AI in SRE

Unlocking Prometheus Insights with Natural Language

David Asamu

Conf42 SRE2025

OUTLINE

01 INTRODUCTION
- The Problem
- Our Solution

02 APPROACH
- Architecture
- Key Components

03 PROMCHAT
- With Node Exporter
- With Custom Exporter

04 RESULT
- Lessons
- Limitations
- Future Plans
- Contributing

05 CLOSING
- Contact
- Thank You

01

Introduction

- The Problem
- Our Solution

Understanding the Problem & Motivation

- The Need for Faster Incident Response
- Democratizing Monitoring Data Access
- The PromQL Learning Curve
- Inspired by "Chat with Your Data"

The Solution: Natural Language to PromQL

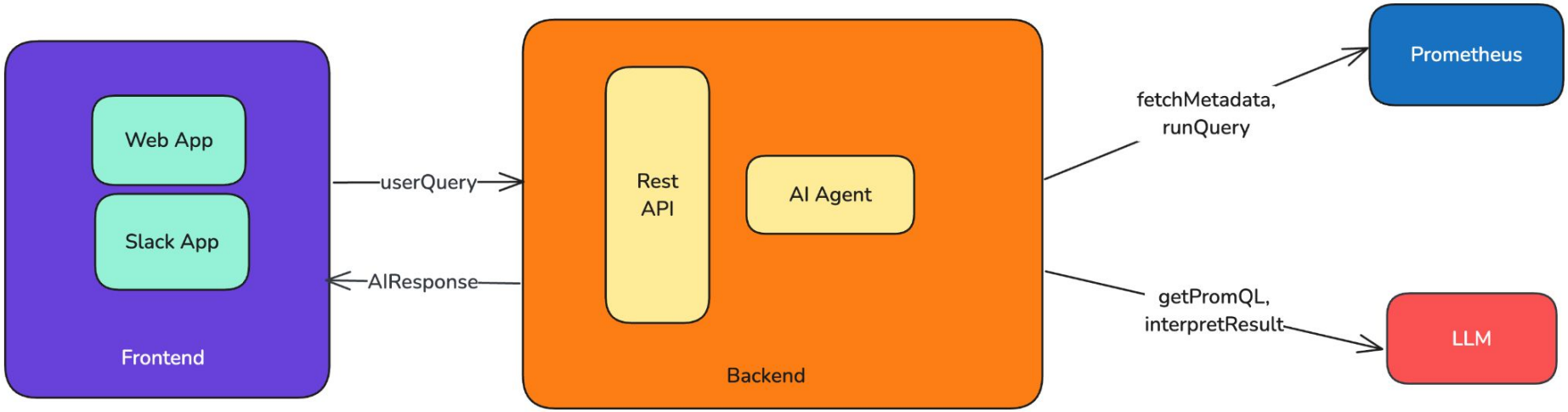
- Ask Questions in Natural Language
- PromQL query generated from Natural Language
- Result fetched from prometheus using generated PromQL Query
- Result presented back in Natural Language

