

The SLayer

your data pipeline needs

26 May 2022 | By Eden Gluska

Hi everyone!



Eden Gluska

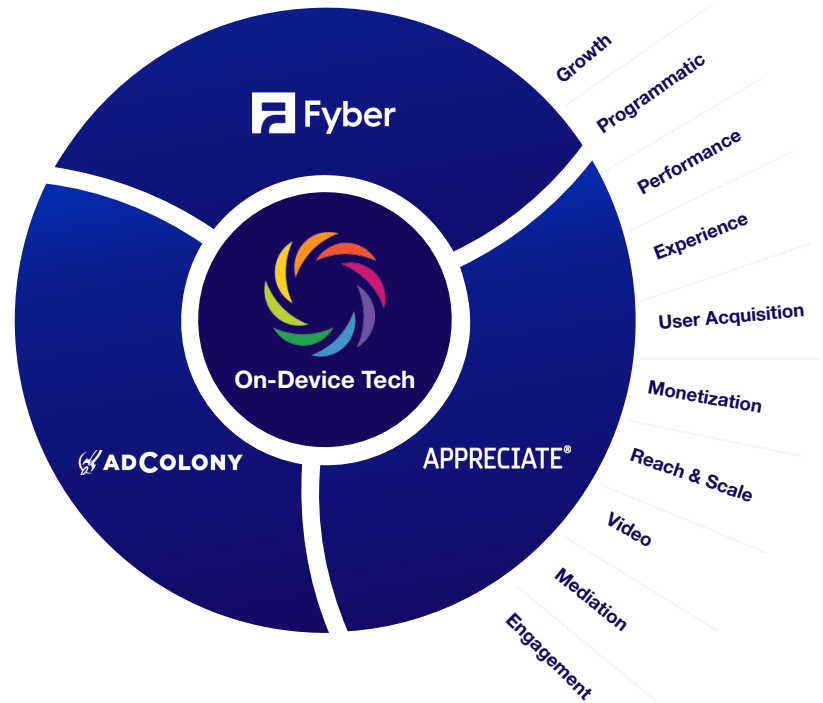
Big Data Developer



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

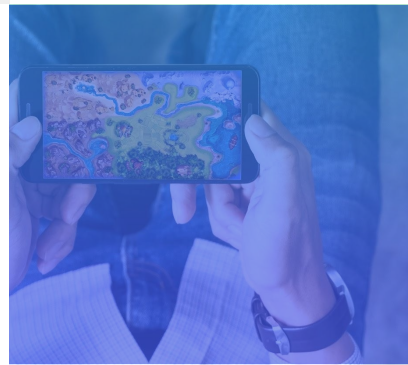
10K+ Apps

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

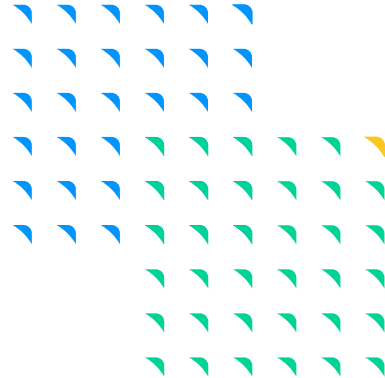
Part 2: Airflow SLA monitoring

03

Part 3: What is SLAyer

04

Part 4: SLAyer Implementation



Completeness



**Our Data
Pipeline
Goals**

Latency



Correctness






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)  
}
```


Airflow SLA miss mails

[airflow] SLA miss on DAG=druidBids_Indexing  Trash x



notify@fyber.com

Sat, Nov 7, 7:30 PM (5 days ago)

Here's a list of tasks that missed their SLAs: bids_daily_rare on 2020-11-07T13:30:00 bids_daily_lite_1 on 2020-11-07T13:30:00 bids_daily_lite on 2020-11-07T13:



notify@fyber.com

Sat, Nov 7, 8:30 PM (5 days ago)

bids_daily_lite_1 on 2020-11-07T14:30:00 bids_daily_lite on 2020-11-07T14:30:00 Blocking tasks: bids_daily_lite on 2020-11-07T14:30:00 bids_daily_lite_1 on 2020



notify@fyber.com

Sat, Nov 7, 9:30 PM (5 days ago)

bids_daily_lite on 2020-11-07T15:30:00 Blocking tasks: bids_daily_lite on 2020-11-07T15:30:00 \ =, . = =, | , - - , | , = =, | " - (: : : :) - " | , = \ \ _ / ^ - , - \ _ _ // ^ - | : :



