

Machine Learning for Manufacturing with Airflow

Devon Peticolas - Oden Technologies

Market

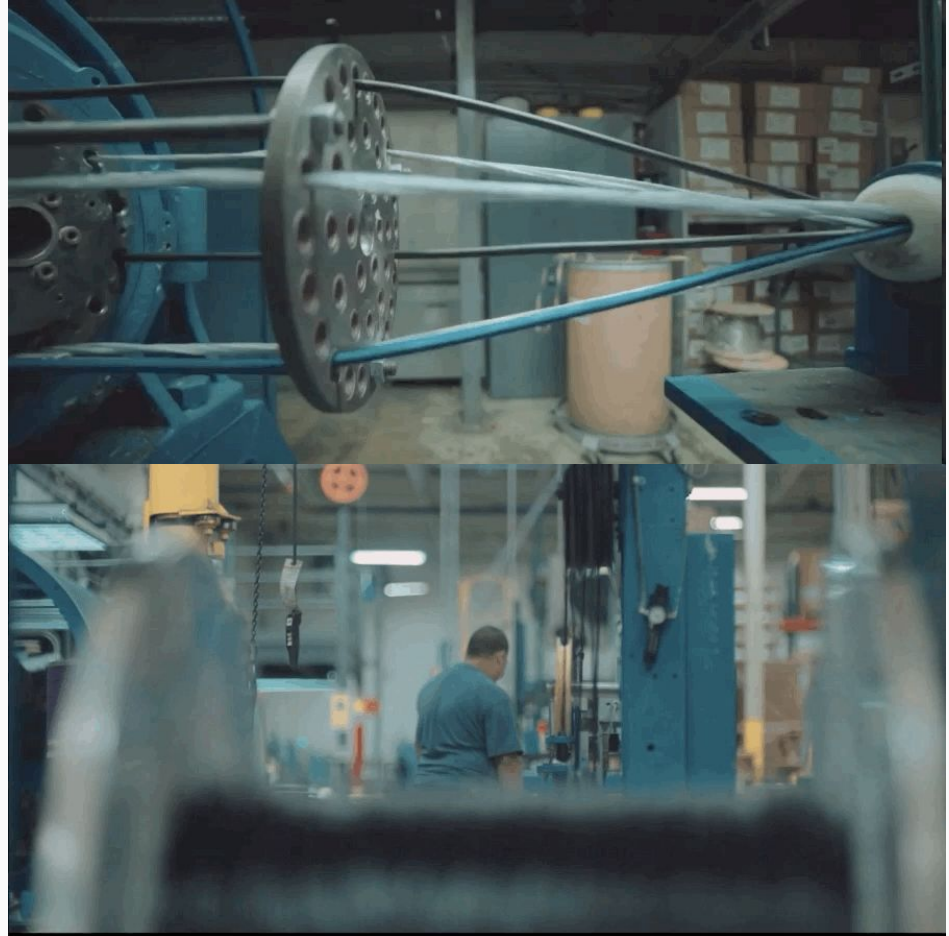


Oden's Customers

Medium to large manufacturers in
plastics extrusion, injection molding,
and metal stamping.

Process and Quality Engineers looking
to centralize, analyze, and act on their
data.

Plant managers who are looking to
optimize logistics, output, and cost.



\$3,200,000,000,000

Projected impact on global GDP from manufacturing IoT over by 2031
~ McKinsey, 2016



\$3.2e12

Projected impact on global GDP from manufacturing IoT over by 2031
~ McKinsey, 2016

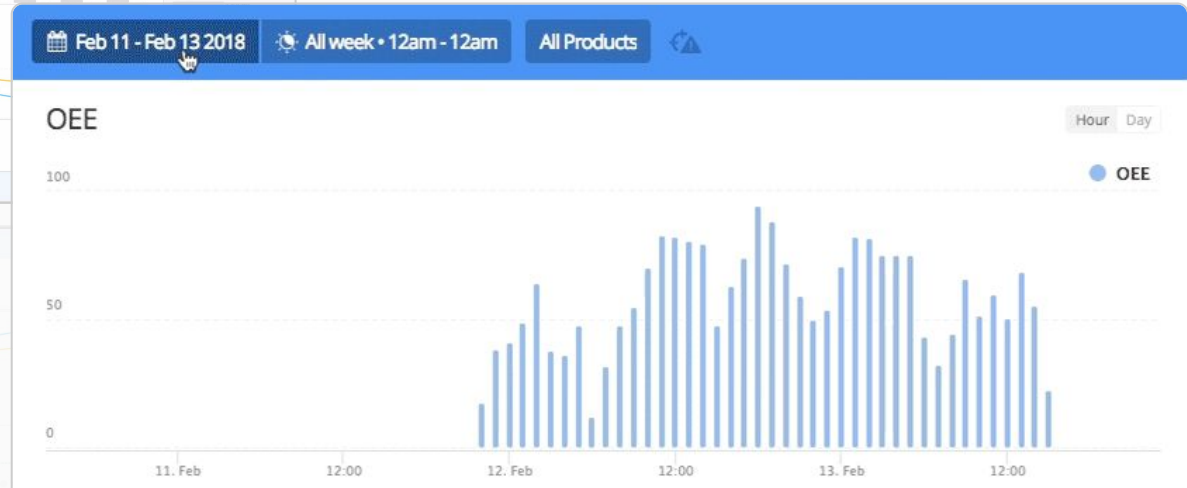
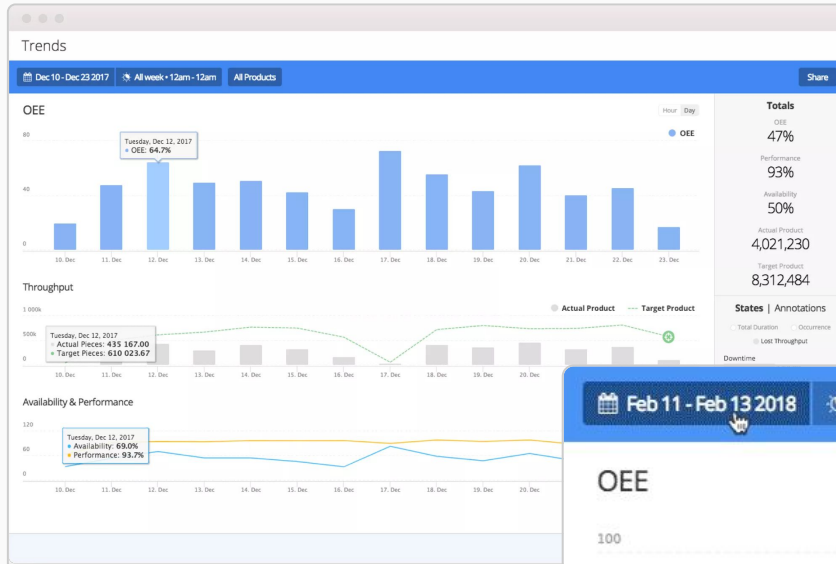


Product



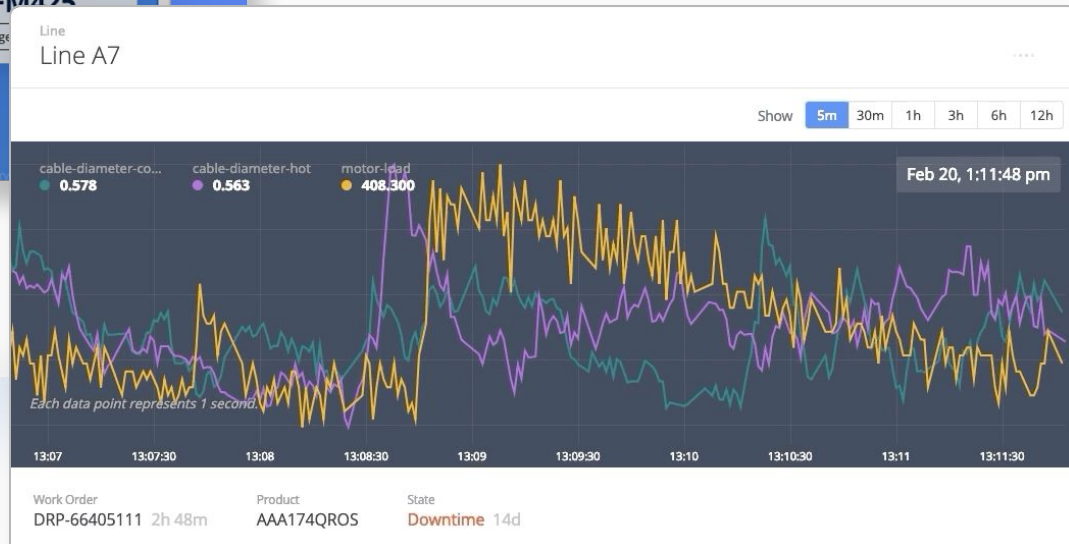
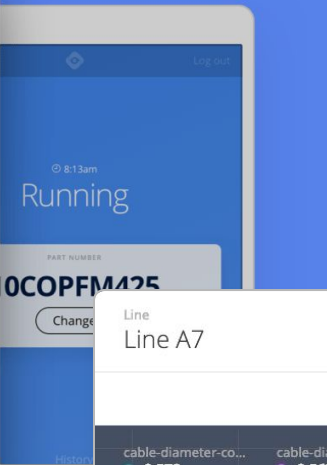
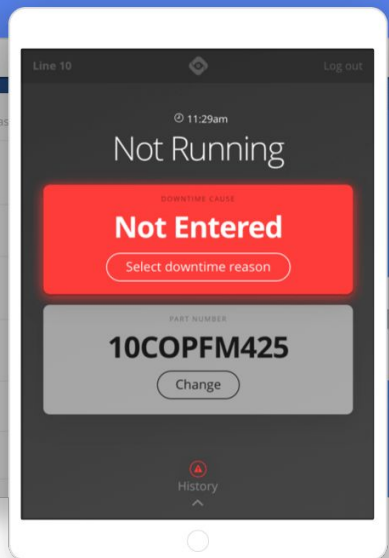
Interactive Time-series Analysis

- Compare performance across different equipment.
- Visualize hourly uptime and key custom metrics.
- Calculations for analyzing and optimizing factory performance.



