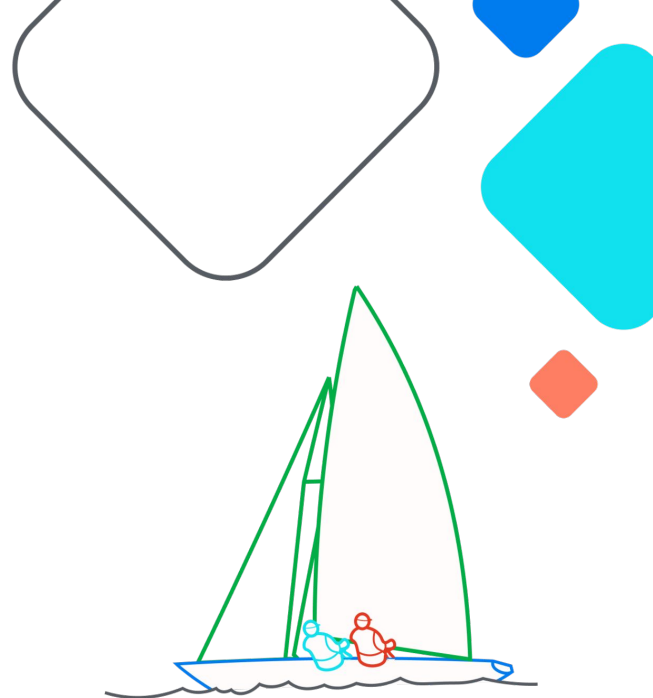


Airflow and OpenTelemetry

Better instrumentation for
improved system visibility

Howard Yoo, Dennis Ferruzzi



 **Airflow Summit**
Let's flow together

September 19-21, 2023,
Toronto, Canada

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

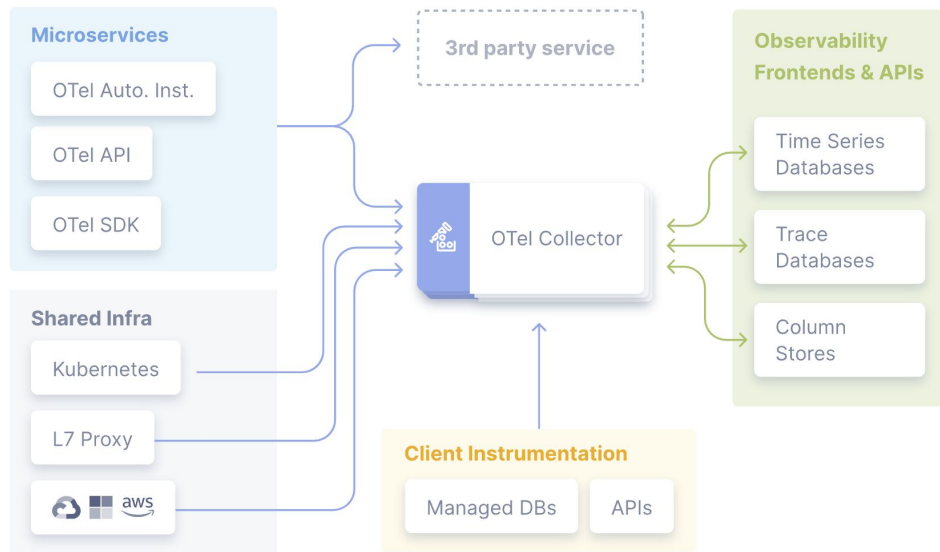
- What is OpenTelemetry (OTel)?
- Why use OTel over StatsD?
- What is currently working?
- What's to come?
- More information

What is OpenTelemetry?



One Telemetry to unite them all...

