



[www.advancinganalytics.co.uk](http://www.advancinganalytics.co.uk)

# Custom Logging within Databricks



@ADVANCINGANALYTICS



@ADVANALYTICSUK



/ADVANCING ANALYTICS



# Who am I



anna-maria-wykes



@annawykes



Senior Advancing Analytics Consultant specializing in Data Engineering & Cloud.

Over 15 years' experience, most recently working within the Microsoft Data Platform, Scala, Kafka and various cloud tech

BSc in Multimedia Computing & Business, and a HND in Visual Communication



@ADVANCINGANALYTICS





@ADVANALYTICSUK



/ADVANCING ANALYTICS




# Meetups and Community Groups





### Subject Data: Bristol

Bristol, United Kingdom  
1,532 members · Public group  
Organized by Subject D. and 5 others

Share:   

Join this group

...

#### About

##### What we're about

Subject Data: Bristol has been formed to offer knowledge sharing and networking opportunities for data professionals, techies, researchers or simply those with an interest in data applications and technologies. We aim to provide an open, friendly environment in which everyone can participate, learn and share their knowledge and experience with presentations covering a wide range of topics.


So, whether, you're an administrator, developer, data scientist, analyst,...

[Read more](#)

#### Upcoming events (5)

WED, MAR 18, 6:00 PM

APIs with AI, and Finance

 Just Fat

#### Events


#### Members

#### Photos

#### Discussions

#### More

#### Organizers


















Subject D. and 5 others

[Message](#)

#### Members (1,532)

[See all](#)







# Custom Logging within Databricks





# What are we going to do

- Logging, why should you care?
- Structured Logging
- Logging/Metrics in Azure
- Logging/Metrics in Databricks
- (Real World) Demo







# Logging, why should you care?



Have you ever got into work at 9am and had to immediately debug an ETL pipeline that failed overnight?





## Why should you care

Logging is common place in Software Engineering, why?

Why is Logging often overlooked in Data Engineering?







# Logging in Software Engineering

Just capture and react to errors right?

Help capture events in a linear order:

- Helping you trace back when errors occur
- You can see how far you process got to, including triggered events and subsequent paths taken
- Output of processes (results, counts, timings)

Help meet **SLAs (Service Level Agreements) / SLOs (Service Level Objectives)**





# Structured logging

**Structured logging** is the practice of implementing a consistent, predetermined message format for application **logs** that allows them to be treated as data sets rather than text

```
Log.log("info", "this is my log message")
```

```
Log.log("info", {"message": "this is my log message", "logEvent": "processStarted", "CreatedAt": "14/04/2021"})
```

