

# Unleashing the Potential of Unstructured Data with LLMs and Databricks

---

North Dakota University System  
Cordell Wagendorf - Software Developer  
Ryan Jockers - Assistant Director of Reporting and Analytics

# Overview



About Us



The Problem



The Journey



Current State of Gen AI Apps



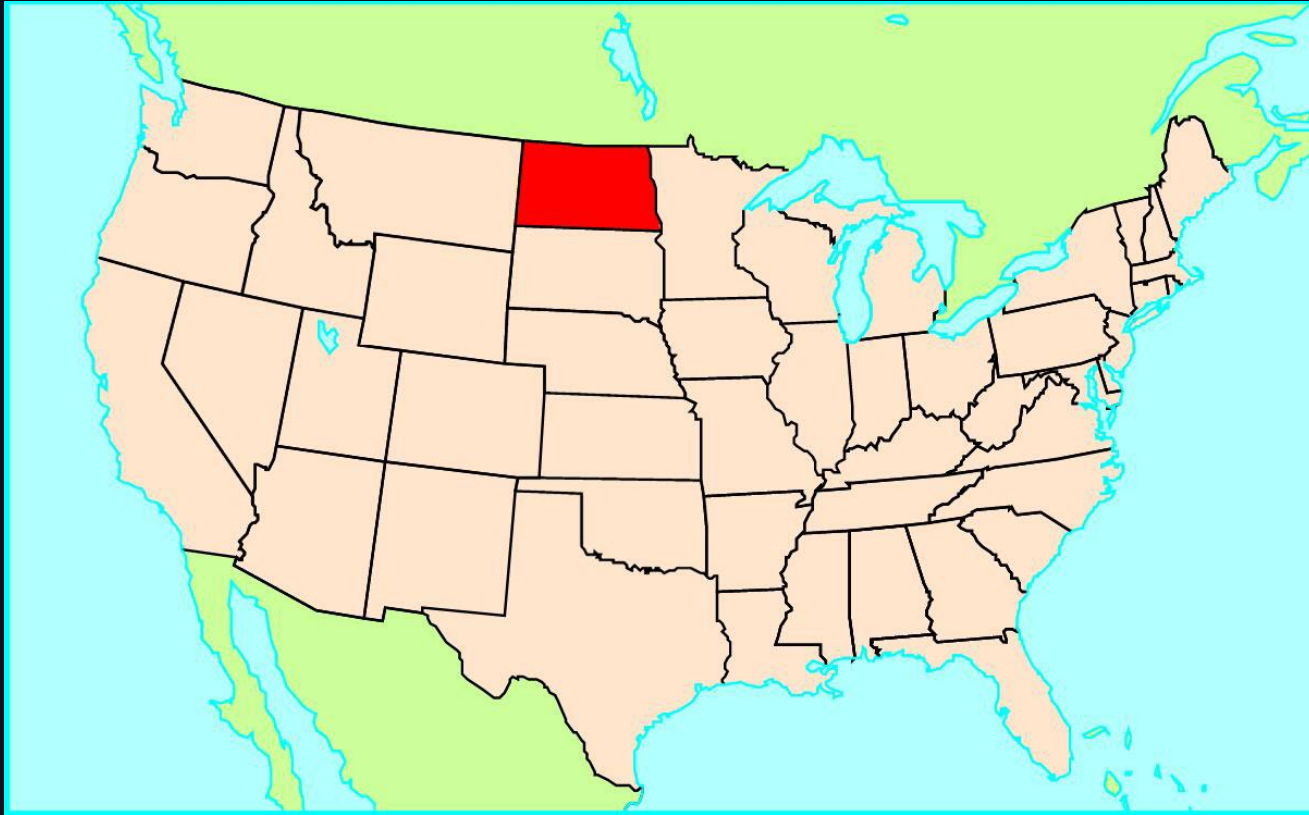
Future Development



# About Us



# Where is North Dakota?



North Dakota  
Population  
Estimate

788,940

# Brief Introduction



- Organized in 1990 and includes two research universities, four regional universities, and five community colleges.
- The mission of the State Board of Higher Education is to enhance the **quality of life** for all those we serve and the economic and social vitality of North Dakota through the **discovery, sharing, and application of knowledge**.