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

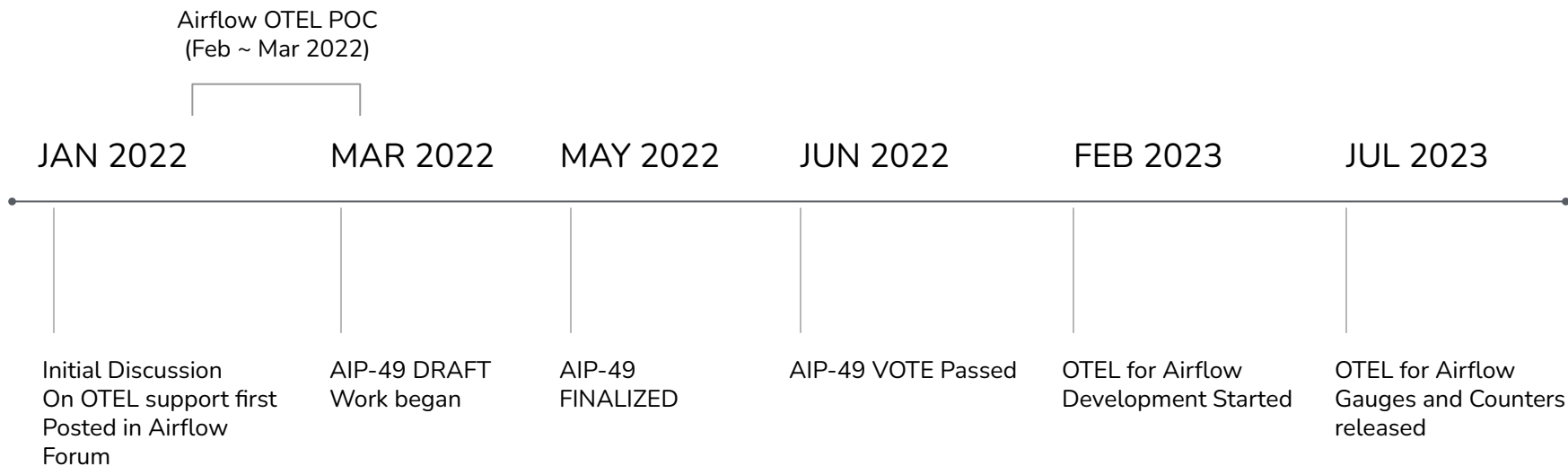


Why use OTel?

- 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



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?

```
local_task_job.task_exit.<job_id>.<dag_id>.<task_id>.<return_code>  
becomes
```