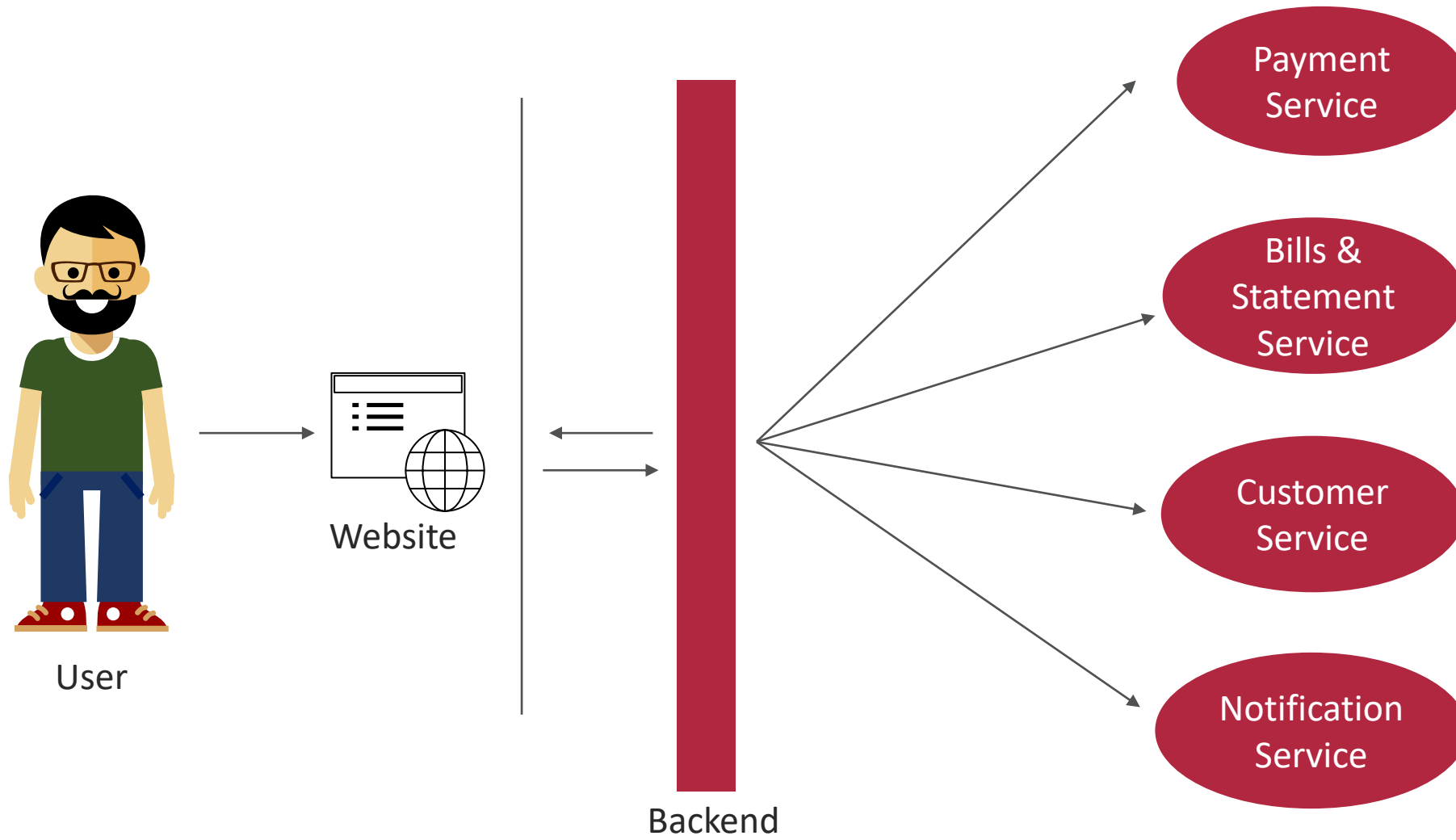


How have I used this before in Software Engineering?





