

NeuBird.ai

Building Agentic Systems



Gou Rao, Co-Founder

Evolution of GenAI in the Enterprise

Enterprise Chat Bots

Software Applications like Salesforce, Mongo etc will add chat bots

Enterprise Search Applications like Glean can enable chats with your enterprise data.

GenAI platforms

LLM platforms from OpenAI, Accenture, Hugging Face allow enterprises to build new AI powered applications to provide enhanced products to their customers



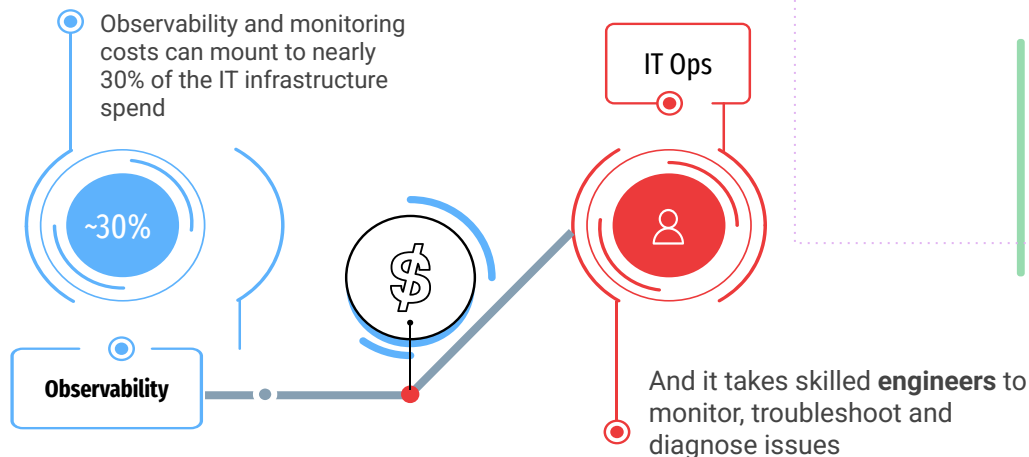
Agentic Systems

Startups are enhancing the enterprise workforce with specific, purpose build reasoning systems for specific job functions



The Business Impact of slow IT Operations

Running **production IT** operations is a hard business



Modern IT stack too **complex** to keep up with

Tickets keep piling up

Impedes Business Productivity

The Current IT Stack is too **Complex**

Too Many Data Sources

Log Files, Traces, Metrics,
Alerts, Infra Config



Extract Transform
and Load into
Analytics DB (Elastic)

Yet Another DB

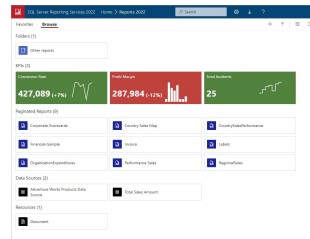
```
sql
SELECT
  books.title,
  authors.author_name,
  COUNT(book_loans.book_id) AS total_loans
FROM
  books
  INNER JOIN
    book_authors ON books.book_id = book_authors.book_id
  INNER JOIN
    authors ON book_authors.author_id = authors.author_id
  LEFT JOIN
    book_loans ON books.book_id = book_loans.book_id
WHERE
  books.publication_year >= 2000
GROUP BY
  books.title, authors.author_name
HAVING
  COUNT(book_loans.book_id) > 10
ORDER BY
  total_loans DESC, authors.author_name ASC;
```

