Intro to Data Engineering

Mils Burasakorn

Burasakorn Sabyeying
(Mils, มิลส์, มิล, มิว)

Mesodiar.com

Data Engineer @ CJ Express (TILDI team)

CJ EXPRESS

Agenda

- **Why** do we need Data Engineer?
- **Who** is Data Engineer
- **What** does Data Engineer do?
- **Where** does Data Engineer stand
- **How** do I become Data Engineer?

Why do we need Data Engineer?

Database

Google Analytics

salesforce

CRM

SAP®

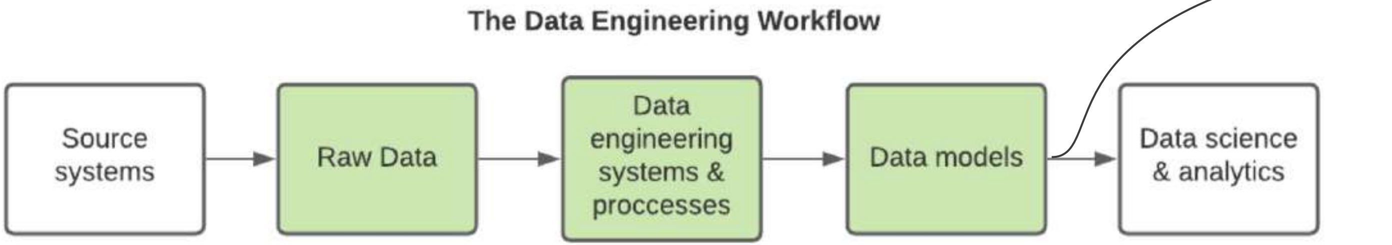
www

and more data sources..

Cr. Data TH.com - Data Science ชิลชิล

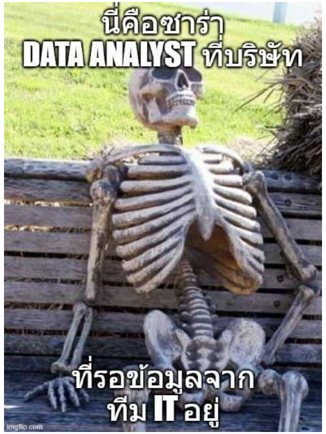
What is Data Engineer?

data engineers set up and operate the **organization’s data infrastructure preparing it for** further analysis by **data analysts and scientists**

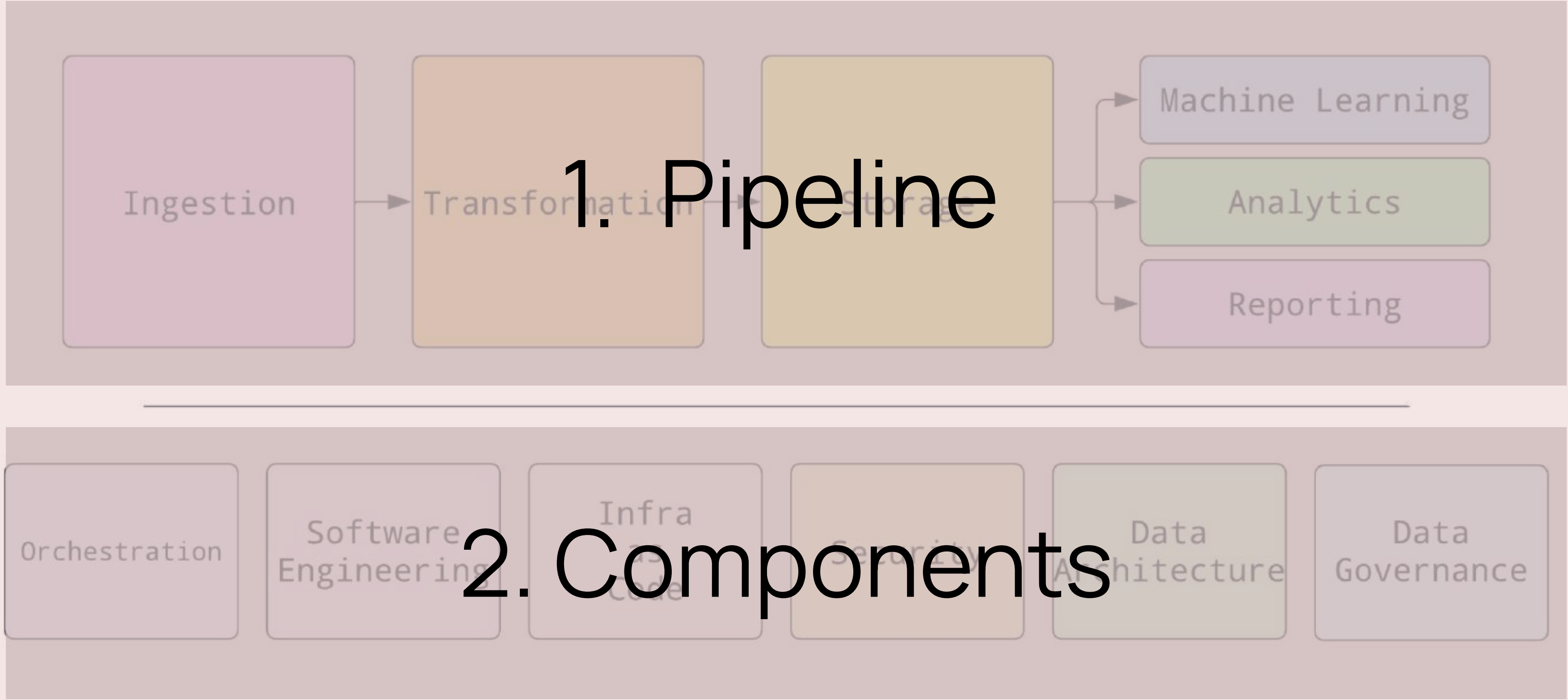


The Data Engineering Workflow

Source systems → Raw Data → Data engineering systems & proccesses → Data models → Data science & analytics