# Logging in Software Engineering





# Payment Service Logging Events

## Custom Logging Events

Payment Service

PaymentReceived

PaymentRejected

PaymentApproved

TimeReceived

TimeProcessed

FailedToCommunicate







## Custom Logging Events

Bills and  
Statement  
Service

BillRequested

Bill Amended

AmendmentApproved

TimeReceived

TimeProcessed

FailedToCommunicate



## Custom Logging Events

Customer  
Service

CustomerInfoRequested

CustomerInfoRetrieved

CustomerNotFound

TimeReceived

TimeProcessed

FailedToCommunicate



## Custom Logging Events

Notification  
Service

NotificationEventTrigger

NotificationSent

TimeReceived

TimeProcessed

FailedToCommunicate

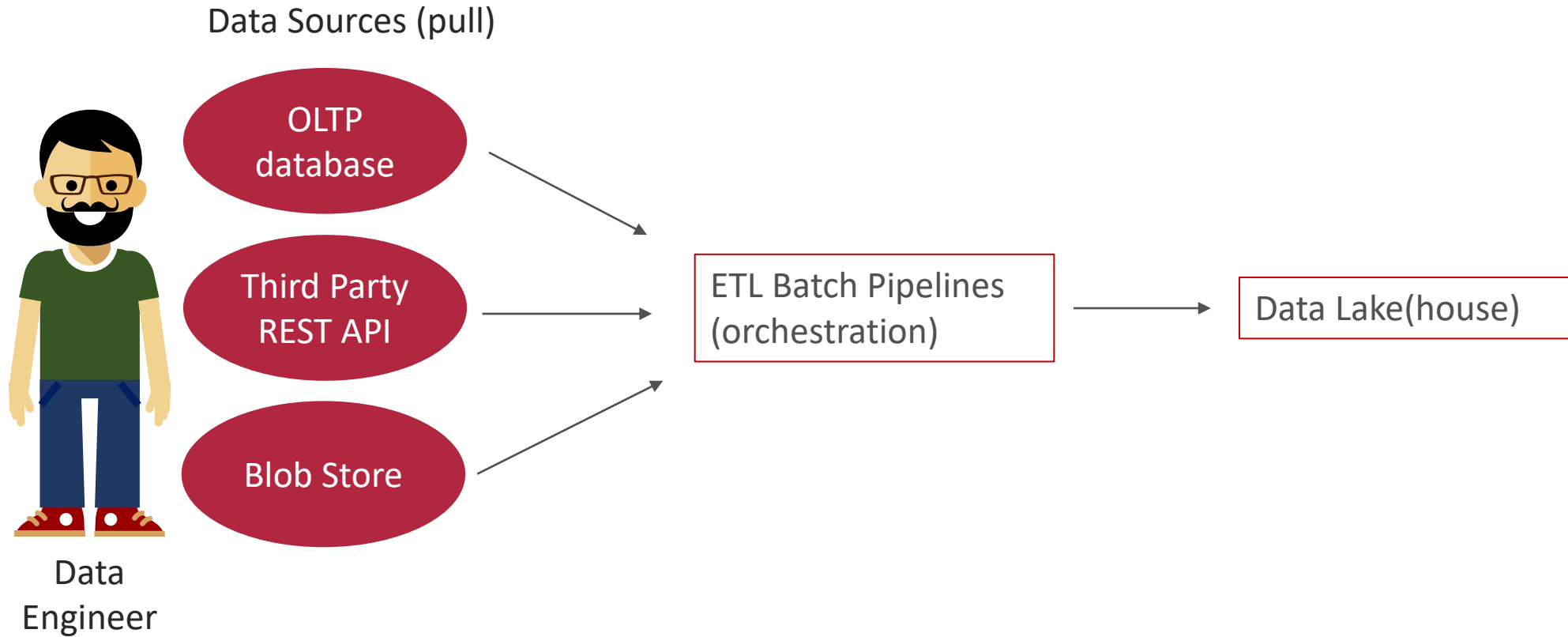


# How can we apply this to Data Engineering?





# Logging in Data Engineering







OLTP database

Third Party REST  
API

Blob Storage

## Custom Logging Events

RowCount

ProcessedRowCount

FailedRowCount