Create complex
queries

Database Engineering

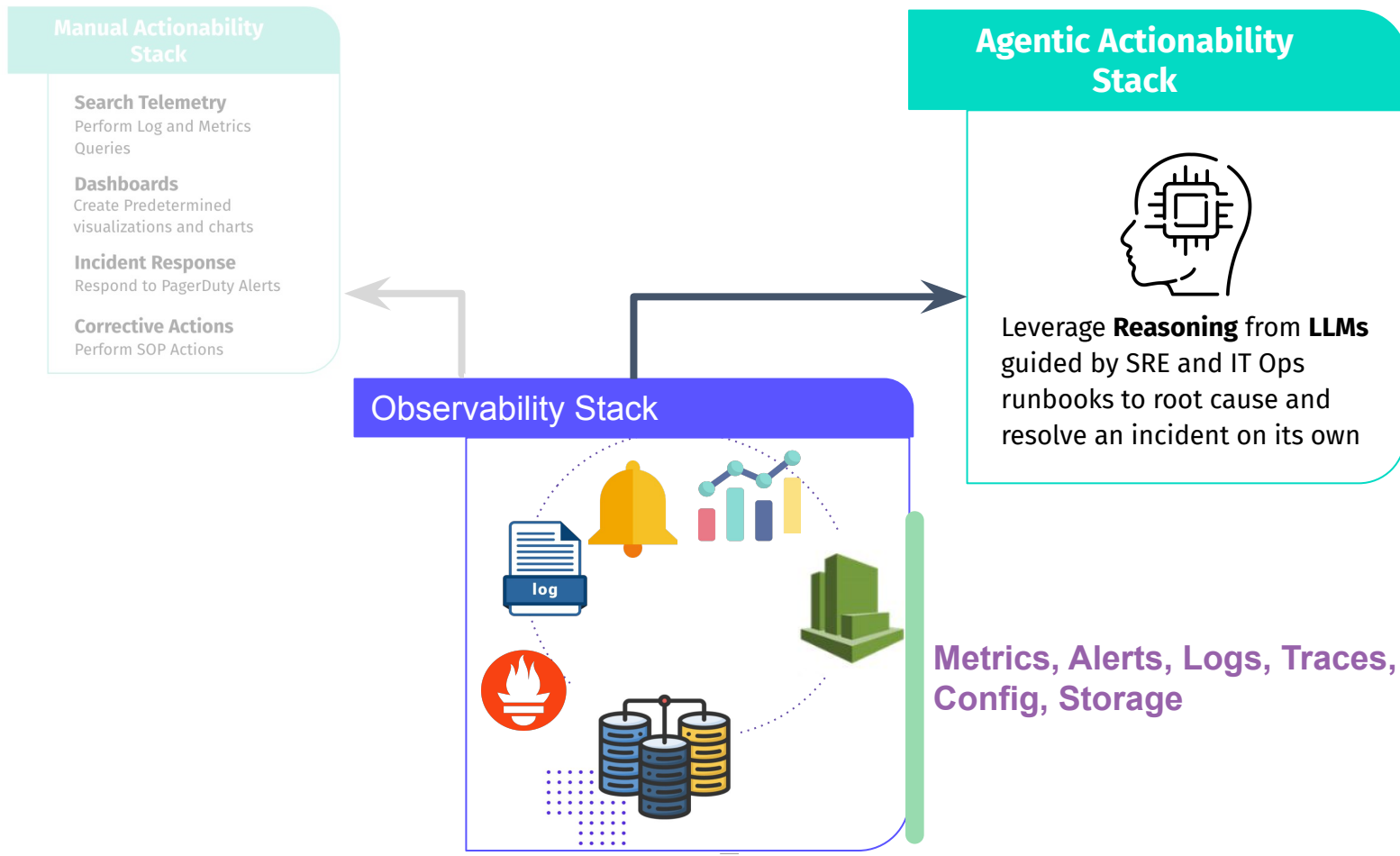
Visualization

Integrate the results
into intuitive
dashboards



Too many layers of **complex**
telemetry

Bringing **Agentic** Workflows to IT Operations

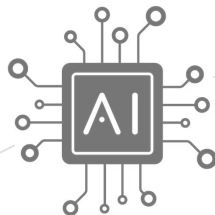
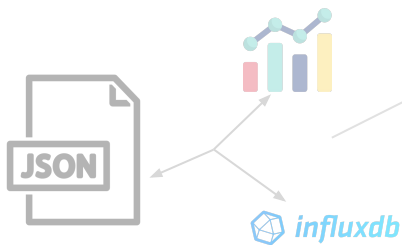


Use **REASONING** from the LLMs to build Agents

Use the **reasoning** capability of LLMs to solve IT operations issues in **real-time**

LLMs can query your observability tools directly

LLMs can mimic an experienced SRE's thought process in just minutes



"AI is an **amplifier** of human intelligence & when people are smarter, better things happen: people are more **productive**, happier & the economy thrives."

- **Yann LeCun**, Head of Meta AI, **Turing Award**

Meet Hawkeye - The First Agentic IT Ops Engineer



Part of your IT workforce

Works alongside IT staff
as an assistant to help
take on mundane tasks

Autonomously or
Collaboratively resolve IT
tickets and issues

01

Predetermined Dashboards

Dashboards and problem scenarios are **prematurely** determined ahead of time

02

Uncorrelated

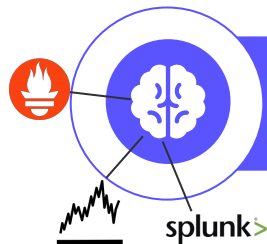
Information is from unrelated data sources - **causality** has to be inferred visually