02

Approach

- Architecture Overview
- Key Components Review

Architecture Overview



Prometheus

Notation

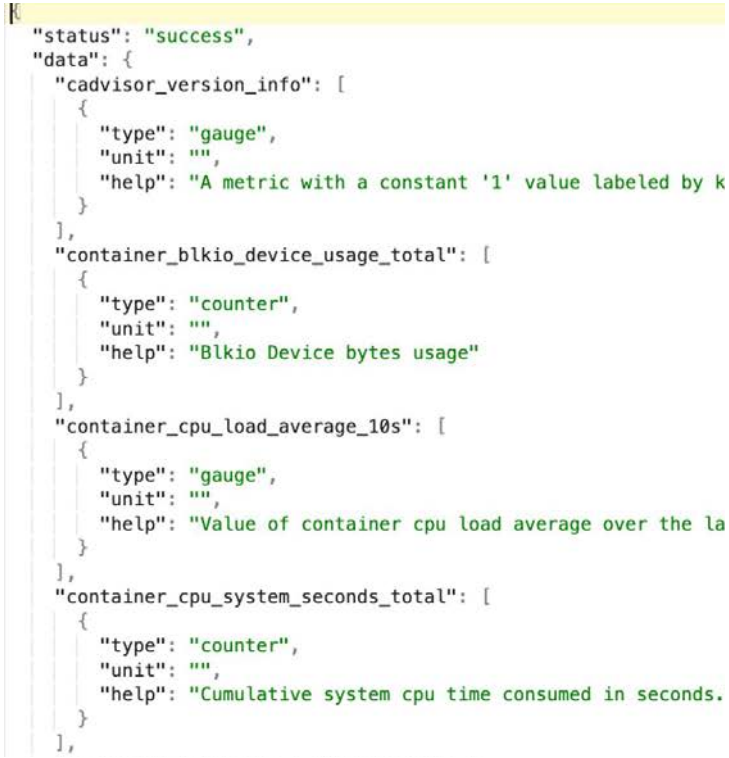
Typically written in the format:

<metric name>{<label name>=<label value>, ...}

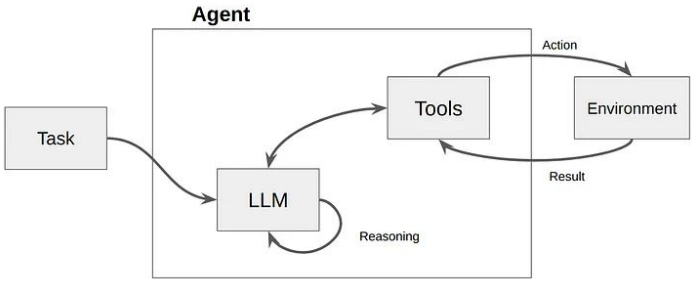
For example:

api_http_requests_total{method="POST", handler="/messages"}

/api/v1/metadata



LLM TOOLs



LLM Tool

```python
def _query_prometheus_tool(self, promql_query, source):
        """Wrapper for query_prometheus to return string output for tool."""
        try:
            data_handler = self.data_handlers.get(source)
            if not data_handler:
                return f"Error: unknown data source: {source}"
            result = data_handler.query_prometheus(promql_query)
            return json.dumps(result) # Convert result to JSON string for tool output
        except Exception as e:
            return f"Error querying Prometheus: {e}"
```

LLM Tool

```python
def _get_metric_metadata_tool(self, source):
        """Wrapper for get_metric_metadata to format output for LLM tool use."""
        try:
            data_handler = self.data_handlers.get(source)
            if not data_handler:
                return f"Error: unknown data source: {source}"
            metadata_description = data_handler.get_metric_metadata_description_for_llm()
            return metadata_description
        except Exception as e:
            return f"Error retrieving Prometheus metric metadata: {e}"
```

03

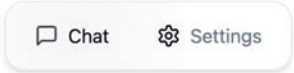
PromChat

- Node Exporter
- Custom Exporter

NODE EXPORTER

Sample metrics and Queries

rate(node_cpu_seconds_total{ mode="system"}[1m])	The average amount of CPU time spent in system mode, per second, over the last minute (in seconds)
node_filesystem_avail_bytes	The filesystem space available to non-root users (in bytes)
rate(node_network_receive_bytes_total[1m])	The average network traffic received, per second, over the last minute (in bytes)



Chat with Metrics

Clear History

Hello! I'm your Prometheus metrics assistant. How can I help you today?

what is the current cpu utilization of the node?

The current CPU utilization of the node is 6.4%.

what about the battery percentage for the node?

The battery percentage for the node is 100%.

what OS is the node running?

The node is running macOS.