Data Engineering Life Cycle



1. Pipeline

Ingestion → Transformation → Storage → Machine Learning, Analytics, Reporting

2. Components

Orchestration

Software Engineering

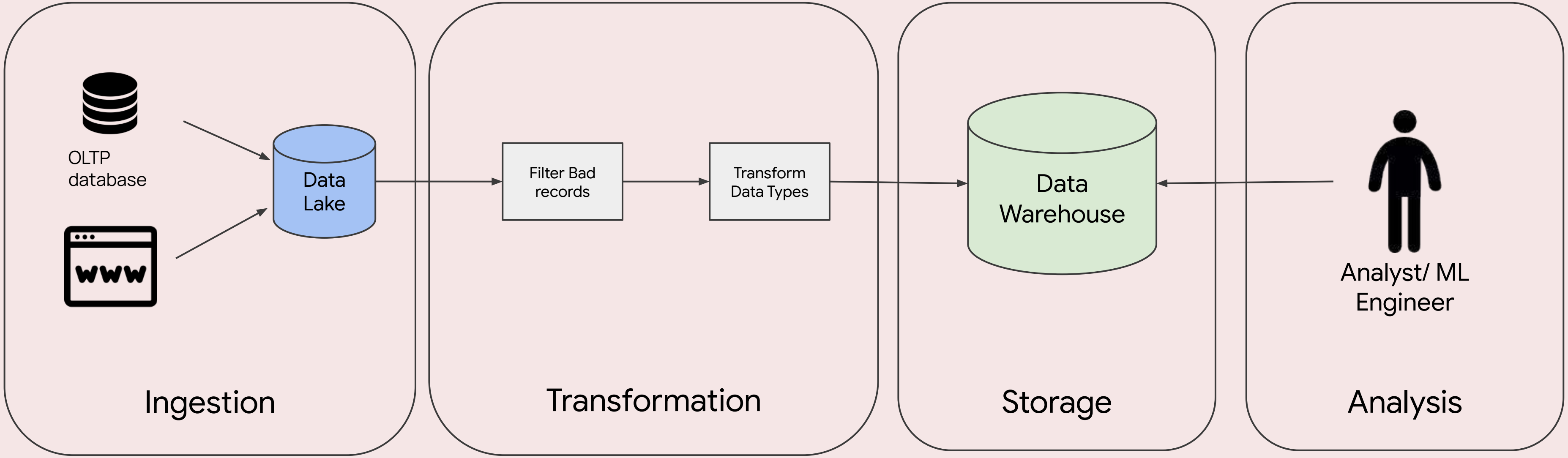
Infra as Code

Security

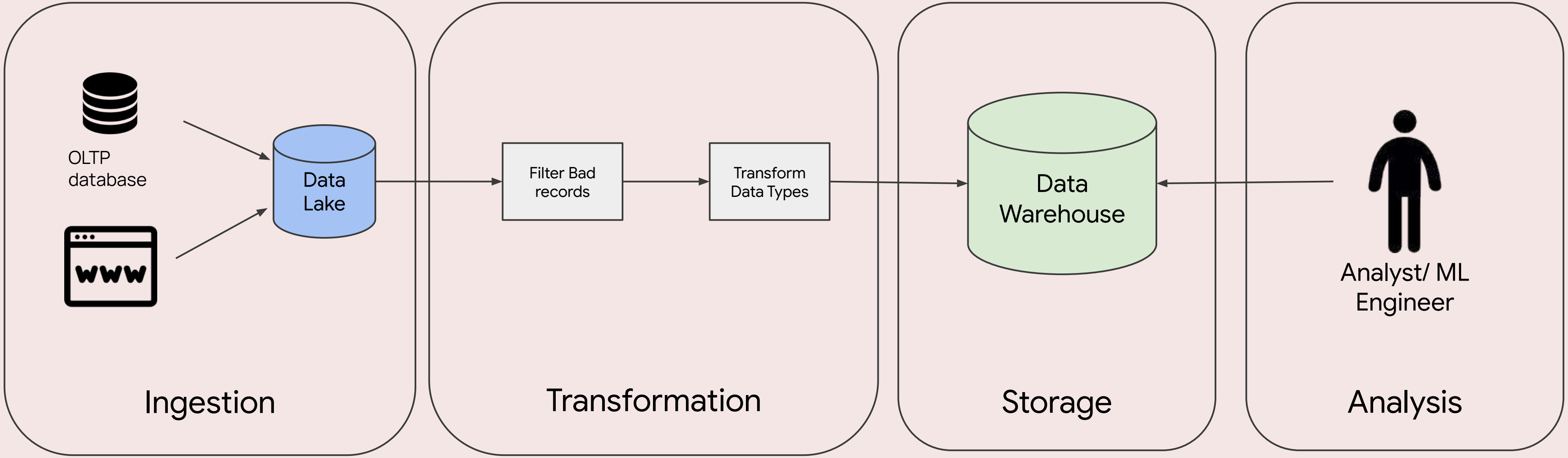
Data Architecture

Data Governance

4 stages in Data Pipeline



Data Lake vs Data Warehouse vs Data Mart



OLTP database

Data Lake

Filter Bad records

Transform Data Types

Data Warehouse

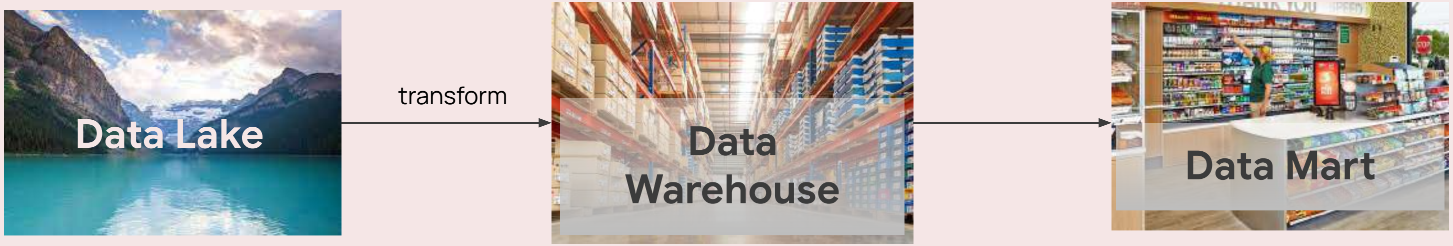
Analyst/ ML Engineer

Ingestion

Transformation

Storage

Analysis



Data Lake

transform

Data Warehouse

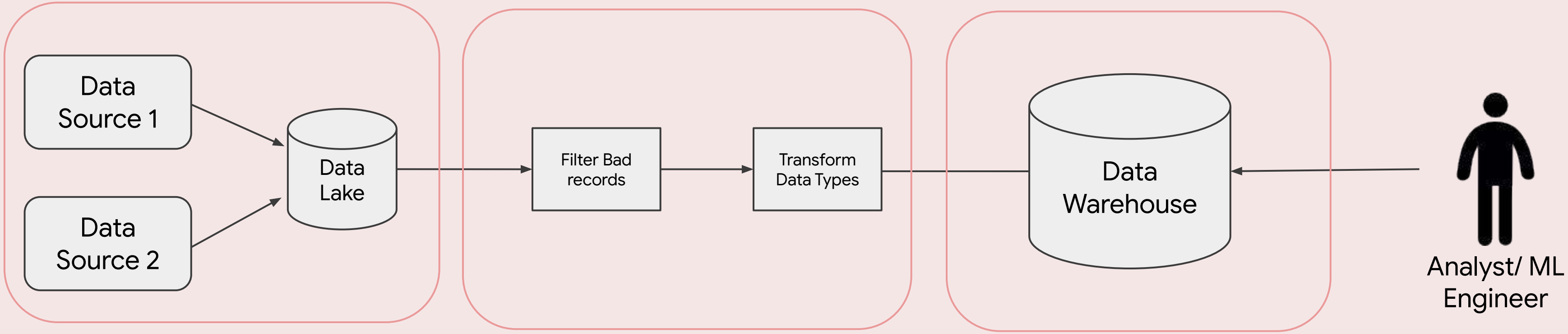
Data Mart

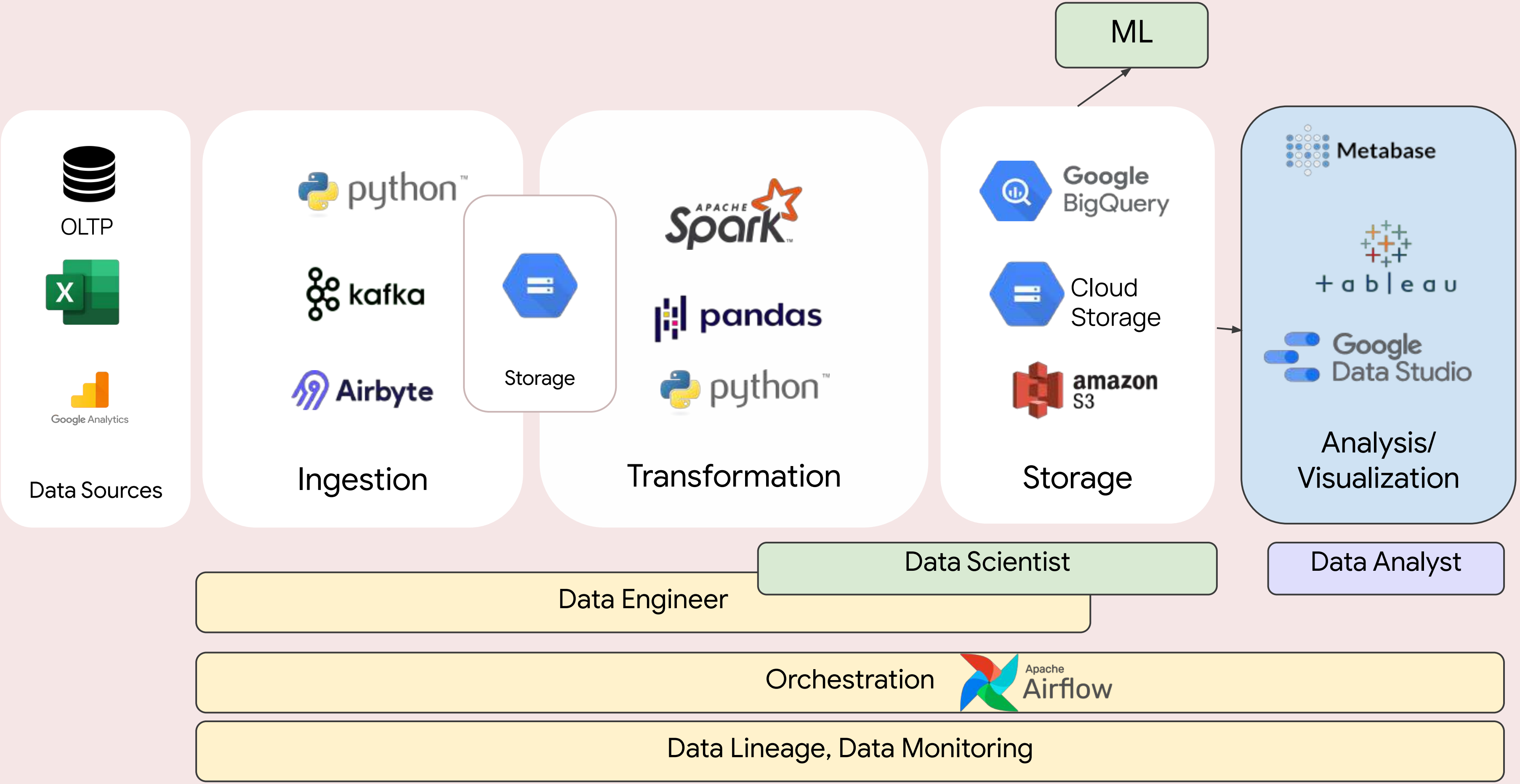
ETL vs ELT

Extract - pull data from all your data sources

Transform - clean and process data

Load - load into storage destination





ML

OLTP

Google Analytics

Data Sources

python

kafka

Airbyte

Ingestion

Storage

Apache Spark

pandas

python

Transformation

Google BigQuery

Cloud Storage

amazon S3

Storage

Metabase

tableau

Google Data Studio

Analysis/ Visualization

Data Scientist

Data Engineer

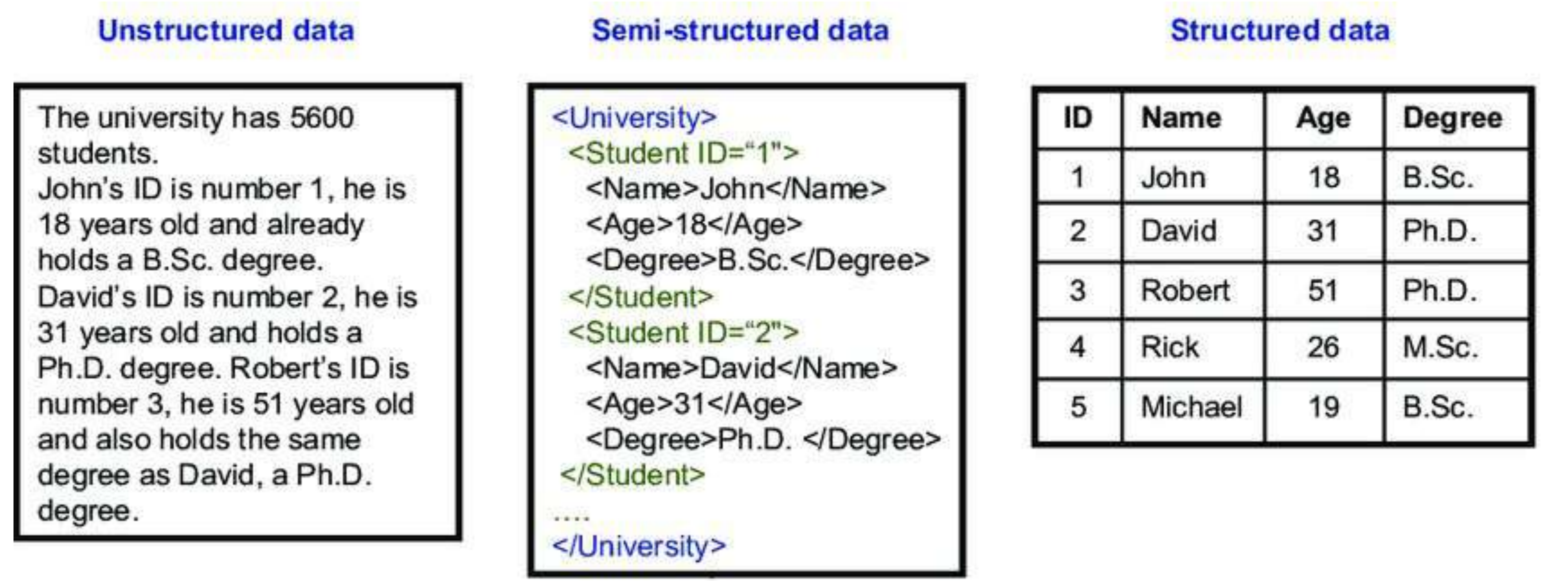
Data Analyst

Orchestration Apache Airflow

Data Lineage, Data Monitoring

https://www.researchgate.net/figure/Unstructured-semi-structured-and-structured-data_fig4_236860222

Structure



Unstructured data

The university has 5600 students. John's ID is number 1, he is 18 years old and already holds a B.Sc. degree. David's ID is number 2, he is 31 years old and holds a Ph.D. degree. Robert's ID is number 3, he is 51 years old and also holds the same degree as David, a Ph.D. degree.

Semi-structured data

```
<University>
 <Student ID="1">
  <Name>John</Name>
  <Age>18</Age>
  <Degree>B.Sc.</Degree>
 </Student>
 <Student ID="2">
  <Name>David</Name>
  <Age>31</Age>
  <Degree>Ph.D. </Degree>
 </Student>
....
</University>
```

Structured data

ID	Name	Age	Degree
1	John	18	B.Sc.
2	David	31	Ph.D.
3	Robert	51	Ph.D.
4	Rick	26	M.Sc.
5	Michael	19	B.Sc.

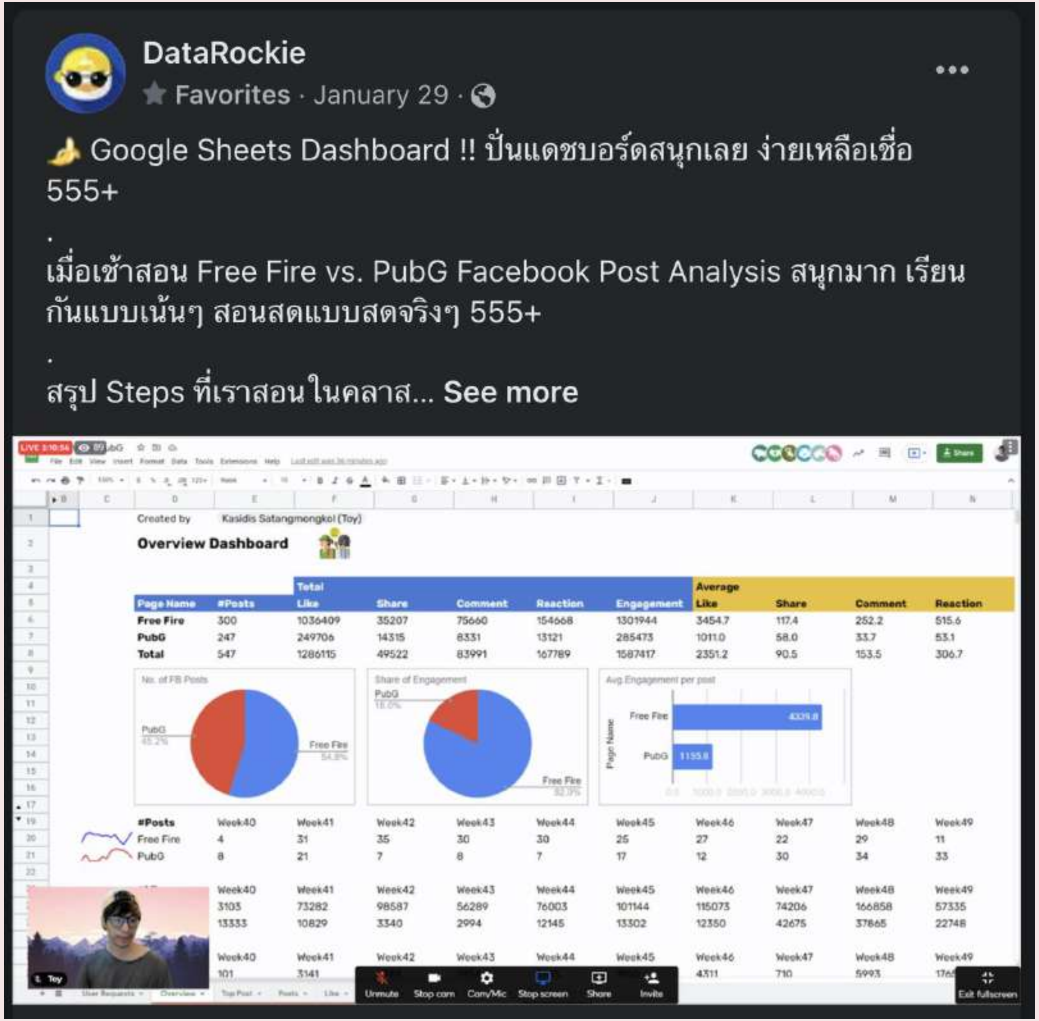
Text, Audio, Video, PDF, Internet of Things (IoT) sensor data

XML, CSV, JSON, Web pages

PostgreSQL, MySQL

Q: Facebook post เป็นข้อมูลแบบไหน?

1. Structured data
2. Semi-structured data
3. Unstructured data



DataRockie

Favorites · January 29

🍌 Google Sheets Dashboard !! ปั่นแดชบอร์ดสนุกเลย ง่ายเหลือเชื่อ 555+

.

เมื่อเช้าสอน Free Fire vs. PubG Facebook Post Analysis สนุกมาก เรียนกันแบบเน้นๆ สอนสดแบบสดจริงๆ 555+

.

สรุป Steps ที่เราสอนในคลาส... See more

Created by Kasidis Satangmongkol (Toy)

Overview Dashboard

		Total					Average			
Page Name	#Posts	Like	Share	Comment	Reaction	Engagement	Like	Share	Comment	Reaction
Free Fire	300	1036409	35207	75660	154668	1301944	3454.7	117.4	252.2	515.6
PubG	247	249706	14315	8331	13121	285473	1011.0	58.0	33.7	53.1
Total	547	1286115	49522	83991	167789	1587417	2351.2	90.5	153.5	306.7

#Posts	Week40	Week41	Week42	Week43	Week44	Week45	Week46	Week47	Week48	Week49
Free Fire	4	31	35	30	30	25	27	22	29	11
PubG	8	21	7	8	7	17	12	30	34	33

Week40	Week41	Week42	Week43	Week44	Week45	Week46	Week47	Week48	Week49
3103	73282	98587	56289	76003	101144	115073	74206	166858	57335
13333	10829	3340	2994	12145	13302	12350	42675	37865	22748

Week40	Week41	Week42	Week43	Week44	Week45	Week46	Week47	Week48	Week49
101	3141					4311	710	5993	176

Databases

Relational(SQL)

Traditional database/ DBMS

Non-relational (NoSQL)

Databases

Relational (SQL)

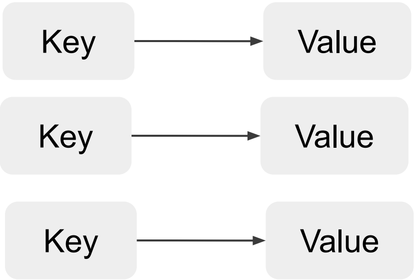
Row - oriented

1	red	square	..
2	blue	square	..
3	yellow	square	..
4	none	triangle	..

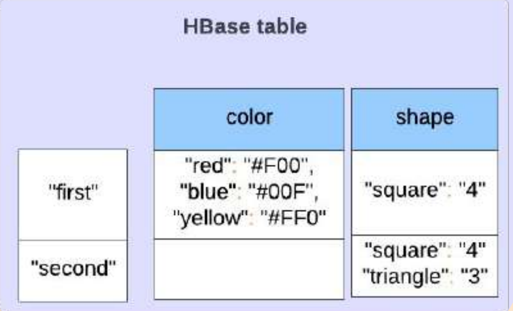
square	4
triangle	3

Non-relational (NoSQL)

Key-Value



Columnar



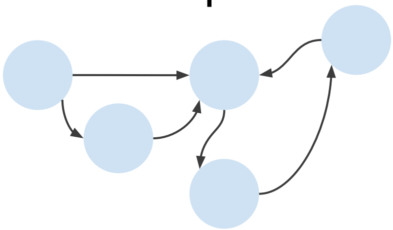
HBase table

	color	shape
"first"	"red": "#F00", "blue": "#00F", "yellow": "#FF0"	"square": "4"
"second"		"square": "4" "triangle": "3"

Document

```
{
  "_id": "ObjectId("d7caskf00010dsa")",
  "firstName": "Burasakorn",
  "lastName": "Sabyeying",
  “nickName”: “Mils”,
   "role": "Data Engineer"
}
```

Graph



Databases

Relational (SQL)

Traditional database/ DBMS

PostgreSQL

MySQL

SQLite

Non-relational (NoSQL)

Key-Value

redis

amazon DynamoDB

Columnar

APACHE HBASE

cassandra

Document

mongoDB

CouchDB relax

Graph

neo4j

OLTP vs OLAP

Online **Transaction** Processing

- captures, **stores**, and processes data from transactions **in real time**
- banking and credit card activity or retail checkout scanning.
- **traditional DBMS**
- Based on INSERT, UPDATE, DELETE commands