Real Time Manufacturing Data

- Streaming second-by-second metrics
- Interactive app that prompts on production state changes and collects user input.



Reporting and Alerting

- Daily summaries on key process metrics from continuous intervals of production work.
- Real-time email and text alerts on target violations and concerning trends.



Daily Run Report

Runs completed 9:00am EST February 12, 2019 – 9:00am EST February 13, 2019

Runs sorted by worst Cpk for Cold OD Avg

SWJNG519-LQ8

Line 10 · 10 Reels · 06:11 2/12 – 11:42 2/12 · 3h 16m uptime

[View run →](#)

METRIC	MEAN	STD DEV	TARGET	NON CON*	Cpk
Cold OD Avg	0.403	0.010	0.391 - 0.411	4.235%	0.274
Feet per min	274.794	194.059	-	-	-

SWHD72Y-R4

Line 10 · 10 Reels · 10:08 2/12 – 12:34 2/13 · 1h 35m uptime

[View run →](#)

METRIC	MEAN	STD DEV	TARGET	NON CON*	Cpk
Cold OD Avg	0.141	0.002	0.135 - 0.145	0.242%	0.782
Feet per min	829.680	492.109	-	-	-



ALERT

Downtime violation on Line 1

As of 12:55pm, Line 1 has been in Downtime for more than 15 minutes.

[View line](#)

Snooze this alert for: [30m](#) [2h](#) [8h](#) [24h](#)

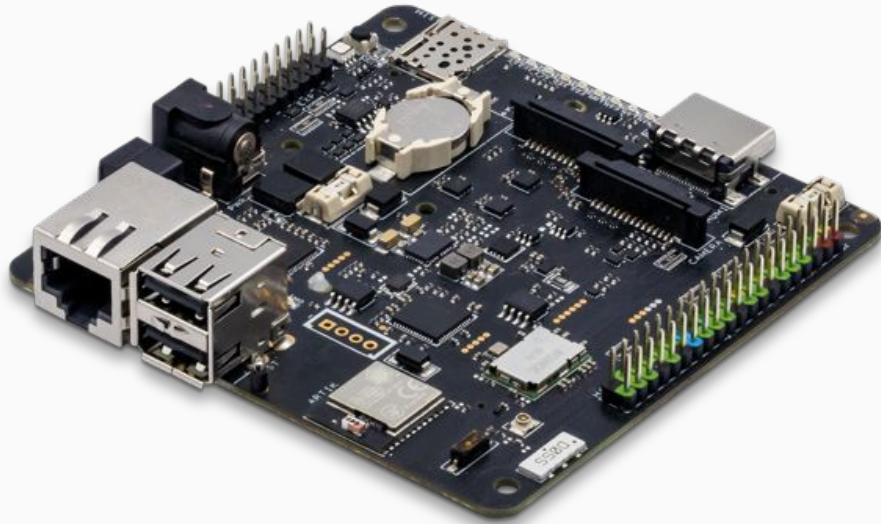
Powered by Oden Technologies

Is this alert useful? [Let us know!](#)