InvalidDataType

TimeReceived

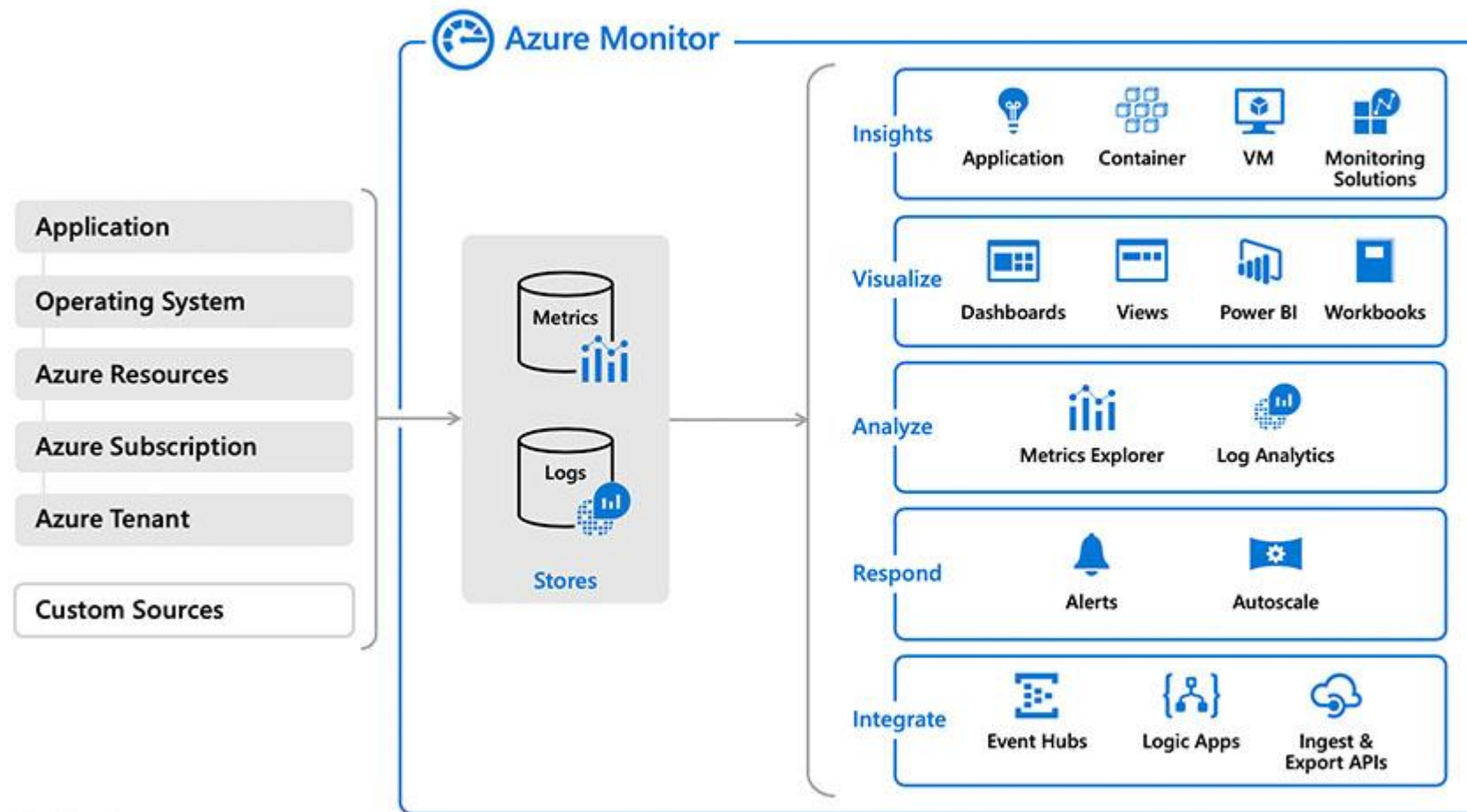
TimeProcessed

FailedToCommunicate



# Logging in Azure





Custom sources can also include AWS, GCP, Local data Centers





# Log Analytics Workspace



**Log Analytics Workspace**

Like a Data Warehouse of logs  
KQL (Kusto Query Language)



**Event Hub**

Good for event pipelines



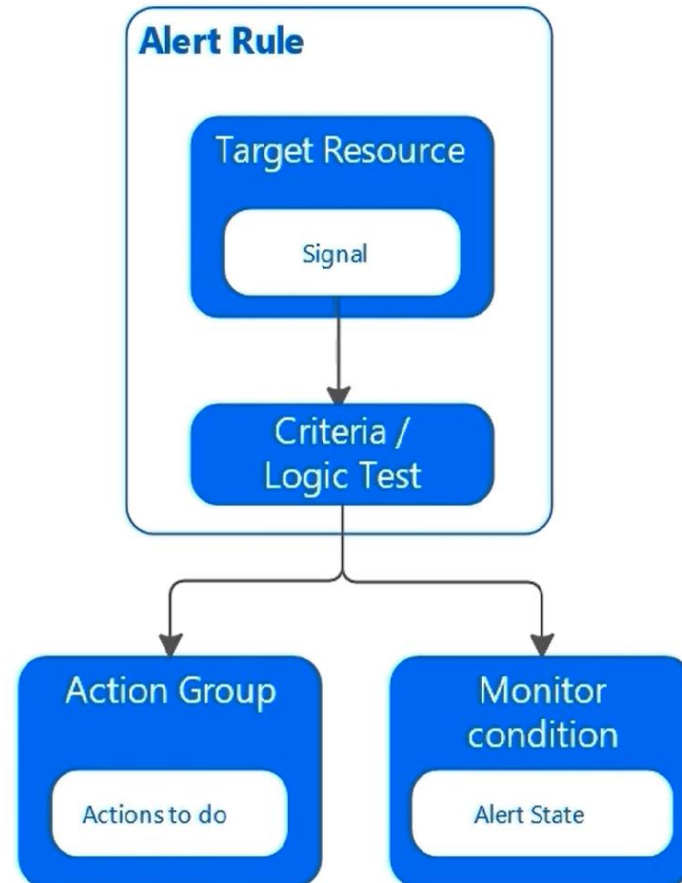
**Storage**

The original one  
Long-term retention





# Azure Monitor Alert Workflow



Great for triggering Email/SMS Alerts, and running processes of the back of alerts







# Logging in Databricks

# What do we already have for Databricks?



**ADVANCING**  
**ANALYTICS**



# Databricks Logging/Metrics Options



<https://docs.microsoft.com/en-us/azure/architecture/databricks-monitoring/dashboards>





# Databricks Logging/Metrics Options



## Spark UI

### ▼ Active Jobs (1)

Page:  1 Pages. Jump to  . Show  items in a page.