MySQL

Online **Analytical** Processing

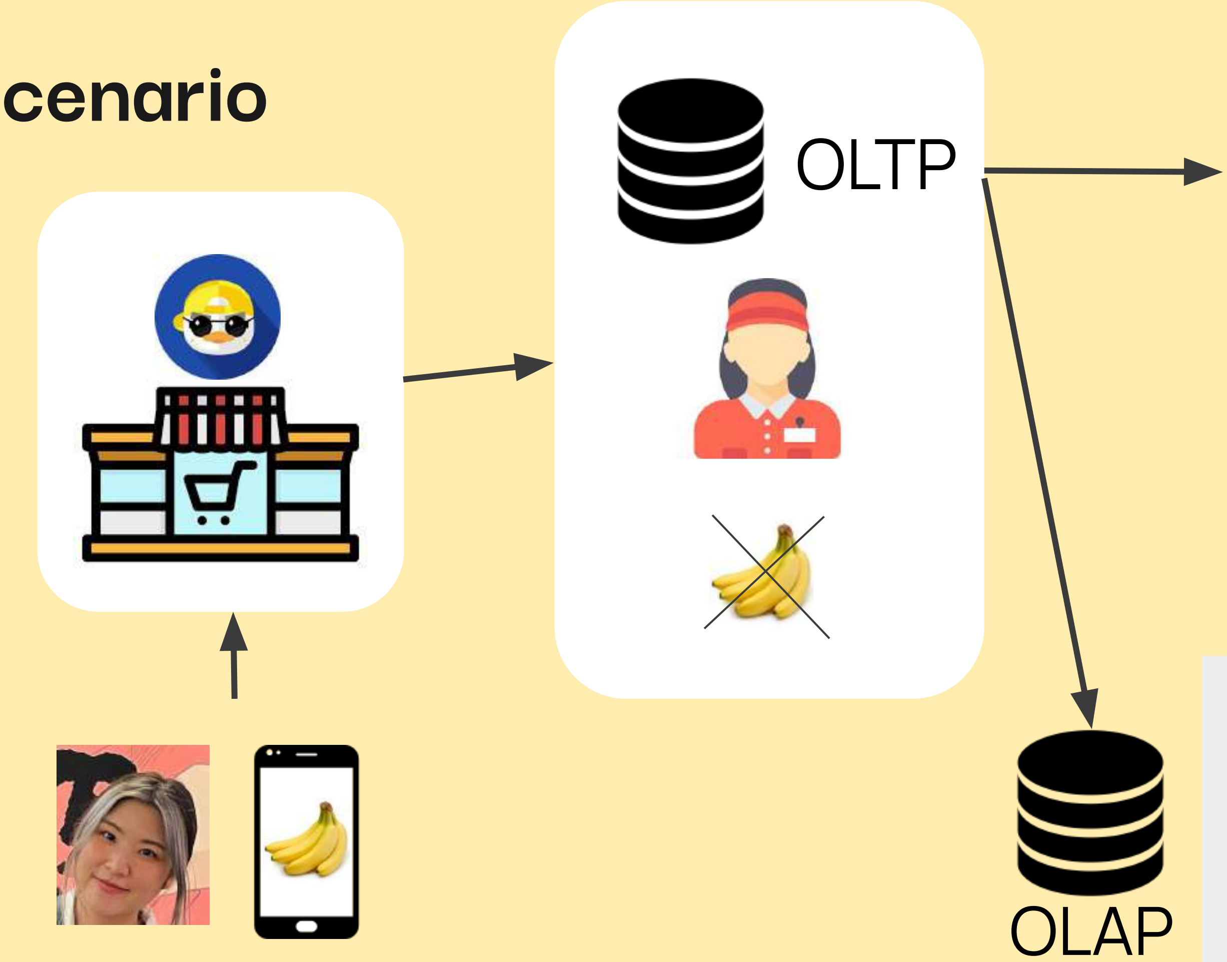
- analyze **aggregated historical data** from OLTP systems.
- designed for use by data scientists, business analysts
- For **data warehouse** and data mart applications
- Based on SELECT commands to aggregate data for reporting

Google BigQuery

Azure Synapse Analytics

amazon REDSHIFT

Scenario



```
SELECT *
FROM table
WHERE date=today
```

```
SELECT
          country.country_name_eng,
          SUM(CASE WHEN call.id IS NOT NULL THEN 1 ELSE
0 END) AS calls,
          AVG(ISNULL(DATEDIFF(SECOND, call.start_time,
call.end_time),0)) AS avg_difference
FROM country
LEFT JOIN city ON city.country_id = country.id
LEFT JOIN customer ON city.id = customer.city_id
LEFT JOIN call ON call.customer_id = customer.id
GROUP BY
          country.id,
          country.country_name_eng
HAVING AVG(ISNULL(DATEDIFF(SECOND, call.start_time,
call.end_time),0)) > (SELECT AVG(DATEDIFF(SECOND,
call.start_time, call.end_time)) FROM call)
ORDER BY calls DESC, country.id ASC;
```

Data Lake, Data Warehouse, Data Lakehouse

+

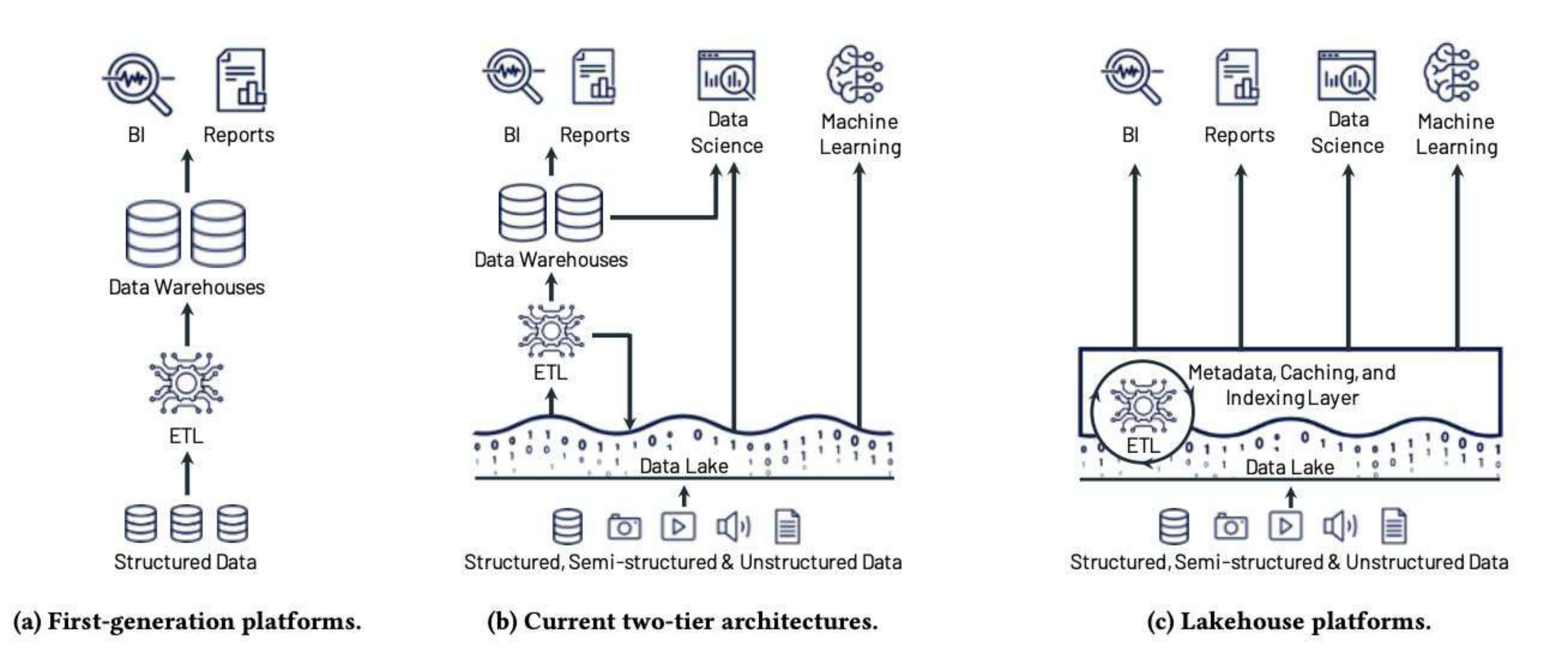
=

Unstructured data

Structured data

Unstructured data

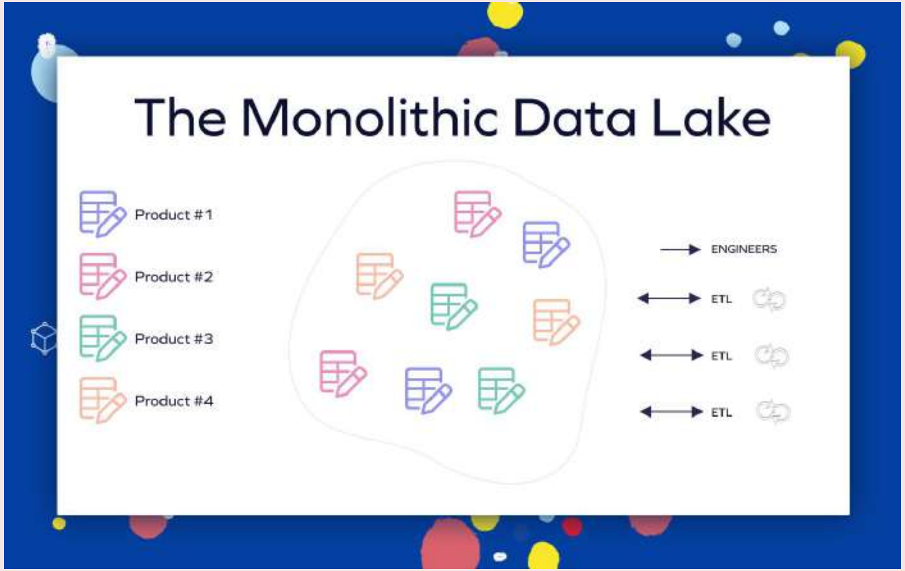
Data Lakehouse



(a) First-generation platforms. **(b) Current two-tier architectures.** **(c) Lakehouse platforms.**

CIDR 2021: **Lakehouse: A New Generation of Open Platforms that Unify Data Warehousing and Advanced Analytics**

Centralized Data vs Decentralized Data



The Monolithic Data Lake

Product #1

Product #2

Product #3

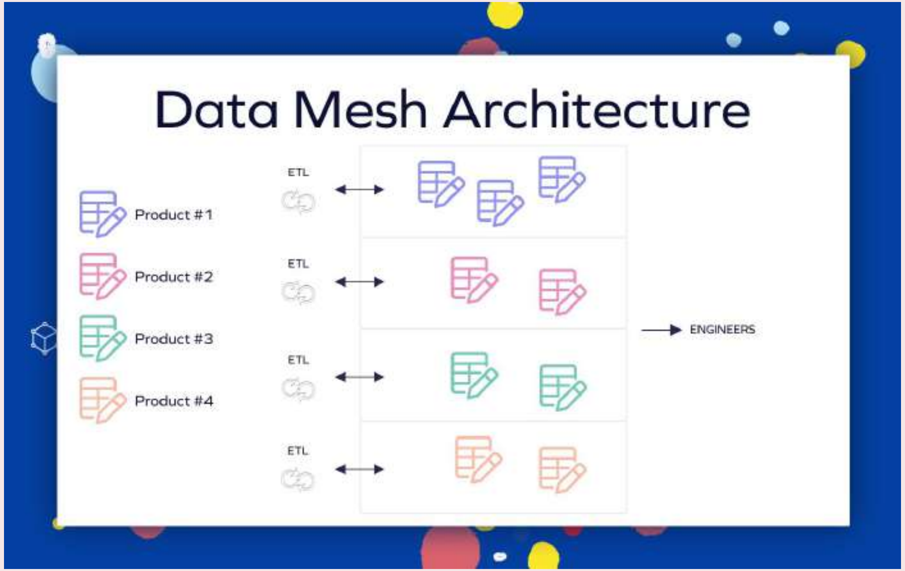
Product #4

ENGINEERS

ETL

ETL

ETL



Data Mesh Architecture

Product #1

Product #2

Product #3

Product #4

ETL

ETL

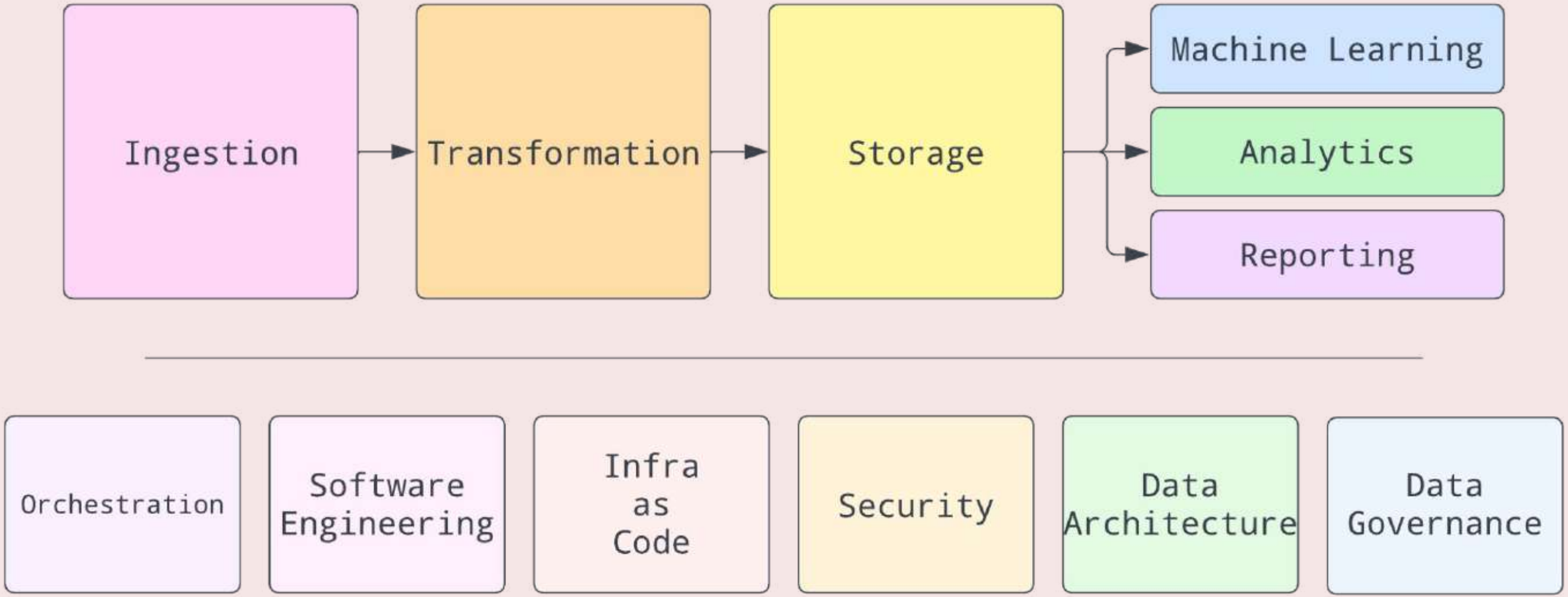
ETL

ETL

ENGINEERS

https://medium.com/yotpoengineering/the-4-data-mesh-principles-to-create-a-data-oriented-rnd-6f2e291bcb5b