# The Problem

# Unstructured Pain



NDUS and its campuses rely on many state-level policy and law documents that are hard to find in many places.

NDUS Policy, Procedures, State Law, and others are used to make decisions every day



# Unstructured Pain

## CHAPTER 43-11 COSMETOLOGISTS

### 43-11-01. Definitions.

In this chapter, unless the context or subject matter otherwise require

1. "Board" means the state board of cosmetology.
2. "Cosmetologist" means an individual licensed under this cosmetology.
3. "Cosmetology" means any one or a combination of practices performed by and known as the occupation of beauty culturists, cosmetologists or hairdressers, or of any other individual holding cosmetology by whatever designation and within the meaning in and upon whatever place or premises.
  - a. The term includes:
    - (1) Hair care, including arranging, dressing, curling, wavy cleansing, cutting, shaving, trimming, singeing, straightening, or similar work, upon the hair of any in or with hands or mechanical or electrical apparatus or
    - (2) Skin care;
    - (3) Manipulation and application of product to eyelids including extensions, design, treatment, tinting, and lip
    - (4) Hair extensions using chemical hair joint agents, silicon keratin bonds, or fusion bonds; and
    - (5) Manicuring.
  - b. The term does not include natural hair braiding or threading
4. "Esthetician" means an individual licensed under this chapter to skin care.
5. "Esthetics" means manipulation and application of product to eyebrows, including extensions, design, treatment, tinting, or ar care.
6. "Independent licensee" is a licensed individual who maintains a and operates independently from the establishment owner in suite or chair is located.
7. "Instructor" means an individual who is a licensed cosmetologist or manicurist who teaches cosmetology, esthetics, manicuring, or the scope of the individual's license, in a duly registered school who has met the requirements of section 43-11-27 and has applied instructor's license.
8. "Invasive care" means any procedure that invades the live tissue means, including cutting, puncturing, burning, insertion of instrument.
  - a. Laser use; and
  - b. Chemical peels using:
    - (1) Thirty percent or higher concentration of alpha hydroxy
    - (2) Twenty percent or higher concentration of beta hydroxy
    - (3) Two percent or higher concentration of resorcinol;
    - (4) Fifteen percent or higher concentration of trichloroacetic
    - (5) Fifteen percent or higher concentration of phenol.
9. "Manicuring" means:
  - a. Cleansing, cutting, shaping, or beautifying nails;
  - b. Massaging from the elbow to the fingertips or knee to toes
  - c. Caring for and treating the cuticles and nails; and

## NORTH DAKOTA STATE BOARD OF HIGHER EDUCATION Policy Manual

**Policy:** 401.1 Academic Freedom  
**Effective:** June 29, 2021

1. **SBHE Policy and Principles.** The State Board of Higher Education in the 1940 Statement of Principles on Academic Freedom adopted Association of University Professors with 1970 Interpretive Comments of academic freedom and responsibility to the institutions under its commitment to ensuring that the institutions of the NDUS shall foster academic community for faculty members, students, and all other who engage in scholarly work.
2. **Academic Freedom.** Academic freedom is the freedom, without in other outside pressure or restraint, to explore any avenues of scholarly creative expression, and to speak or write on matters of public concern related to professional duties and the functioning of the NDUS and are entitled the freedom in designing and teaching their assigned course shall be free to involve interested students or other professionals in research and to pursue funding from internal or external sources to : this principle is the toleration of the conflict of ideas and the opportunity of diverse points of view. Faculty members and other NDUS employees engaged in scholarly work shall be subject to the full protections of speech and students under SBHE Policy 503.1 and 503.3.
3. **Academic Responsibility.** Academic responsibility implies the fulfillment of professional duties and obligations, the recognition of the demands of enterprise, and the candor to make clear that when one is speaking of interest, one is not speaking for the NDUS or any of its institutions. Faculty, work, faculty and NDUS employees acknowledge the importance of accuracy in research and publication, and upholding the principles of open exchange of ideas, even where those ideas conflict.
4. **Classroom Speech and Expression.** Faculty at institutions under this shall generally adhere to the 1940 Statement of Principles on Academic with 1970 Interpretive Comments adopted by the American Association of Professors, which provides that "Teachers are entitled to freedom in the classroom in discussing their subject, but they should be careful not to introduce into their teaching controversial matter which has no relation to their subject." As a result, no faculty member