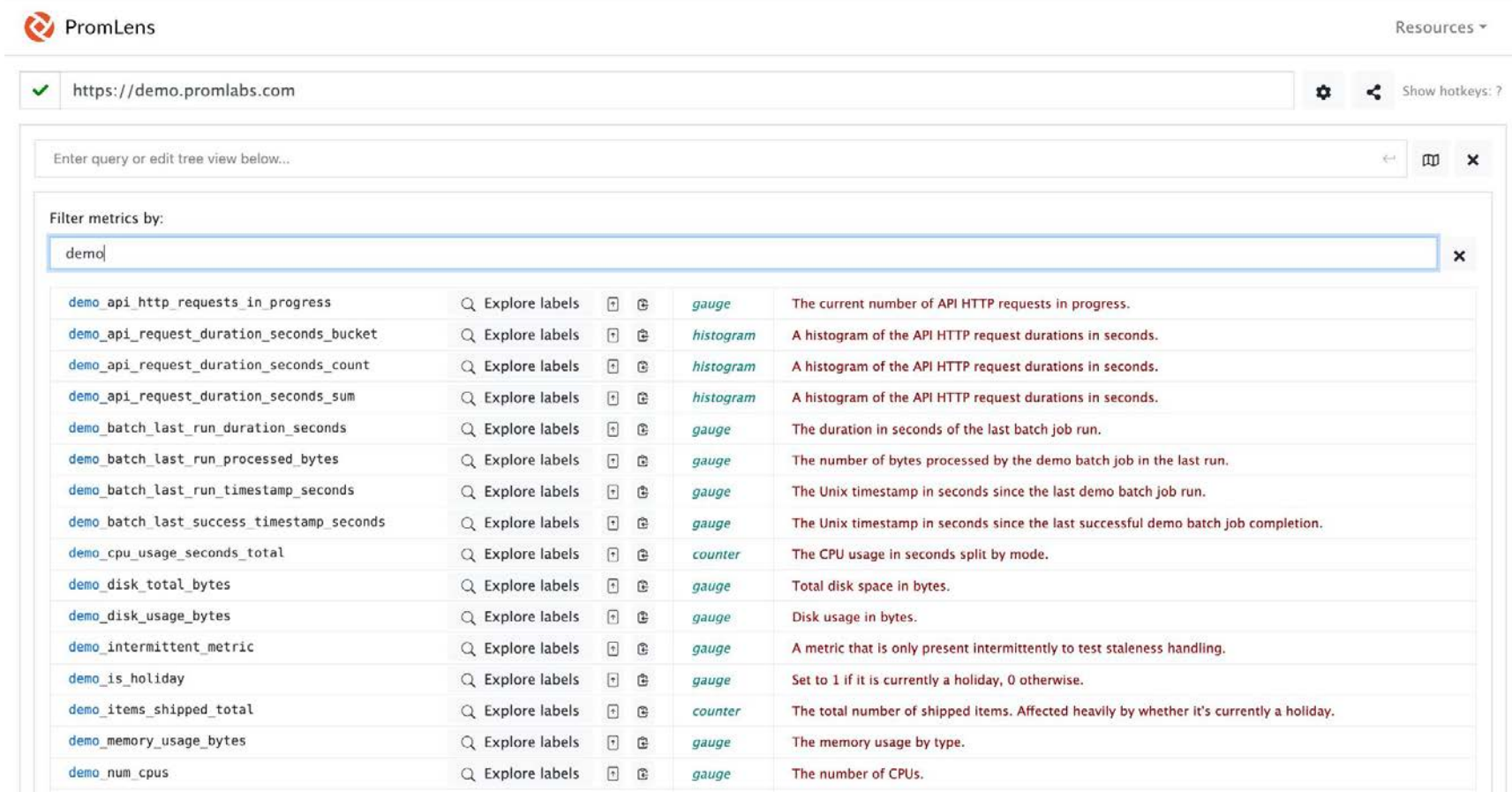
how much free disk space is available on the node?

