Scheduled	🟢🟢🟢🟢🟢
☆ ml_risk_stratification_end_to_end	devops	Paused - Scheduled	— — — — —
☆ ml_risk_stratification_inference	devops	Scheduled	🟢🟢🟢🟢🟢
☆ outputs_reports	devops	Scheduled	🟢🟢🟢🟢🟢
☆ rosters_stream_to_end_to_end	devops	Scheduled	🟢🟢🟢🟢🟢
☆ smartmeter_end_to_end	devops	Scheduled	🟢🟢🟢🟢🟢
☆ zoho_stream_to_end_to_end	devops	Paused - Scheduled	— — — — 🟢



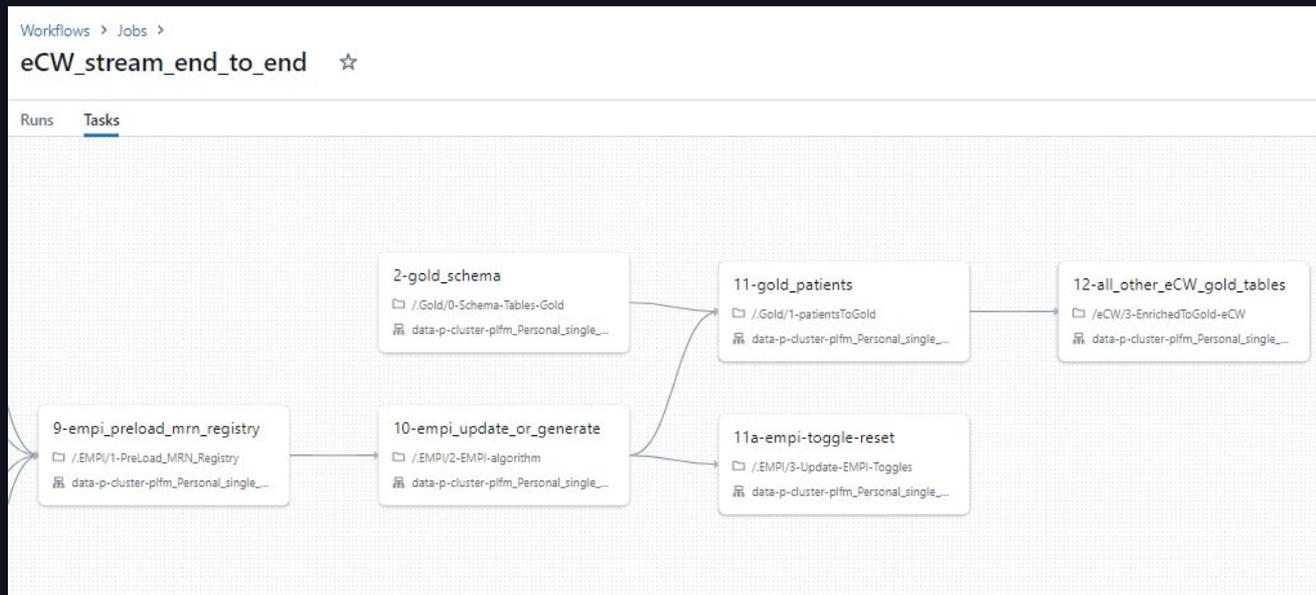
END-TO-END ORCHESTRATION

Workflows / Jobs artifacts inside Databricks



END-TO-END ORCHESTRATION

Workflows / Jobs artifacts inside Databricks



END-TO-END ORCHESTRATION WITH DEVOPS

Azure DevOps Integration

Pipelines		
Recent	All	Runs
Recently run pipelines		
Pipeline	Last run	
 Data_Streams_eCW Data-Streams	#20240424.1 • Merged PR 2763: Procedure fix Individual CI for  feature/develop_v1	
 Data_Streams_EMPI Data-Streams	#20240424.1 • Merged PR 2760: query change Individual CI for  feature/develop_v1	
 Data_Outputs_reports Data-Outputs	#20240423.1 • Merged PR 2744: Entra modification changes for tracking Individual CI for  feature/develop_v1	
 Data_Streams_ActiveDirectory Data-Streams	#20240416.1 • Merged PR 2691: added new enriched tables to Active Directory Individual CI for  feature/develop_v1	
 Data_Streams_Zoho Data-Streams	#20240415.4 • Merged PR 2677: fixed state column Individual CI for  feature/develop_v1	
 Data_Streams_HHAExchange Data-Streams	#20240405.2 • Merged PR 2587: adding patient status and C to caregivers Manually triggered for  feature/develop_v1	
 Data_Streams_Gold Data-Streams	#20240401.1 • Like P Individual CI for  feature/develop_v1	
 Data_Streams_Healthix Data-Streams	#20240329.2 • infra time change Manually triggered for  feature/develop_v1	
 Data_Outputs_risk_stratification Data-Outputs	#20240329.2 • time changes Manually triggered for  feature/develop_v1	

Recent	All	Runs
	Data_Streams_Library Data-Streams	#20240325.1 • Merged PR 2511: fixed output report and modified AD workflow Individual CI for  feature/develop_v1
	Data_Streams_AthenaHealth Data-Streams	#20240321.2 • distinct Individual CI for  feature/develop_v1
	Data_Streams_SmartMeter Data-Streams	#20240228.2 • Merged PR 2254: Added all readmes Manually triggered for  feature/develop_v1
	Data_Healthviewx_Streams Data-Streams	#20240228.1 • Merged PR 2254: Added all readmes Individual CI for  feature/develop_v1
