Scaling Industrial IoT in the Upstream Energy Industry

Sanjay Paranji, Former CTO, Anadarko Petroleum
Nick Losier, Data & Analytics Consultant, Google Cloud

Google Cloud

Agenda

IoT provides a novel source high fidelity, high-resolution data enabling us to rapidly and automatically detect changes to operational conditions. Update oil and gas data scientists are using this data to derive new insights regarding equipment health and asset performance.

Developers are using GCP native stack to build applications that monitor drilling, completions, and facility activities providing real-time performance measurement and advisory control recommendations to field operators.

1. **A MACRO VIEW**
   - Why Energy Matters
   - IoT in Action at Upstream Operators
2. **IoT BLUEPRINT FOR TIME SERIES DATA**
   - Overview
   - Data Processing Pipeline
   - Machine Learning Pipeline
   - Future Direction
3. **Q&A**

Google Cloud

A Macro View

01.1 Why Energy Matters
01.2 IoT in Action at Upstream Operators

Google Cloud

By 2040, global energy demand is set to grow by 25%

IEA, World Energy Outlook, 2018

How do we plan to keep up?

Google Cloud

Digital Solutions

Exploration

Identifying sweet spots where well performance is high and land entry costs are low can generate significant value.

Development

Selecting the optimal well design - which involves choices in numerous areas such as completion size and well spacing - requires predicting the performance for each candidate design.

Operations - Fleet Performance Optimization

Monitoring and understanding asset behavior through the life-cycle of construction to production.

Google Cloud

Real-Time Operations

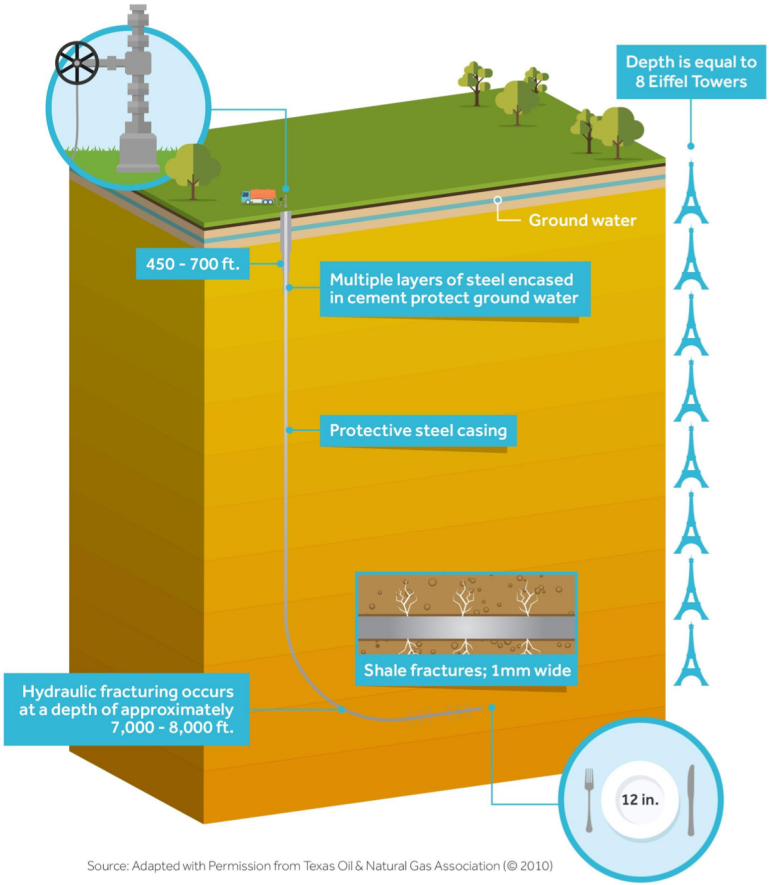
Improve Safety and Environmental Performance

Optimize Operations & Standardize Repetitive Tasks

Reduce Costs

- Process real-time streaming log data & other non-streaming data
- Derive operational KPIs at very high resolution
- Runs online 24/7 with continuous monitoring and uptime