03

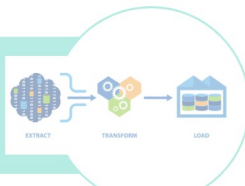
Decisions Not in Real Time

Data has to be manually correlated and analyzed increasing the time to response

Traditional Observability



AI



Generative IT Ops

01

Inplace and Interactive Analytics

Data is analyzed at the source, in **real time**

02

Generative AI assisted

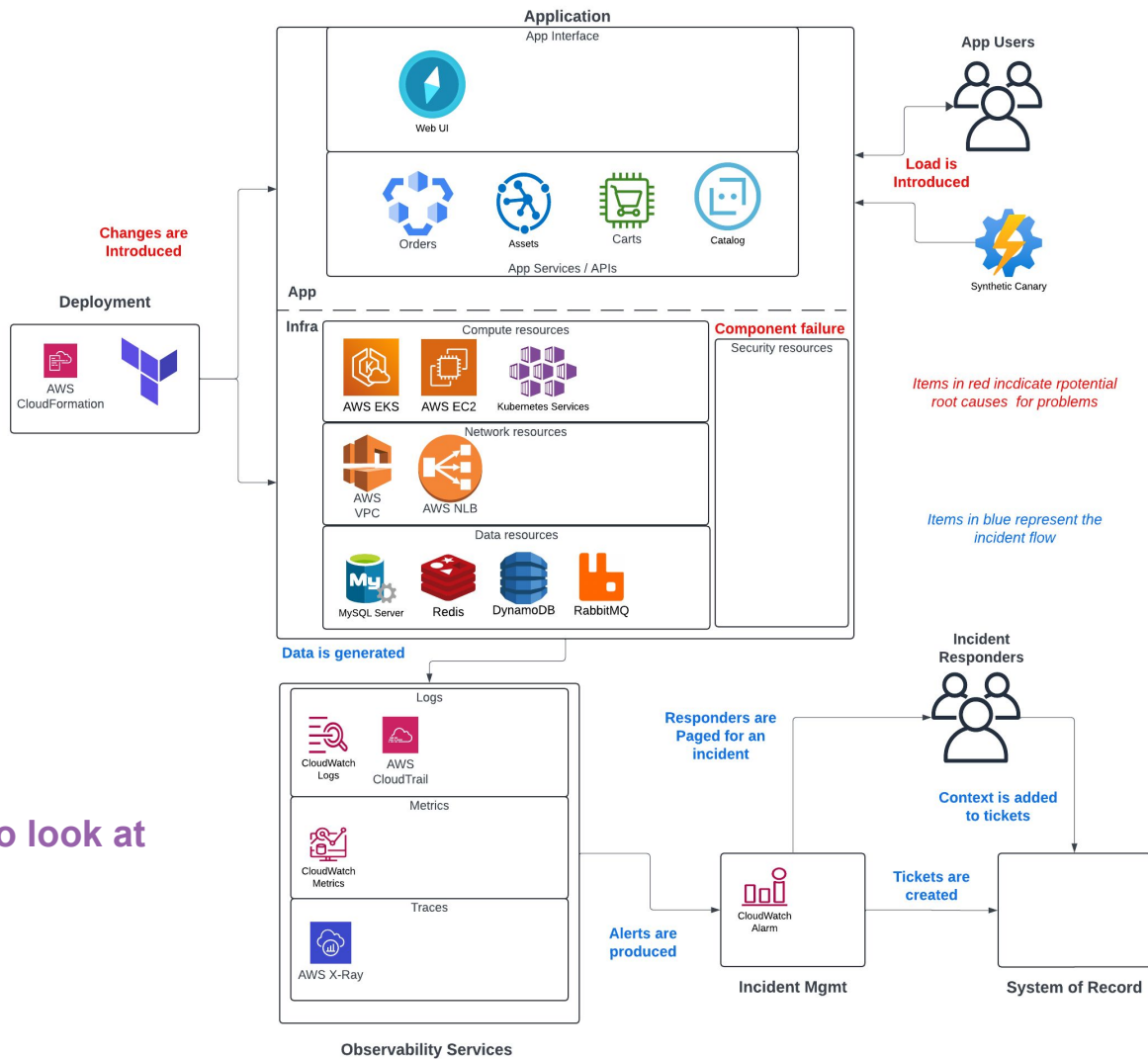
Schematic inferences and queries are **cognitively** generated - with visual feedback

03

Multi Modal Capable

AI generated queries are federated across multiple data sources and the results are further combined by **LLMs**

Demo



Take a moment to look at the environment

Example - E-Commerce SRE @ ACME.com

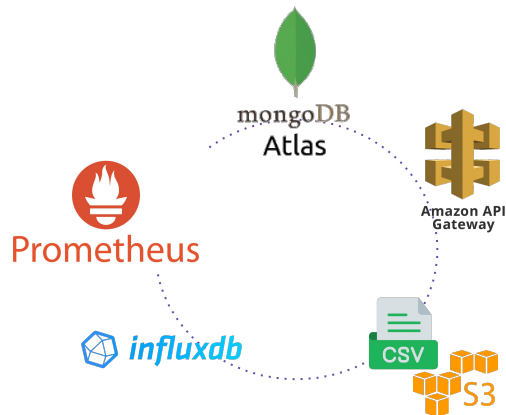
Problem: SRE notices that the ACME.com website is **performing** poorly on populating shopping cart suggestions

Causality: AI recognizes increased Daily Active Users (DAU)