	Data_Streams_Rosters Data-Streams	#20240228.1 • Merged PR 2254: Added all readmes Individual CI for  feature/develop_v1
	Data_Streams_GapsInCare Data-Streams	#20240228.1 • Merged PR 2254: Added all readmes Individual CI for  feature/develop_v1
	Data_Streams_Empeon Data-Streams	#20240228.1 • Merged PR 2254: Added all readmes Individual CI for  feature/develop_v1
	Data_Outputs_ny_health_home_potentials Data-Outputs	#20231125.1 • ML changes Individual CI for  feature/develop_v1

HEALTHCARE ORGANIZATIONS CAN LEAPFROG CLOUD COMPUTING AND GENERATIVE AI IN TANDEM

Custom Risk Stratification Algorithm



Percept Health

MLOPS and AUTOML FOR CUSTOM ALGORITHM

AutoML artifacts inside Databricks

The screenshot shows the Databricks AutoML experiment interface. At the top, the experiment name is "v1.0-manual-training-set2023-11-26 06:50:26.391797". Below this, there are tabs for "Configure" and "Train", both of which are active. The "Overview" tab is selected, showing a progress bar and a "Warnings (4)" link. The training dataset is identified as a "Dataframe in notebook". The target column is "Label", and the evaluation metric is "val_f1_score". The timeout is set to 120 minutes. The AutoML evaluation is complete, and all runs have been added to a table. The table lists several runs with their respective metrics and tags. The best model is identified as the one with the highest val_f1_score.

Experiments >
v1.0-manual-training-set2023-11-26 06:50:26.391797

Experiment ID: 237825330089832 Artifact Location: dbfs/databricks/mlflow-tracking/237825330089832

AutoML

Configure Train

Overview Warnings (4)

Training dataset: Dataframe in notebook (see notebook that launched this AutoML experiment)
Target column: Label
Evaluation metric: val_f1_score
Timeout: 120 minutes

AutoML Evaluation complete
All runs have completed, and have been added to the table below. Click a specific run to view details.

Model with best val_f1_score
The model is ready to be registered and deployed. Or, access the source code for the model training to make modifications by clicking a notebook under the Source column in the table below.

View notebook for best model View data exploration notebook

AutoML generated notebooks are now saved as MLflow artifacts. Click here to learn more.