## Office of Management and Budget

Travel

August 1, 2021

### Policy 516 – Miscellaneous Charges on Travel Vouchers (Reimbursable and Non-reimbursable Expenses)

On travel vouchers, whenever a charge is made under the "Miscellaneous" column, it requires detailed explanation, and a receipt must be attached if the item is greater than \$10.

Examples of non-reimbursable expenses are:

1. Alcoholic beverages
2. Entertainment
3. Late check-out charges
4. Parking tickets or other traffic tickets
5. Laundry

Exceptions to this policy may be made by an agency director for unusual or extenuating circumstances such as international travel or travel extending ten days or longer.

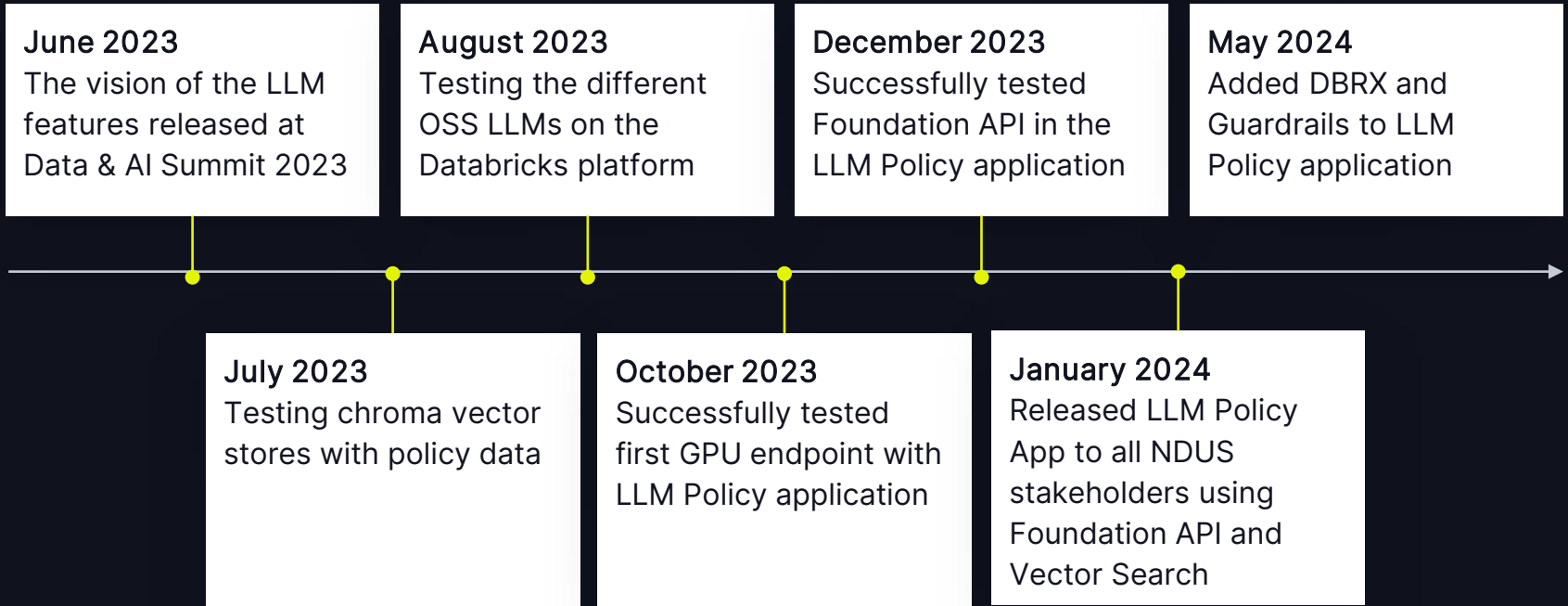
Over 3000 pdfs.....