Data Engineering Life Cycle



Ingestion

Transformation

Storage

Machine Learning

Analytics

Reporting

Orchestration

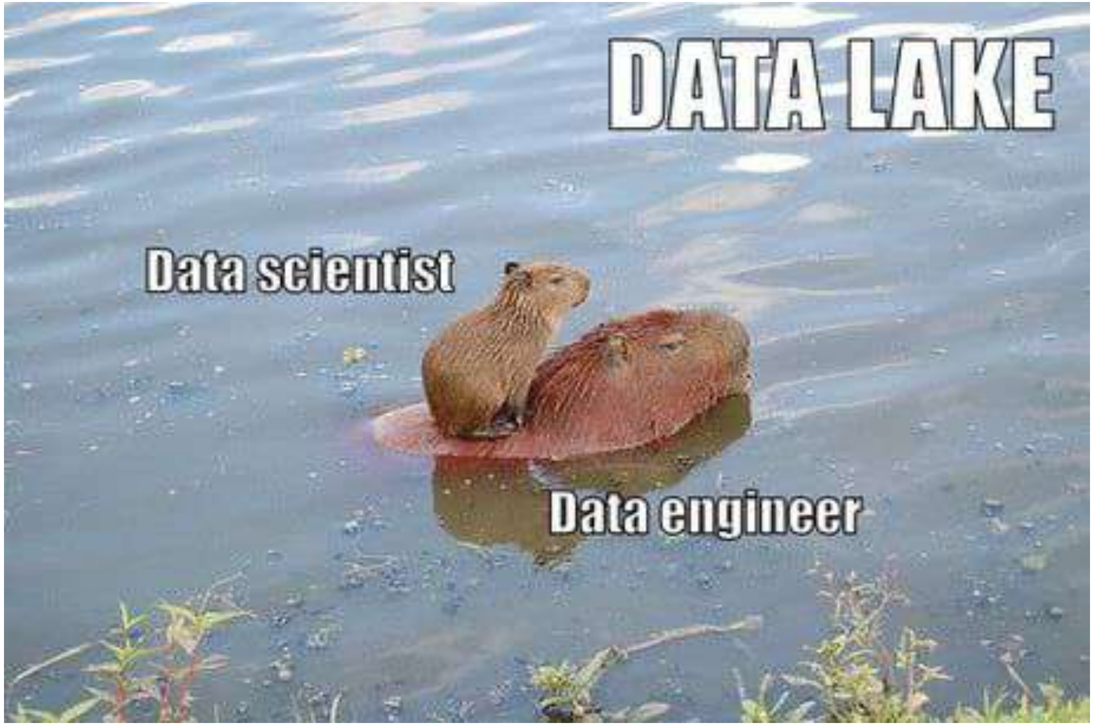
Software Engineering

Infra as Code

Security

Data Architecture

Data Governance

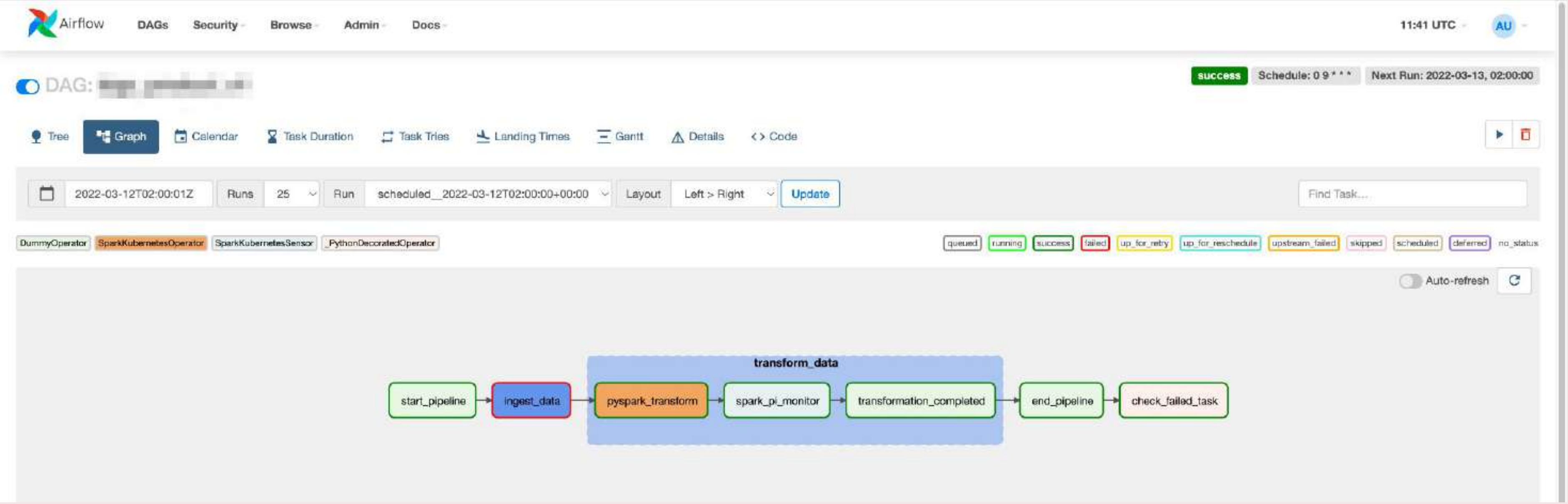

DATA LAKE
Data scientist
Data engineer

Orchestration

https://www.primoartists.com/news/gemma-new-receives-praise-for-milwaukee-symphony-debut

Orchestration

process of coordinating many jobs to run efficiently



DAG: [redacted]

success

Schedule: 0 9 * * *

Next Run: 2022-03-13, 02:00:00

2022-03-12T02:00:01Z

Runs 25

Run scheduled__2022-03-12T02:00:00+00:00

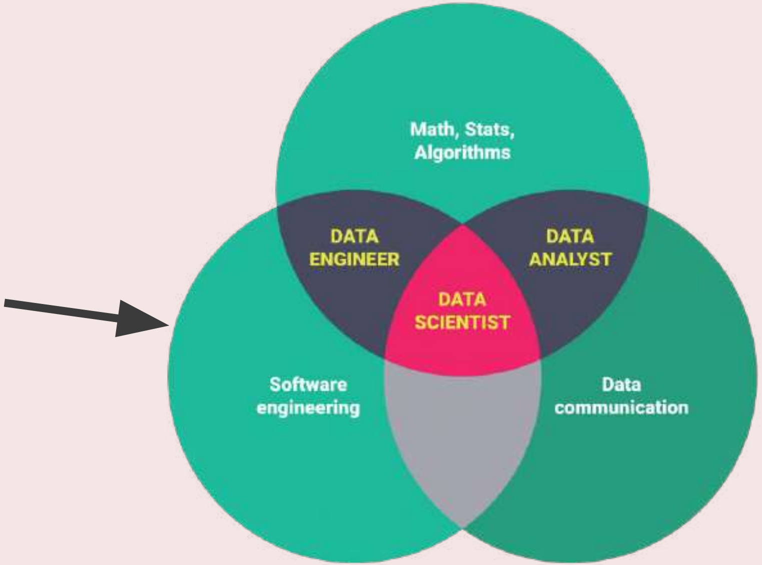
Layout Left > Right

Apache Airflow

Google Cloud Composer

Managed service

Software Engineer



- Web Scraping
- Get data through API

Google Analytics

1 page (1-100 rows)
Total: 300 pages
300 API calls

- Process data

  Pyspark, Pandas

- Pipeline as code

https://www.techiexpert.com/close-look-data-scientist-vs-data-engineer/

Which language should I know?

Which language should I know?

```
SELECT * FROM Customers;
```

SQL

pandas

NumPy

Apache Airflow

Python

Apache Spark

JVM languages (Java, Scala)

Apache Spark

Bash

```
ubuntu@ubuntu:~$ sudo
```

https://www.freecodecamp.org/news/linux-ln-how-to-create-a-symbolic-link-in-linux-example-bash-command/

Infrastructure as a Code

Containers

**important

docker

kubernetes

Provisioning

HashiCorp
Terraform

Version control

**important

git

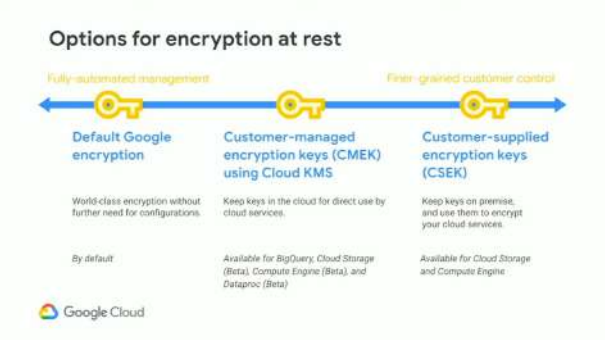
Security

In both Data and System

Who should see the data?
Owner
Editor
Viewer

Key management

Encryption



Options for encryption at rest

Default Google encryption

World-class encryption without further need for configurations.

By default

Customer-managed encryption keys (CMEK) using Cloud KMS

Keep keys in the cloud for direct use by cloud services.

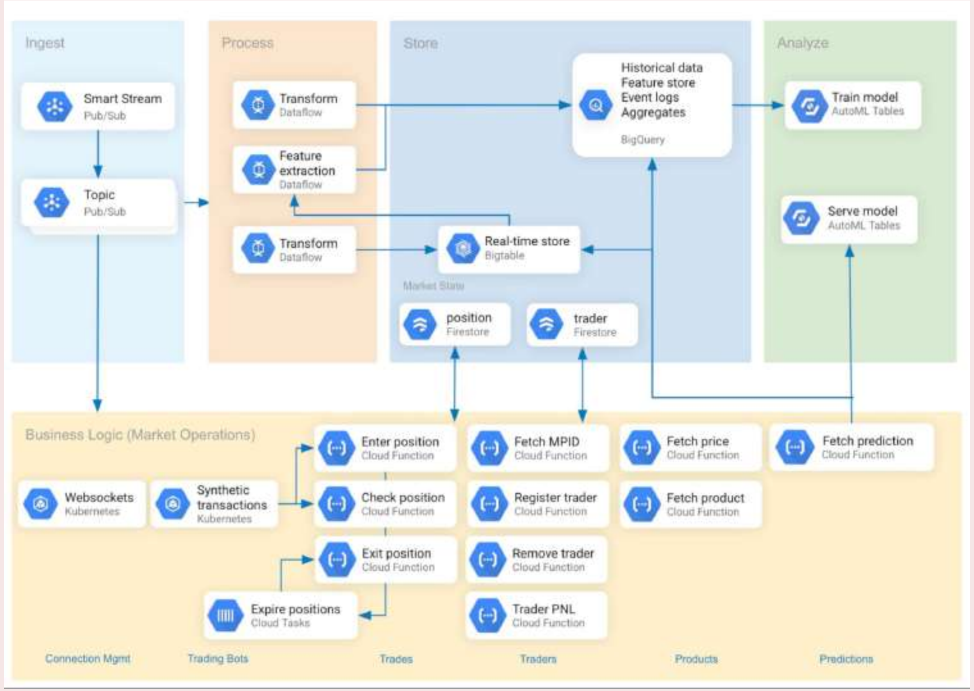
Available for BigQuery, Cloud Storage (Beta), Compute Engine (Beta), and Dataproc (Beta)

Customer-supplied encryption keys (CSEK)

Keep keys on premise, and use them to encrypt your cloud services.

Available for Cloud Storage and Compute Engine

Data Architecture



Architecture != Tools

ต้องเข้าใจ

- **business requirement**
- Nature การเกิดของข้อมูล
- Nature การใช้ของข้อมูล

แล้วจะปรับสิ่งพวกนี้มา design ในการ serve data ยังไง

On premises vs Cloud

On premises/ On-prem

= purchase hardware/ data centers they own

- Still default for established companies
- install/ maintain/ upgrade by their own
- Direct control over configuration, management, security

Cloud

= **Cloud provider** (AWS, Azure, Google Cloud, etc)

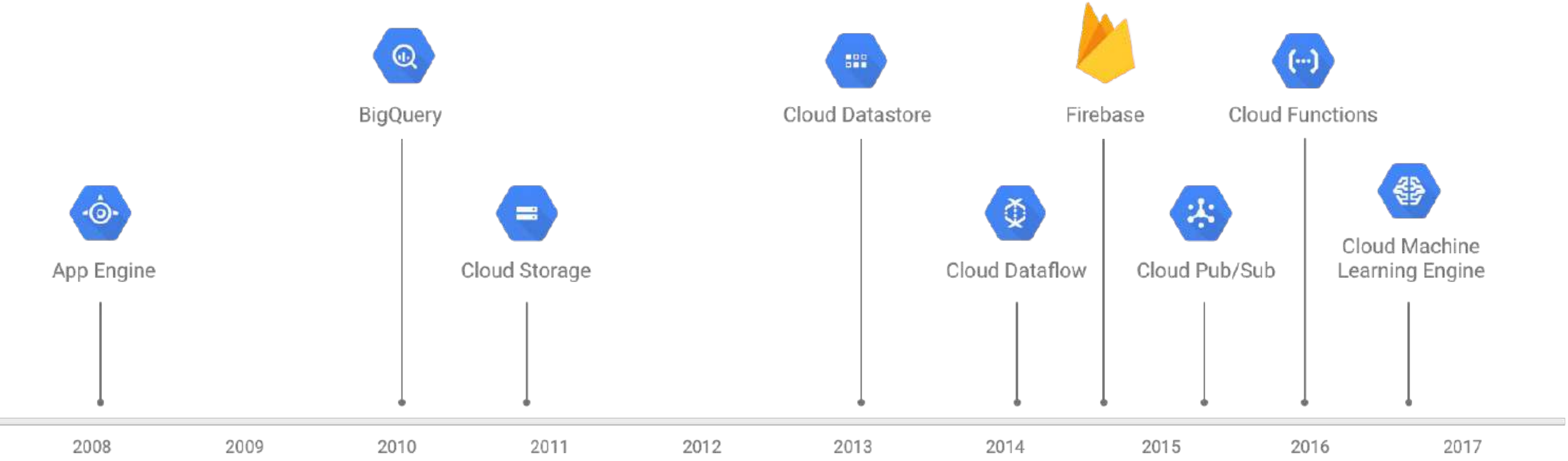
- Infrastructure as a Service(IaaS)
- **Serverless products** and managed service
- Billed on **pay-as-you-go**
- Unpredictable scale requirements

Serverless

cloud-native development model that allows developers to build and run applications **without having to manage servers**

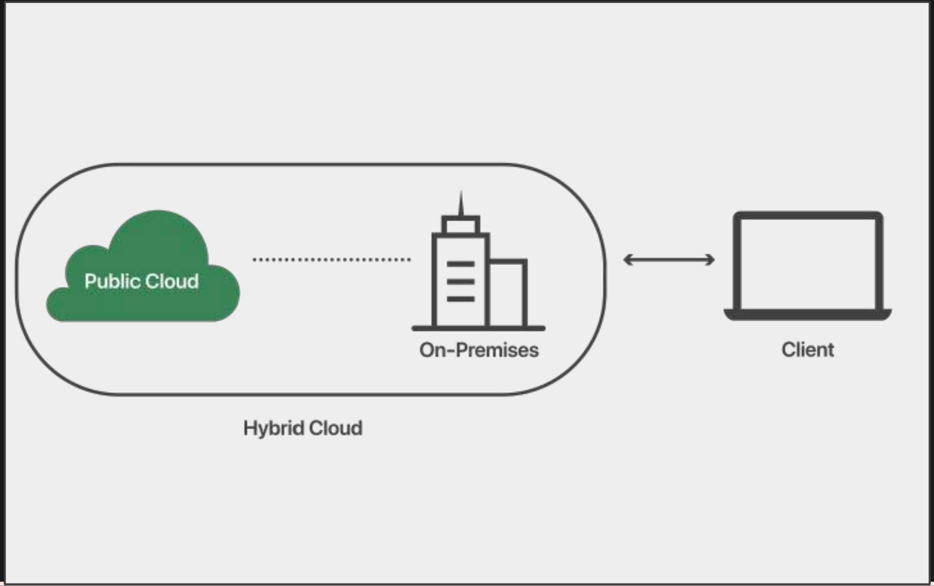
Serverless = Still have server

Google Cloud Platform (GCP) serverless products



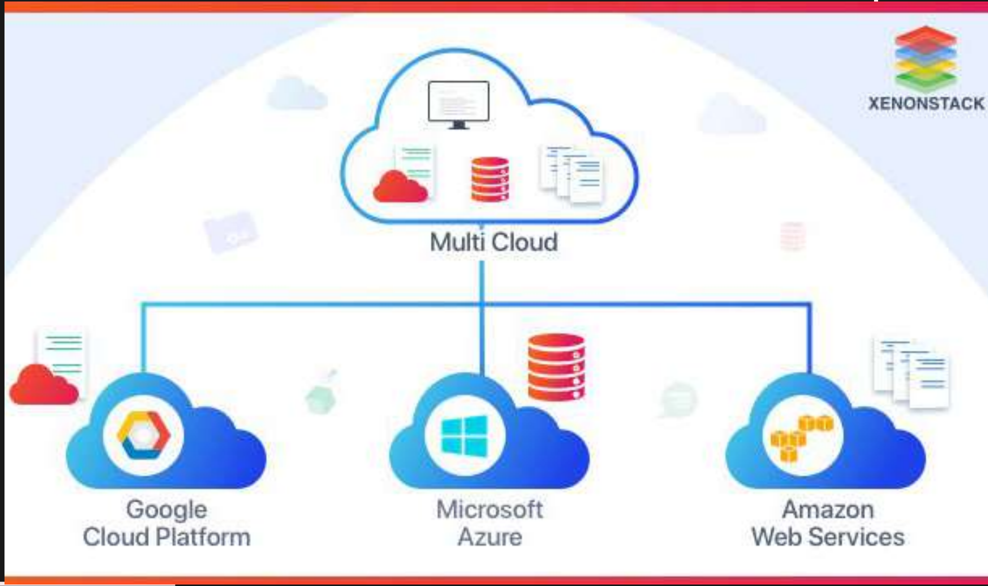
Hybrid and multi-cloud

Hybrid



Cr. cloudflare.com

Multi-cloud



Cr. xenonstack.com

Data Architecture

Best Data Pipeline ?