Description Edit

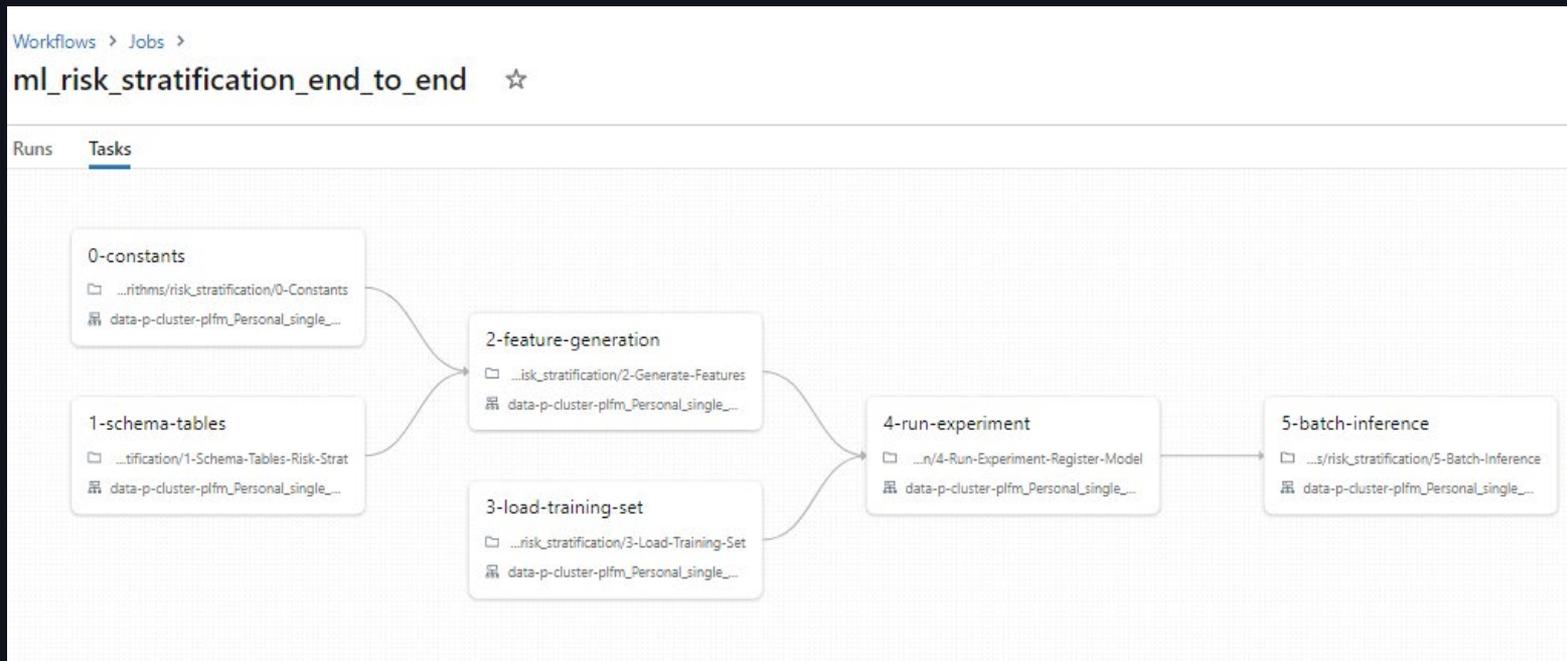
Q metrics,rmsc < 1 and params.model = "tree" Time created State: Active Datasets Sort: val_f1_score Columns Expand rows

Run Name	Created	Dataset	Duration	Source	Models	Metrics	Tags
unique-seal-890	15 days ago	dataset (5ce3b5a0) Test, data... +2	6.4min	Noteboo...	risk_strat.../1	0.94521812...	lightgbm_cl...
big-duck-417	15 days ago	dataset (5ce3b5a0) Test, data... +2	5.7min	-	sklearn	0.94454335...	lightgbm_cl...
enthused-hen-848	15 days ago	dataset (670b0262) Training, data... +2	24.0min	-	sklearn	0.93926836...	lightgbm_cl...
abrasive-dove-524	15 days ago	dataset (becaf19d) Train, data... +2	1.9min	-	sklearn	0.93051194...	xgboost_ca...
thoughtful-roo-714	15 days ago	dataset (9c587c5) Val, data... +2	1.2min	-	sklearn	0.92710993...	xgboost_ca...
dashing-mule-49	15 days ago	dataset (becaf19d) Train, data... +2	1.9min	-	sklearn	0.89564514...	xgboost_ca...
bemused-snai-963	15 days ago	dataset (670b0262) Training, data... +2	3.9min	-	sklearn	0.855876382...	xgboost_ca...