Technology

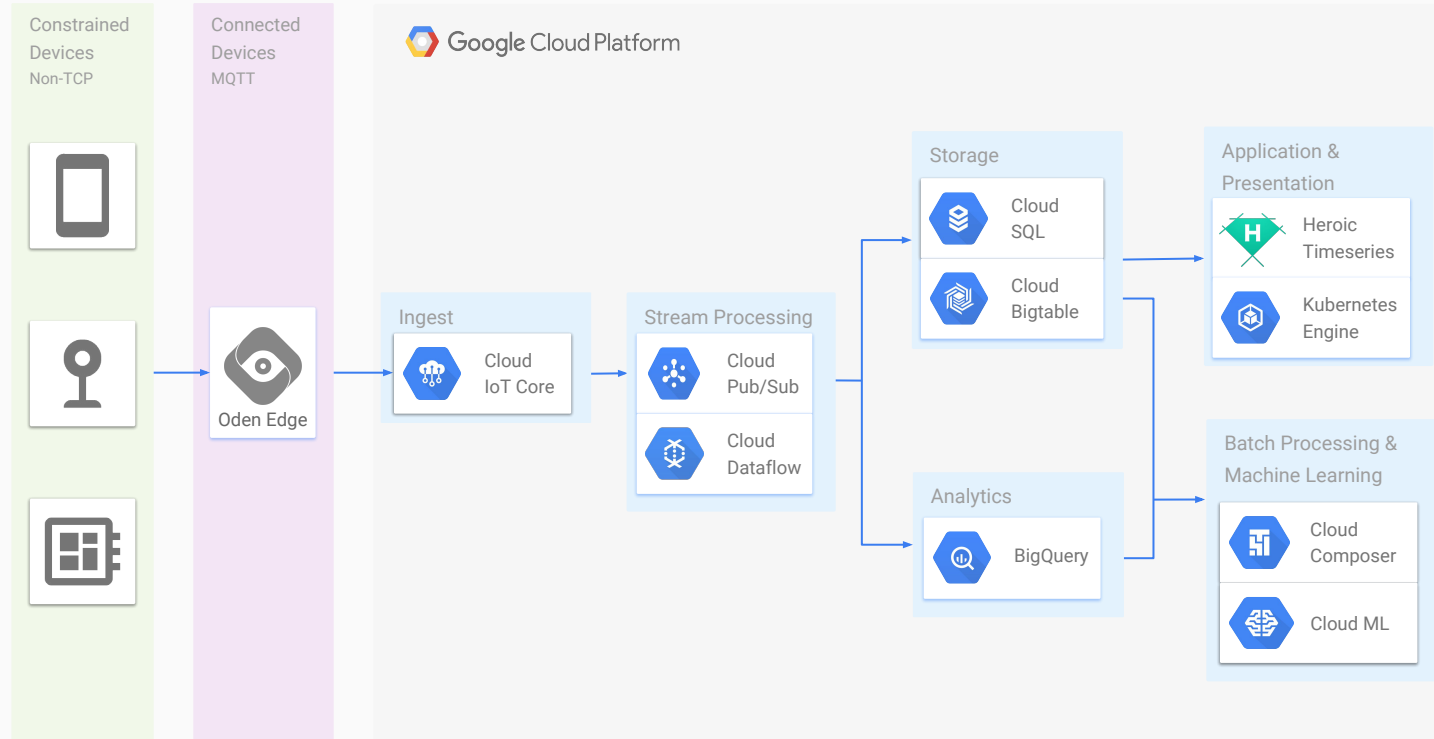




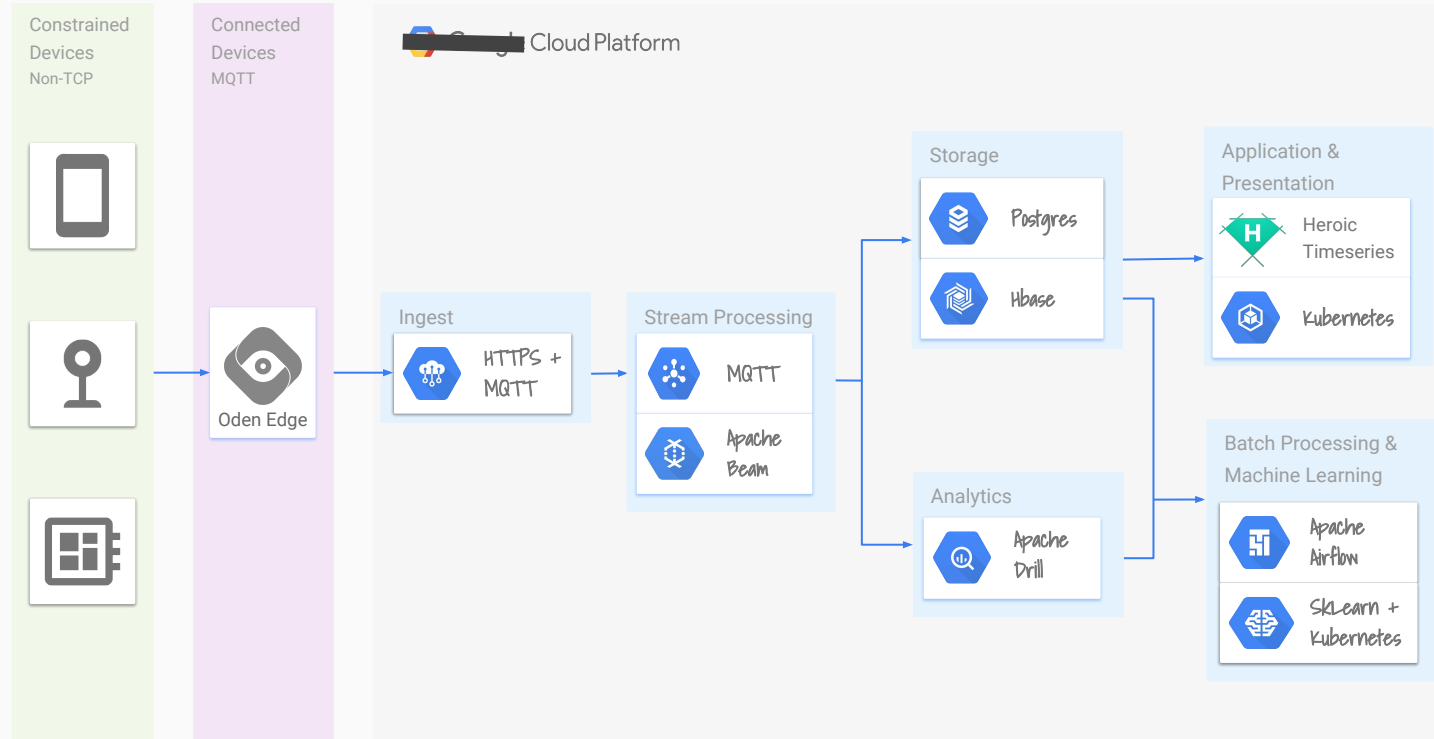
Oden Edge Device

- Embedded Linux device
- Python and Go docker containers that interact with industrial protocols over serial and ethernet
- Connects to cloud via Google IoT and wired, wifi, and cellular networks

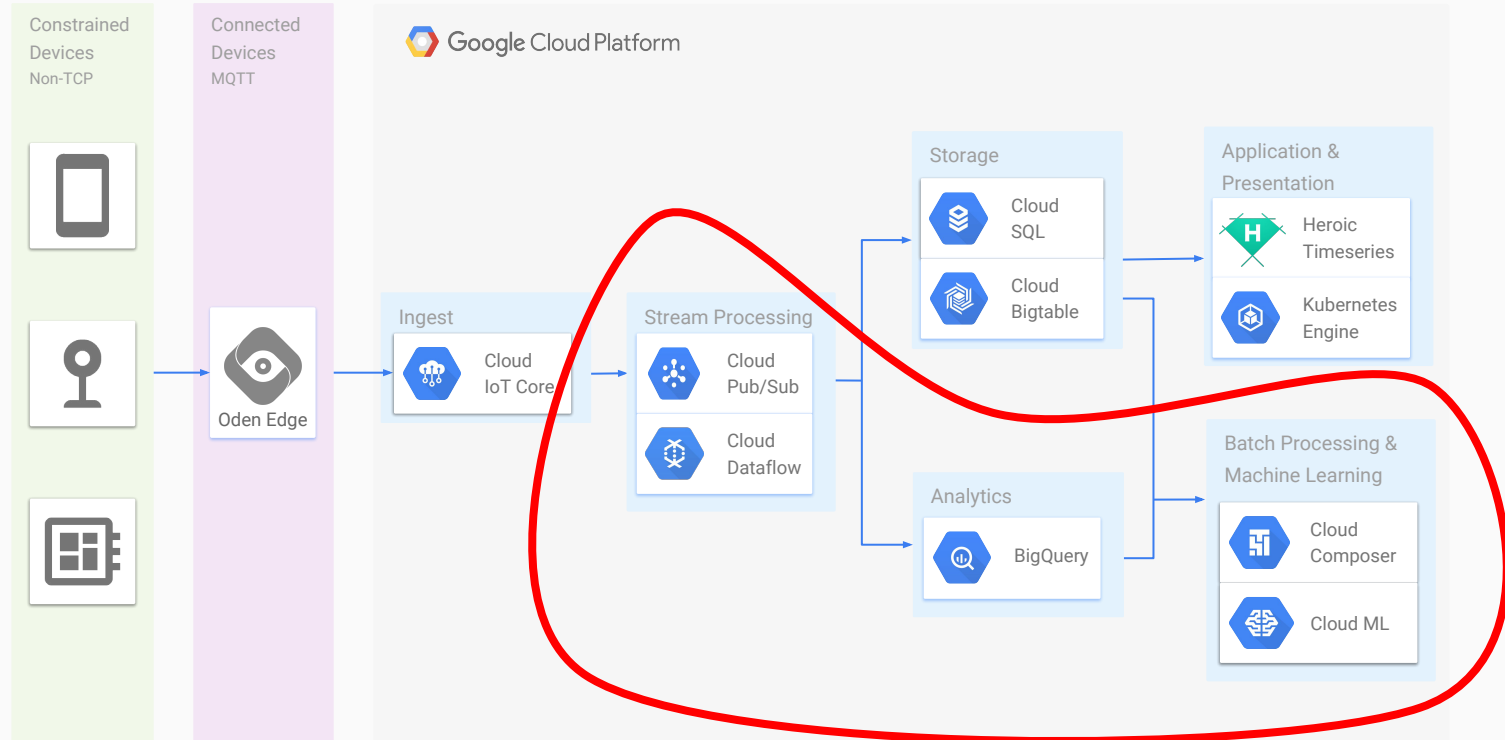
Technology - Architecture



Technology - Architecture (ungoogled)



Technology - Architecture



Airflow At Oden

- Run on a VM, transitioning to Cloud Composer
- Development is done **locally** using *puckel/docker-airflow*
- Used for all cron-like jobs, some “on-demand” jobs
- Primarily interacts with:
 - Bigquery
 - CloudSQL
 - GCS
 - Apache Beam
 - Kubernetes

Airflow Operators

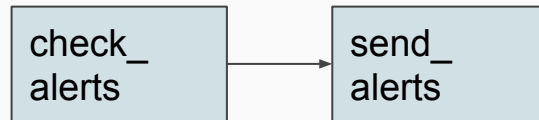
- Bigquery/GCS
 - BigqueryOperator
 - BigqueryDeleteOperator
 - PostgresToCloudStorageOperator
 - BigqueryCreateExternalTableOperator
- Machine Learning
 - KubernetesPodOperator
 - MLEngineTrainingOperator
 - DataflowTemplateOperator
- Internal APIs
 - PythonOperator + HTTPHook
 - Custom Plugins

Alerting DAG

```
dag = DAG(  
    schedule_interval="* * * * *"  
    default_args={  
        "depends_on_past": False,  
    }  
)
```

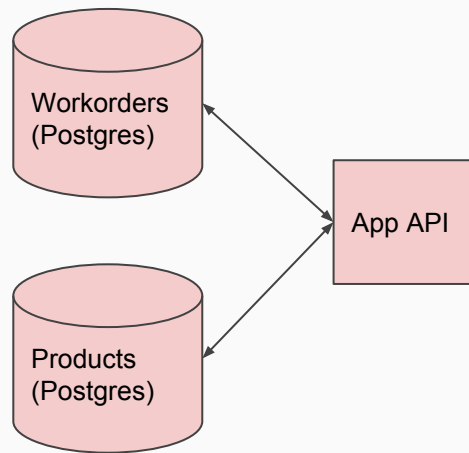
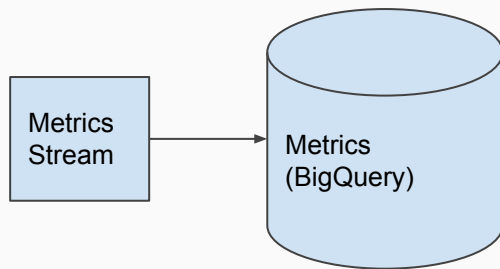
```
check_alerts = SQLSensor(...)  
send_alerts = PythonOperator(...)
```

```
check_alerts >> send_alerts
```