notify@fyber.com

Sat, Nov 7, 11:31 PM (5 days ago)

bids_daily_rare on 2020-11-07T17:30:00 bids_daily_lite_1 on 2020-11-07T17:30:00 bids_daily_lite on 2020-11-07T17:30:00 bids_daily_rare_1 on 2020-11-07T17:30:00



notify@fyber.com

Sun, Nov 8, 1:30 AM (4 days ago)

bids_daily_rare on 2020-11-07T19:30:00 bids_daily_lite_1 on 2020-11-07T19:30:00 bids_daily_lite on 2020-11-07T19:30:00 bids_daily_rare_1 on 2020-11-07T19:30:00



notify@fyber.com

Sun, Nov 8, 3:30 AM (4 days ago)



to eran.shemesh, noc, me, amit.menashe ▾

Here's a list of tasks that missed their SLAs:

bids_daily_rare on 2020-11-07T21:30:00
bids_daily_lite_1 on 2020-11-07T21:30:00



SCREW YOU GUYS

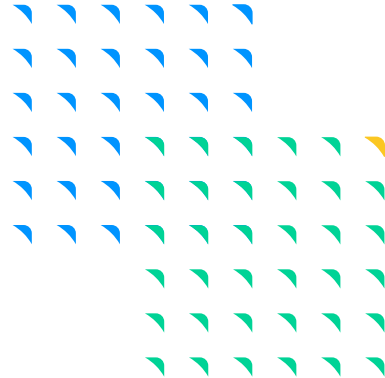


IM GOING HOME

MEME.LOLOLS.COM

Latency monitoring goals:

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



Monitoring tools

Prometheus



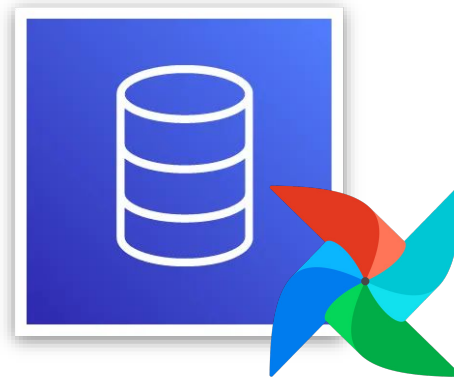
Grafana



The SLayer



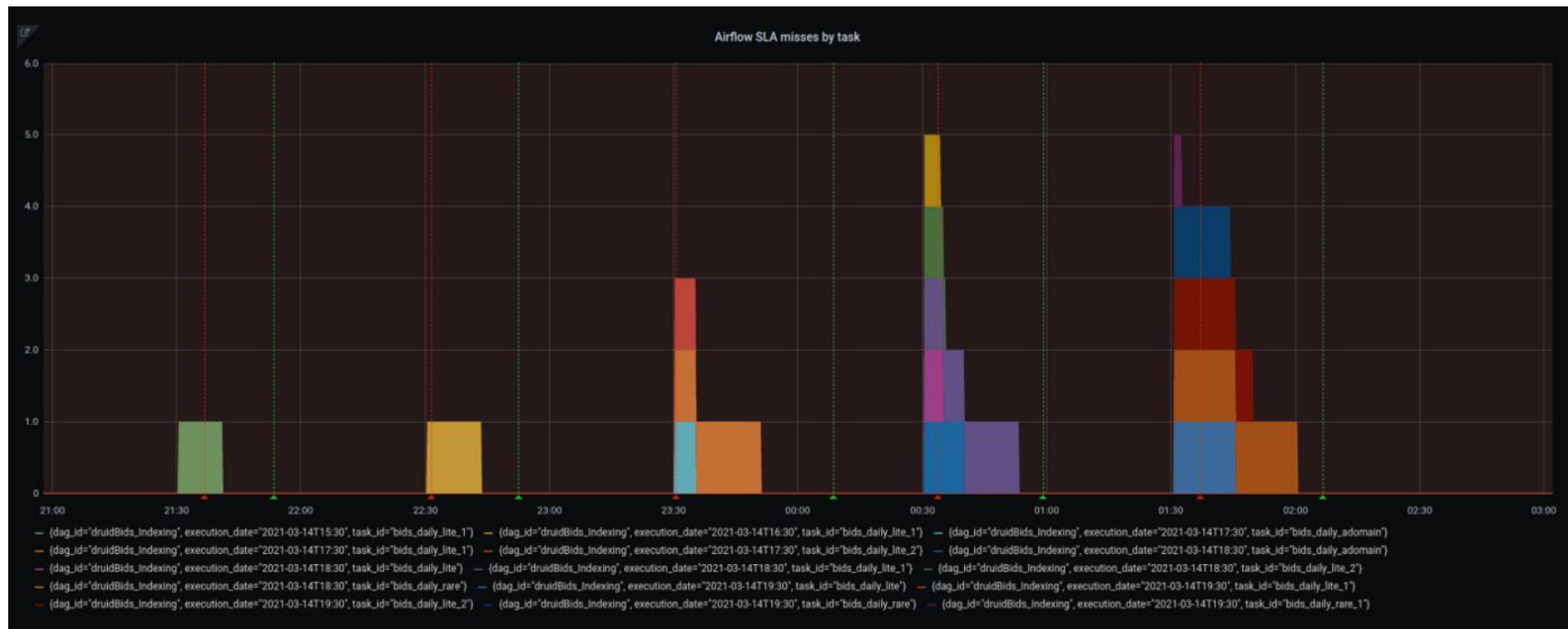
Query



Reporting SLA metrics
{ dag_id , task_id , execution_date }



Grafana metric



Grafana alert

Gmail interface showing an email from Grafana <inn.monitoring.alerts@inner-active.com> received at 12:59 AM (8 hours ago).

[OK] Airflow SLA misses by task alert

