Next-Gen Observability:

Leveraging AI and Data Pipelines to reduce Cost and MTTR

About US

Sarav Jagadeesan

Director
Infrastructure, Observability and AI/MLOps
Informatica

Kirti Ranjan Parida

Architect
Observability and AI/MLOps
Informatica

Agenda

1

Problem Statement

Existing issues with our infrastructure

2

Telemetry Pipeline

How telemetry pipeline helped to get control over observability data

3

AI Pipelines

How to leverage AI to build automated pipelines with less human intervention

4

RCA without AI

How incident investigation was taking high time due to telemetry volume

5

AI Based RCA

How AI helped to analyze different signals and summarize the root cause

6

Outcome

Benefits : Reduces MTTR and reduced Cost

Infrastructure Scale @ Informatica

4
Cloud Providers

20+
Regions

600+
Kubernetes Clusters

10000+
Virtual Machines

100+
Applications

~60 TB
Log ingestion Per day

3.8 Trillion
Documents Ingested/ month

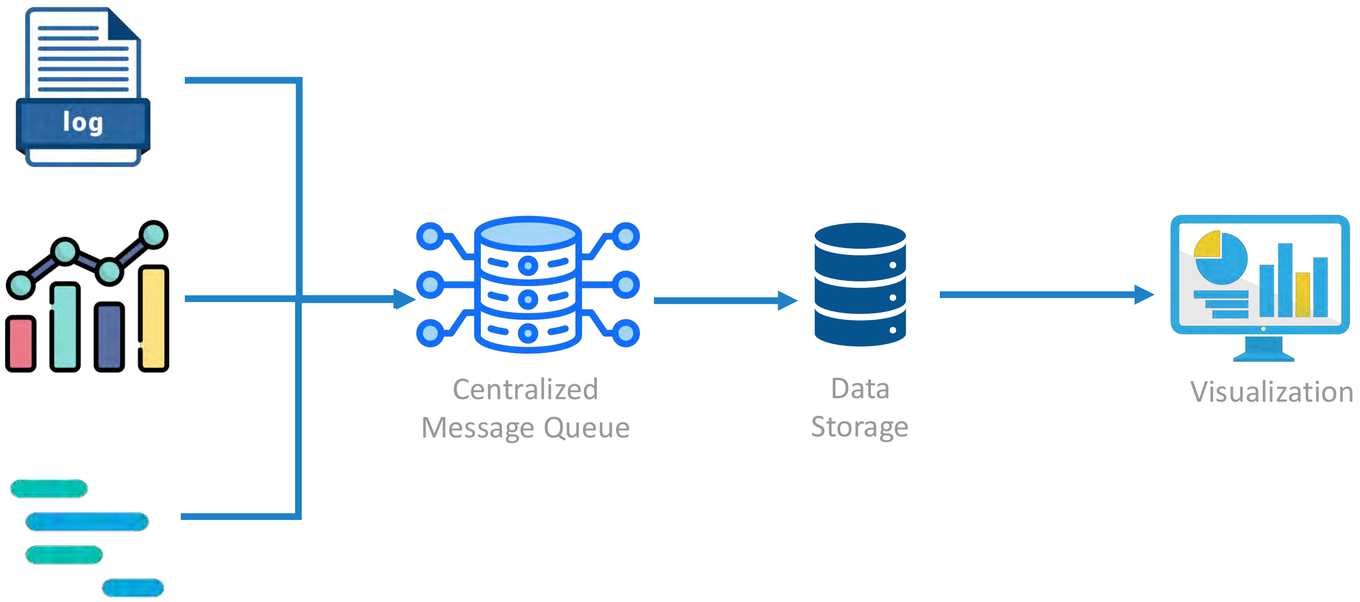
1.5 Billion
Api Calls Per day

400 Mil
Bytes/second ingestion in Kafka

15 PB
S3 Storage(standard & Glacier)

2.5 PB
provisioned In Cluster Storage

Problem Statement



Multi region data transfer cost

Sending data across regions, AZ's are expensive

Higher Storage and Backup Cost

It is expensive to store all the telemetry data and backup for long period to meet Compliance