Collaborative: AI suggests Mongo Atlas is a problem

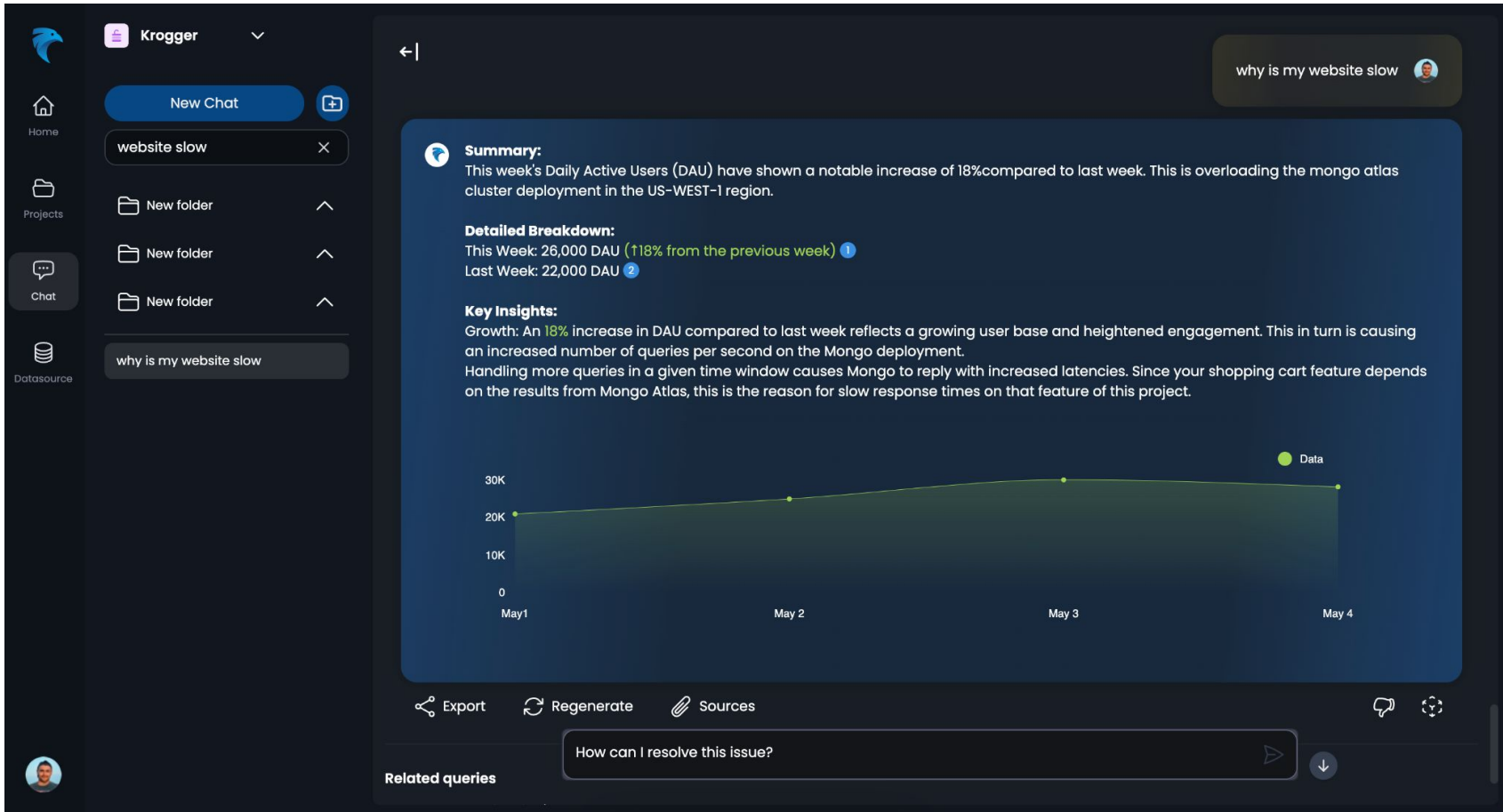
Conclusive: AI provides a possible solution

Data Sources Consulted:




Customer Example:

SRE notices slow website responses



Customer Example:

Wants to know how they can resolve the issue



Krogger

Home

Projects

Chat

Datasource

New Chat

website slow

New folder

New folder

New folder

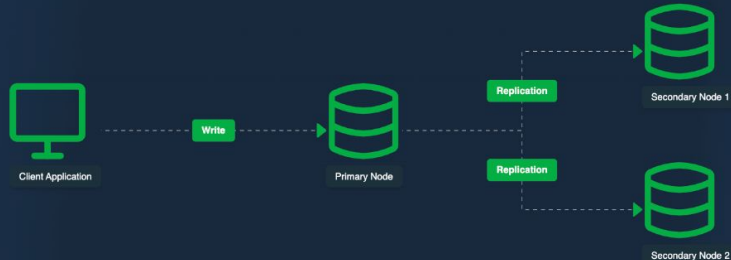
why is my website slow

Export Regenerate Sources

How can I resolve this issue?

Summary:
After analyzing the atlas logs, it seems that the most active queries to Atlas happens between 3PM to 5PM PST. You need to increase the number of replication nodes to be able to service an increased number of queries per minute during these hours. You can utilize auto-scaling rules to add and remove replicas.

