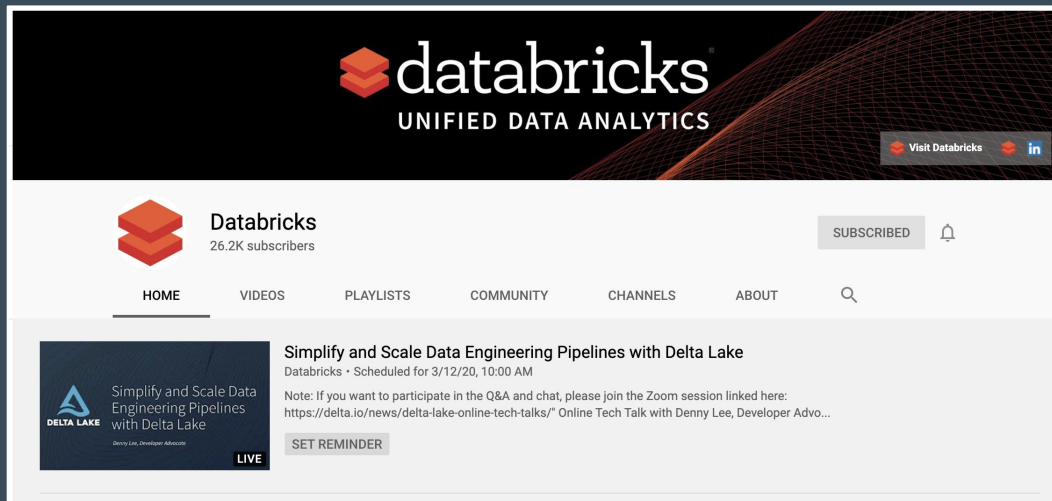




# Predictive Maintenance on IoT Data for Early Fault Detection with Delta Lake

Anindita Mahapatra and Denny Lee

# Subscribe Today!



<https://dbricks.co/youtube>

# Who are we?



- Solutions Architect – Databricks  
*“If Data has arrived, it better be served!”*
- Past Experience in Big Data
  - Teradata/Think Big Analytics
  - Nokia/Microsoft
- MS in Computer Science - Boston University
- Master of Liberal Arts & Management - Harvard University