78 matching runs

RISK STRATIFICATION ALGORITHM INFERENCE

End to end workflow to do feature engineering, and batch inference



AUTOML CODE SNIPPET

Databricks allows low-code ML algorithm development with ease

PYTHON

```
# 1. Run Experiment

from databricks import automl
from datetime import datetime
summary = automl.classify(train_df, target_col="label", experiment_name = ml_experiment_prefix +
str(datetime.now()), timeout_minutes=120, exclude_cols=["file_name","creation_date","version","empi"])

# 2. register usign the latest run id

import mlflow
mlflow.set_registry_uri(ml_registry_uri)
mlflow.register_model(f"runs:{mlflow_run_id}/model", f"{ml_schema}.{ml_model_name}")
```

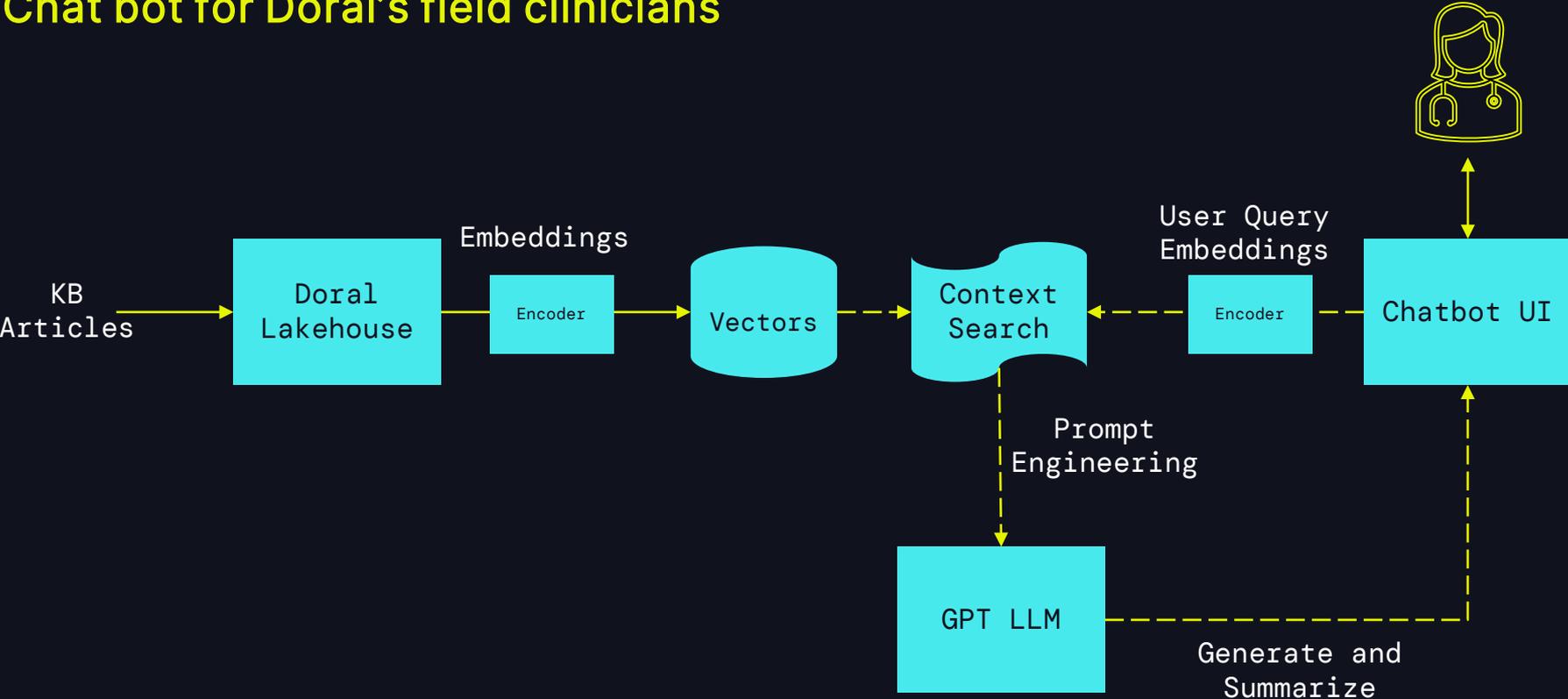
Generative AI Use Case



Percept Health

Custom Generative AI Algorithm

Chat bot for Doral's field clinicians



Custom Generative AI Algorithm

Chat bot for Doral's field clinicians

PYTHON

```
# 1. Embeddings
```

```
from sentence_transformers import SentenceTransformer
```

```
...
```

```
doral_kb.apply(lambda x: func["content"], axis =1)
```

```
...
```

```
# 2. Prompt Engineering
```

```
question = "how do i signup for payment?"
```

```
context = " ".join([f"#{str(i)}" for i in results["documents"][0]])
```

```
prompt_template = f"Relevant context: {context}\n\n The user's question: {question}"
```

```
...
```

```
lm_response = llm_pipe(prompt_template)
```

Healthcare Data Platform without Paying \$Millions



Percept Health

BUILD YOUR OWN HEALTHCARE DATA PLATFORM

Doral's data intelligence platform development facts:

45,000+

Patients

14

Diverse sources

2,000+

Medical devices

10M+

Records

USE CASES WE ARE ENABLING

Care management team using patient markers for risk stratification

The screenshot shows a web interface for 'Risk Stratification - Care Mangement'. At the top right is the 'Health & Wellness' logo with the tagline 'Care for Our Community's Health'. The main title is 'Risk Stratification - Care Mangement'. Below the title are several filter sections:

- Predicted Risk Score:** All
- Program:** All
- Facility (Location):** All

Search Results Matching the Criteria: **46,090** # of Patients

Smart meter abnormal values:

- BMI Category:** All
- BP Category:** All
- Is BP Abnormal?:** All
- Pulse Category:** All
- SpO2 Category:** All

Upcoming Visits:

- Athena - Next 7 Days:** All
- eCW Next 7 Days:** All

24-hour Visits:

- Acute Visits:** All
- ER Visits:** All

48-hour Visits:

- Acute Visits:** All
- ER Visits:** All

72-hour Visits:

- Acute Visits:** All
- ER Visits:** All

Search Result Details:

Full Name	Athena MRN	eCW MRN	Age	Gender	City	Zip	Cell	Email	Race	Language	Medicaid indicator	90d fut appts	Fut appts	Latest Enc	# of Meds	# of Probs	12m Offi
...

USE CASES WE ARE ENABLING

Analytics team using a lot more patient markers for patient search

The screenshot displays a 'Patient Search' interface with the following sections:

- Demographics:** Gender, Zip, Age, Race, Language.
- Overall Visits:** Latest E&M Visit, Latest Wellness Visit, Latest Visit.
- Current Rosters and Programs:** Roster, Program, Attr Prov, Gaps, Suspect Cond #.
- Predicted Risk:** Current Risk Level, 7d Risk Level Diff, Next 2 Days, Next 7 Days, Next 30 Days.
- Future Appointments:** (Empty section)
- Current Chronic Conditions:** # of Chronics, Asthma, Behav Health, Cancer, CKD, Cardiac, Depression, Diabetes, Hematologic, Hypertension, Infections, Lung, Metabolic, Neurologic, Psychiatric, Rheumatologic, Stroke, Substance Use, Vascular.
- Home Health Hours:** Weekly Home Health Hours, Monthly Home Health Hours, Yearly Home Health Hours.
- 24m Visit Stats:** Office Visits, Acute Visits, ER Visits.
- 12m Visit Stats:** Office Visits, Acute Visits, ER Visits.
- 24m Vitals Stats:** Min BP Systolic, Min BP Diastolic, Min BMI, Max BP Systolic, Max BP Diastolic, Max BMI.
- 30d Visit Stats:** Office Visits, Acute Visits, ER Visits.
- Search Results Matching the Criteria:** # of Patients: 45,166.