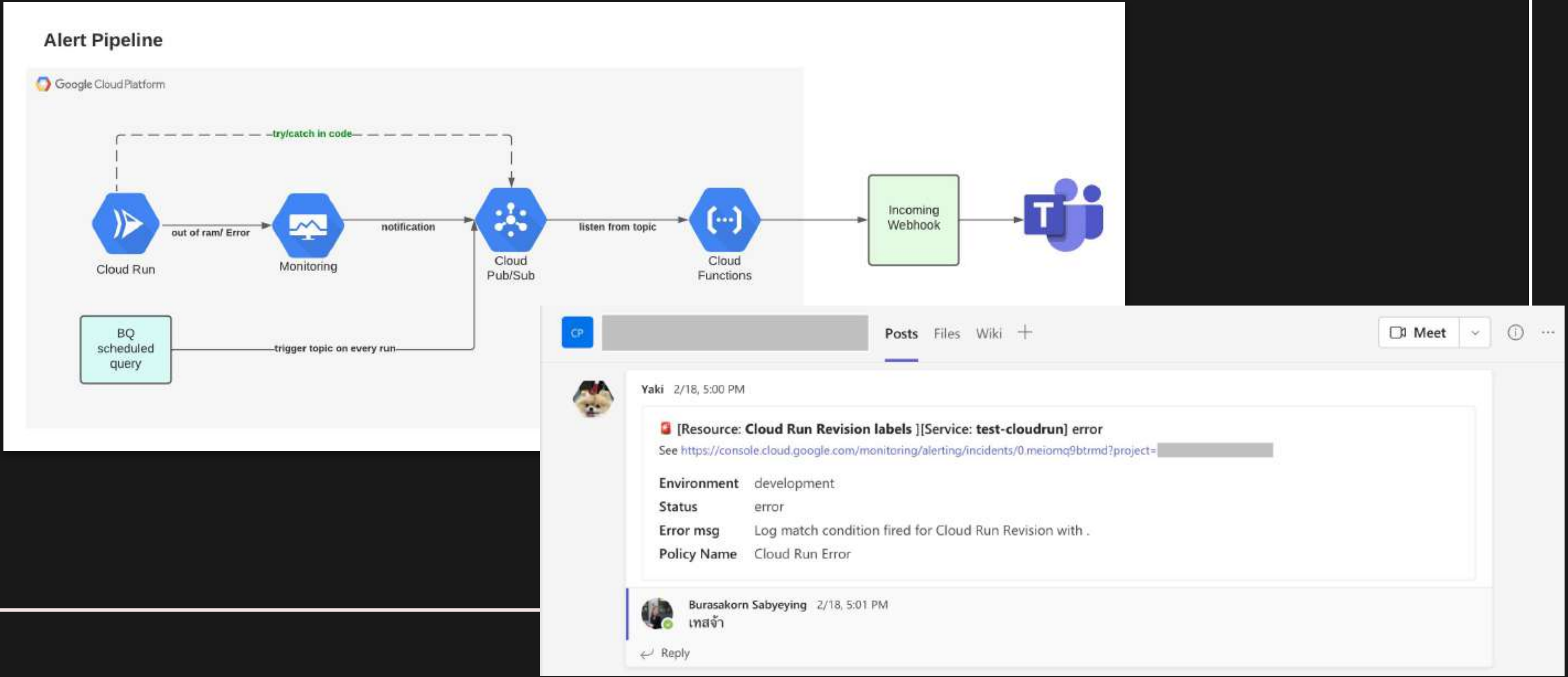
cr. Richard Burlton on Unsplash

Data Governance

Data Governance: The Definitive Guide -
“Data governance is, first and foremost, a **data management function** to ensure the quality, integrity, security, and usability of the data collected **by an organization**.”

1. Data Monitoring
2. Data Discovery & Data Catalog
3. Data Lineage
4. Data Quality

Data Monitoring



[redacted]

Yaki 2/18, 5:00 PM

[Resource: **Cloud Run Revision labels**][Service: **test-cloudrun**] error

See https://console.cloud.google.com/monitoring/alerting/incidents/0.meiomq9btrmd?project=[redacted]

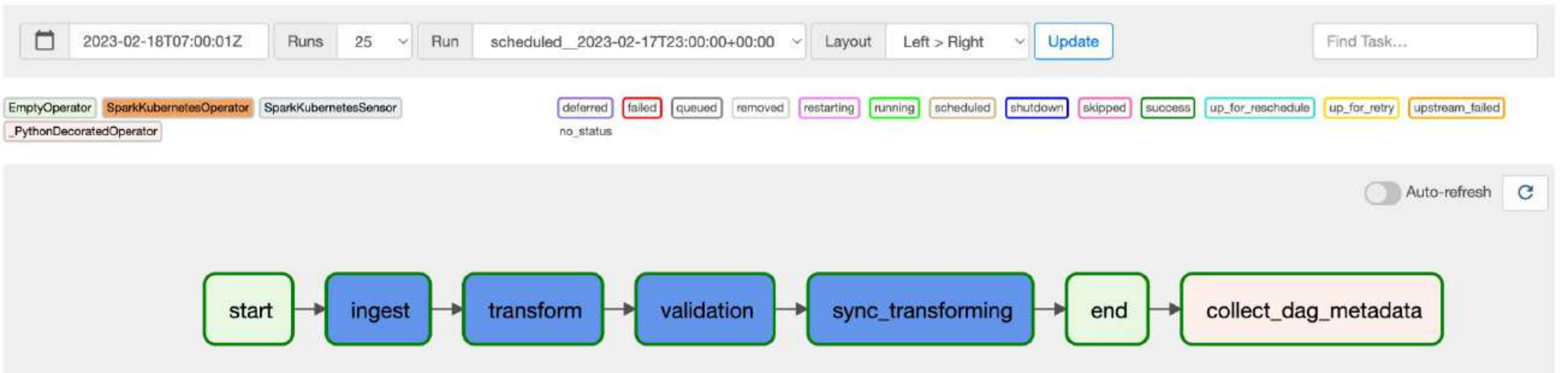
Environment	development
Status	error
Error msg	Log match condition fired for Cloud Run Revision with .
Policy Name	Cloud Run Error

Burasakorn Sabyeying 2/18, 5:01 PM

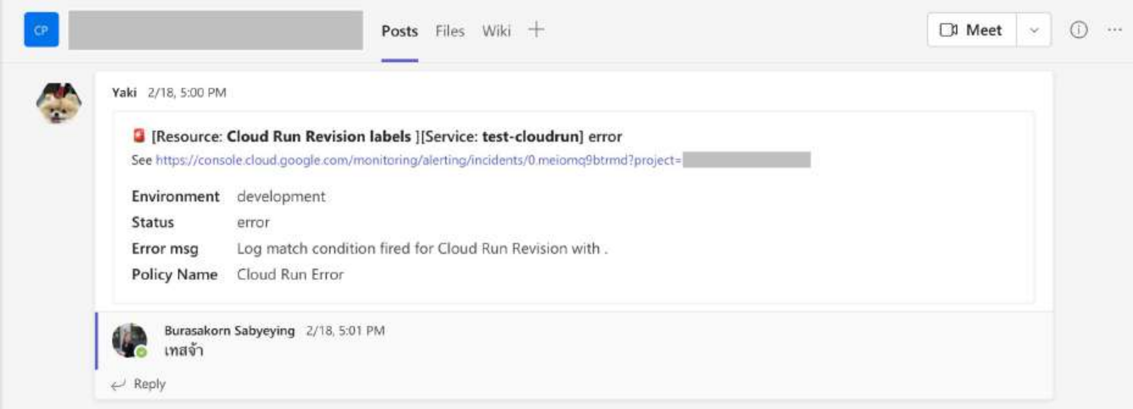
เทสจ้า

Reply

Data Monitoring



2023-02-18T07:00:01Z Runs 25 Run scheduled__2023-02-17T23:00:00+00:00 Layout Left > Right



Yaki 2/18, 5:00 PM

[Resource: **Cloud Run Revision labels**][Service: **test-cloudrun**] error

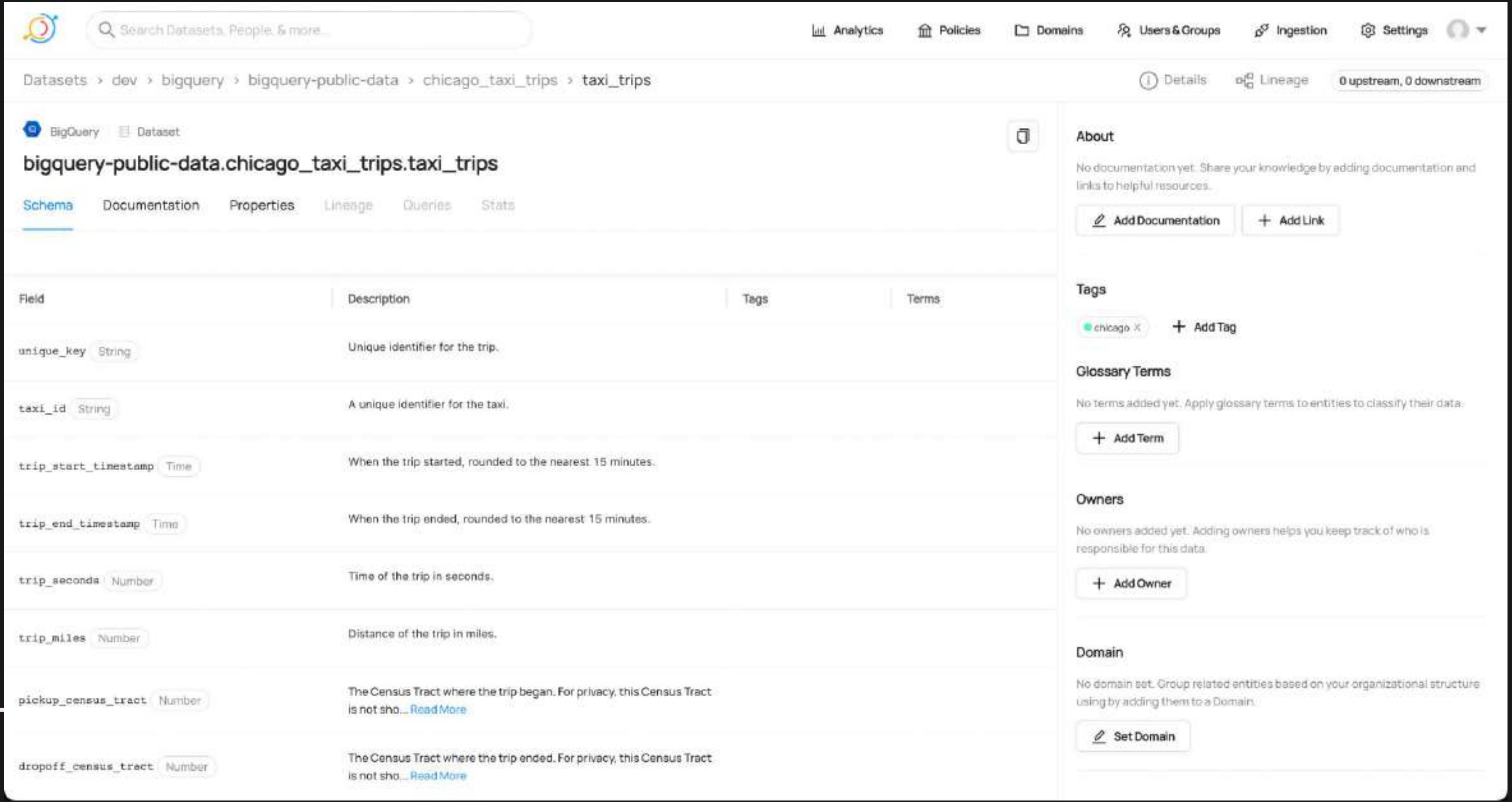
See https://console.cloud.google.com/monitoring/alerting/incidents/0.meiomq9btrmd?project=[redacted]

Environment	development
Status	error
Error msg	Log match condition fired for Cloud Run Revision with .
Policy Name	Cloud Run Error

Burasakorn Sabyeying 2/18, 5:01 PM

เทสจ้า

Data Discovery & Data Catalog



bigquery-public-data.chicago_taxi_trips.taxi_trips

Field	Description	Tags	Terms
unique_key String	Unique identifier for the trip.		
taxi_id String	A unique identifier for the taxi.		
trip_start_timestamp Time	When the trip started, rounded to the nearest 15 minutes.		
trip_end_timestamp Time	When the trip ended, rounded to the nearest 15 minutes.		
trip_seconds Number	Time of the trip in seconds.		
trip_miles Number	Distance of the trip in miles.		
pickup_census_tract Number	The Census Tract where the trip began. For privacy, this Census Tract is not sho... Read More		
dropoff_census_tract Number	The Census Tract where the trip ended. For privacy, this Census Tract is not sho... Read More		

About

No documentation yet. Share your knowledge by adding documentation and links to helpful resources.

Tags

Glossary Terms

No terms added yet. Apply glossary terms to entities to classify their data.