# Who are we?

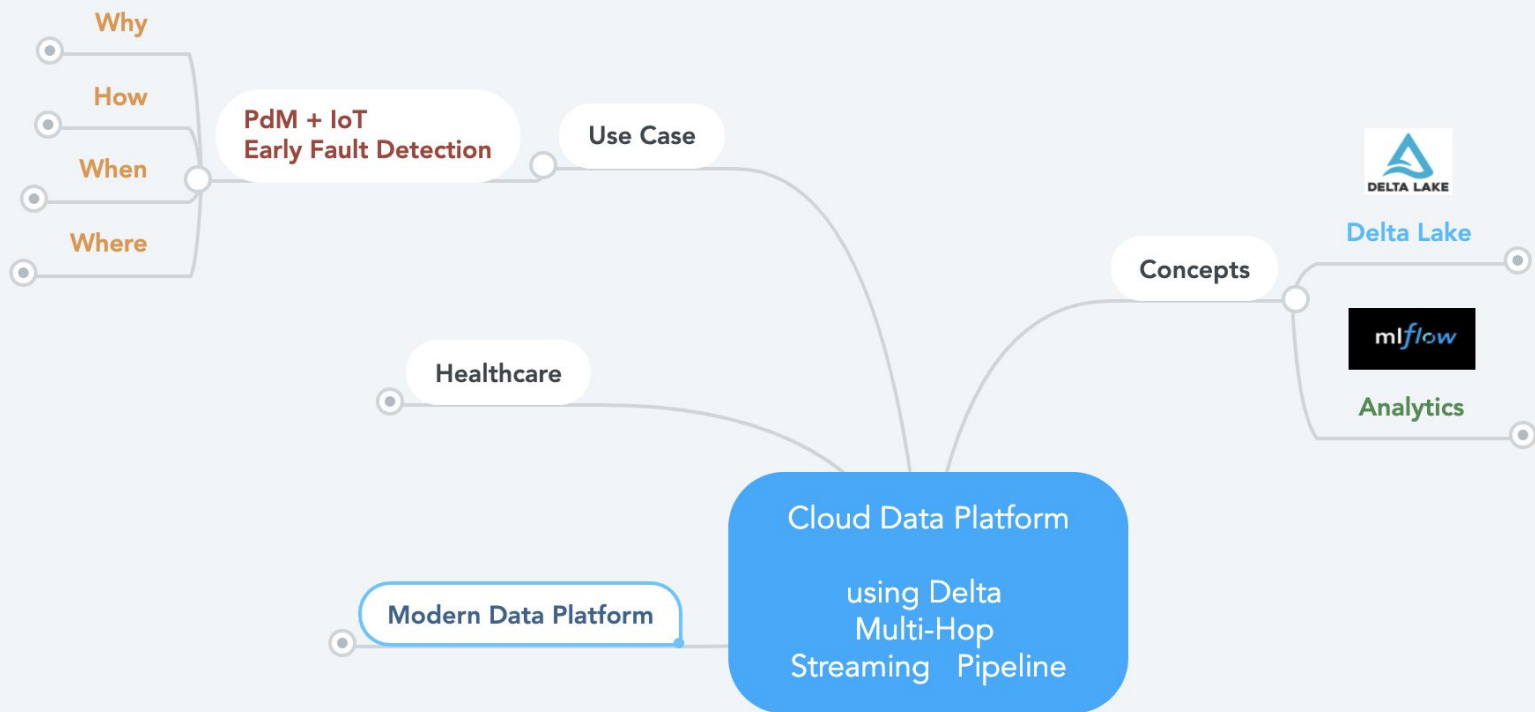


**databricks**

- Staff Developer Advocate – Databricks
- Working with Apache Spark™ since v0.6
- Former Senior Director Data Science Engineering at Concur
- Former Microsoftie: Cosmos DB, HDInsight (Isotope)
- Masters Biomedical Informatics - OHSU
- BS in Physiology - McGill



# Agenda

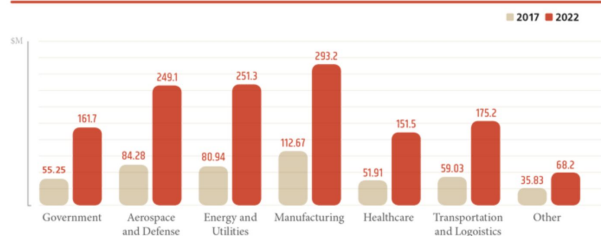


# Use Case

Increase the reliability and extend the useful life

## EXHIBIT 1

North America Predictive Maintenance Market by Vertical (2017-2022)



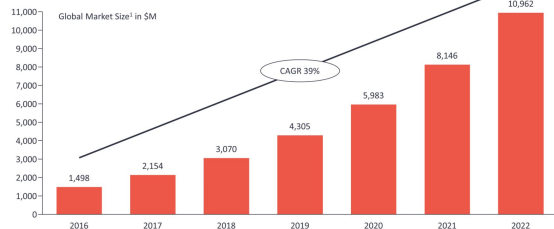
## 2 Market Size & Outlook

### 2.1 Total Market

#### Total PdM Market

Predictive Maintenance a \$1,498M market in 2016, expected to grow to \$10,962M by 2022

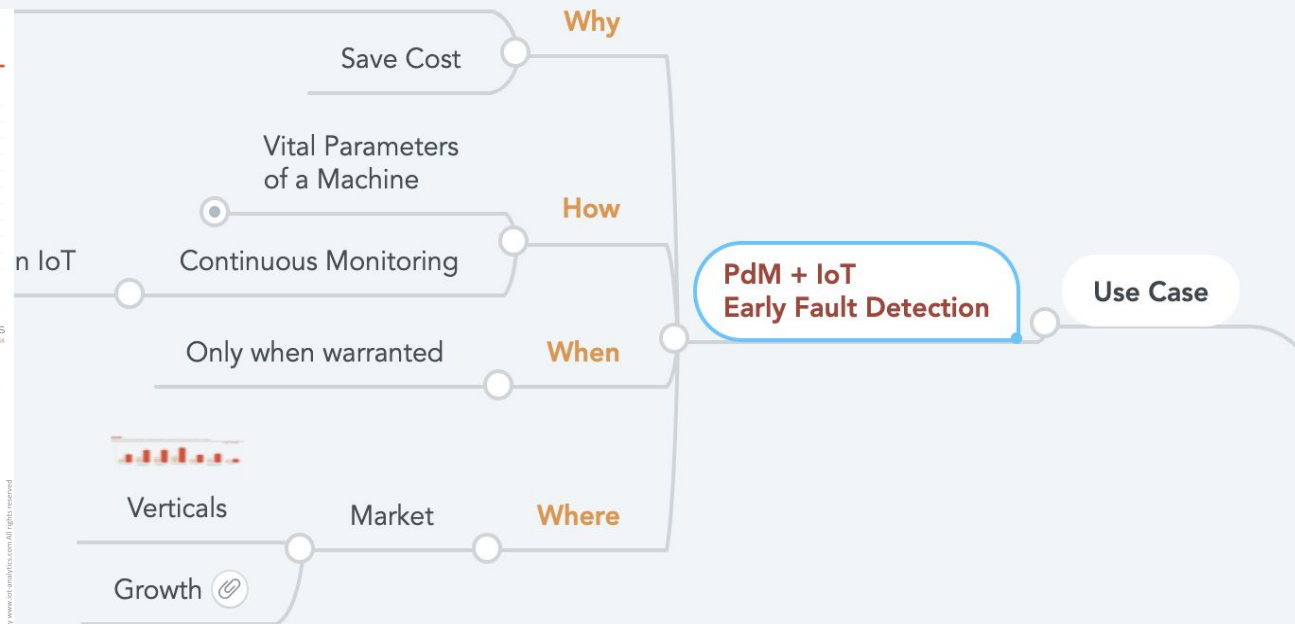
#### Global Market Development (PdM)