Job Id ▼	Description	Submitted	Duration	Stages: Succeeded/Total	Tasks (for all stages): Succeeded/Total
7	count at <console>:26 count at <console>:26 (kill)	2019/08/10 17:50:13	17 s	0/2	<div><div>0/5 (4 running)</div></div>

Page:  1 Pages. Jump to  . Show  items in a page.

### ▼ Completed Jobs (7)

Page:  1 Pages. Jump to  . Show  items in a page.

Job Id ▼	Description	Submitted	Duration	Stages: Succeeded/Total	Tasks (for all stages): Succeeded/Total
6	show at <console>:26 show at <console>:26	2019/08/10 17:49:30	0.4 s	1/1	<div><div>1/1</div></div>
5	show at <console>:28 show at <console>:28	2019/08/10 17:48:32	0.8 s	3/3	<div><div>9/9</div></div>
4	show at <console>:28 show at <console>:28	2019/08/10 17:47:40	2 s	3/3	<div><div>9/9</div></div>



# Databricks Logging/Metrics Options



Azure Log Analytics







# Log Analytics for Databricks: What do we get out the box?



Azure Log Analytics

custom-logging-in-databricks | Logs ⚙️ ...  
Log Analytics workspace

Search (Ctrl+/)

- Overview
- Activity log
- Access control (IAM)
- Tags
- Diagnose and solve problems

Settings

- Locks
- Agents management
- Agents configuration
- Computer Groups
- Linked storage accounts
- Network Isolation
- Advanced settings

General

- Workspace summary
- Workbooks
- Logs**
- Solutions

New Query 1 x +

custom-logging-in... [Select scope](#)

Tables Queries Filter <<

Search

Filter Group by: Solution ▾

Collapse all

**Favorites**

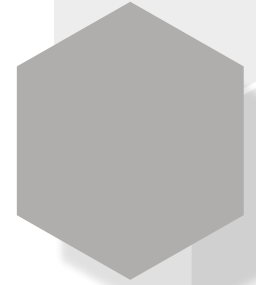
You can add favorites by clicking on the ☆ icon

**LogManagement**

- Alert
- DatabricksAccounts
- DatabricksClusters
- DatabricksDBFS
- DatabricksNotebook
- DatabricksSecrets
- DatabricksSSH
- LAQueryLogs
- Usage

**Custom Logs**

**Functions**





# Log Analytics REST API

<https://docs.microsoft.com/en-us/rest/api/loganalytics/>

Manage Log Analytics Resources

Send Custom Log Data to Log Analytics

Access Log Analytics Data





[www.advancinganalytics.co.uk](http://www.advancinganalytics.co.uk)

# (Real World) Demo

# What will we do?

- Go through set-up of Log Analytics for Databricks
- Create a Python Wheel containing Log Analytics REST API logic to create our logs
- Write some Custom Logs via a Databricks Notebook
- Explore our Custom logs, in Log Analytics
- Create Azure Monitor Alerts from our Custom Logs
- Create Azure Monitor Dashboard from our Custom Logs



# Useful links



## Code Repository :

<https://github.com/AnnaWykes/custom-logging-in-databricks>



## Log Analytics Getting Started

<https://docs.microsoft.com/en-us/azure/azure-monitor/logs/log-analytics-tutorial>



## Log Analytics REST API

<https://docs.microsoft.com/en-us/rest/api/loganalytics/>



## Grafana Databrick Metrics Monitoring with Log Analytics

<https://docs.microsoft.com/en-us/azure/architecture/databricks-monitoring/dashboards>



# THANK YOU



**anna-maria-wykes-31939454**



**@annawykes**