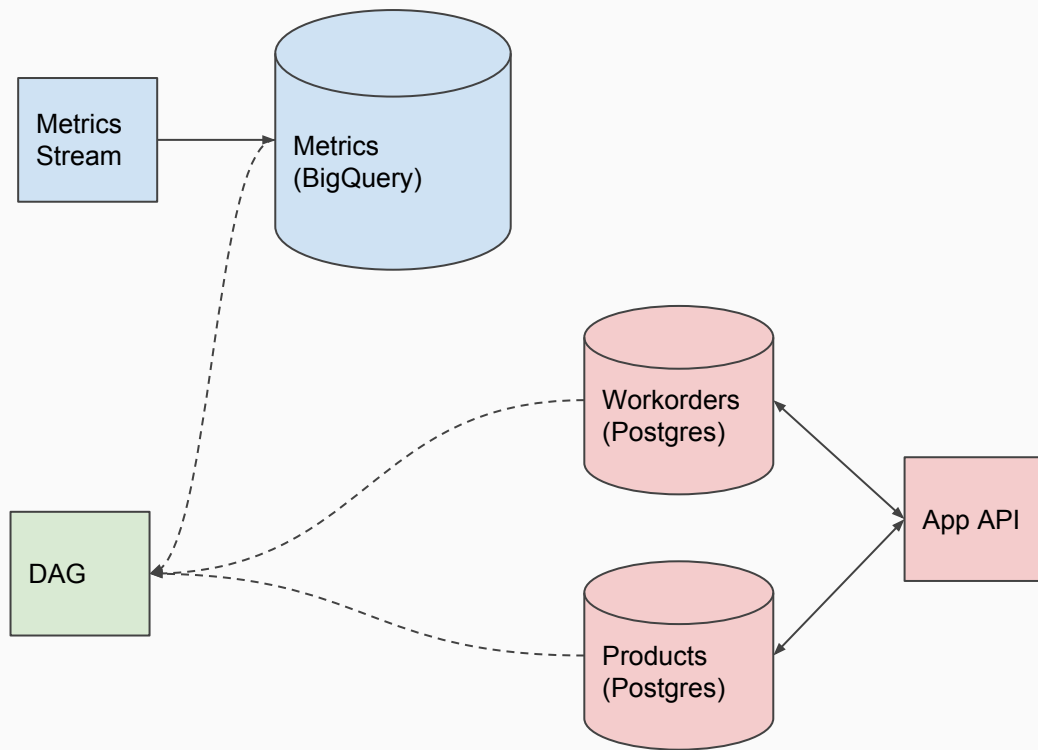
Postgres, GCS, and BigQuery

- Primary “big” data such as second-by-second metrics are stored in BigQuery.
- Customer “business” data such as what they were making lives in Postgres.



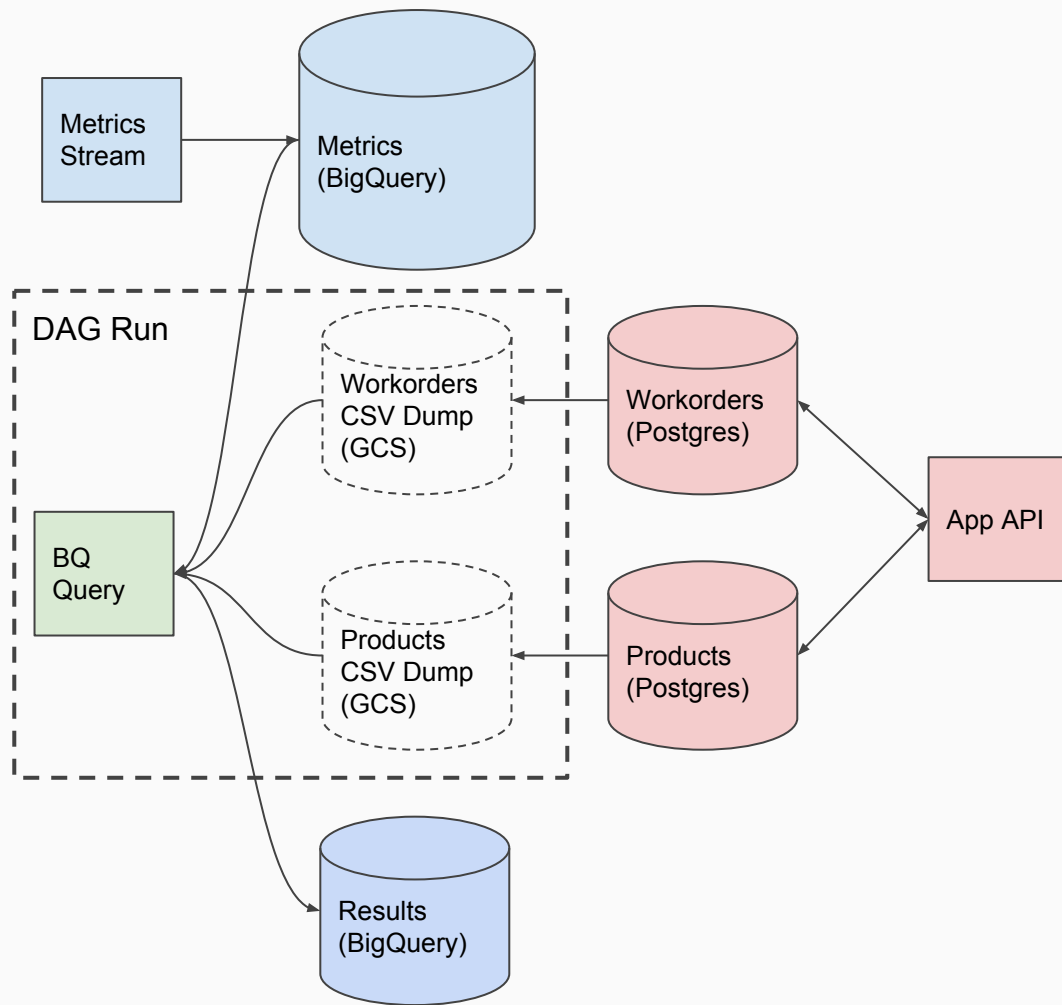
Postgres, GCS, and BigQuery

- Primary “big” data such as second-by-second metrics are stored in BigQuery.
- Customer “business” data such as what they were making lives in Postgres.
- Product need to join “business” and “big” data.



Postgres, GCS, and BigQuery

- Primary “big” data such as second-by-second metrics are stored in BigQuery.
- Customer “business” data such as what they were making lives in Postgres.
- Product need to join “business” and “big” data.
- *PostgresToGoogleCloudStorage-Operator* and *BigQuery External Tables* allows joining of postgres data with bigquery data in DAGs

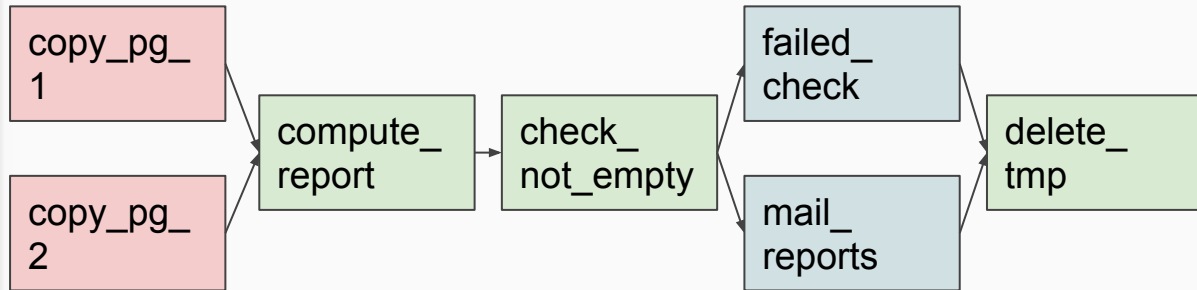


Simple Report Operator DAG

```
dag = DAG(
    schedule_interval="0 0 * * *"
)
TMP_TABLE = "{{macros.tmp_table_name()}}"

copy_pg_1 = PgToGGCSOperator(...)
copy_pg_2 = PgToGGCSOperator(...)
compute_report = BigQueryOperator(...)
check_not_empty = BigQueryCheckOperator(...)
failed_check = DummyOperator(
    trigger_rule=TriggerRule.ONE_FAILED)
mail_reports = PythonOperator(...)
delete_tmp = BigQueryTableDeleteOperator(...)
```

```
copy_pg_1 >> compute_report
copy_pg_2 >> compute_report
Compute_report >> check_not_empty
check_not_empty >> mail_reports >> delete_tmp
check_not_empty >> failed_check >> delete_tmp
```