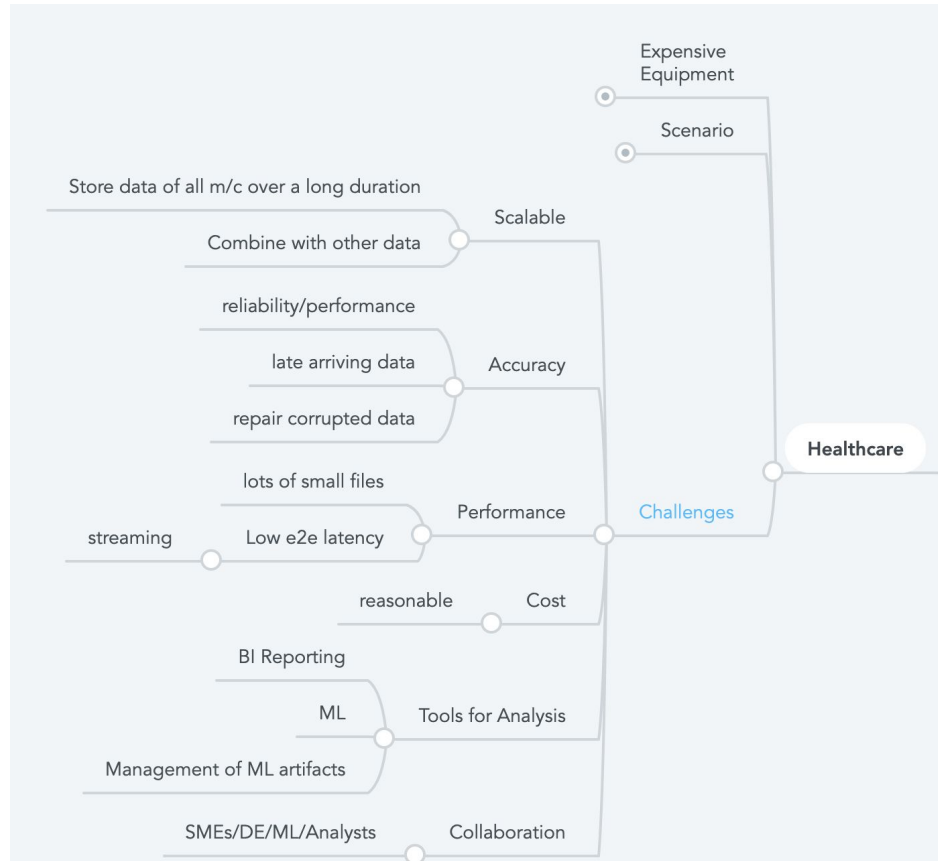
1. Market Size defined as entire "Annual PdM technology spend" by companies implementing PdM solutions in USD  
Source: IoT Analytics Research