Detailed Breakdown:
This Week: 26,000 DAU (↑18% from the previous week)
Last Week: 22,000 DAU



```
graph LR;
    CA[Client Application] -- Write --> PN[(Primary Node)];
    PN -. Replication .-> SN1[(Secondary Node 1)];
    PN -. Replication .-> SN2[(Secondary Node 2)];
```

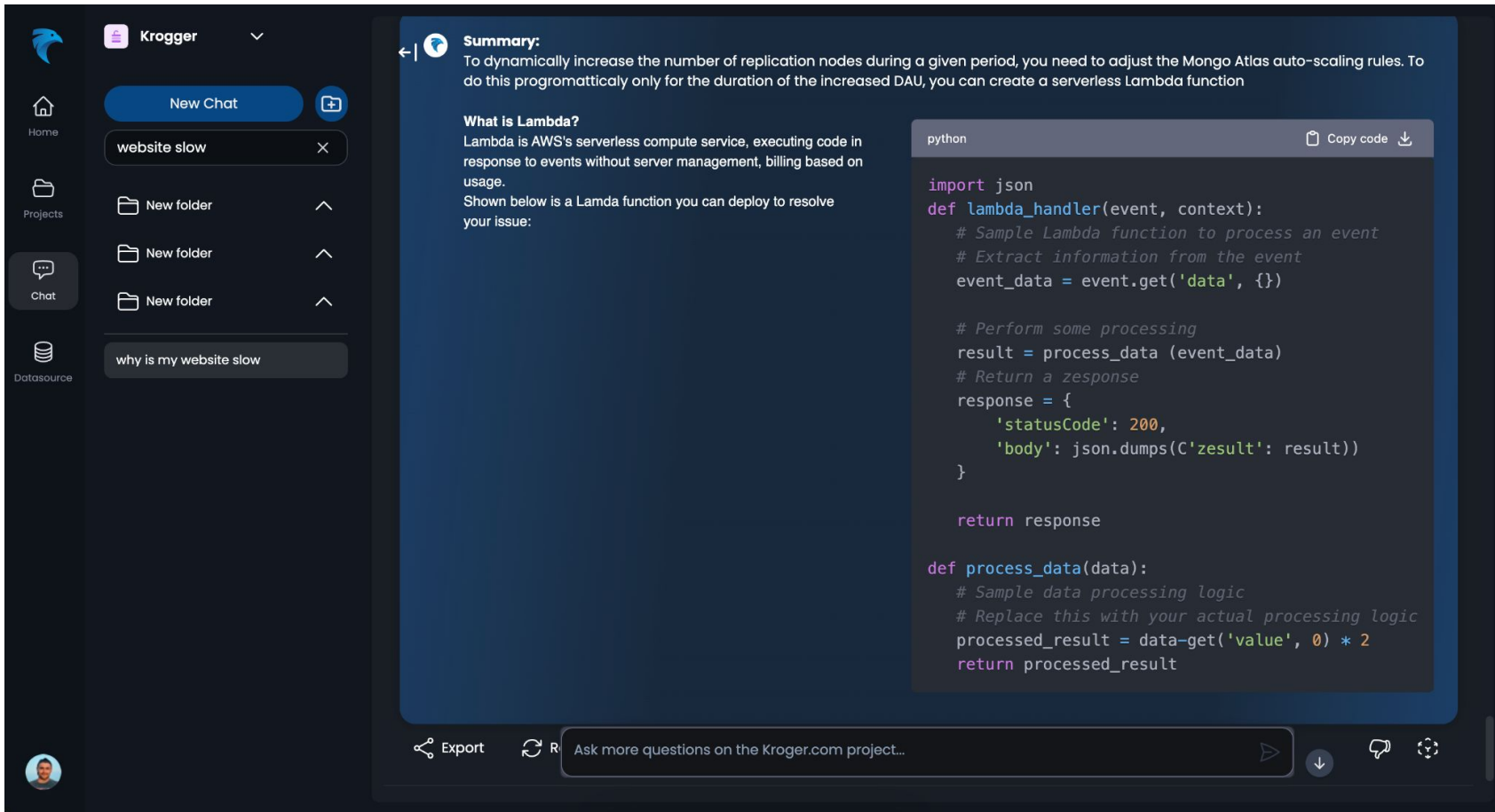
Export Regenerate Sources

show me how to setup autoscaling for that period

Related queries

Customer Example:

And gets a real time solution



The screenshot displays the Neubird AI interface. On the left sidebar, there are navigation icons for Home, Projects, Chat, and Datasource. The 'Krogger' project is selected, showing a 'New Chat' button and a list of chat sessions. One session titled 'website slow' is active, with a summary of the issue: 'why is my website slow'. The main chat area shows a 'Summary' section explaining that to dynamically increase the number of replication nodes during a given period, one needs to adjust the Mongo Atlas auto-scaling rules. It also provides a brief explanation of Lambda as AWS's serverless compute service. Below the summary, a Python code snippet for a Lambda handler is displayed, which processes data and returns a response with a status code of 200. The code is as follows:

```
python
import json
def lambda_handler(event, context):
    # Sample Lambda function to process an event
    # Extract information from the event
    event_data = event.get('data', {})

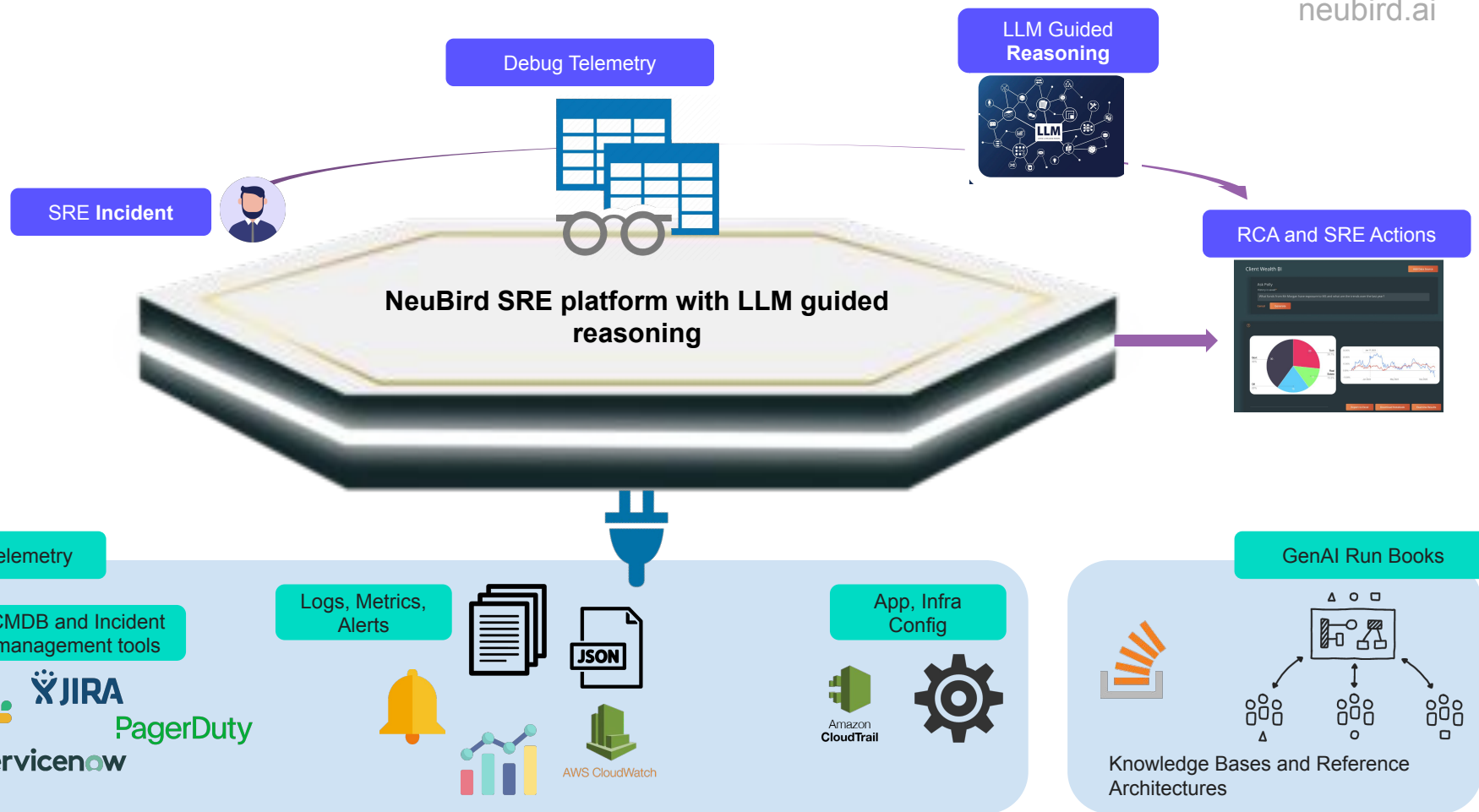
    # Perform some processing
    result = process_data(event_data)
    # Return a response
    response = {
        'statusCode': 200,
        'body': json.dumps(C'zresult': result))
    }

    return response

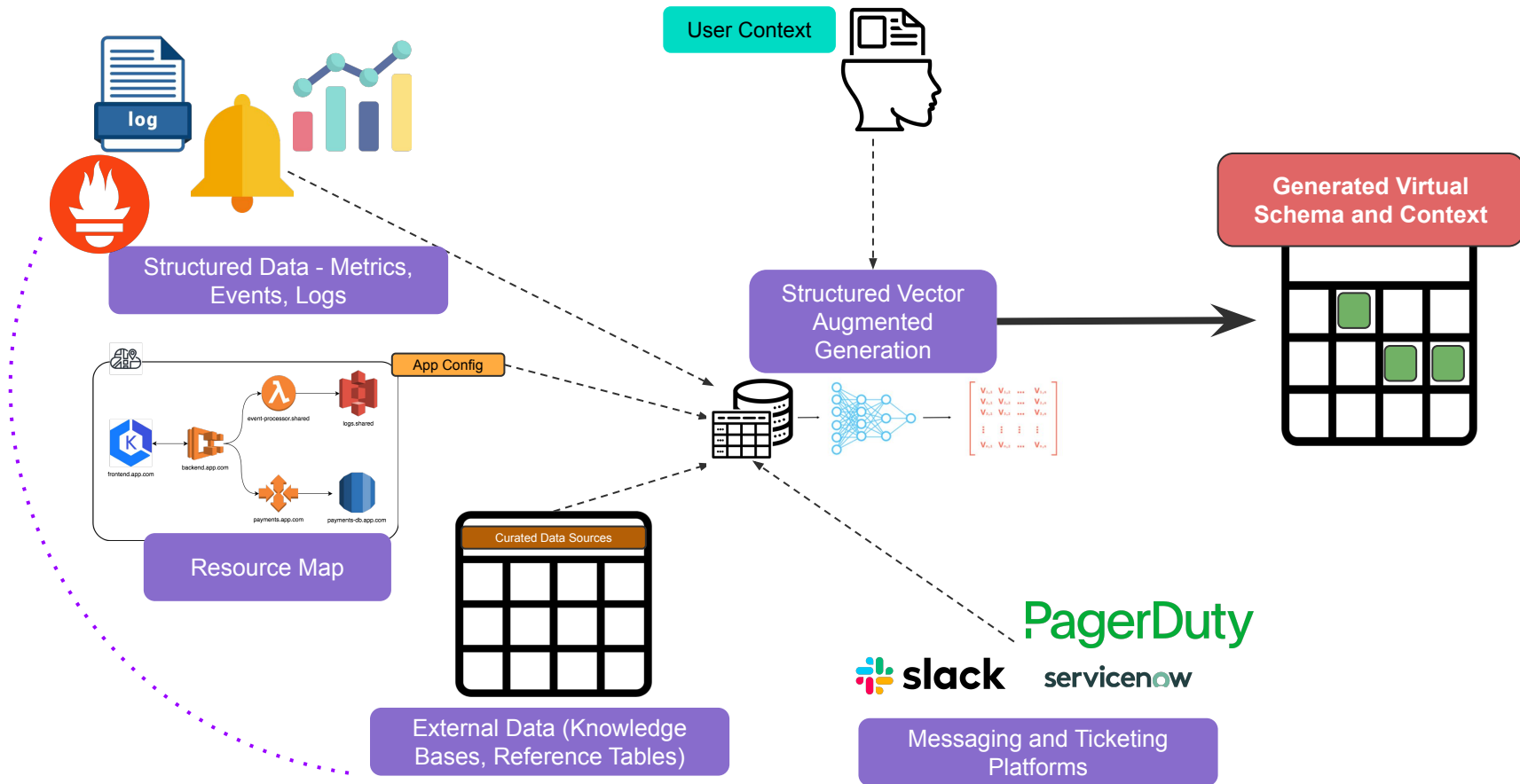
def process_data(data):
    # Sample data processing logic
    # Replace this with your actual processing logic
    processed_result = data.get('value', 0) * 2
    return processed_result
```