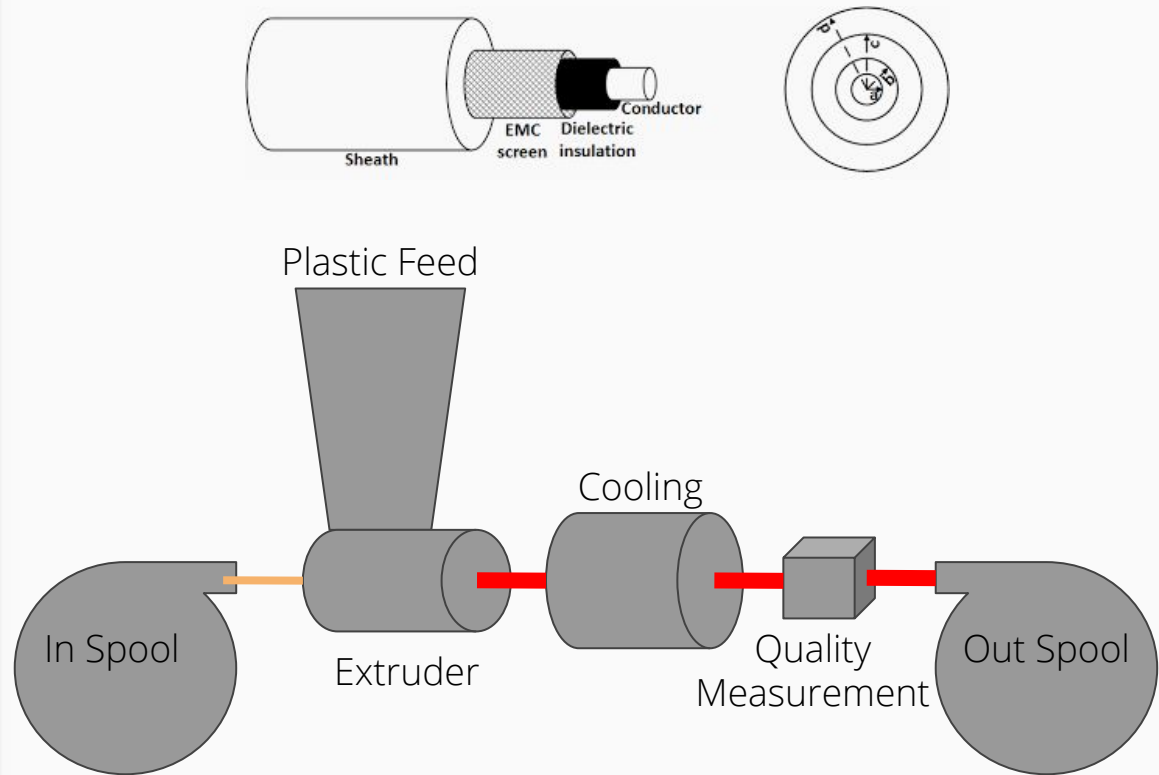


Using Machine Learning for Prediction



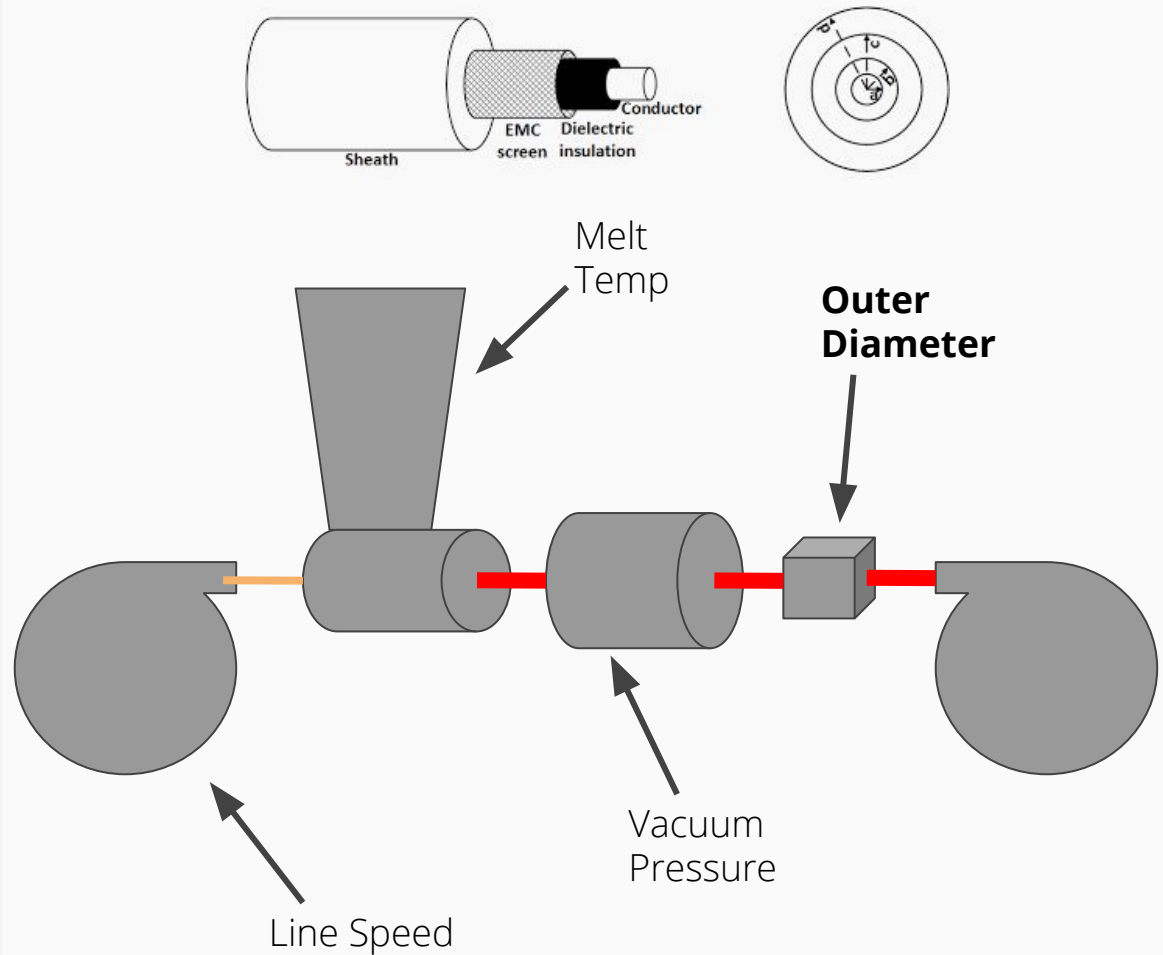
Cable Manufacturing

- Copper is pulled from an in-spool into an extruder.
- Plastic is melted over the copper to make wire.
- Wire is cooled.
- Wire is pulled into an out-spool.



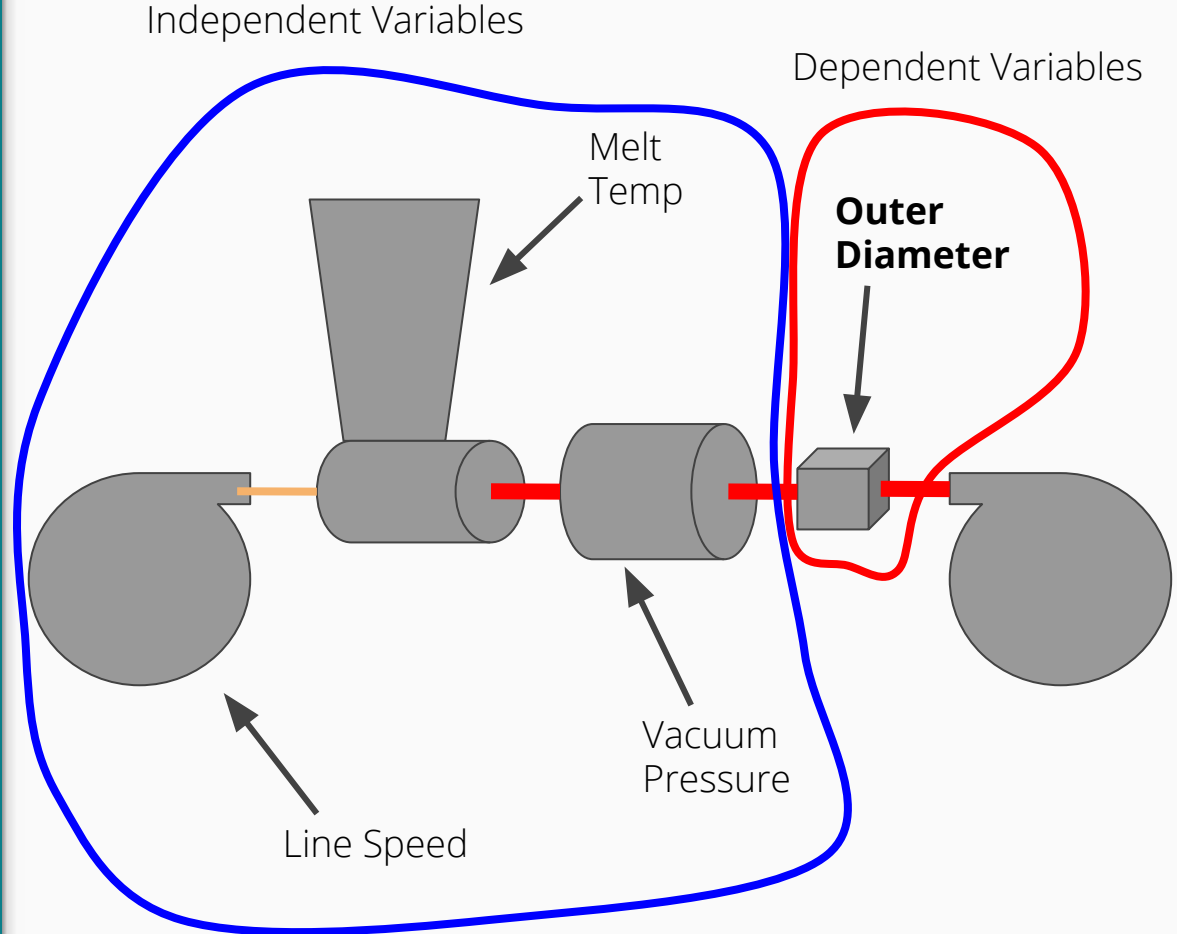
Cable Manufacturing

- Copper is pulled from an in-spool into an extruder.
 - Plastic is melted over the copper to make wire.
 - Wire is cooled.
 - Wire is pulled into an out-spool.
-
- A laser measures the diameter of the wire to monitor its closeness to spec.