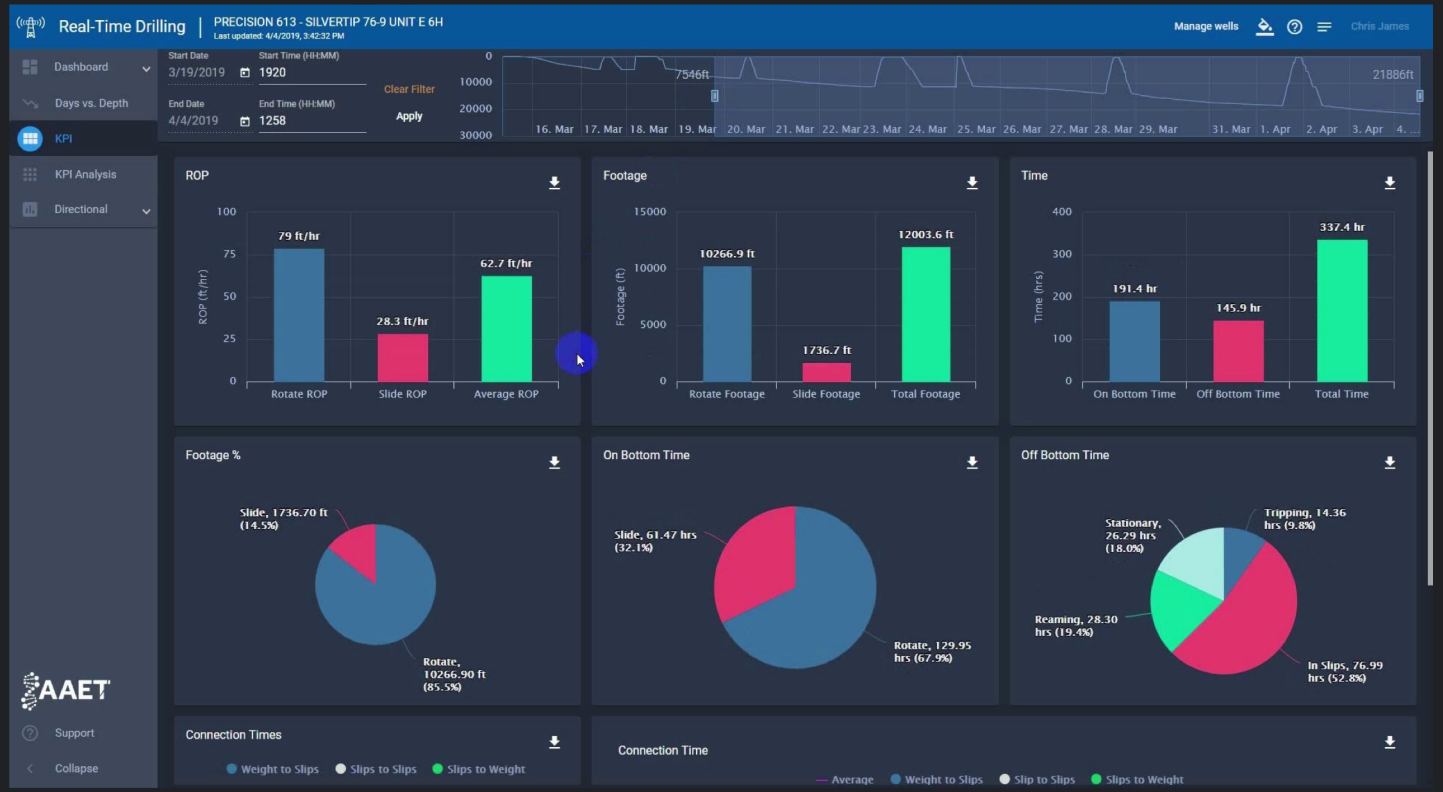
Google Cloud



Source: Adapted with Permission from Texas Oil & Natural Gas Association (© 2010)

Real-Time Operations

Demo & KPI Analysis



Google Cloud

IoT Blueprint for Time Series Data

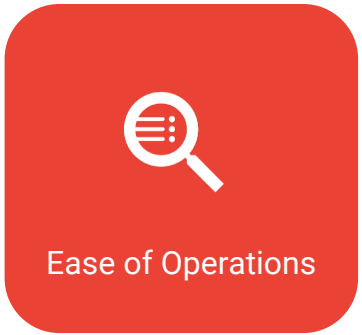
02.1 Overview
02.2 Data Processing Pipeline
02.3 Machine Learning Pipeline
02.4 Future Direction

Google Cloud

GCP Enabled Solution

Adoption Benefits

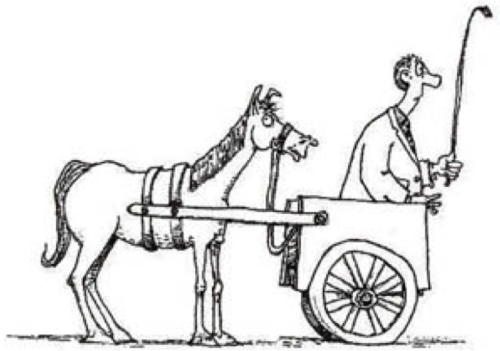
Accelerating Speed to Market



Google Cloud

Cloud First

- Built Proof of Concept using existing data collection infrastructure
- Enabled rapid iterations - fail fast with lower risk
- Allows Digital twin applications