Owners

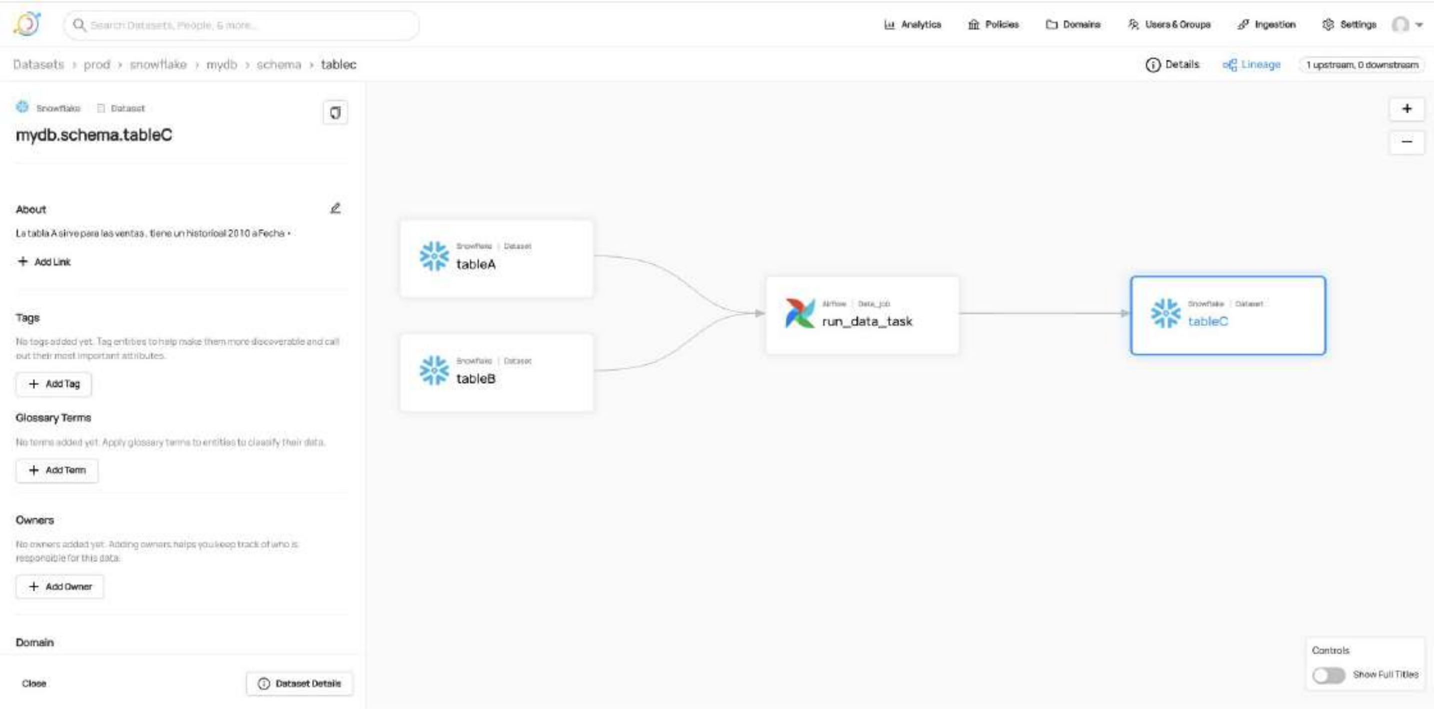
No owners added yet. Adding owners helps you keep track of who is responsible for this data.

Domain

No domain set. Group related entities based on your organizational structure using by adding them to a Domain.

Data Lineage

allows you to know where that data is stored and its dependencies.



mydb.schema.tableC

About

La tabla A sirve para las ventas , tiene un historical 2010 a Fecha •

Tags

No tags added yet. Tag entities to help make them more discoverable and call out their most important attributes.

Glossary Terms

No terms added yet. Apply glossary terms to entities to classify their data.

Owners

No owners added yet. Adding owners helps you keep track of who is responsible for this data.

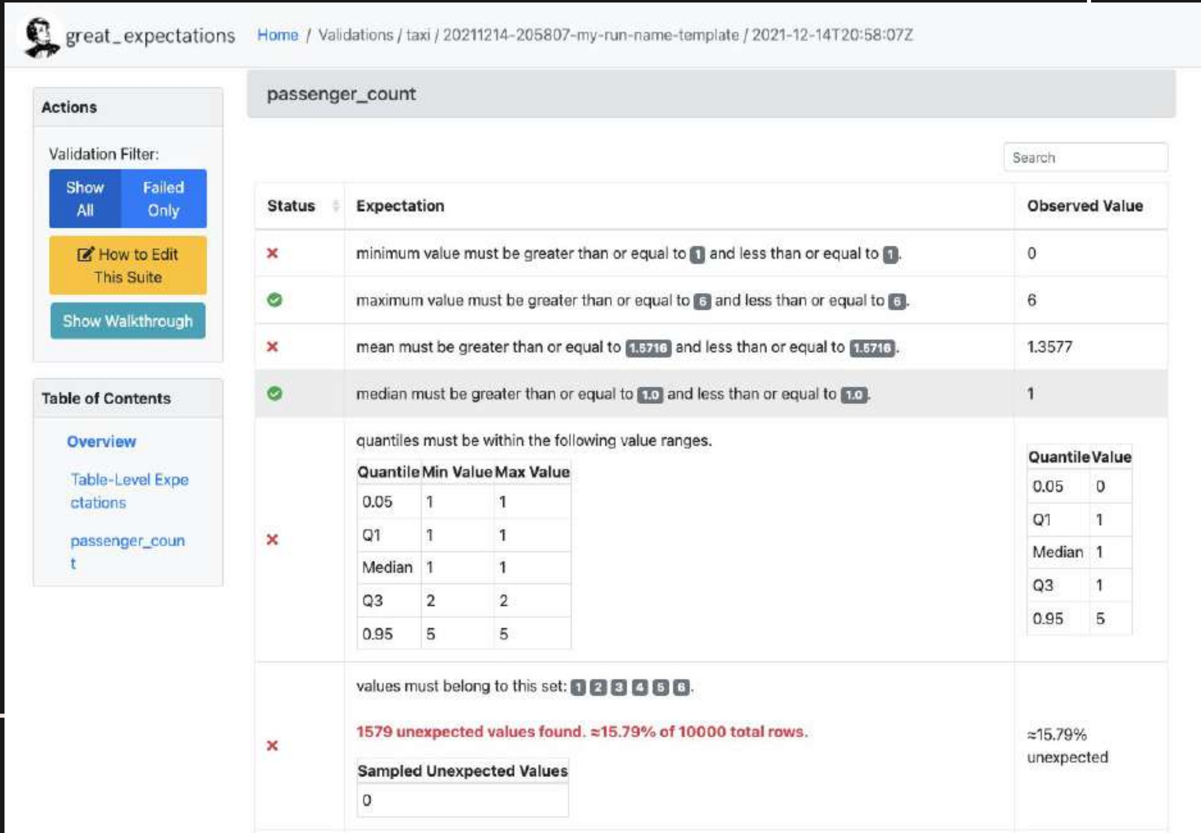
Data Quality* (!important)

According to *Data Governance: The Definitive Guide*, data quality is defined by **three main characteristics**

Accuracy
Is the collected data factually correct? Are there duplicate values? Are the numeric values accurate?

Completeness
Are the records complete? Do all required fields contain valid values?

Timeliness
Are records available in a timely fashion?



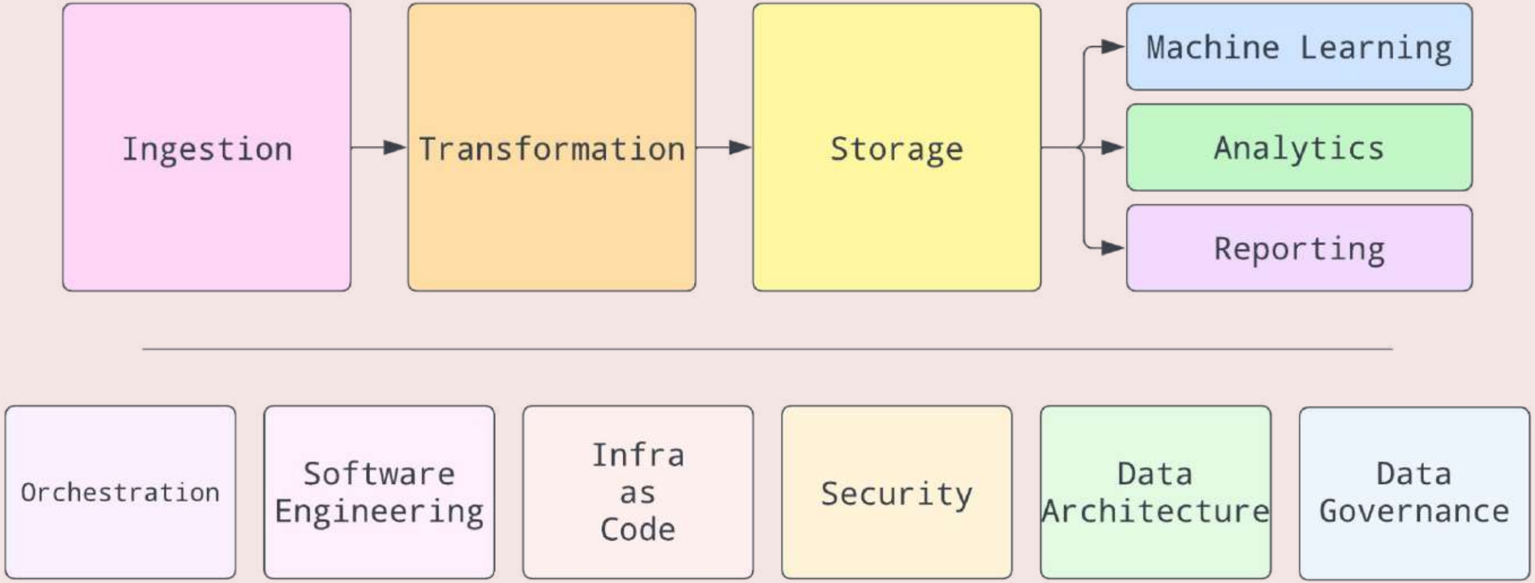
Home / Validations / taxi / 20211214-205807-my-run-name-template / 2021-12-14T20:58:07Z

passenger_count

<table>
<tr><th>Status</th><th>Expectation</th><th>Observed Value</th></tr>
<tr><td>✗</td><td>minimum value must be greater than or equal to 1 and less than or equal to 1.</td><td>0</td></tr>
<tr><td>✓</td><td>maximum value must be greater than or equal to 6 and less than or equal to 6.</td><td>6</td></tr>
<tr><td>✗</td><td>mean must be greater than or equal to 1.5716 and less than or equal to 1.5716.</td><td>1.3577</td></tr>
<tr><td>✓</td><td>median must be greater than or equal to 1.0 and less than or equal to 1.0.</td><td>1</td></tr>
<tr><td>✗</td><td>quantiles must be within the following value ranges.
<table>
<tr><th>Quantile</th><th>Min Value</th><th>Max Value</th></tr>
<tr><td>0.05</td><td>1</td><td>1</td></tr>
<tr><td>Q1</td><td>1</td><td>1</td></tr>
<tr><td>Median</td><td>1</td><td>1</td></tr>
<tr><td>Q3</td><td>2</td><td>2</td></tr>
<tr><td>0.95</td><td>5</td><td>5</td></tr>
</table></td><td>
<table>
<tr><th>Quantile</th><th>Value</th></tr>
<tr><td>0.05</td><td>0</td></tr>
<tr><td>Q1</td><td>1</td></tr>
<tr><td>Median</td><td>1</td></tr>
<tr><td>Q3</td><td>1</td></tr>
<tr><td>0.95</td><td>5</td></tr>
</table></td></tr>
<tr><td>✗</td><td>values must belong to this set: 1 2 3 4 5 6.
1579 unexpected values found. ≈15.79% of 10000 total rows.
<table>
<tr><th>Sampled Unexpected Values</th></tr>
<tr><td>0</td></tr>
</table></td><td>≈15.79% unexpected</td></tr>
</table>

สรุป

Data Engineering Life Cycle



Ingestion → Transformation → Storage → Machine Learning / Analytics / Reporting

Orchestration

Software Engineering

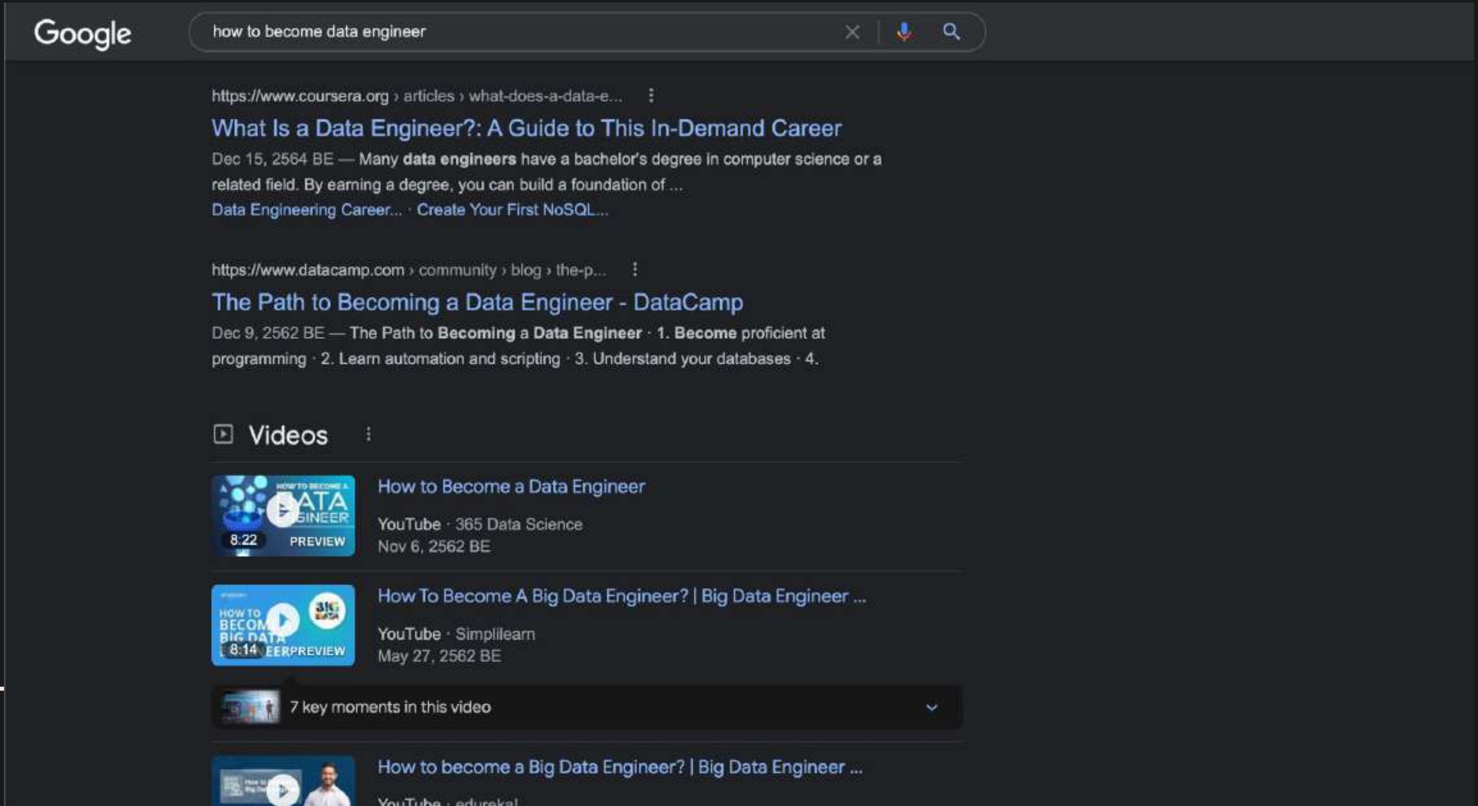
Infra as Code

Security

Data Architecture

Data Governance

How do I become Data Engineer?



how to become data engineer

https://www.coursera.org › articles › what-does-a-data-e...

What Is a Data Engineer?: A Guide to This In-Demand Career

Dec 15, 2564 BE — Many **data engineers** have a bachelor's degree in computer science or a related field. By earning a degree, you can build a foundation of ...

Data Engineering Career... · Create Your First NoSQL...

https://www.datacamp.com › community › blog › the-p...