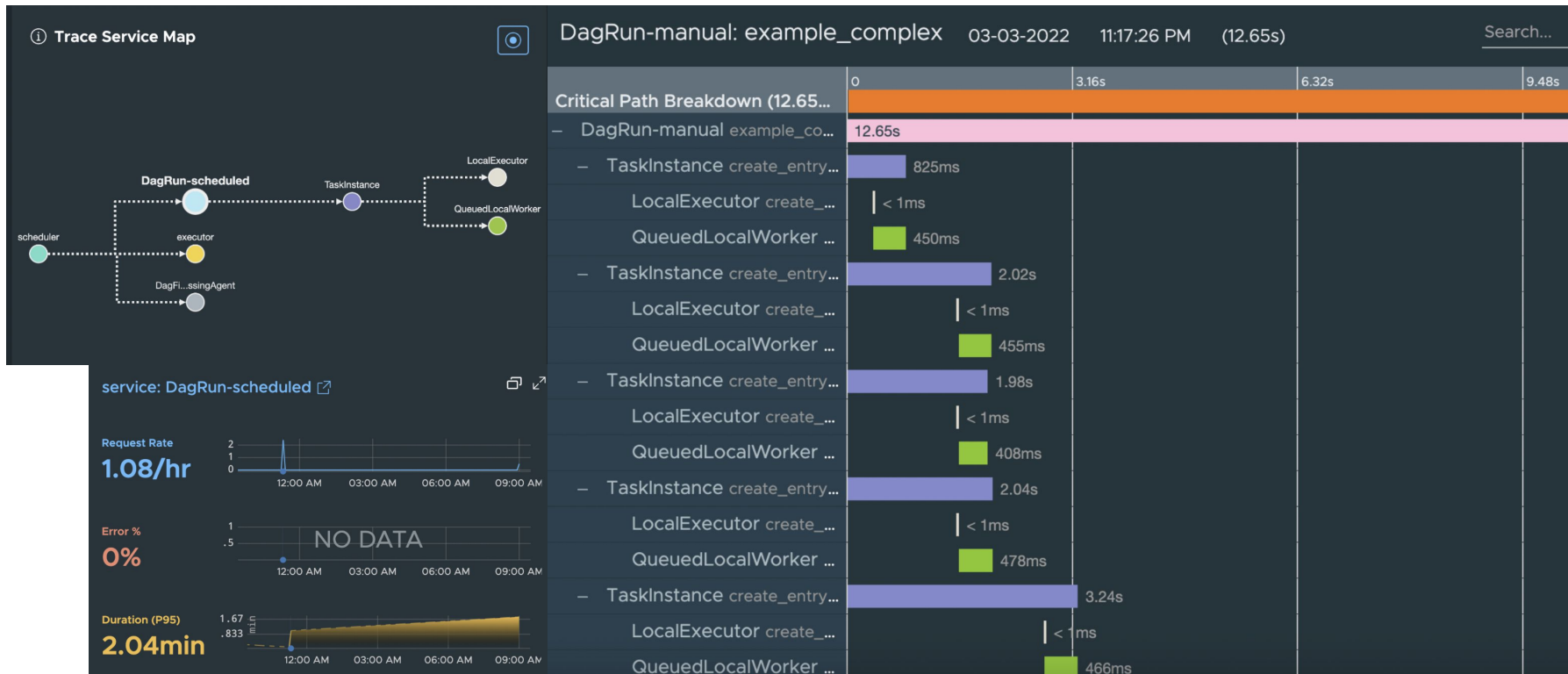
```
local_task_job.task_exit
```

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

Applications > Service Dashboard > Traces

</>

Traces (100): Sort By: Longest First

Matching Spans

- ✓ `_run_scheduler_loop`
2022-03-04 09:01:13 AM (1.67s)
- ✓ `_run_scheduler_loop`
2022-03-04 09:01:45 AM (1.13s)
- ✓ `_run_scheduler_loop`
2022-03-04 09:08:33 AM (1.1s)
- ✓ `_run_scheduler_loop`
2022-03-04 09:01:35 AM (1.1s)
- ✓ `_run_scheduler_loop`
2022-03-04 09:02:17 AM (1.1s)
- ✓ `_run_scheduler_loop`
2022-03-04 09:02:13 AM (1.09s)
- ✓ `_run_scheduler_loop`
2022-03-04 09:01:36 AM (1.08s)
- ✓ `_run_scheduler_loop`
2022-03-04 09:08:49 AM (1.08s)

scheduler: `_run_scheduler_loop` 03-04-2022 09:01:13 AM (1.67s) Search...

Critical Path Breakdown (1.67s)

Operation	Duration
scheduler._run_scheduler_loop	1.67s
scheduler._do_scheduling	476ms
scheduler._create_dagrun	47ms
scheduler._create_dagrun	10ms
scheduler._start_queue	128ms
scheduler._get_next_task	4ms
scheduler._get_next_dagrun	< 1ms
scheduler._schedule_dagrun	75ms
scheduler._verify_inlets	15ms
scheduler._send_dagrun	1ms
scheduler._send_slaves	1ms
scheduler._critical_section	84ms
scheduler._execute_dagrun	83ms
scheduler._get_task_instances	19ms
scheduler._enqueue_task_instances	1ms
executor.queue...	1ms
executor.heartbeat	1.0s

