Airflow and OpenTelemetry

Better instrumentation for improved system visibility

Howard Yoo, Dennis Ferruzzi
Hello,

Howard Yoo  
Sr. Solutions Architect at Cerebral Blue

Howard Yoo loves modern application observability, open source projects, and has been contributing to projects like Apache Airflow, OpenTelemetry, OpenLineage, and Marquez. He lives in Austin, TX, USA.

Dennis Ferruzzi  
OSS Developer - AWS

Dennis is an AWS Software Developer and one of the founding members of the Amazon OSS team focused on the continual growth and success of Apache Airflow.
Overview

- What is OpenTelemetry (OTel)?
- Why use OTel over StatsD?
- What is currently working?
- What’s to come?
- More information
What is OpenTelemetry?

One Telemetry to unite them all...

- Open Observability Framework
- Vendor Agnostic
- Tool Agnostic
- Cloud Native Computing Foundation (CNCF) Project

https://opentelemetry.io
Why use OTel?

- OTel provides common specification, protocol, API, and libraries to collect Metrics, Logs, and Traces.
- OTel is designed to work especially well in distributed computing environment.
- OTel is supported by major vendors in the Observability space, and continues to evolve.
Getting OTEL into Airflow

**TIMELINE**

- JAN 2022: Initial Discussion On OTEL support first Posted in Airflow Forum
- MAR 2022: AIP-49 DRAFT Work began
- MAY 2022: AIP-49 FINALIZED
- JUN 2022: AIP-49 VOTE Passed
- FEB 2023: OTEL for Airflow Development Started
- JUL 2023: OTEL for Airflow Gauges and Counters released

Airflow OTEL POC (Feb ~ Mar 2022)
What Currently Works?

- Feature parity with StatsD
  - All existing metrics are implemented as Counters and Gauges with tagging/attributes for improved granularity
What’s Coming?

- Drop redundant metric names
  - Why are they even there?

  \[
  \text{local\_task\_job.task\_exit.<job\_id>.<dag\_id>.<task\_id>.<return\_code>}
  \]
  becomes

  \[
  \text{local\_task\_job.task\_exit}
  \]
  with the tags:

  \{
  "job\_id": job\_id,  "dag\_id": dag\_id,  "task\_id": task\_id,  "return\_code": return\_code
  \}

- Logging
  - Why move off the current native pylogger?

- Spans and Traces
  - What are they and why do we care?
Traces in Airflow
Traces in Airflow
Airflow With OTel

Worker  Scheduler  Executor  Metadata  Ext. System

OTel

Metrics  Logs  Traces
Airflow With OTel

Worker

Scheduler

Executor

Metadata

Ext. System

RUN_ID : xyz
Airflow With OTel

Worker

Scheduler

Executor

Metadata

Ext. System
Airflow With OTel
Improved System Visibility of Connecting Metrics, Logs, and Traces