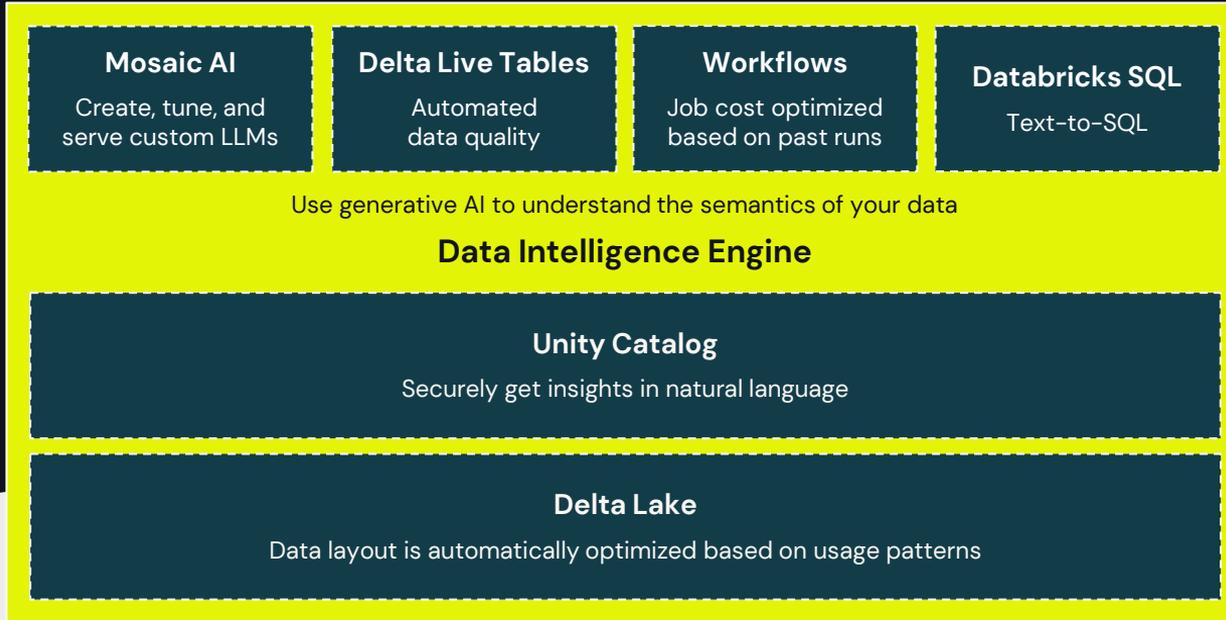
USE CASES WE ARE ENABLING

Lot more use cases are possible for healthcare and life sciences

- Risk Stratification
- Care Management
- Risk Adjustment
- Quality Improvement
- Patient Engagement
- Suspect Diagnoses NLP
- Clinical Summary

DATA INTELLIGENCE PLATFORM

Databricks' vision and industry-leading platform enablers



Open Data Lake

All Raw Data
(Logs, Texts, Audio, Video, Images)

**BUILD YOUR DATA
ASSETS AND NOT
GIVE AWAY
\$MILLIONS TO
RIGID PLATFORM
COMPANIES**

CHALLENGES, TIPS AND LEARNINGS

Every healthcare client can build their own platform, with the help of domain and databricks experts

- Healthcare data domain knowledge is irreplaceable
- Think DevOps first
- Think “platform,” “data streams,” and “Devops pipelines”
- Unity Catalog and Schemas across DEV and PROD
- Permissions and service principals
- Workflows management
- Auto ML orchestration

Q&A