The Path to Becoming a Data Engineer - DataCamp

Dec 9, 2562 BE — The Path to **Becoming** a **Data Engineer** · **1. Become** proficient at programming · 2. Learn automation and scripting · 3. Understand your databases · 4.

Videos

How to Become a Data Engineer

YouTube · 365 Data Science
Nov 6, 2562 BE

How To Become A Big Data Engineer? | Big Data Engineer ...

YouTube · Simplilearn
May 27, 2562 BE

7 key moments in this video

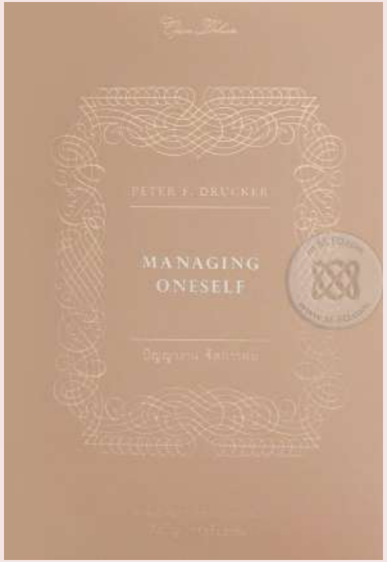
How to become a Big Data Engineer? | Big Data Engineer ...

YouTube · edureka!

~~How to become Data Engineer ?~~

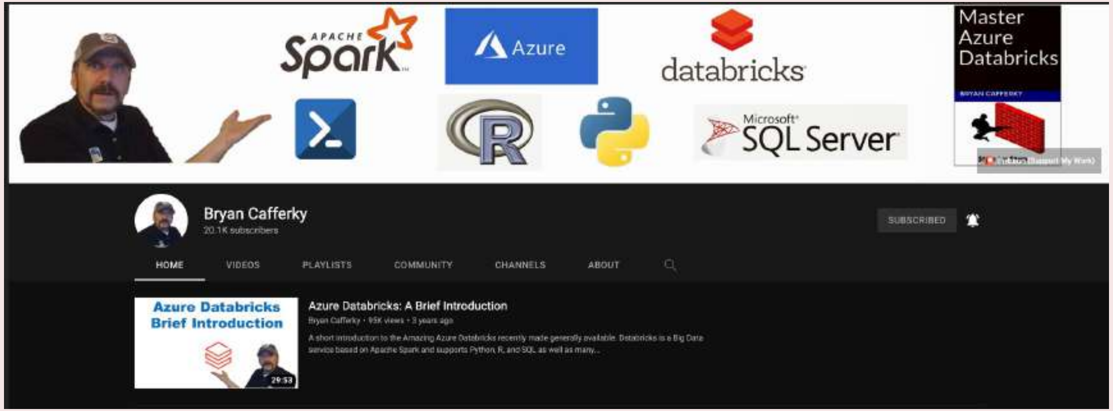
how I learn

- Data Engineer Roadmap: https://github.com/datastacktv/data-engineer-roadmap
- Compare tools
- Compare services https://comparecloud.in/



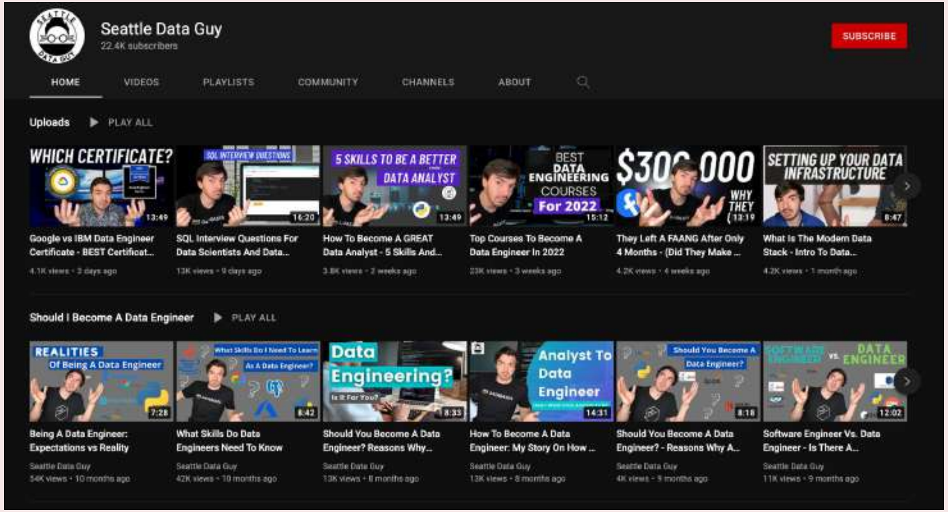
Listen

random**



Bryan Cafferky
20.1K subscribers

Azure Databricks: A Brief Introduction
Bryan Cafferky • 95K views • 3 years ago
A short introduction to the Amazing Azure Databricks recently made generally available. Databricks is a Big Data service based on Apache Spark and supports Python, R, and SQL as well as many...


SUBSCRIBE
HOME
VIDEOS
PLAYLISTS
COMMUNITY
CHANNELS
ABOUT
PLAY ALL
WHICH CERTIFICATE?
13:49
SQL INTERVIEW QUESTIONS
16:20
5 SKILLS TO BE A BETTER DATA ANALYST
13:49
BEST DATA ENGINEERING COURSES For 2022
15:12
$300,000 WHY THEY
13:19
SETTING UP YOUR DATA INFRASTRUCTURE
8:47
PLAY ALL
REALITIES Of Being A Data Engineer
7:28
What Skills Do I Need To Learn As A Data Engineer?
8:42
Data Engineering? Is It For You?
8:33
Analyst To Data Engineer
14:31
Should You Become A Data Engineer?
8:18
SOFTWARE ENGINEER VS. DATA ENGINEER
12:02

Seattle Data Guy
22.4K subscribers

Uploads

Google vs IBM Data Engineer Certificate - BEST Certificat...
4.1K views • 3 days ago

SQL Interview Questions For Data Scientists And Data...
13K views • 9 days ago

How To Become A GREAT Data Analyst - 5 Skills And...
3.8K views • 2 weeks ago

Top Courses To Become A Data Engineer In 2022
23K views • 3 weeks ago

They Left A FAANG After Only 4 Months - (Did They Make ...
4.2K views • 4 weeks ago

What Is The Modern Data Stack - Intro To Data...
4.2K views • 1 month ago

Should I Become A Data Engineer

Being A Data Engineer: Expectations vs Reality
Seattle Data Guy
54K views • 10 months ago

What Skills Do Data Engineers Need To Know
Seattle Data Guy
42K views • 10 months ago

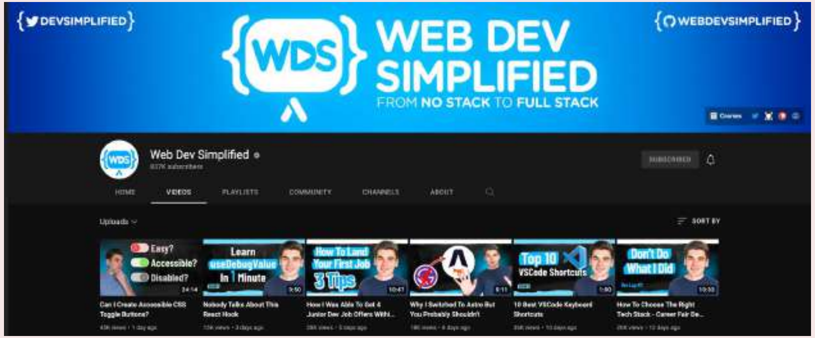
Should You Become A Data Engineer? Reasons Why...
Seattle Data Guy
13K views • 8 months ago

How To Become A Data Engineer: My Story On How ...
Seattle Data Guy
13K views • 8 months ago

Should You Become A Data Engineer? - Reasons Why A...
Seattle Data Guy
4K views • 9 months ago

Software Engineer Vs. Data Engineer - Is There A...
Seattle Data Guy
11K views • 9 months ago

PyData
130K subscribers



Web Dev Simplified
837K subscribers

Uploads

Can I Create Accessible CSS Toggle Buttons?

Nobody Talks About This React Hook

How I Was Able To Get 4 Junior Dev Job Offers Withi...

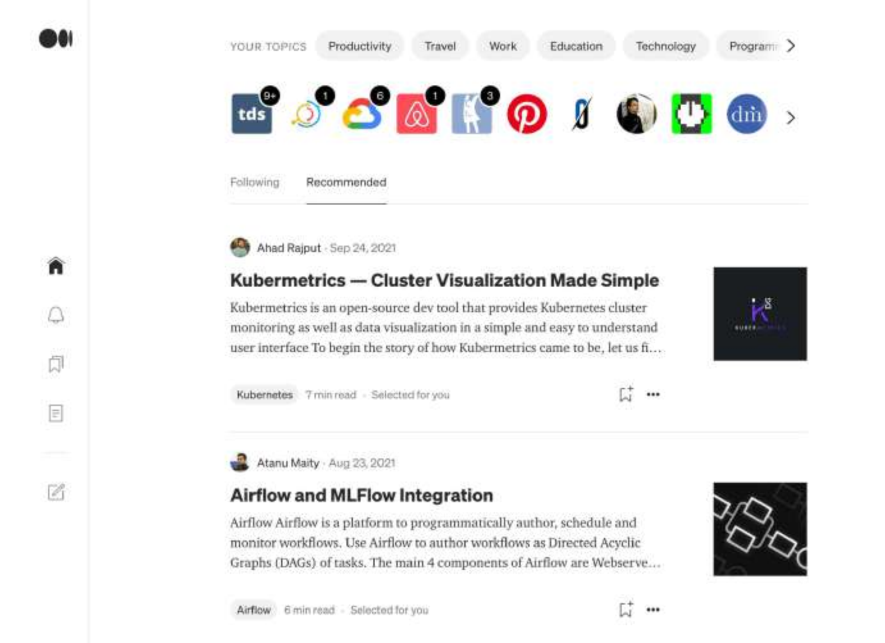
Why I Switched To Astro But You Probably Shouldn't

10 Best VSCode Keyboard Shortcuts

How To Choose The Right Tech Stack - Career Fair De...

READ !

medium.com



Ahad Rajput · Sep 24, 2021

Kubermetrics — Cluster Visualization Made Simple

Kubermetrics is an open-source dev tool that provides Kubernetes cluster monitoring as well as data visualization in a simple and easy to understand user interface To begin the story of how Kubermetrics came to be, let us fi…

Kubernetes 7 min read · Selected for you

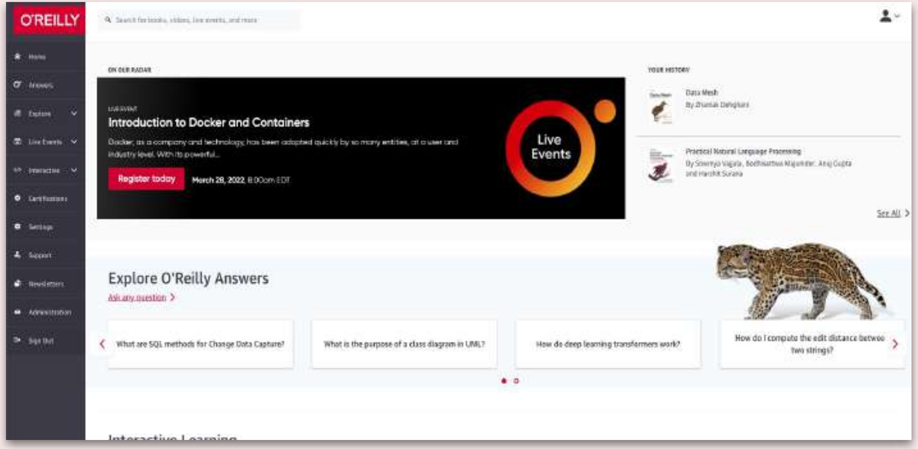
Atanu Maity · Aug 23, 2021

Airflow and MLFlow Integration

Airflow Airflow is a platform to programmatically author, schedule and monitor workflows. Use Airflow to author workflows as Directed Acyclic Graphs (DAGs) of tasks. The main 4 components of Airflow are Webserve…

Airflow 6 min read · Selected for you

learning.oreilly.com



ON OUR RADAR

LIVE EVENT

Introduction to Docker and Containers

Docker, as a company and technology, has been adopted quickly by so many entities, at a user and industry level. With its powerful…

March 28, 2022, 8:00am EDT

YOUR HISTORY

Data Mesh
By Zhamak Dehghani

Practical Natural Language Processing
By Sowmya Vajjala, Bodhisattwa Majumder, Anuj Gupta and Harshit Surana

Explore O'Reilly Answers

Ask any question

What are SQL methods for Change Data Capture?

What is the purpose of a class diagram in UML?

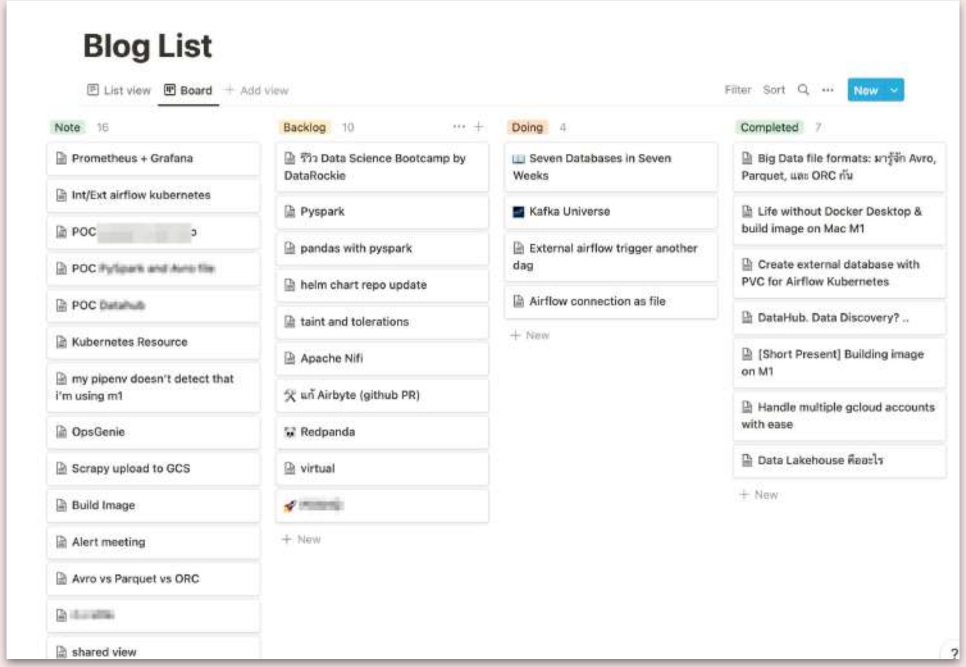
How do deep learning transformers work?

How do I compute the edit distance betwee two strings?

Interactive Learning

Write !

