The SLAyer
your data pipeline needs

26 May 2022 | By Eden Gluska
Hi everyone!

Eden Gluska
Big Data Developer

Using Airflow since 2019
Digital Turbine - The largest independent mobile growth platform powered by On-Device technology
How big is our Big Data?

1,450B Auctions Per Month

33B Impressions Per Month

53.7T Bid Requests Per Month

10K+ Apps

300TB Generated Monthly

540M Unique Users Per Month

80 Reported Dimensions (on real-time reporting)

65 Reported Metrics
Agenda

01 Part 1: Data pipeline Goals

02 Part 2: Airflow SLA monitoring

03 Part 3: What is SLAyer

04 Part 4: SLAyer Implementation
S
Service

L
Level

A
Agreement
Airflow SLA defining

You can set an SLA for all tasks in your DAG by defining 'sla' as a default argument.

default_args = {
    'owner': 'airflow',
    'sla': timedelta(hours=2)
}
| notify@fyber.com | Sat, Nov 7, 7:30 PM (5 days ago) |
| notify@fyber.com | Sat, Nov 7, 8:30 PM (5 days ago) |
| notify@fyber.com | Sat, Nov 7, 9:30 PM (5 days ago) |
| notify@fyber.com | Sat, Nov 7, 11:31 PM (5 days ago) |
| notify@fyber.com | Sun, Nov 8, 1:30 AM (4 days ago) |
| notify@fyber.com | Sun, Nov 8, 3:30 AM (4 days ago) |

| notify@fyber.com | Sat, Nov 7, 3:30 AM (5 days ago) |
SCREW YOU GUYS

IM GOING HOME
Latency monitoring goals:

- Delay and completion failures?
- Monthly/Weekly Occurrence
- Average delay?
- Affected Tasks?
- Frequency?
Monitoring tools

Prometheus

Grafana
The SLAyer

kube

Query

Reporting SLA metrics
{ dag_id, task_id, execution_date }
Grafana metric
Grafana alert
Implementation

1. SLA miss metric info.
2. Prometheus custom exporter.
3. Dockerizing the exporter.
4. Deploying on Kubernetes.
SLA miss metric info
Prometheus Custom Exporter

```python
def collect(self):

def _define_metric_family(self):
```
FROM python:3.6

ADD slayer /slayer

RUN pip install -r

/slayer/requirements.txt

WORKDIR /slayer

ENV PYTHONPATH '/slayer/'

CMD ['"python"",

'/slayer/prometheus_exporter.py']
Deploying on Kubernetes

Deployment

Service
Prometheus Operator
SLAyer Over all

Kubernetes

- Deployment
- SLAyer pod
- Service
- ServiceMonitor
- Prometheus Operator

Prometheus Operator
Q & A
Thank you!

SLAyer Blog:
Medium - Eden Gluska - The slayer your data pipeline needs

SLAyer github repository:
fyber github - SLAyer

For more QA:
Linkedin - Eden Gluska