In case you should reindex data (because any kind of reason) : <https://docs.google.com/document/d/1icJWVRjuDP7szoaLB-FF7y-tidXOf2UCT5XOYxDY8KM/edit#bookmark=id.gr11chiyn3q>

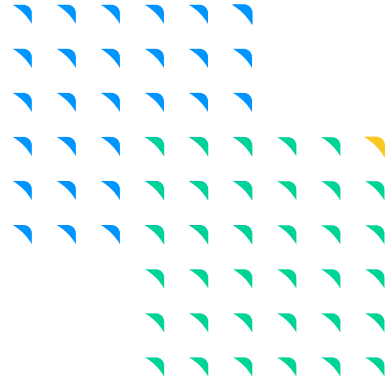
Airflow SLA misses by task

Task	Execution Date	Task ID	Misses
(dag_id="druidBids_Indexing", execution_date="2021-03-14T15:30", task_id="bids_daily_lite_1")	2021-03-14T15:30	bids_daily_lite_1	1
(dag_id="druidBids_Indexing", execution_date="2021-03-14T16:30", task_id="bids_daily_lite_1")	2021-03-14T16:30	bids_daily_lite_1	1
(dag_id="druidBids_Indexing", execution_date="2021-03-14T17:30", task_id="bids_daily_adomain")	2021-03-14T17:30	bids_daily_adomain	1
(dag_id="druidBids_Indexing", execution_date="2021-03-14T17:30", task_id="bids_daily_lite_1")	2021-03-14T17:30	bids_daily_lite_1	1
(dag_id="druidBids_Indexing", execution_date="2021-03-14T17:30", task_id="bids_daily_lite_2")	2021-03-14T17:30	bids_daily_lite_2	1
(dag_id="druidBids_Indexing", execution_date="2021-03-14T18:30", task_id="bids_daily_adomain")	2021-03-14T18:30	bids_daily_adomain	1

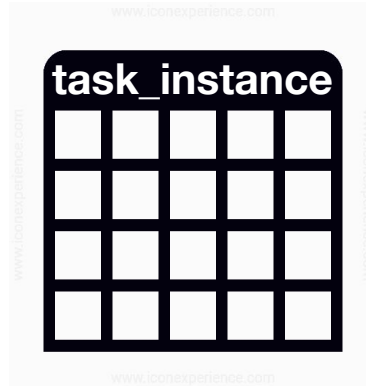
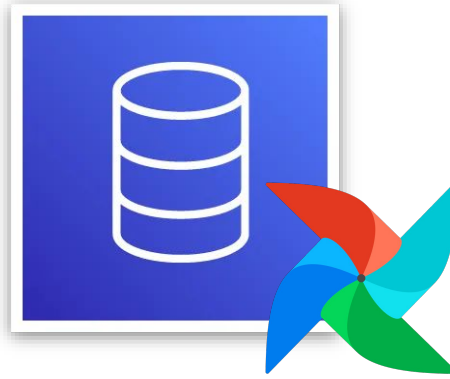


Implementation

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



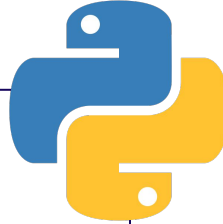
SLA miss metric info



Prometheus Custom Exporter