Our insights empower you to understand IoT markets

Copyright © 2017 by www.iot-analytics.com All rights reserved

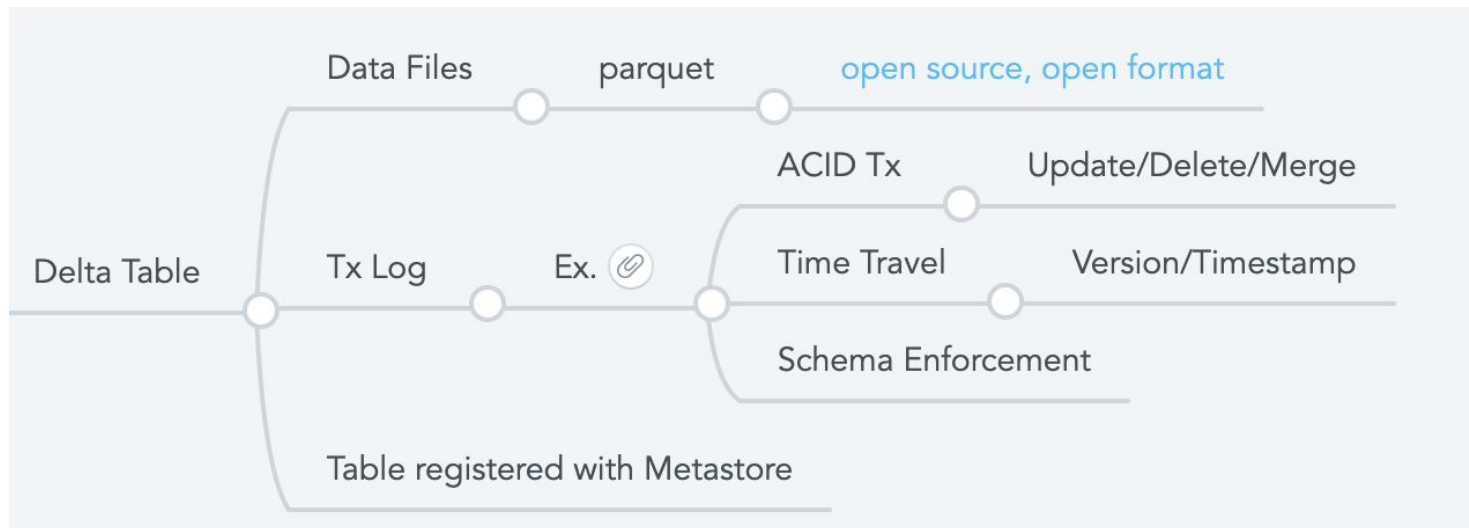
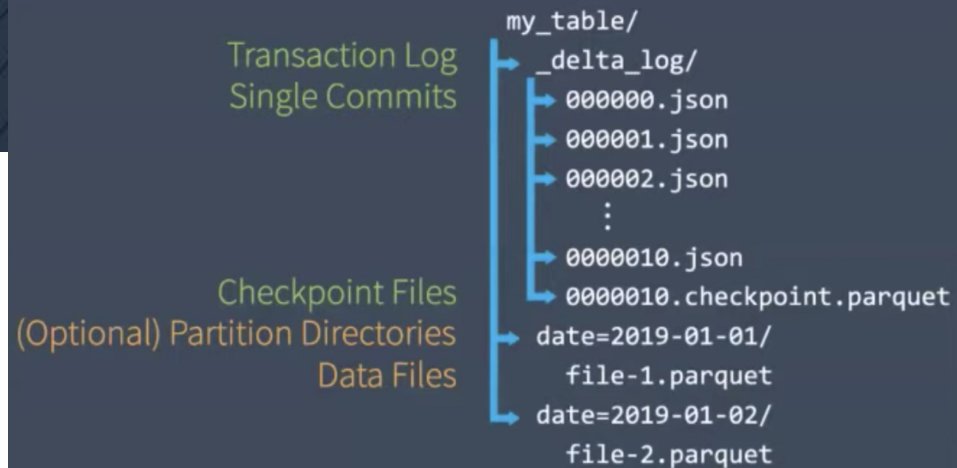