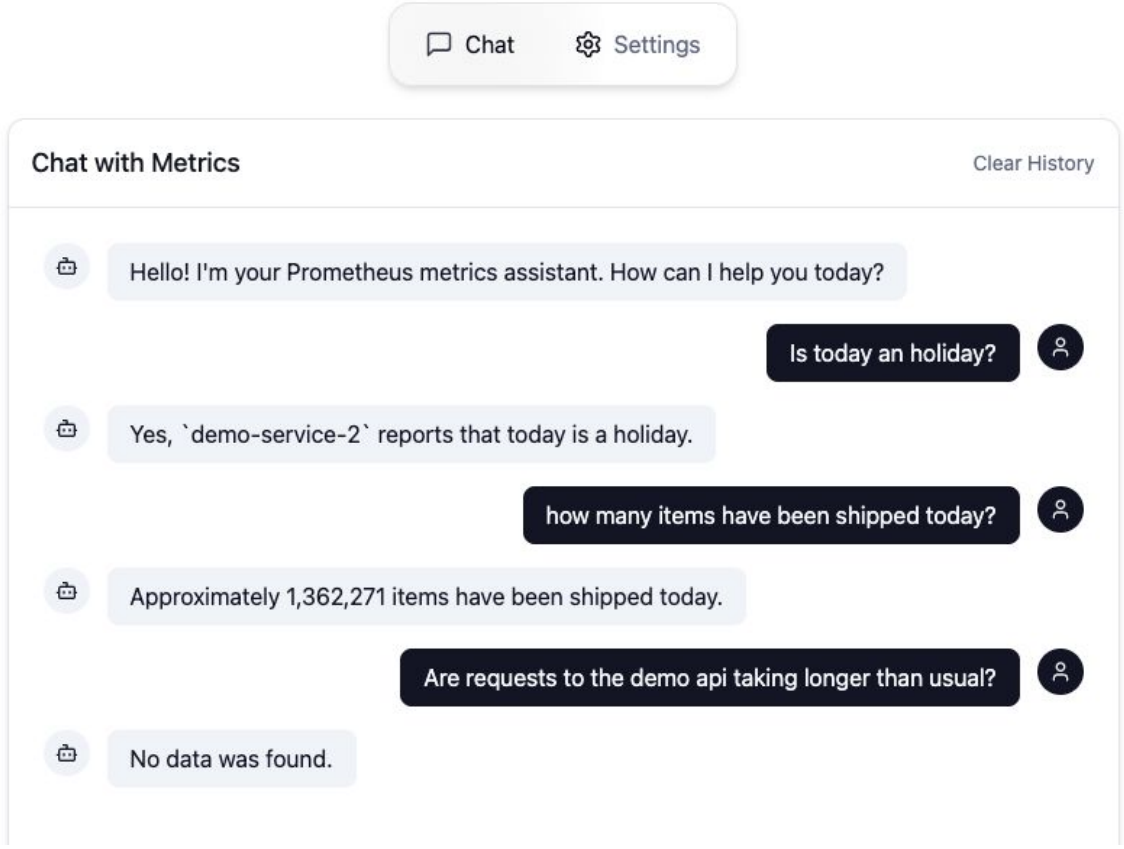
The total free disk space across all nodes is 2.38 TB.

what is the current memory utilization of the node?

The current memory utilization of the node davids_macbook_pro is 99.64%.

CUSTOM EXPORTER




Chat
Settings
Chat with Metrics
Clear History
Hello! I'm your Prometheus metrics assistant. How can I help you today?
Is today an holiday?
Yes, `demo-service-2` reports that today is a holiday.
how many items have been shipped today?
Approximately 1,362,271 items have been shipped today.
Are requests to the demo api taking longer than usual?
No data was found.