At the bottom of the interface, there is an input field with the text 'Ask more questions on the Krogger.com project...' and buttons for 'Export', 'R', and a search icon.

SRE as a Service using all your IT telemetry and observability tools



The Science Behind it: **Cognitive** Views for IT Telemetry



RAEL: A purpose built IT coding language for Agentic Systems

```
from: aws_cloudwatch
filter: "service=EKS AND logGroup=NodeStatus AND severity >= ERROR"
time_range: "last 15 minutes"
}

# Query Splunk for application error rates
query app_error_rate {
  from: splunk_observability
  filter: "app_name=my_microservice AND metric=error_rate AND status_code >= 500"
  time_range: "last 5 minutes"
}

# Query Datadog for high CPU usage
query cpu_alerts {
  from: datadog
  filter: "metric=cpu.utilization AND value > 90"
  time_range: "last 10 minutes"
}

# Decision logic
if eks_node_health.errors > 10 then
  action "Send alert to SRE team: EKS nodes experiencing frequent errors."

if app_error_rate.value > 5 then
  action "Auto-scale microservice: Increase instance count by 2."
```

Grammar and Syntax

```
PROGRAM ::= { SOURCE_DEFINITION | QUERY_DEFINITION | DECISION_BLOCK } EOF

SOURCE_DEFINITION ::= "source" IDENTIFIER "{"
  "type:" STRING
  "region:" STRING?
  "role_arn:" STRING?
  "endpoint:" STRING?
  "token:" STRING?
  "api_key:" STRING?
  "app_key:" STRING?
  "}"

QUERY_DEFINITION ::= "query" IDENTIFIER "{"
  "from:" IDENTIFIER
  "filter:" STRING
  "time_range:" STRING
  "}"

DECISION_BLOCK ::= "if" CONDITION "then" ACTION_BLOCK

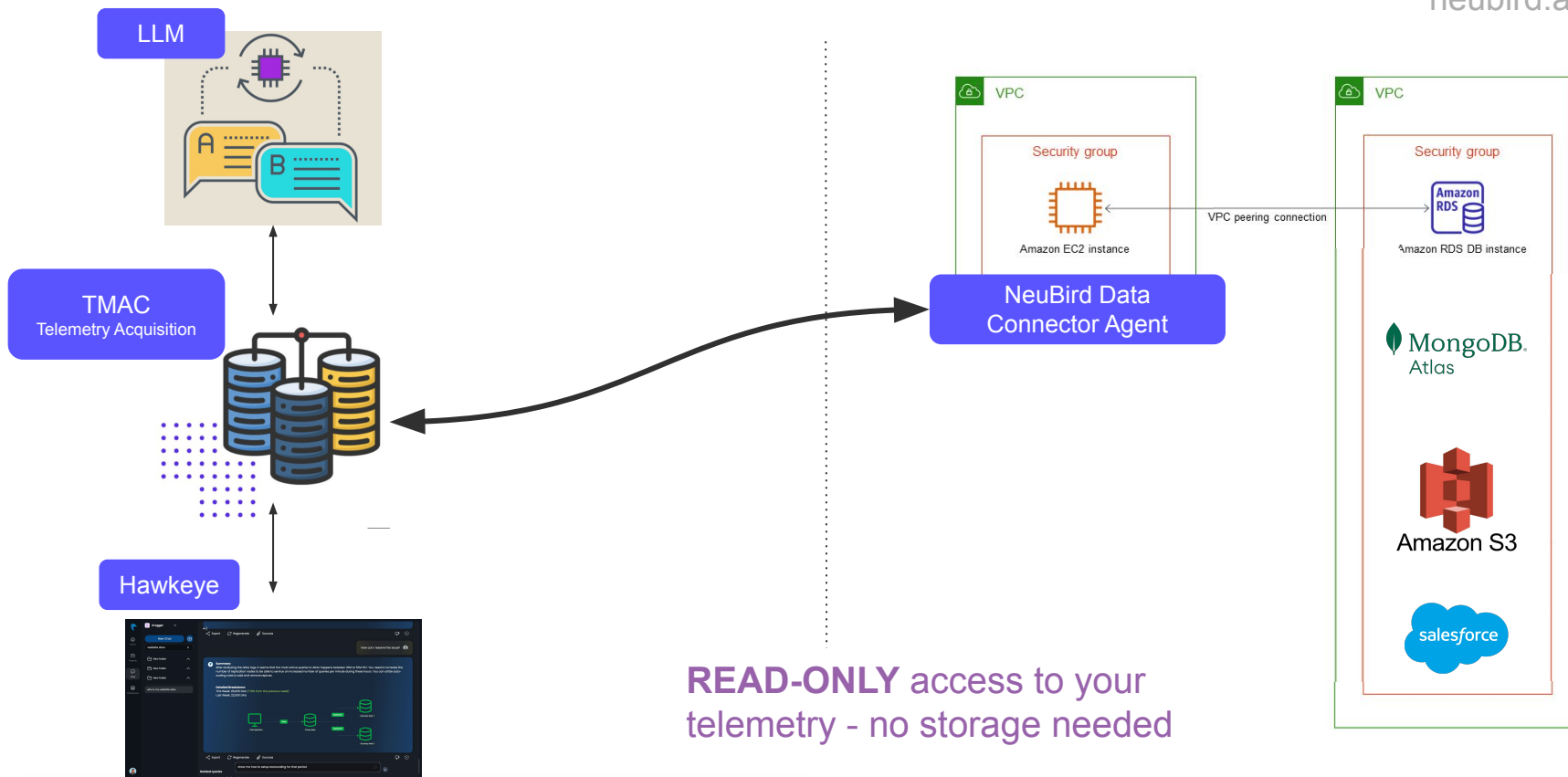
CONDITION ::= QUERY_RESULT "." FIELD COMPARATOR VALUE
  | QUERY_RESULT "." FIELD LOGICAL_OP CONDITION
  | CONDITION LOGICAL_OP CONDITION

COMPARATOR ::= ">" | "<" | ">=" | "<=" | "==" | "!="
LOGICAL_OP ::= "and" | "or"

QUERY_RESULT ::= IDENTIFIER

FIELD ::= IDENTIFIER
```


And It's **Simple** to Use: Plug and Play... on your telemetry



NeuBird.ai

Reimagine



IT Ops

Generative {IT SRE}.Ops

Manual Actionability Stack

Search Telemetry

Perform Log and Metrics Queries

Dashboards

Create Predetermined visualizations and charts

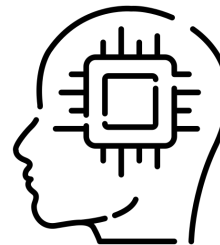
Incident Response

Respond to PagerDuty Alerts

Corrective Actions

Perform SOP Actions

Agentic Actionability Stack



Leverage **Reasoning** from **LLMs** guided by SRE and IT Ops runbooks to root cause and resolve an incident on its own