# Challenges



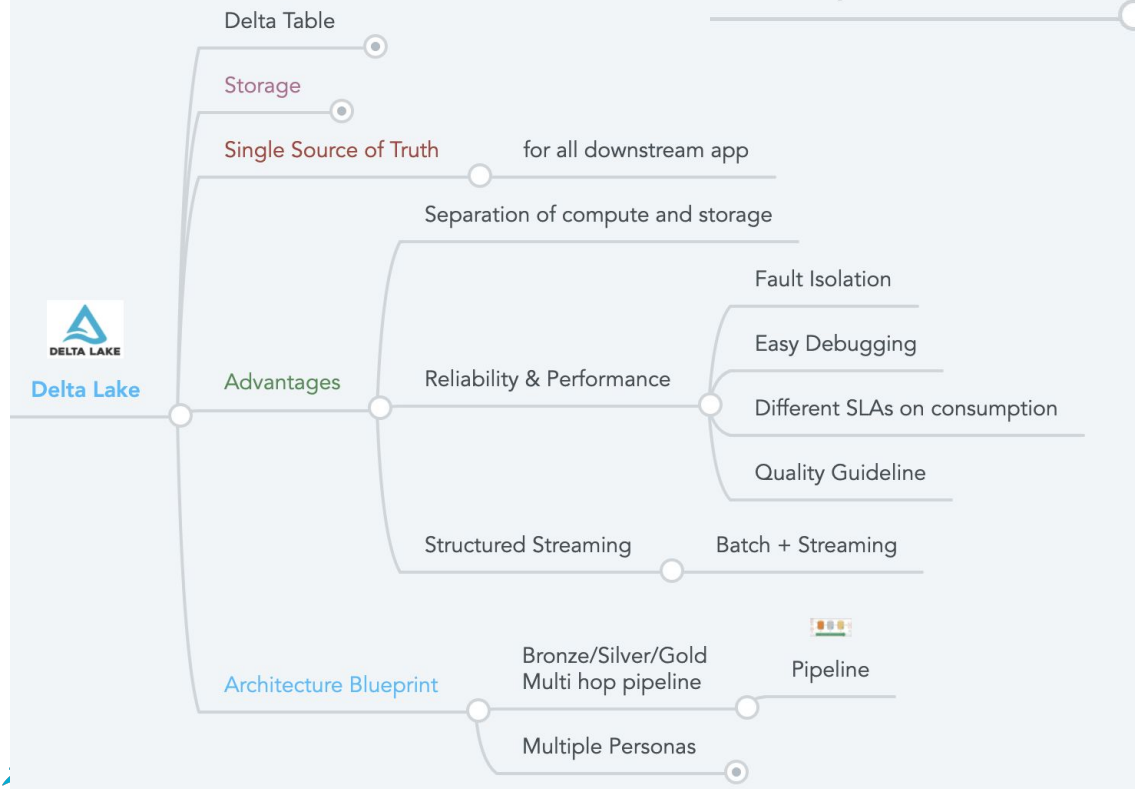
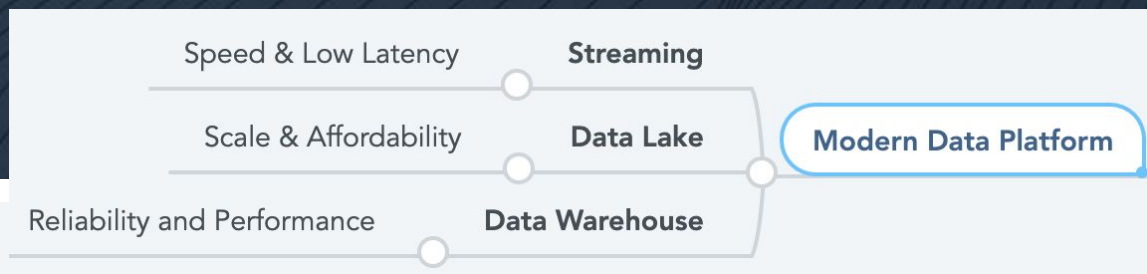