Behind the Scene

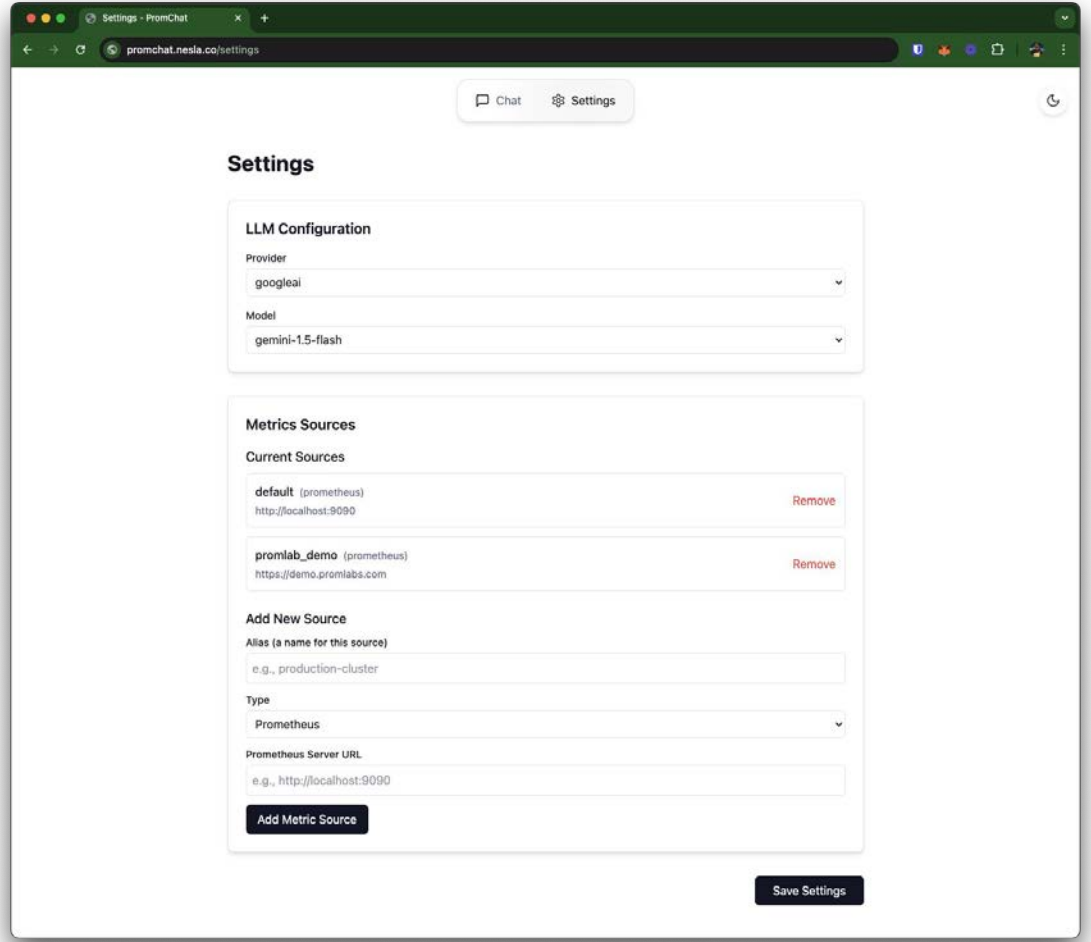
INFO:main:**User query: Is today an holiday?**
INFO:agent.agent:**Generated PromQL Query: demo_is_holiday**
INFO:agent.agent:**Prometheus Result: [{"metric": {"__name__": "demo_is_holiday", "instance": "demo-service-2:10002", "job": "demo"}, "value": [1744152905.074, "1"]}]**
INFO:main:**Agent response: Yes, `demo-service-2` reports that today is a holiday**