```
def collect(self):
```

```
def _define_metric_family(self):
```



Push Metric



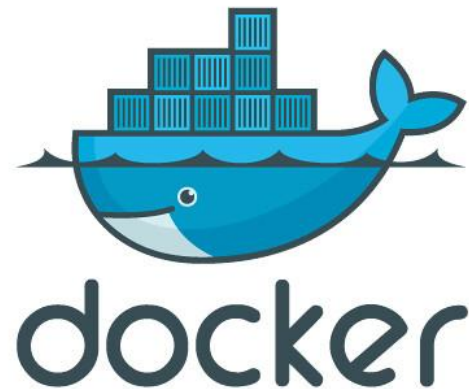
Dockerizing the exporter

```
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



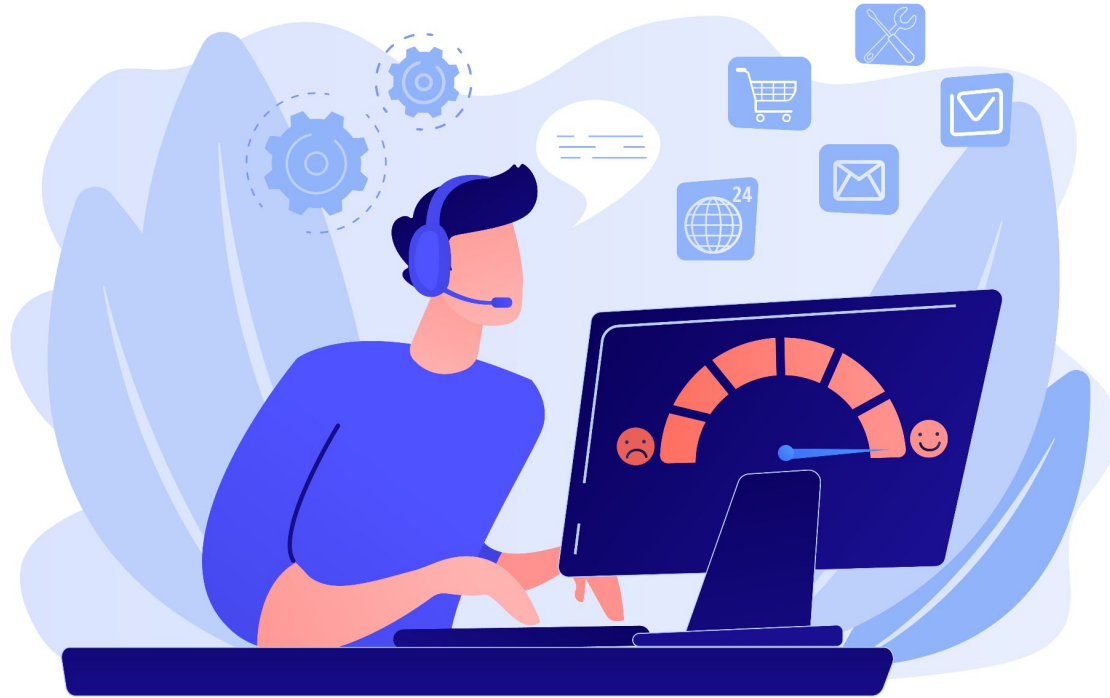
Deployment



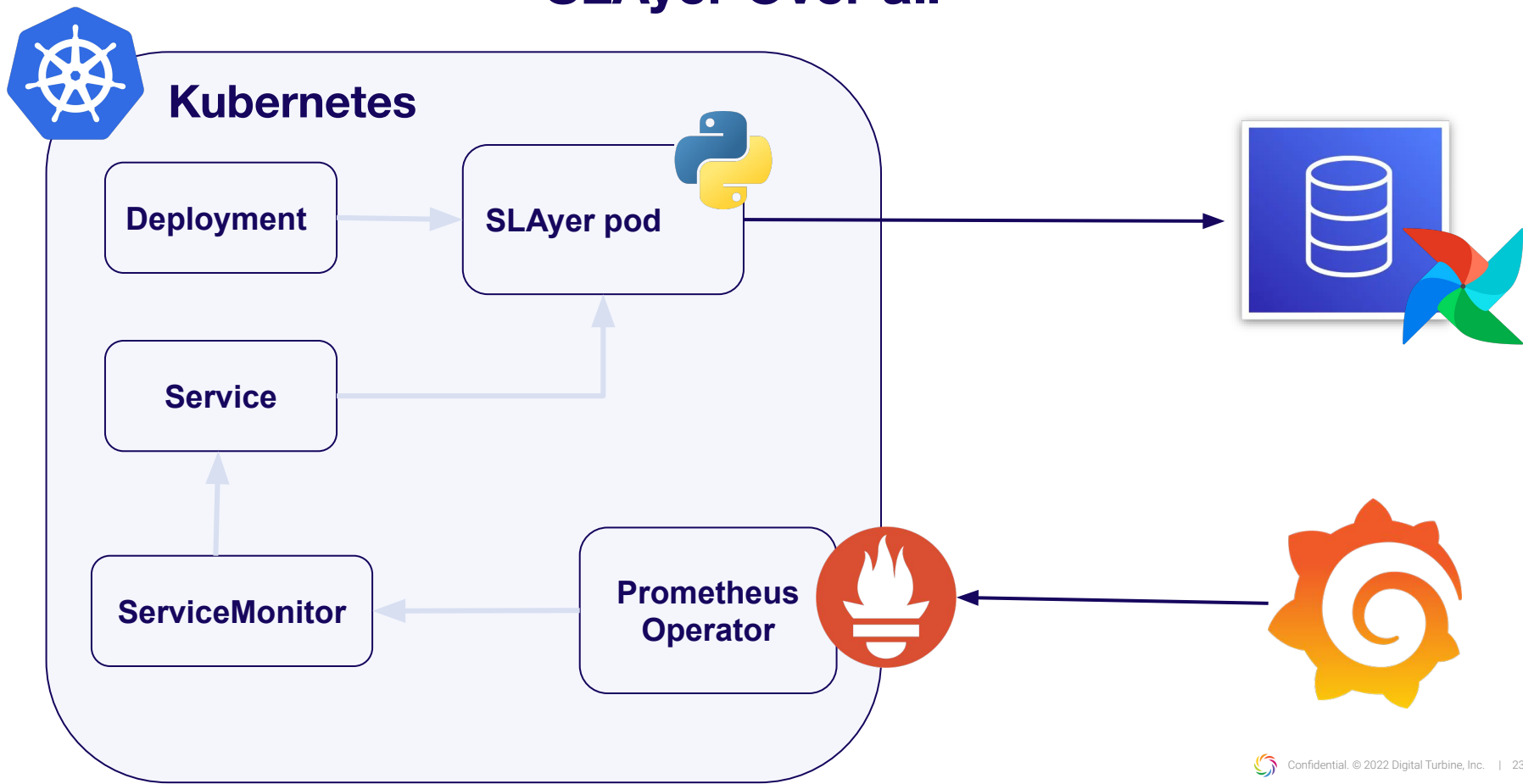
Service



Prometheus Operator



SLayer Over all



Q & A

SLayer Blog :

Medium - Eden Gluska -
The slayer your data
pipeline needs

SLayer github repository :

fyber github - SLayer

For more QA :

Linkedin- Eden Gluska

Thank you!



2022 © Digital Turbine, Inc. All rights reserved.

Digital Turbine is a trademark and/or registered trademark of Digital Turbine, Inc. or its affiliates. Other product names and logos mentioned herein are trademarks and/or registered trademarks of their respective holders.

This document contains confidential information of Digital Turbine and may not be reproduced, disclosed or otherwise used except as authorized by express written permission by Digital Turbine.