# Delta





# Delta Lake

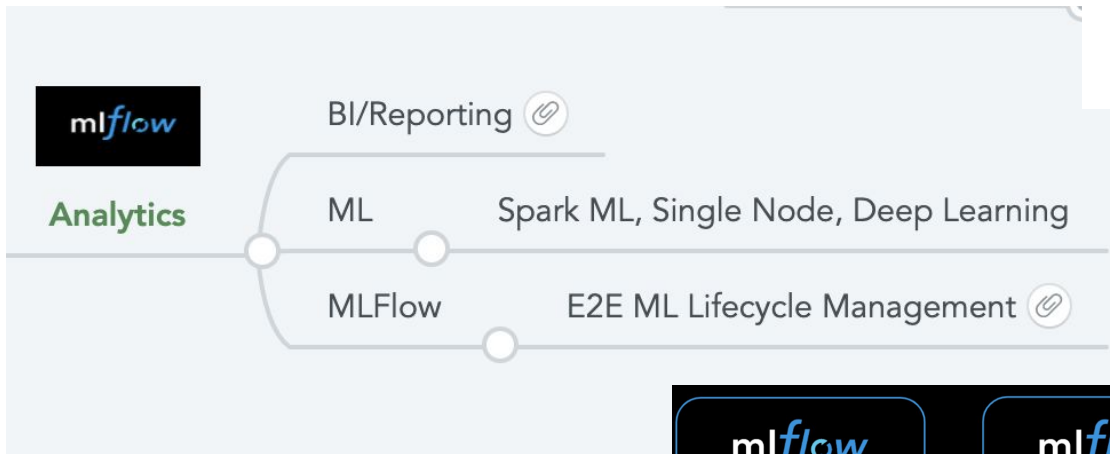


# MLflow

## BI INTEGRATIONS



Access all your data



### mlflow Tracking

Record and query experiments: code, data, config, results

### mlflow Projects

Packaging format for reproducible runs on any platform

### mlflow Models

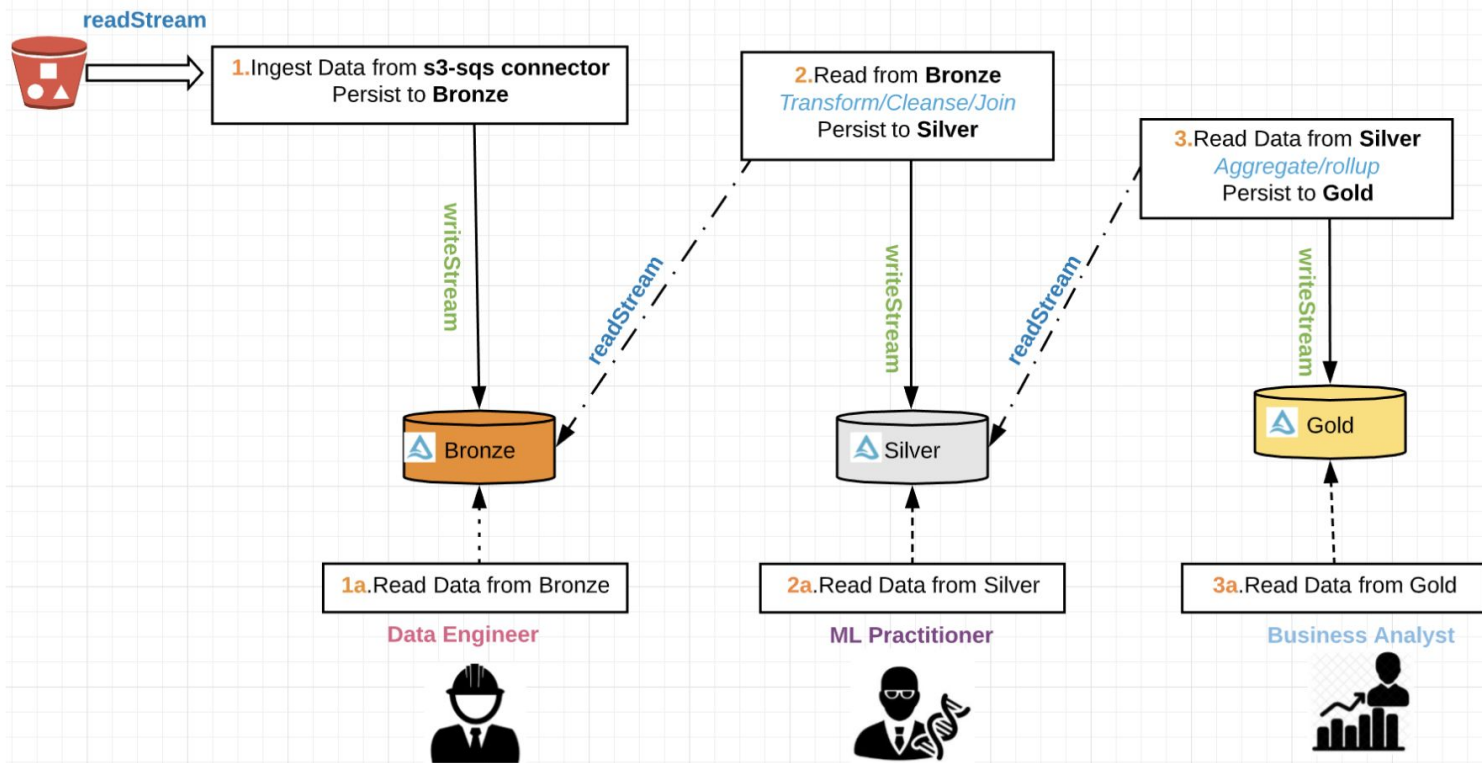
General model format that supports diverse deployment tools