# The Journey

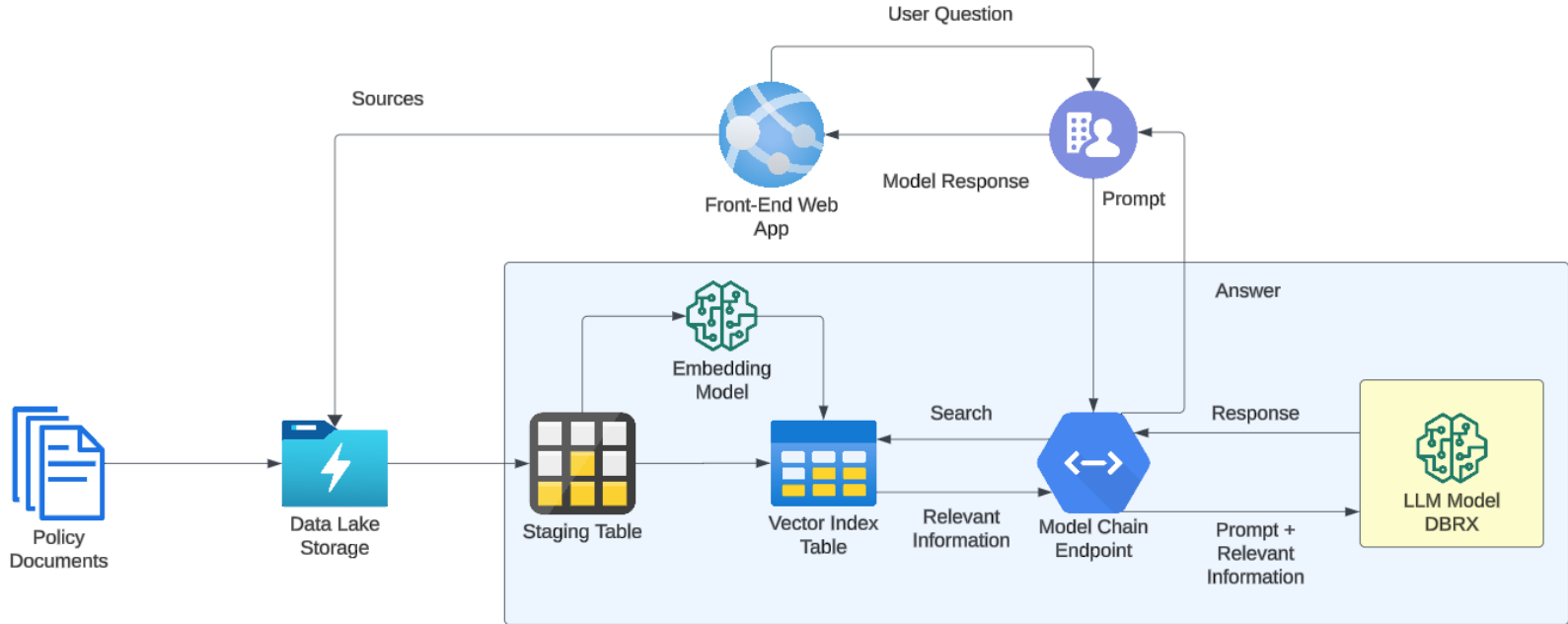


# The NDUS LLM Journey



# Current State of NDUS GenAI Apps

# NDUS RAG Architecture



# Why Retrieval-Augmented Generation (RAG)?



**Reduced Hallucinations:** RAG helps mitigate the factually incorrect or irrelevant responses generated by LLMs.

**Knowledge Cutoff:** RAG extends the knowledge of LLMs by referencing external sources.

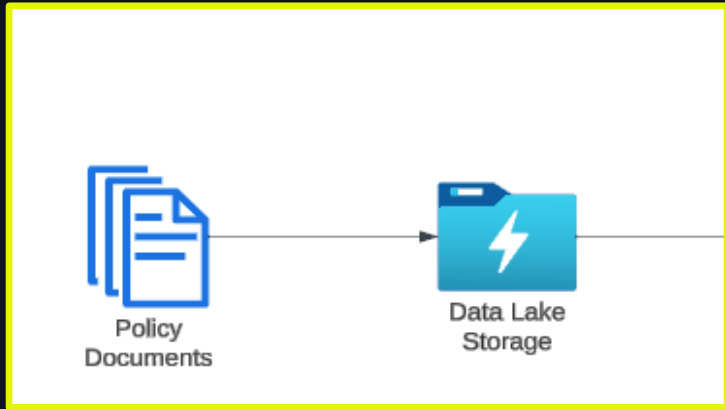
**Cost-Effective:** RAG extends the capabilities of LLMs without the need to retrain the model, making it a cost-effective solution.