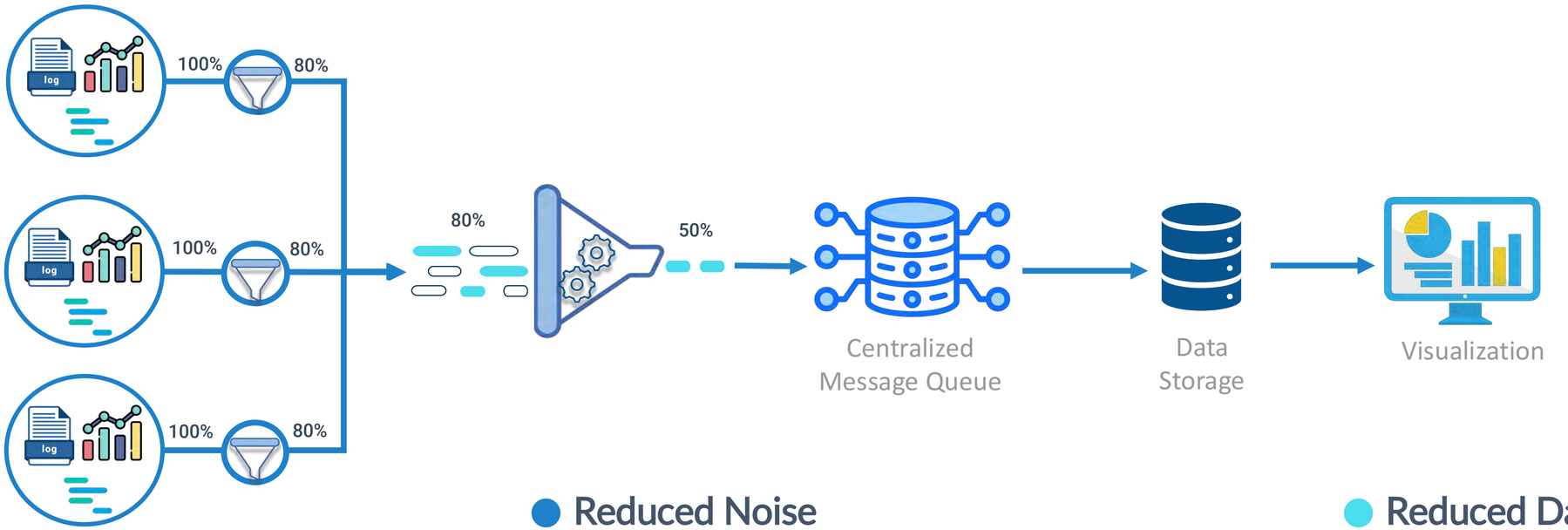
Slow Dashboards

High volume telemetry data makes dashboard loading very slow with longer periods

No audit on noisy telemetry signals

There was no way to check if any rouge data coming in which has no value. For e:g each successful health check logs

Telemetry Pipeline



Reduced Noise

Filtering only required data reduces all noise going through the observability system

Better Control over Data

We can control how we need to write data in destination.
Masking sensitive data or retrieving metrics out of expensive logs

Faster Dashboards

Less data to render means faster performant dashboards

Reduced Data transfer Cost

Adding filter at edge and with centralized Data pipeline we drop most unused data resulting less data to travel across network

Reduced Live Storage Cost

Filtering all the noisy data helps in savings less data means less disk space required

Reduced Backup Cost

As we have to deal with less amount of data to backup over a period backup cost also gets reduced

AI Telemetry Pipeline

Challenges (Traditional Pipeline)

- Creating telemetry pipeline is time consuming
- Traditional pipelines are complex and manual
- AI Copilots guide but don't automate
- Dedicated resource required to manage pipelines

How AI Agent will Help

- Interactive personal assistant
- Create New pipelines
- Manage Existing pipelines
- Proactive Recommendation Engine

Create a pipeline from S3 of type syslog

Add a PII Mask to Pipeline

Show top IP's based on data volume

Unified Observability

Single Unified OTEL based Agent

Single OTEL based agent to collect all metrics, logs and traces

Centralized Fleet Management

Deployed fleet server to manage all agents from single place

Context Propagation

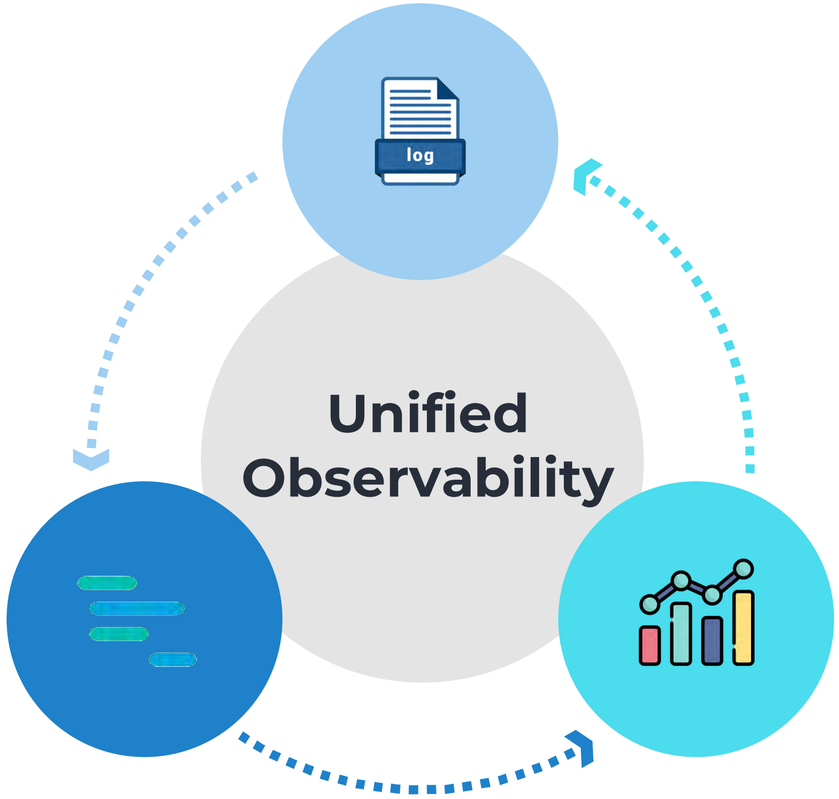
All services are instrumented to propagate trace context from service to service

Service Map & Correlation

Service Map to show how the services are connected

Anomaly detection

ML models to detect anomaly based on the metrics



AI Assistant for Observability

Observability
AI Assistant

Search and provide results, visualization from NLP

Query Assistant. No need to learn specific query language

Decode error messages and suggest remediation

Runbook Automation on known alerts

Thank You