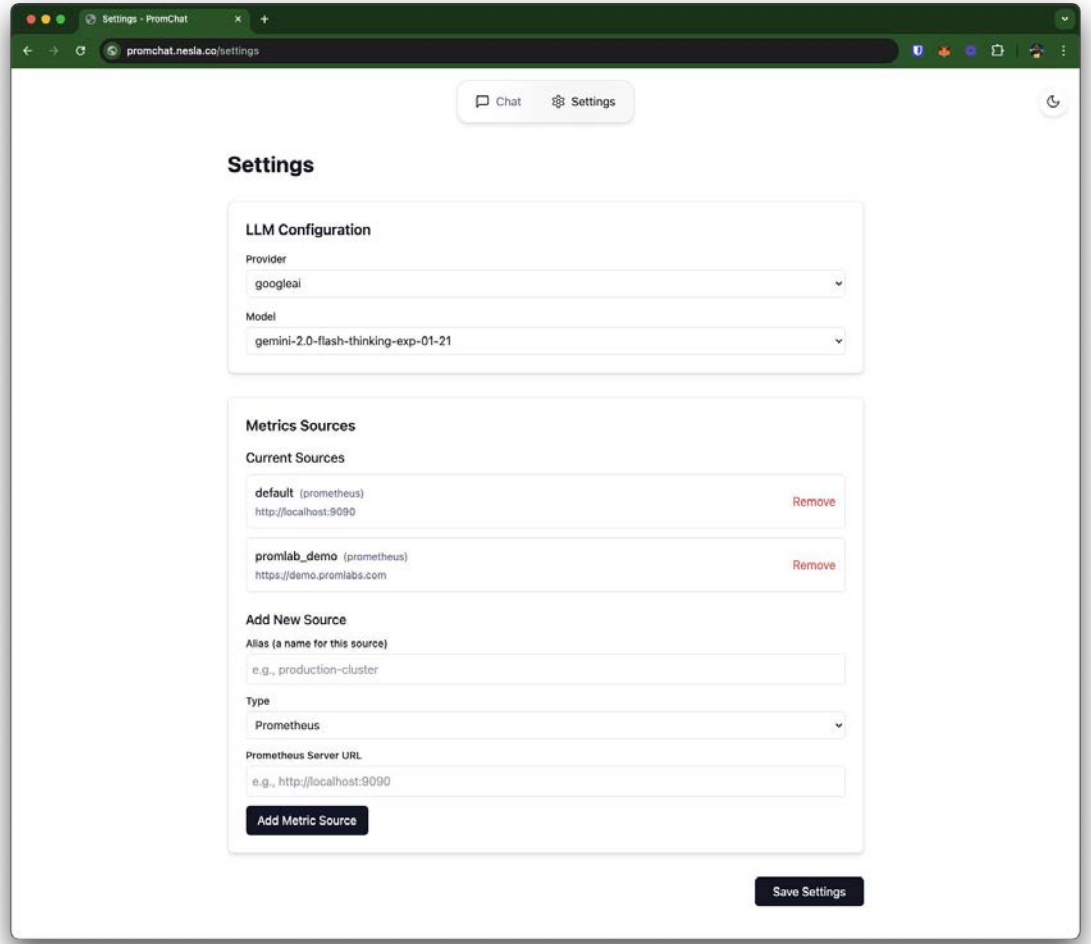
Behind the Scene

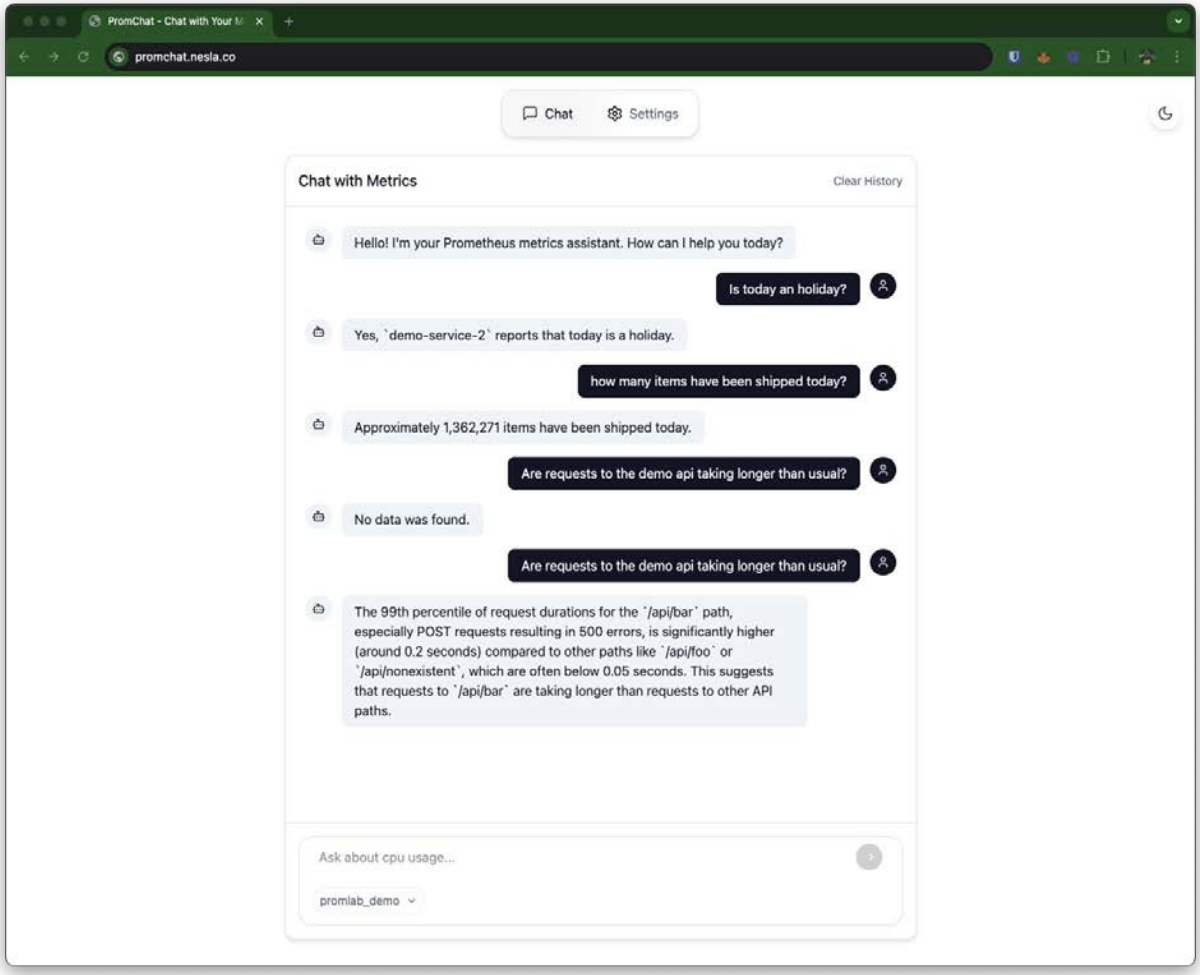
INFO:main:**User query: how many items have been shipped today?**
INFO:agent.agent:**Generated PromQL Query: increase(demo_items_shipped_total[1d])**
INFO:agent.agent:**Prometheus Result: [{"metric": {"instance": "demo-service-0:10000", "job": "demo"}, "value": [1744152970.506, "455360.05556520226"]}, {"metric": {"instance": "demo-service-1:10001", "job": "demo"}, "value": [1744152970.506, "453045.6537593332"]}, {"metric": {"instance": "demo-service-2:10002", "job": "demo"}, "value": [1744152970.506, "453865.79614516406"]}]**
INFO:main:**Agent response: Approximately 1,362,271 items have been shipped today**

Behind the Scene

INFO:main:**User query: Are requests to the demo api taking longer than usual?**
INFO:agent.agent:**Generated PromQL Query:** **increase(demo_api_request_duration_seconds_sum[5m]) >** **increase(demo_api_request_duration_seconds_count[5m]) *** **avg_over_time(demo_api_request_duration_seconds_bucket[5m])**
INFO:agent.agent:**Prometheus Result: [ ]**
INFO:main:**Agent response: No data was found**