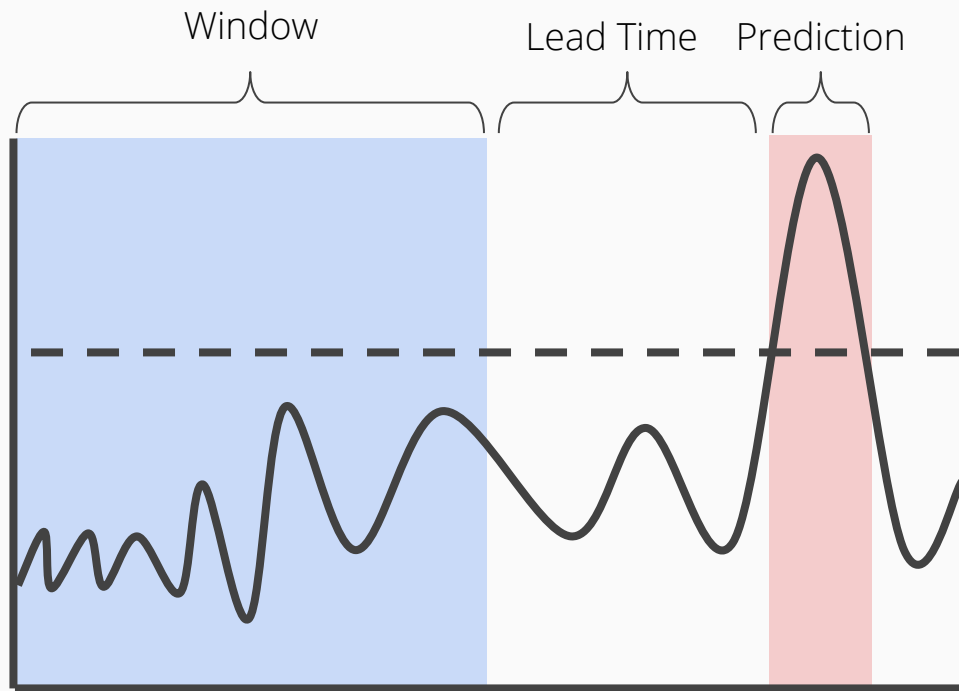
Cable Manufacturing

- Copper is pulled from an in-spool into an extruder.
 - Plastic is melted over the copper to make wire.
 - Wire is cooled.
 - Wire is pulled into an out-spool.
-
- A laser measures the diameter of the wire to monitor its closeness to spec.

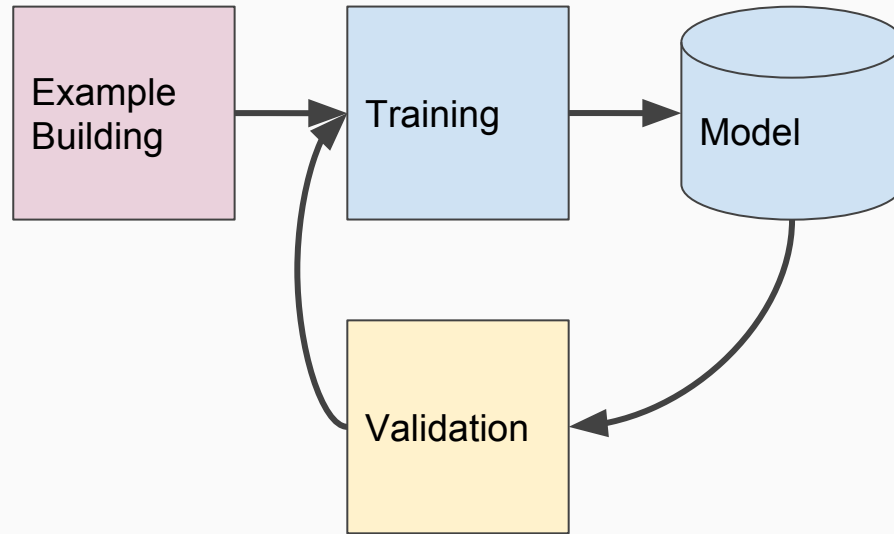


Real Time Stability Prediction

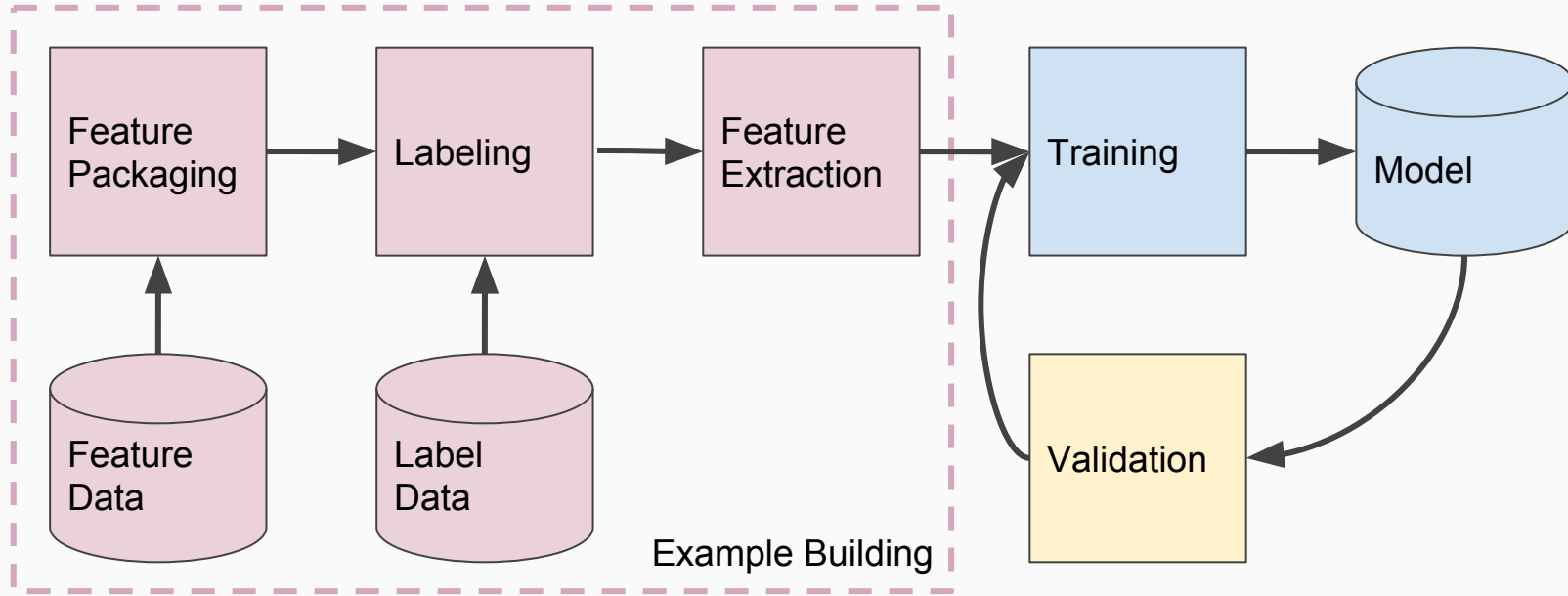
- Given a the set of metrics for a line over some *window*, with some *lead time*, make a *prediction* of a future metric value.