Airflow With OTel



Worker

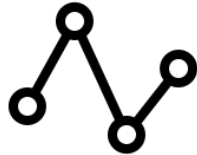
Scheduler

Executor

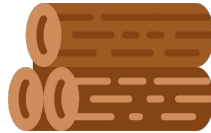
Metadata

Ext. System

OTel



Metrics



Logs



Traces

Airflow With OTel



Worker

RUN_ID : xyz



Scheduler

RUN_ID : xyz



Executor

RUN_ID : xyz



Metadata

Ext. System

RUN_ID : xyz



Airflow With OTel



Worker

RUN_ID : xyz



RUN_ID : xyz



Scheduler

RUN_ID : xyz



RUN_ID : xyz



Executor

RUN_ID : xyz



RUN_ID : xyz



Metadata

Ext. System

RUN_ID : xyz



RUN_ID : xyz



Airflow With OTel



Worker

RUN_ID : xyz



RUN_ID : xyz



RUN_ID : xyz



Scheduler

RUN_ID : xyz



RUN_ID : xyz



RUN_ID : xyz



Executor

RUN_ID : xyz



RUN_ID : xyz



RUN_ID : xyz



Metadata

RUN_ID : xyz



Ext. System

RUN_ID : xyz



RUN_ID : xyz



RUN_ID : xyz



Worker

Scheduler

Executor

Metadata

Ext. System

RUN_ID : xyz



RUN_ID : xyz



RUN_ID : xyz



RUN_ID : xyz



RUN_ID : xyz



RUN_ID : xyz



RUN_ID : xyz



RUN_ID : xyz



RUN_ID : xyz



RUN_ID : xyz



RUN_ID : xyz



RUN_ID : xyz



RUN_ID : xyz



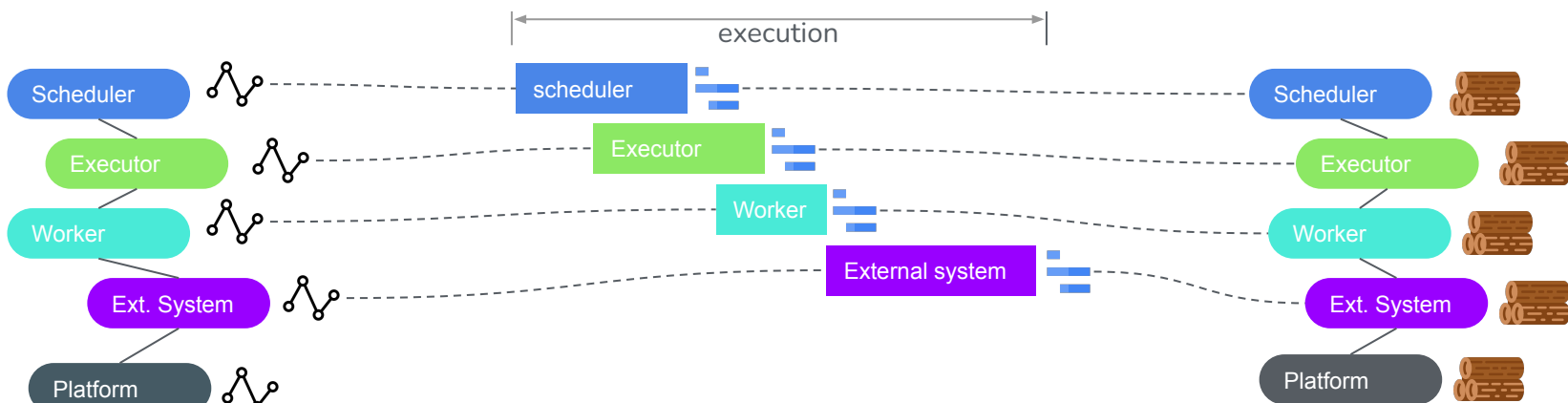
OTLP

OTEL Collector





RUN_ID : xyz



Improved System Visibility of Connecting Metrics, Logs, and Traces

Airflow Docs - OpenTelemetry Setup



Airflow Observability with OpenTemeletry - Guide/Blog