### mlflow Model Registry

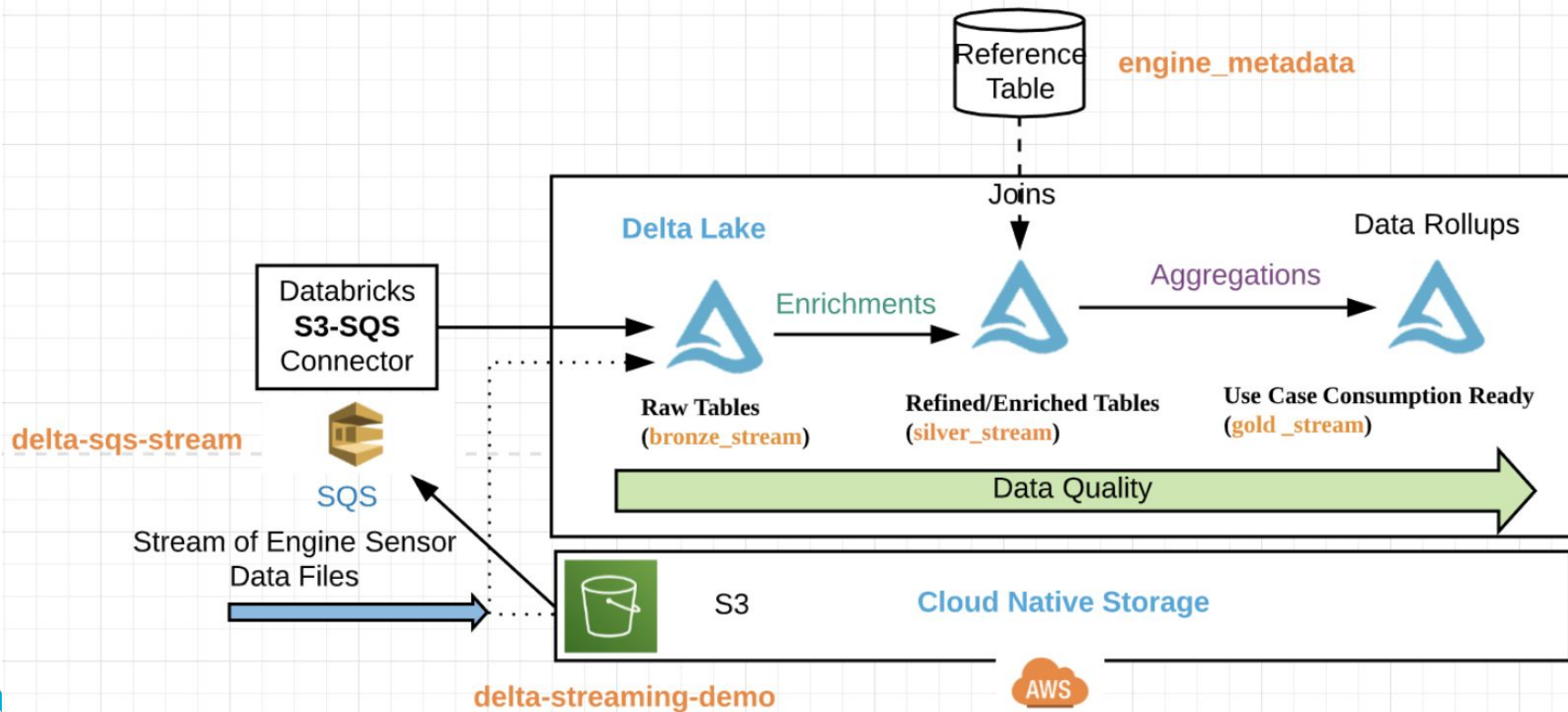
Centralized and collaborative model lifecycle management