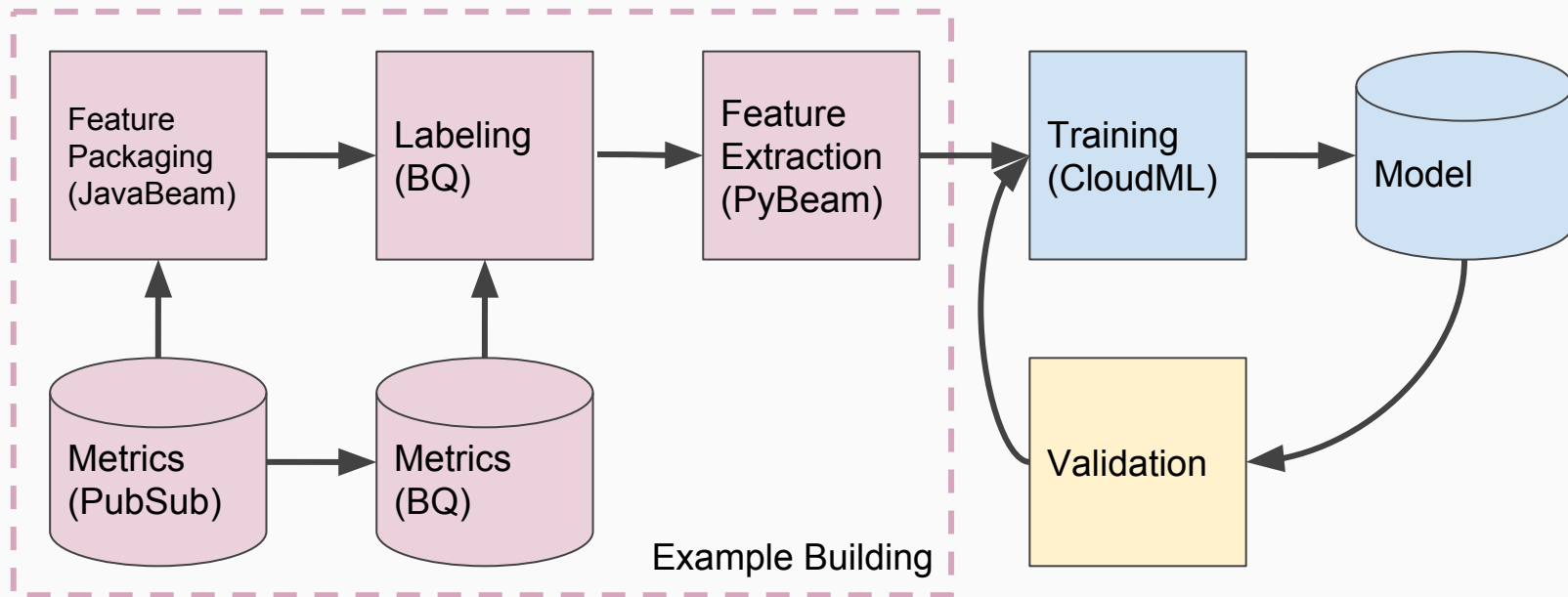
Machine Learning At Oden



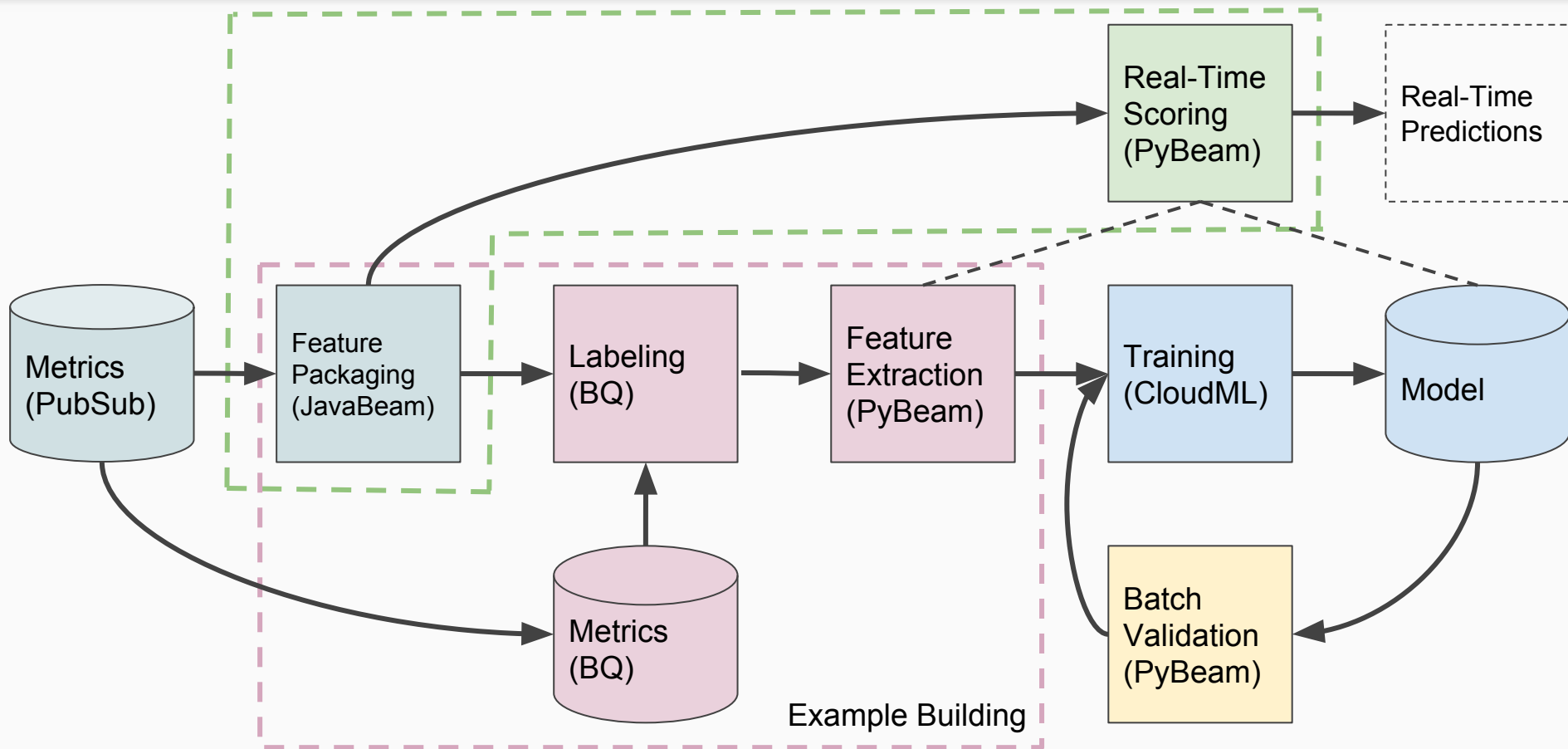
Machine Learning At Oden



Machine Learning at Oden - Real-Time



Real-Time Diameter Prediction



Apache Beam + Dataflow

- Apache Beam is a framework for **batch or stream** processing.
- Logic written in **Java, Python, or Go** (experimental).
- Beam jobs are runner agnostic, runners include Spark, Flink, **Dataflow**
- Beam Templates are parameterizable deployable language agnostic json blobs (storable and runnable from GCS w/ Dataflow)
- Airflow ships with a DataflowTemplateRunner

Example Building DAG

```
dag = DAG(schedule_interval="0 0 * * *")
```

```
FP_TMP_TABLE = "{{macros.tmp_table_name()}}"
```

```
EX_TMP_TABLE = "{{macros.tmp_table_name()}}"
```

```
create_ext_fp_tmp = \
    BigQueryCreateExternalTableOperator(
        destination=FP_TMP_TABLE, ...)
```

```
label_fp = \
    BigQueryOperator(
        sql="SELECT FROM "+FP_TMP_TABLE...,
        destination=EX_TMP_TABLE, ...)
```

```
examples_to_gcs = \
    BigQueryToCloudStorageOperator(
        source=EX_TMP_TABLE, ...)
```

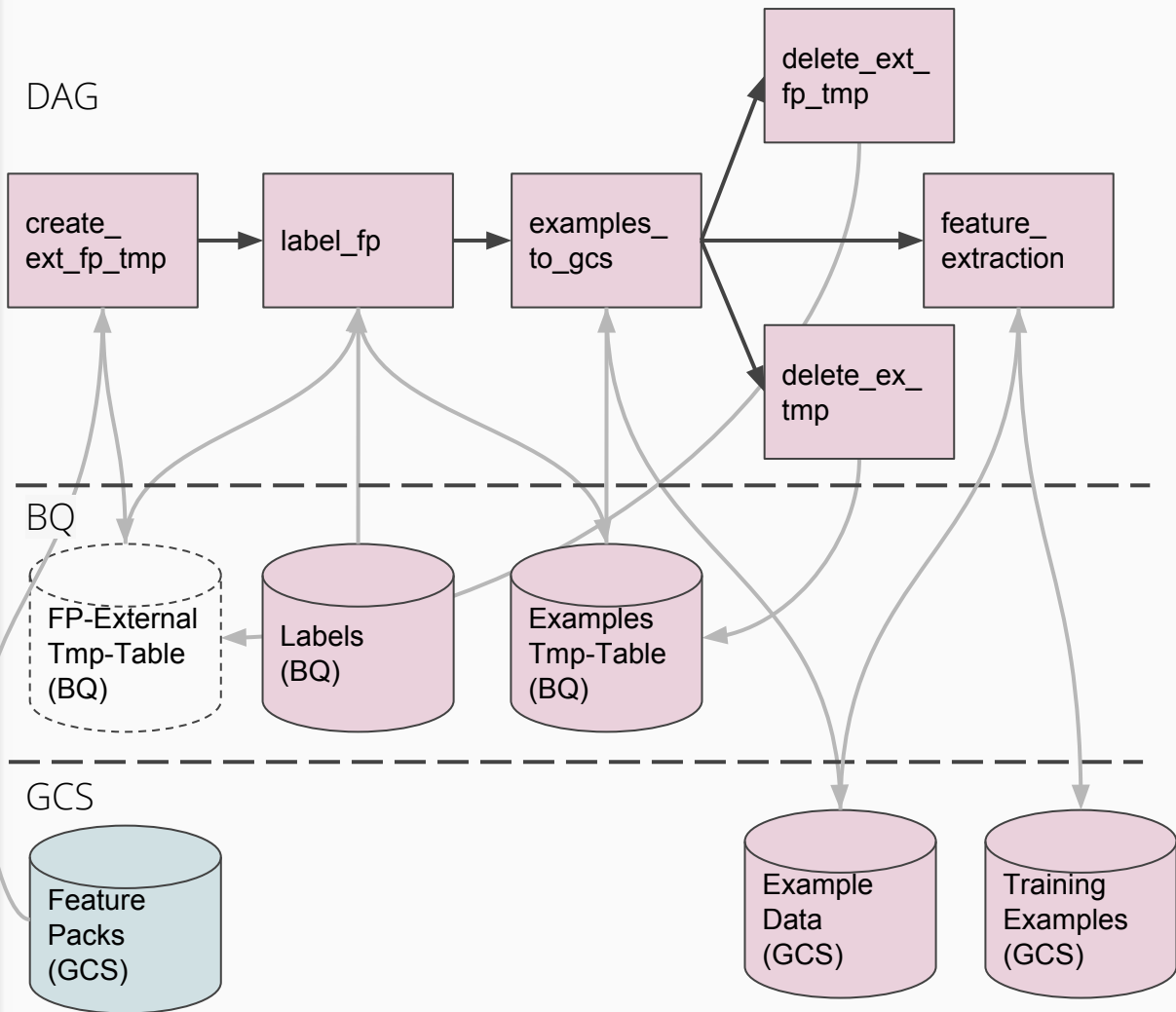
```
delete_fp_tmp = \
    BigQueryTableDeleteOperator(
        table=FP_TMP_TABLE, ...)
```

```
delete_ex_tmp = \
    BigQueryTableDeleteOperator(
        table=EX_TMP_TABLE, ...)
```

```
feature_extraction = \
    DataflowTemplateOperator(...)
```

```
create_ext_fp_tmp >> label_fp
label_fp >> examples_to_gcs
examples_to_gcs >> delete_ex_tmp
examples_to_gcs >> delete_fp_tmp
examples_to_gcs >> feature_extraction
```