# Data Movement and Storage

## Resources:

Azure Data Lake Storage (ADLS) Gen 2  
Databricks External Volumes



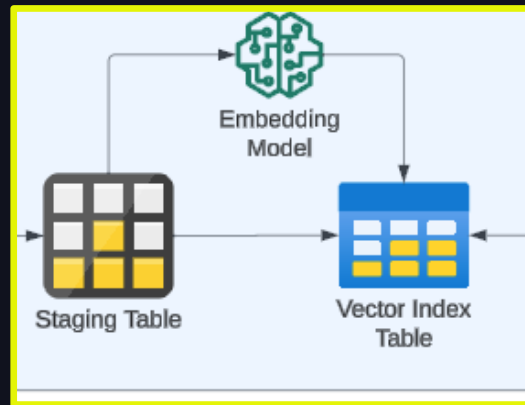
- Policy documents are updated daily. Changes made to the major versions in SharePoint libraries will be captured using Power Automate to copy necessary files.
- All external sources are extracted from the public web URLs and updated as needed.
- All data lands in an ADLS Gen 2 storage account for staging.

# Vector Table and Index

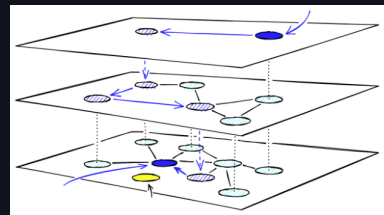
- Databricks extracts and processes the files from the ADLS account into a staging table and saves it as a Delta table in Unity Catalog
- A Vector Index Table is created from the staging table, adding on the embedding vectors using the `bge-large` embedding model in Databricks.

## Resources:

- Azure Data Lake Storage (ADLS) Gen 2
- Access Connector for Azure Databricks
- Databricks Vector Search, Foundation API, External Volumes



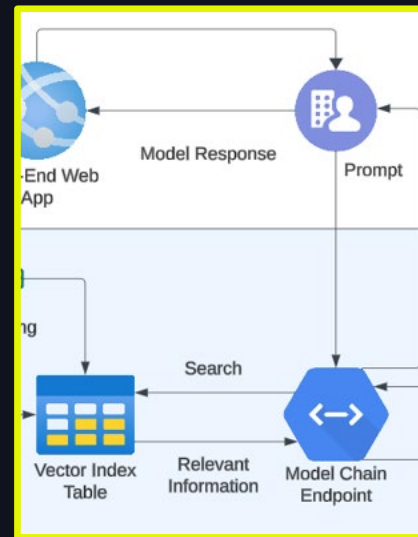
# Similarity Search



- Using the prompt from the user, a similarity search is conducted into the relevant Vector Index Table using Hierarchical Navigable Small Worlds (HNSW)
- HNSW is a method for finding similar vectors in large datasets and creates a graph of nodes and edges with multiple layers.
- Starting from the top layer, it follows edges to the closest nodes, moving down to the next layer until it finds the most similar vectors in the bottom layer.

## Resources:

- Databricks Vector Search





# Model Chain Endpoint



## Model Chain

- Custom MLFlow Pyfunc model class
- Model version stored in Unity Catalog
- Main components include retrievers, LLM selector from Foundation API, prompt builder, and predict output.