Settings - PromChat
promchat.nesla.co/settings
Chat
Settings
Settings
LLM Configuration
Provider
googleai
Model
gemini-1.5-flash
Metrics Sources
Current Sources
default (prometheus)
http://localhost:9090
Remove
promlab_demo (prometheus)
https://demo.promlabs.com
Remove
Add New Source
Alias (a name for this source)
e.g., production-cluster
Type
Prometheus
Prometheus Server URL
e.g., http://localhost:9090
Add Metric Source
Save Settings


Settings - PromChat
promchat.nesla.co/settings
Chat
Settings
Settings
LLM Configuration
Provider
googleai
Model
gemini-2.0-flash-thinking-exp-01-21
Metrics Sources
Current Sources
default (prometheus)
http://localhost:9090
Remove
promlab_demo (prometheus)
https://demo.promlabs.com
Remove
Add New Source
Alias (a name for this source)
e.g., production-cluster
Type
Prometheus
Prometheus Server URL
e.g., http://localhost:9090
Add Metric Source
Save Settings


PromChat - Chat with Your M
promchat.nesla.co
Chat
Settings
Chat with Metrics
Clear History
Hello! I'm your Prometheus metrics assistant. How can I help you today?
Is today an holiday?
Yes, `demo-service-2` reports that today is a holiday.
how many items have been shipped today?
Approximately 1,362,271 items have been shipped today.
Are requests to the demo api taking longer than usual?
No data was found.
Are requests to the demo api taking longer than usual?
The 99th percentile of request durations for the `/api/bar` path, especially POST requests resulting in 500 errors, is significantly higher (around 0.2 seconds) compared to other paths like `/api/foo` or `/api/nonexistent`, which are often below 0.05 seconds. This suggests that requests to `/api/bar` are taking longer than requests to other API paths.
Ask about cpu usage...
promlab_demo

Behind the Scene

INFO:main:**User query: Are requests to the demo api taking longer than usual?**

INFO:agent.agent:**Generated PromQL Query: histogram_quantile(0.99, rate(demo_api_request_duration_seconds_bucket[5m]))**

INFO:agent.agent:**Prometheus Result: [{"metric": {"instance": "demo-service-2:10002", "job": "demo", "method": "GET", "path": "/api/nonexistent", "status": "404"}, "value": [1744153995.69, "0.00099000000000000001"]}, {"metric": {"instance": "demo-service-1:10001", "job": "demo", "method": "POST", "path": "/api/bar", "status": "500"}, "value": [1744153995.69, "0.21946694418525692"]}, ...]**

INFO:main:**Agent response: The 99th percentile of request durations for the `/api/bar` path, especially POST requests resulting in 500 errors, is significantly higher (around 0.2 seconds) compared to other paths like `/api/foo` or `/api/nonexistent`, which are often below 0.05 seconds. This suggests that requests to `/api/bar` are taking longer than requests to other API paths.**

04

Result

- Lessons
- Limitations
- Future Plans
- Contributing

RESULT SUMMARY

Lessons from implementing the PoC

LESSONS

- LLMs are capable of writing correct PromQL without the need for additional training or fine tuning
- One short prompting to avoid response validation
- Thinking models write better PromQL

FUTURE IMPROVEMENTS

- Support more complex queries
- Support more metrics sources
- Learn from user interactions

LIMITATIONS

- Labels missing in metadata documentation
- Inconsistency on the generated queries

CONTRIBUTING

- Source code is available on github at https://github.com/neslahq/promchat
- Issues, PRs are welcomed.
- Live demo available at: https://promchat.nesla.co

THANK YOU

Mail: hello@davidasamu.com

Blog: https://doa.sh

X: @atechbruv

LinkedIn: https://www.linkedin.com/in/david-asamu/

David Asamu