DAG



Google Cloud ML Engine

- Serverless training for SKLearn, XGBoost, Keras, and Tensorflow
- Reads training data from GCS
- Writes models back to GCS
- Is capable of prediction (scoring) as well as training via simple json API.
- Airflow ships with MLEngineTrainingOperator

Training DAG

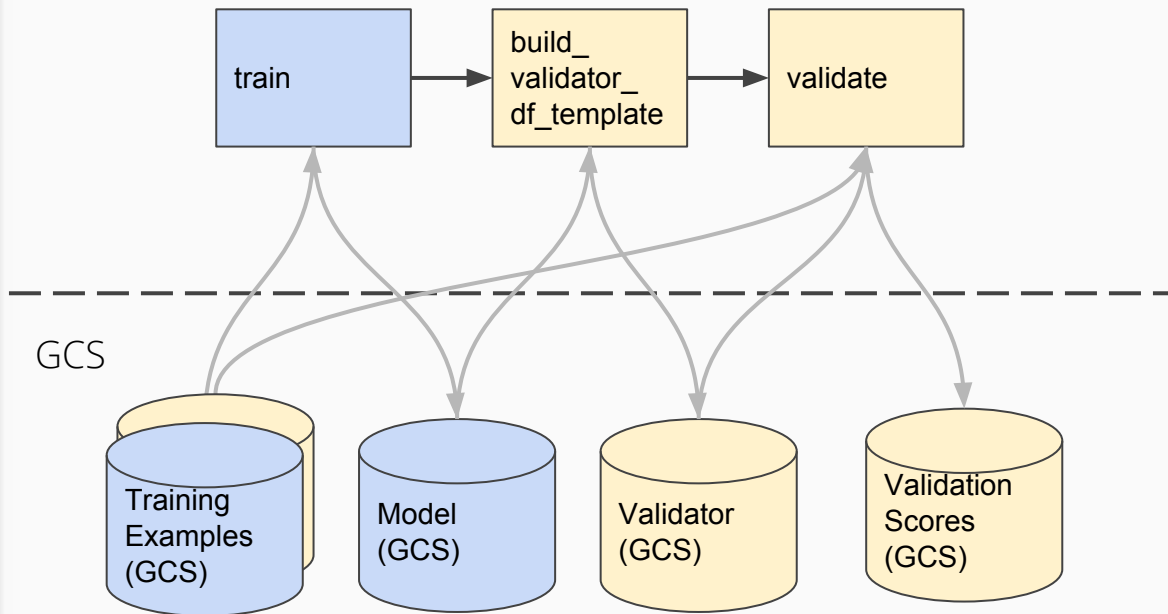
```
dag = DAG()
```

```
EXAMP_GCS = "..."  
TRAIN_GCS = EXAMP_GCS + "/2018*.ndjson"  
VALID_GCS = EXAMP_GCS + "/2019*.ndjson"  
MODEL_GCS = ".../model-{{dag.run_id}}"  
VALIDATOR = ".../validator-{{dag.run_id}}"  
SCORE_GCS = ".../scores-{{dag.run_id}}"
```

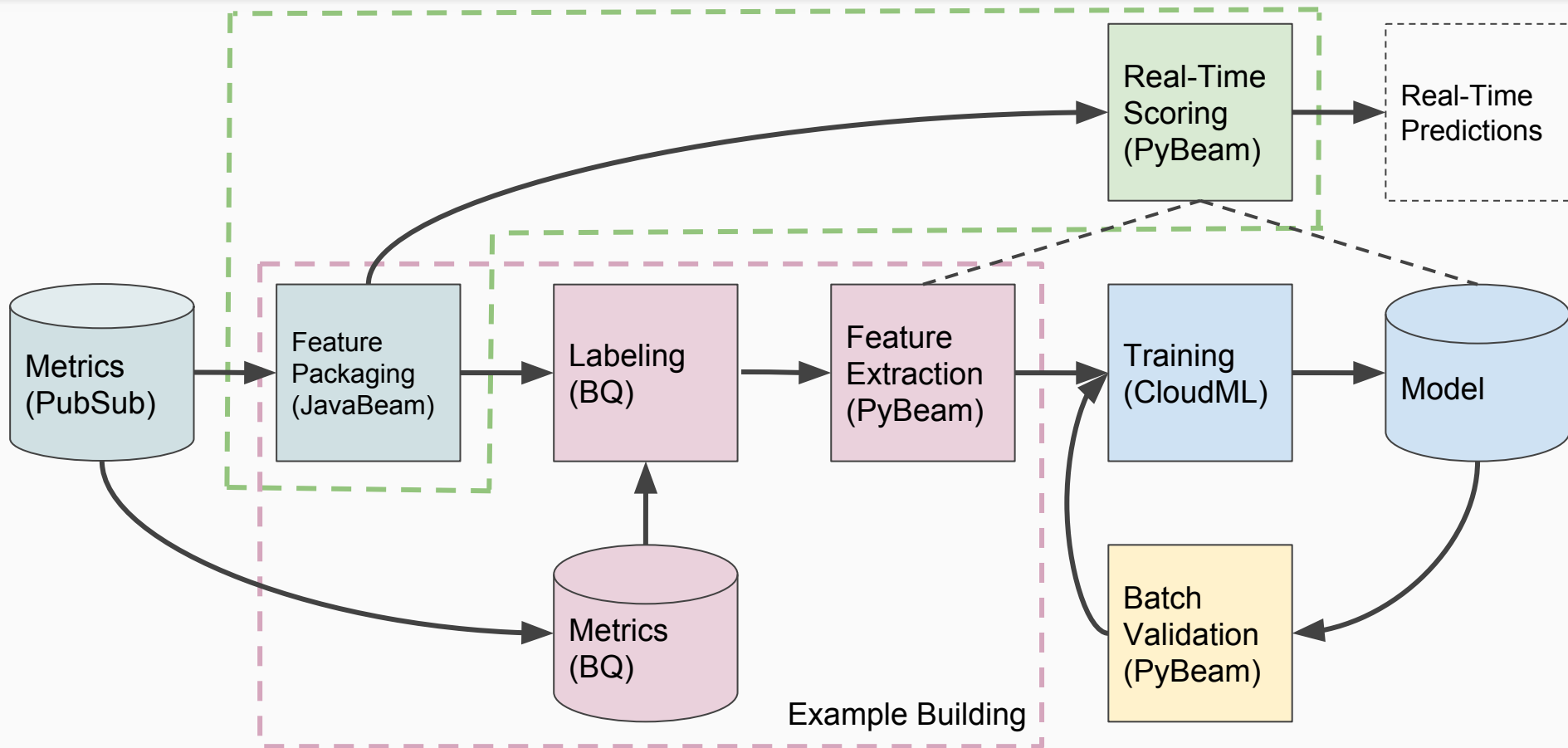
```
train = MLEngineTrainingOperator(  
    ...+"--training_examples_uri"+TRAIN_GCS+  
    "--modeluri="+MODEL_GCS)  
build_validator_df_template = \  
    KubernetesPodOperator(  
        ...+"--modeluri="+MODEL_GCS,  
        "--validatoruri="+VALIDATOR)  
validate = \  
    DataflowTemplateOperator(  
        ...VALID_GCS...VALIDATOR...SCORE_GCS)
```

```
train >> build_validator_df_template >> \  
validate
```

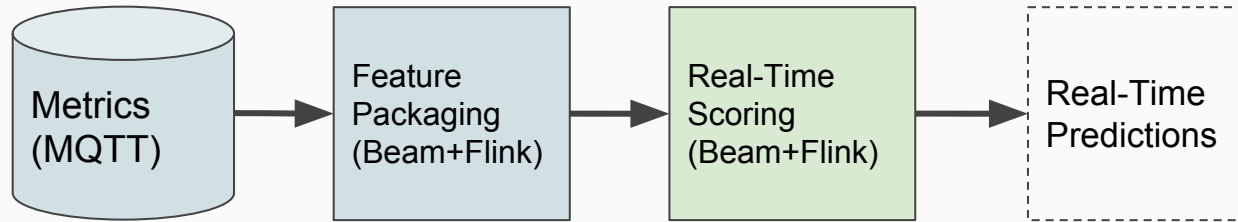
DAG



Real-Time Diameter Prediction



Machine Learning At Oden - Real-Time Diameter Prediction



Takeaways

- GCP has invested heavily in making their infrastructure airflow compatible.
- Airflow, BQ, Dataflow, and CloudML make for easy to manage serverless machine learning infrastructure.
- Apache Beam is an efficient way to deploy streaming ML.

Thank You

We're hiring! oden.io/jobs