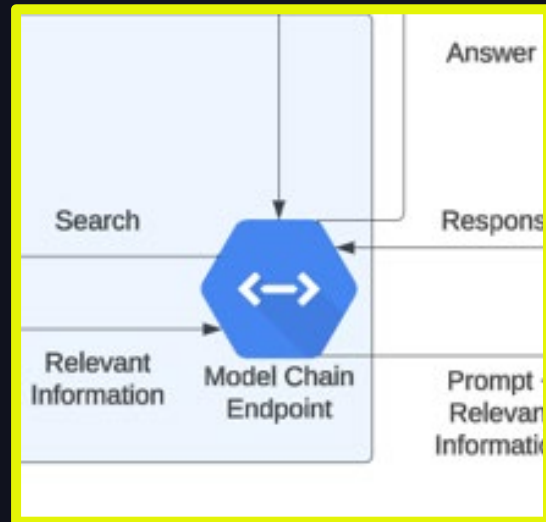


## Endpoint Configuration

- Set as CPU endpoint
- Scaled to need based on volume
- Environment variables using host URL and key with DB secrets

## Resources:

- Databricks Model Serving



# Large Language Model (LLM)

## Resources:

- Databricks Foundation API

## Prompt sent to LLM

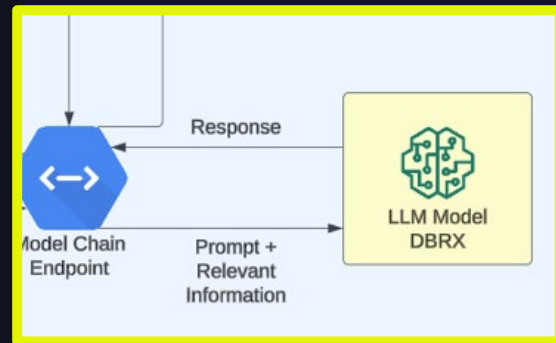
- After constructing the prompt, including relevant information from the similarity search, the Model Chain Endpoint sends this information to the LLM to craft a response.

## Results sent to Endpoint

- The response is returned to the Model Chain Endpoint with any relevant metadata, such as content filter results.

## LLM

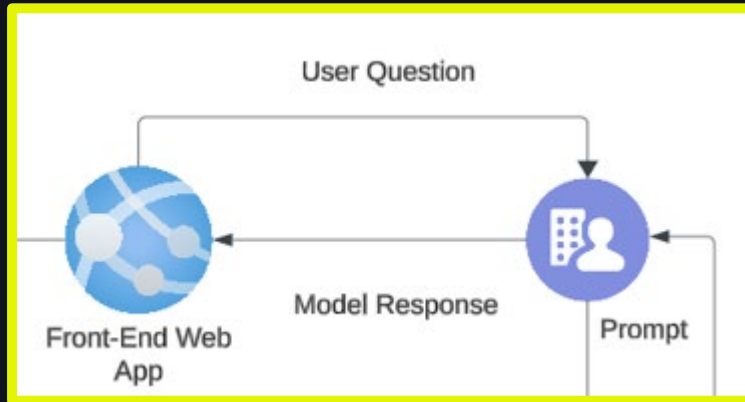
- The LLM is DBRX Instruct from Foundation API