Note in Notion



Blog List

List view · Board · + Add view · Filter · Sort · ··· · New

Note 16

- Prometheus + Grafana
- Int/Ext airflow kubernetes
- POC [redacted]
- POC [redacted]
- POC [redacted]
- Kubernetes Resource
- my pipenv doesn't detect that i'm using m1
- OpsGenie
- Scrapy upload to GCS
- Build Image
- Alert meeting
- Avro vs Parquet vs ORC
- [redacted]
- shared view

Backlog 10

- รีวิว Data Science Bootcamp by DataRockie
- Pyspark
- pandas with pyspark
- helm chart repo update
- taint and tolerations
- Apache Nifi
- 🛠 แก้ Airbyte (github PR)
- Redpanda
- virtual
- 🚀 [redacted]
- + New

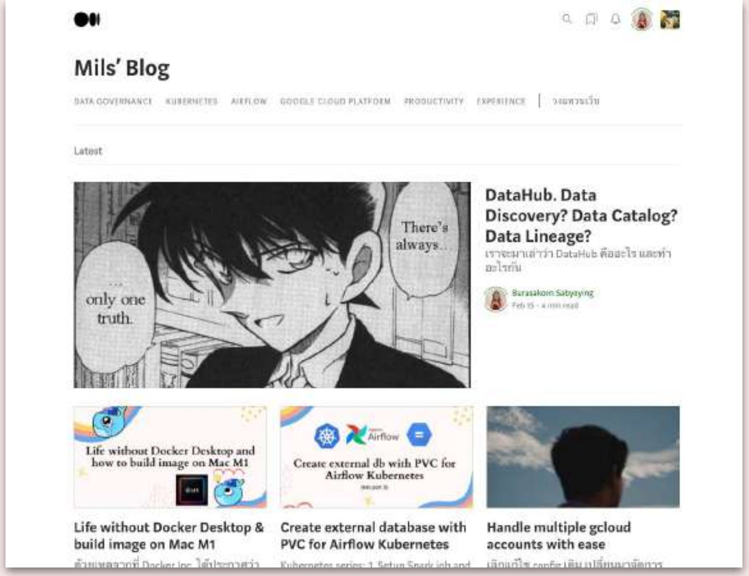
Doing 4

- 📖 Seven Databases in Seven Weeks
- Kafka Universe
- External airflow trigger another dag
- Airflow connection as file
- + New

Completed 7

- Big Data file formats: มารู้จัก Avro, Parquet, และ ORC กัน
- Life without Docker Desktop & build image on Mac M1
- Create external database with PVC for Airflow Kubernetes
- DataHub. Data Discovery? ..
- [Short Present] Building image on M1
- Handle multiple gcloud accounts with ease
- Data Lakehouse คืออะไร
- + New

Mesodiar.com (blog on medium)



Mils' Blog

DATA GOVERNANCE · KUBERNETES · AIRFLOW · GOOGLE CLOUD PLATFORM · PRODUCTIVITY · EXPERIENCE | รวมพวงเว็บ

Latest

There's always...
... only one truth.

DataHub. Data Discovery? Data Catalog? Data Lineage?

เราจะมาเล่าว่า DataHub คืออะไร และทำอะไรกัน

Burasakorn Sabyeying
Feb 15 · 4 min read

Life without Docker Desktop and how to build image on Mac M1
Create external db with PVC for Airflow Kubernetes
Airflow

Life without Docker Desktop & build image on Mac M1

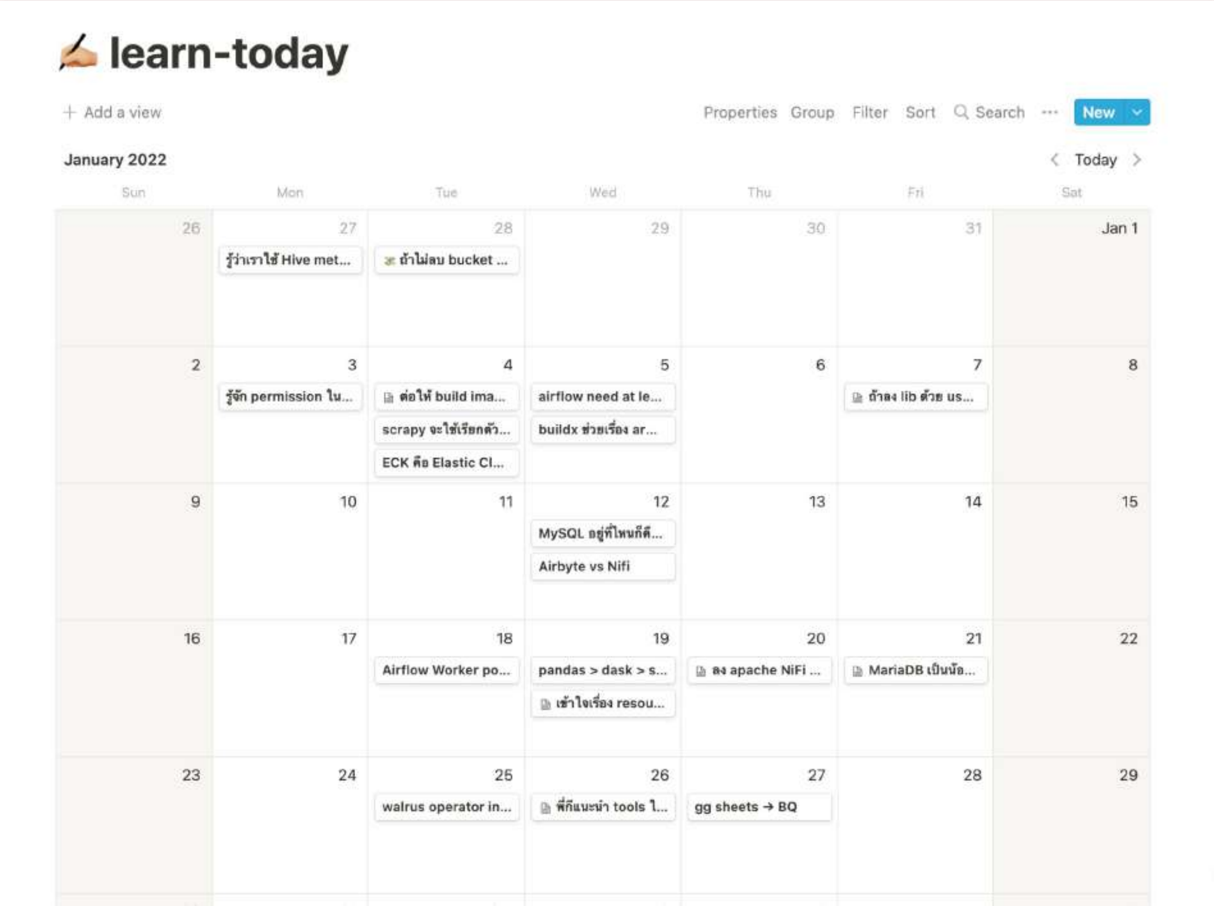
ด้วยเหตุจากที่ Docker Inc. ได้ประกาศว่า

Create external database with PVC for Airflow Kubernetes

Kubernetes series: 1 Setup Spark job and

Handle multiple gcloud accounts with ease

เลิกแก้ไข config เดิม เปลี่ยนมาจัดการ



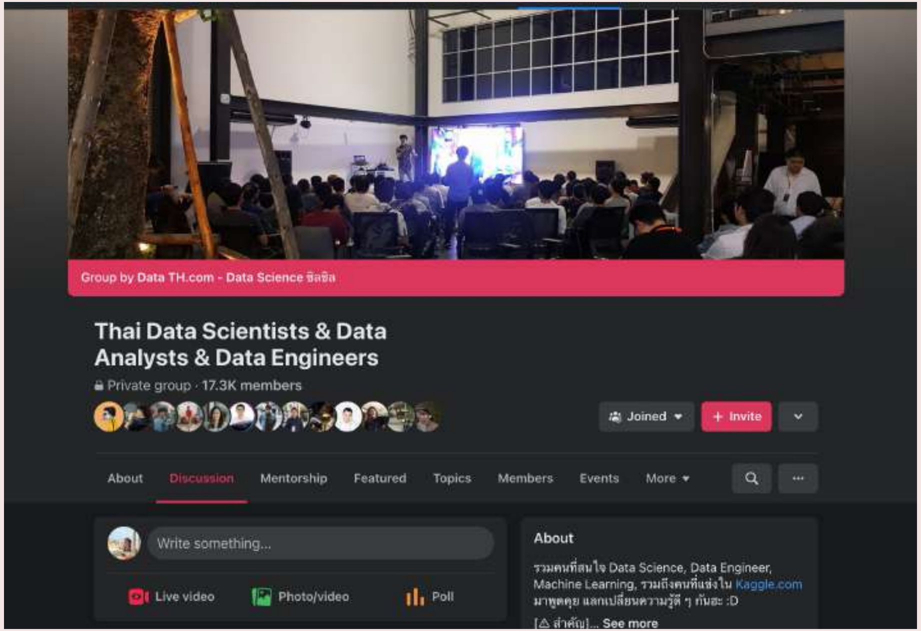
✍️ learn-today

January 2022

Sun	Mon	Tue	Wed	Thu	Fri	Sat
26	27 รู้ว่าเราใช้ Hive met...	28 ถ้าไม่ลบ bucket ...	29	30	31	Jan 1
2	3 รู้จัก permission ใน...	4 ต่อให้ build ima... scrapy จะใช้เรียกตัว... ECK คือ Elastic Cl...	5 airflow need at le... buildx ช่วยเรื่อง ar...	6	7 ถ้าลง lib ด้วย us...	8
9	10	11	12 MySQL อยู่ที่ไหนก็ดี... Airbyte vs Nifi	13	14	15
16	17	18 Airflow Worker po...	19 pandas > dask > s... เข้าใจเรื่อง resou...	20 ลง apache NiFi ...	21 MariaDB เป็นน้อ...	22
23	24	25 walrus operator in...	26 พี่กีแนะนำ tools ไ...	27 gg sheets → BQ	28	29

โน้ตว่าวันนี้เราเรียนรู้อะไร จะได้ไม่ท้อ !

Facebook Group



Group by Data TH.com - Data Science ชิลชิล

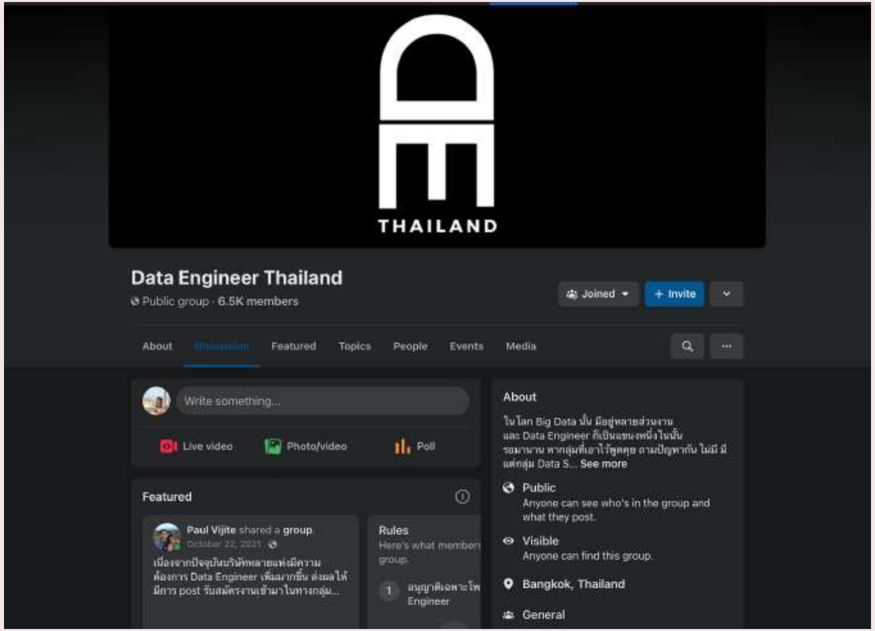
Thai Data Scientists & Data Analysts & Data Engineers

Private group · 17.3K members

About

รวมคนที่สนใจ Data Science, Data Engineer, Machine Learning, รวมถึงคนที่แข่งใน Kaggle.com มาพูดคุย แลกเปลี่ยนความรู้ดี ๆ กันฮะ :D

[⚠ สำคัญ]... See more



THAILAND

Data Engineer Thailand

Public group · 6.5K members

Featured

Paul Vijite shared a group.
October 22, 2021

เนื่องจากปัจจุบันบริษัทหลายแห่งมีความต้องการ Data Engineer เพิ่มมากขึ้น ส่งผลให้มีการ post รับสมัครงานเข้ามาในทางกลุ่ม...

Rules
Here's what member
group.

1 อนุญาติเฉพาะโพ
Engineer

About

ในโลก Big Data นั้น มีอยู่หลายส่วนงาน และ Data Engineer ก็เป็นแขนงหนึ่งในนั้น รอมานาน หากลุ่มที่เอาไว้พูดคุย ถามปัญหากัน ไม่มี มีแต่กลุ่ม Data S... See more

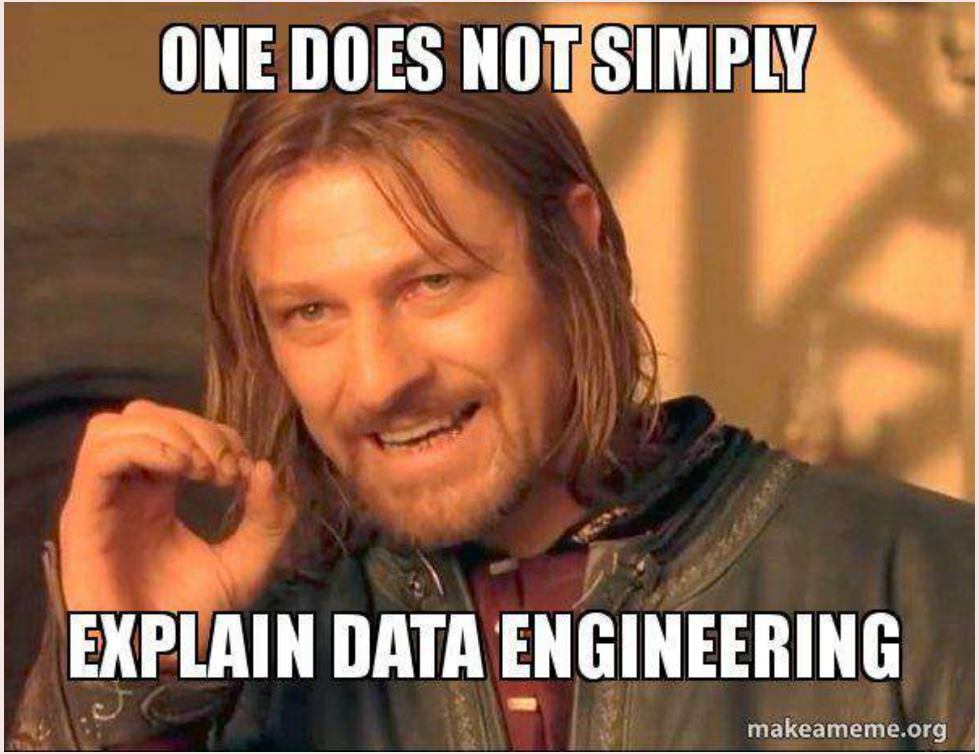
Public
Anyone can see who's in the group and what they post.

Visible
Anyone can find this group.

Bangkok, Thailand

General

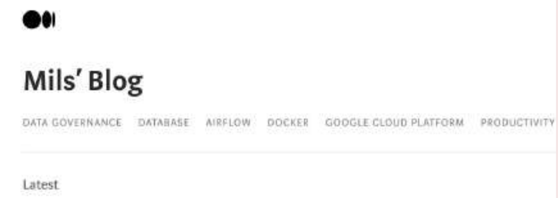
เราอยากเป็นแบบไหน จงเอาตัวเองเข้าไปอยู่วงนั้น



Mesodiar.com

@mesodiar · Education website

FB: Mesodiar.com
Medium: www.mesodiar.com


Mils' Blog

DATA GOVERNANCE DATABASE AIRFLOW DOCKER GOOGLE CLOUD PLATFORM PRODUCTIVITY

Latest