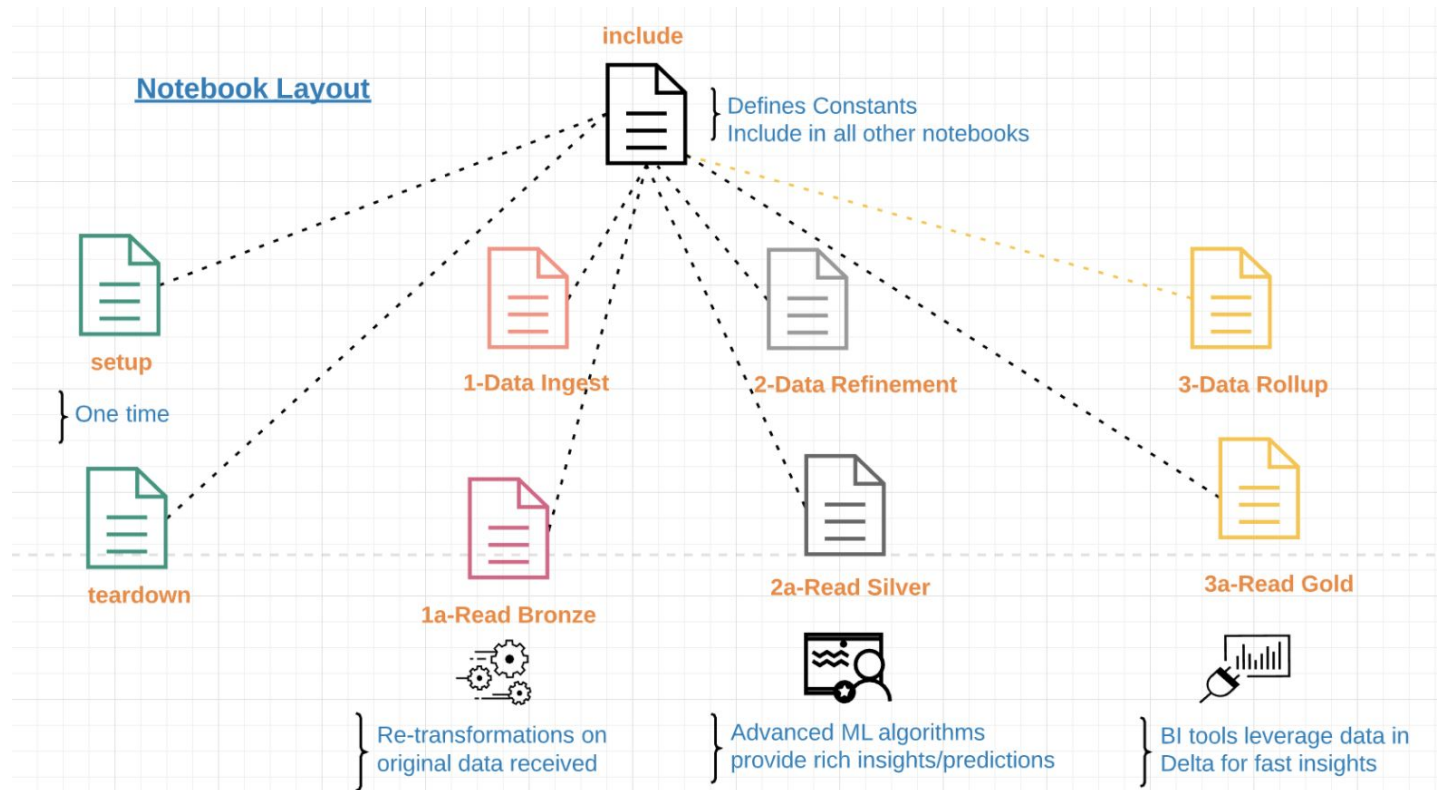
# Delta Multi-hop Pipeline



# Incremental Data Ingestion



# Notebook Layout



# Access based on file; file notification; auto loader

Amazon S3 > delta-streaming-demo

Overview Properties Permissions Manage

Search: Type a prefix and press Enter to search. Press ESC to clear.

Upload Create folder Download Actions

Name

sensor

SQS Queue selected

Details Permissions Redrive Policy Monitoring Tags Encryption Lambda

Add a Permission Edit Policy Document (Advanced) What's an SQS Queue Access Policy?

Effect	Principals	Actions	Conditions
Allow	Everybody (*)	SQS:SendMessage	ArnLike aws:SourceArn: "arn:aws:s3:*:*:delta-streaming-demo"

Events

+ Add notification Delete Edit

Name	Events	Filter	Type
publish-events			

Name

publish-events

Events

☐ PUT

☐ POST

☐ COPY

☐ Multipart upload completed

☒ All object create events

☐ Object in RRS lost

☐ Permanently deleted

☐ Delete marker created

☐ All object delete events

☐ Restore initiated

☐ Restore completed

☐ Replication time missed threshold

☐ Replication time completed after threshold

☐ Replication time not tracked

☐ Replication time failed

Prefix

sensor

Suffix

e.g. .jpg

Send to

SQS Queue

SQS

delta-sqs-stream

1 Active notifications

Cancel Save

# References

- Delta Lake: Reliable Data Lakes at Scale <https://delta.io/>
- MLflow: End to End ML Lifecycle Management <https://mlflow.org/>
- Optimized Amazon S3 Source with Amazon SQS  
<https://docs.databricks.com/spark/latest/structured-streaming/sqs.html>





# Delta Lake Connectors

Standardize your big data storage with an open format accessible from various tools



# Delta Lake Partners and Providers

More and more partners and providers are jumping working with Delta Lake



Google Dataproc



Privacera



Informatica



WANDisco



Qlik



Streamsets



# Users of Delta Lake

Tencent 腾讯



VIACOM

ciena.



Booz | Allen | Hamilton®



CONDÉ NAST

TILTING POINT



upwork



DOLLAR SHAVE CLUB