# DevOps Flow for Application

## Resources:

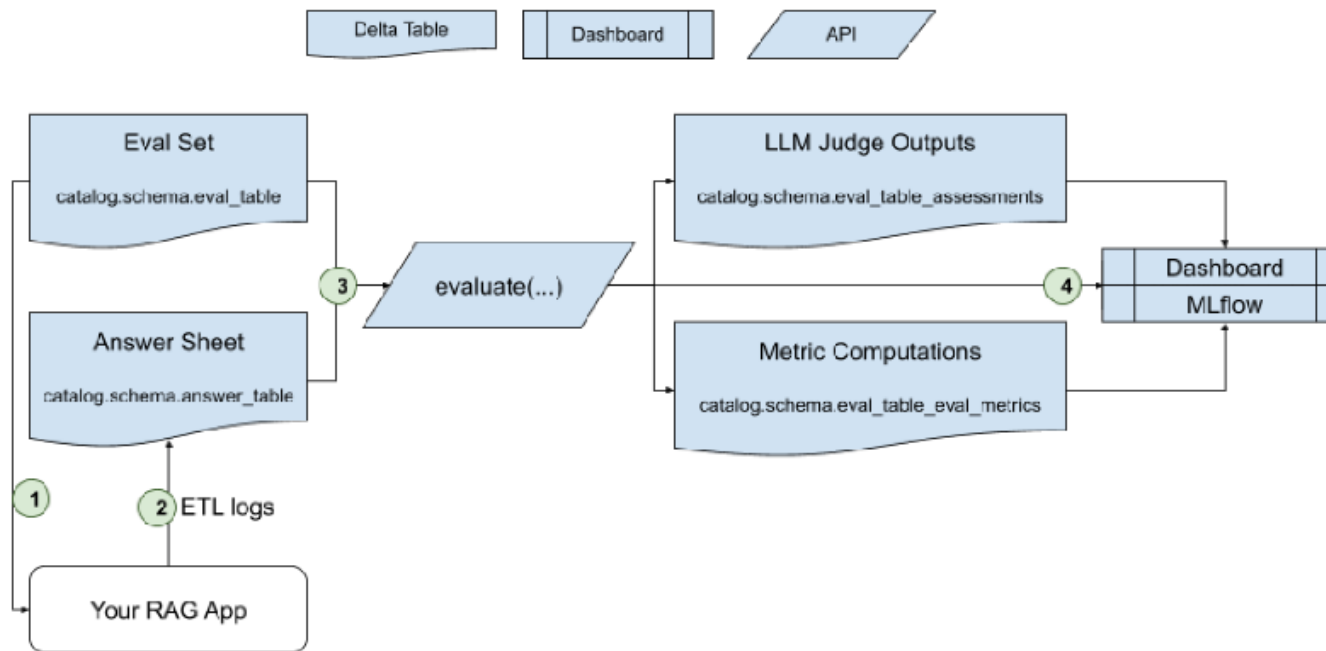
- Azure App Services



- The front-end web application allows users to interact with the Model Endpoint
- User question is sent to the model endpoint for processing. Response object comes back containing the response, rank source w/ reference paths to documents
- Source reference paths are prefixed with site information for source links
- The front-end is a Python Flask app running in Azure App Services

# Evaluation with RAG Studio

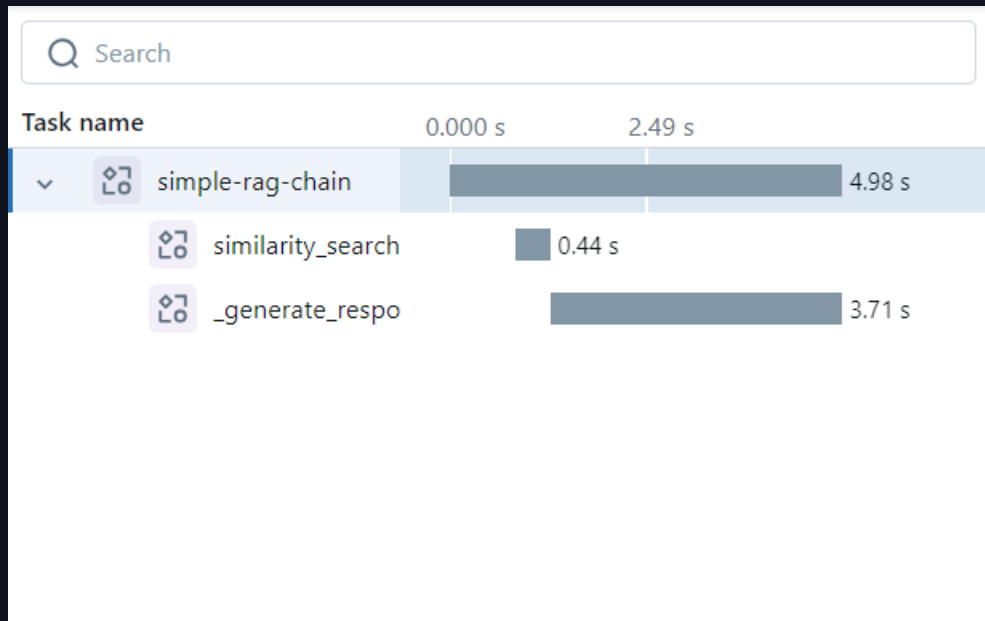
Visualization of the flow:



# Evaluation with RAG Studio

## Evaluation and Tracing

- Using LLM as a judge in conjunction with human feedback to evaluate, log, and process outcomes along with relevant metadata
- Combination of `@mlflow.trace()` decorator and `mlflow.trace()` wrapper within custom pyfunc model



## Artificial Intelligence (AI) Resources

### NDUS AI App Links



Microsoft Copilot



NDUS Policy Assistant



NDUS Audit Assistant



Lightcast Resume Optimizer

### NDUS AI App Guides



Public Affairs Prompt Guide | AI Policy Assistant



Finance and Admin Prompt Guide | AI Policy Assistant



HR Prompt Guide | AI Policy Assistant

### External AI Resources



State of North Dakota AI Policy



ICO Policies and procedures



US Gov AI Bills of Rights



National Artificial Intelligence Policies



Univ of Michigan Gen AI

## You said:

What is the process for a new employee moving reimbursement?

To summarize, the process for a new employee moving reimbursement in the North Dakota University System (NDUS) involves several policies and procedures. According to the OMB Fiscal and Admin Policy, the relocation of permanent employees is limited to \$5,000 by statute, and verification for expenses will be a paid receipt from a licensed moving company, highway mileage between duty stations, meal, and lodging receipts. For new employees, payment of moving expenses is not specifically addressed by statute, but it is recognized that market conditions for some higher-level, professional positions may require payment of moving expenses as a part of the employment agreement.

The ND Century Code allows for the payment of moving expenses for a permanent employee who has been employed in that department, institution, or agency not less than one year when that employee is transferred from one city in the state to another for the purpose of assuming a new duty assignment of a permanent nature within that department, institution, or agency. The amount must not exceed one month's salary, unless a greater amount is approved by the president or their designee.

The SBHE Policies and Procedures state that an institution may pay the moving expenses of a benefited employee hired from outside the institution, subject to definitions in SBHE Policy 703.2, by using the following: a lump-sum payment to the employee through payroll based on the amount approved by the president or designee, or compensating the employee for moving expenses through accounts payable.

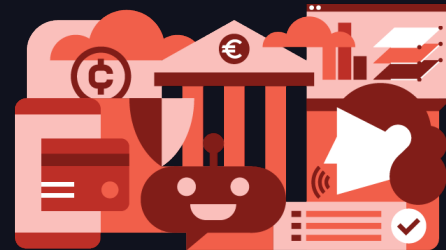
In summary, the process for a new employee moving reimbursement in the NDUS involves several policies and procedures, including the OMB Fiscal and Admin Policy, the ND Century Code, and the SBHE Policies and Procedures. The relocation of permanent employees is limited to \$5,000 by statute, and verification for expenses will be a paid receipt from a licensed moving company, highway mileage between duty stations, meal, and lodging receipts. For new employees, payment of moving expenses is not specifically addressed by statute, but it is recognized that market conditions for some higher-level, professional positions may require payment of moving expenses as a part of the employment agreement. The amount must not exceed one month's salary, unless a greater amount is approved by the president or their designee.

[1. OMB FiscalAdminPolicy](#) [2. Title44Chapter08](#) [3. Payment or Reimbursement of Meals and Other Travel or Institutional Expenses-Procedure](#) [4. Moving Expenses-Policy](#)

Enter prompt here...

Send 

# Challenges and Impact



## Challenges and Barriers

- Small AI team (2 people)
- Limited financial resources
- Limited on-prem compute resources



## Solution Impact

- 10-20x Faster using a comprehensive LLM for policy search and response
- 2x Decrease in time to market from 1 year to 6 months
- 100% Reduced procurement time and costs by leveraging Databricks on Azure



# Future Development

# AI Roadmap

- Internal Audit and Policy Assistant Enhancements (Ongoing)
- Automated NDUS News Pipeline powered by LLMs (July 2024)
- Data Request Assistant for Student, Financial, and HR data (Jan 2025)



# DATA+AI SUMMIT

Thank you:

NDUS Leadership

Databricks Team

Databricks Assistant



Questions?