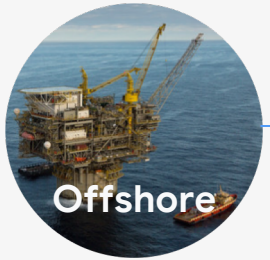


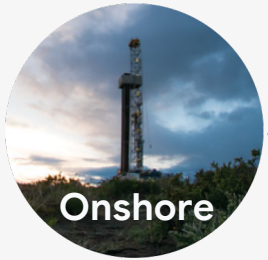
Working Backwards Towards the Solution

IoT Blueprint: **Fleet Performance Optimization on GCP**

Architecture: Internet of Things > Sensor stream ingest and processing

Upstream Producer





Google Cloud Platform

Data Movement
- Cloud Pub/Sub
- Cloud Dataflow
- Apache Beam

Data Repository
- Cloud Bigtable
- OpenTSDB
- BigQuery
- Cloud Storage

Orchestration
- Cloud Composer
- GitLab

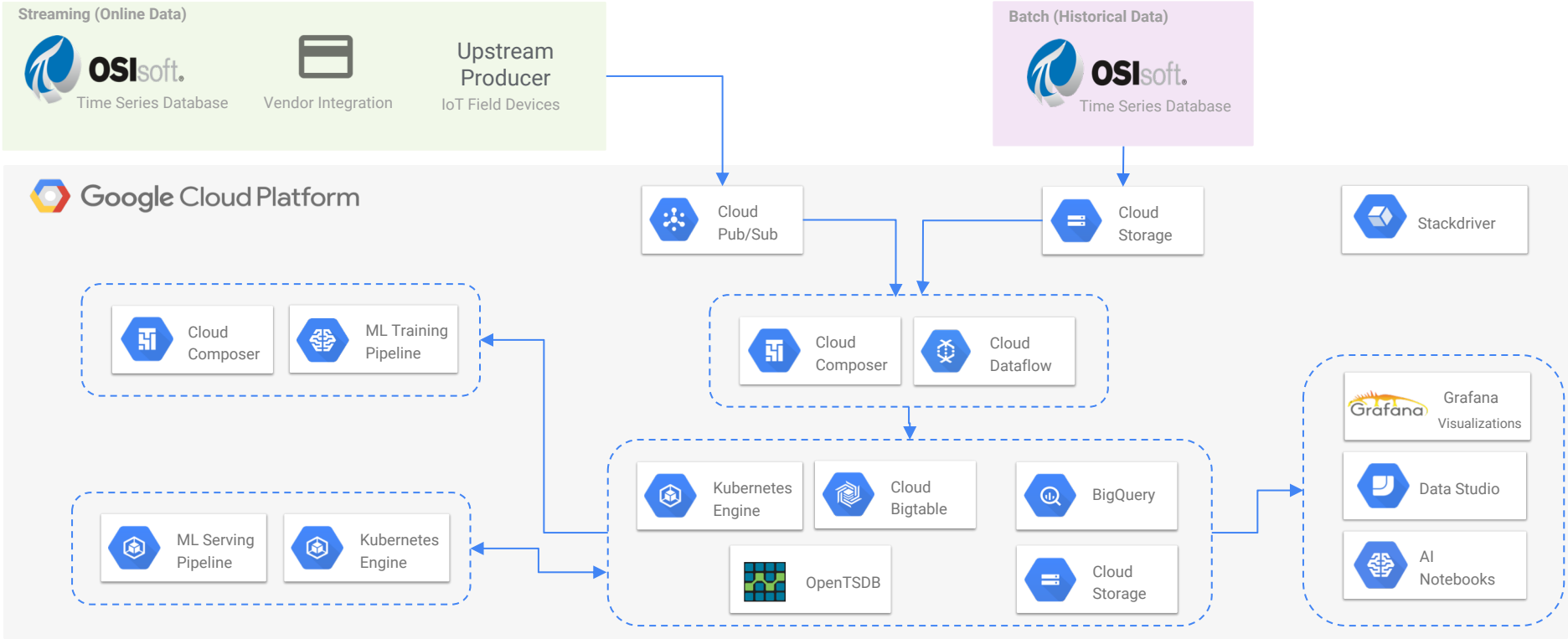
Compute
- AI Engine Jobs
- Kubernetes Engine
- TensorFlow

Insights Tools
- AI Notebooks
- Data Studio
- Grafana

Google Cloud

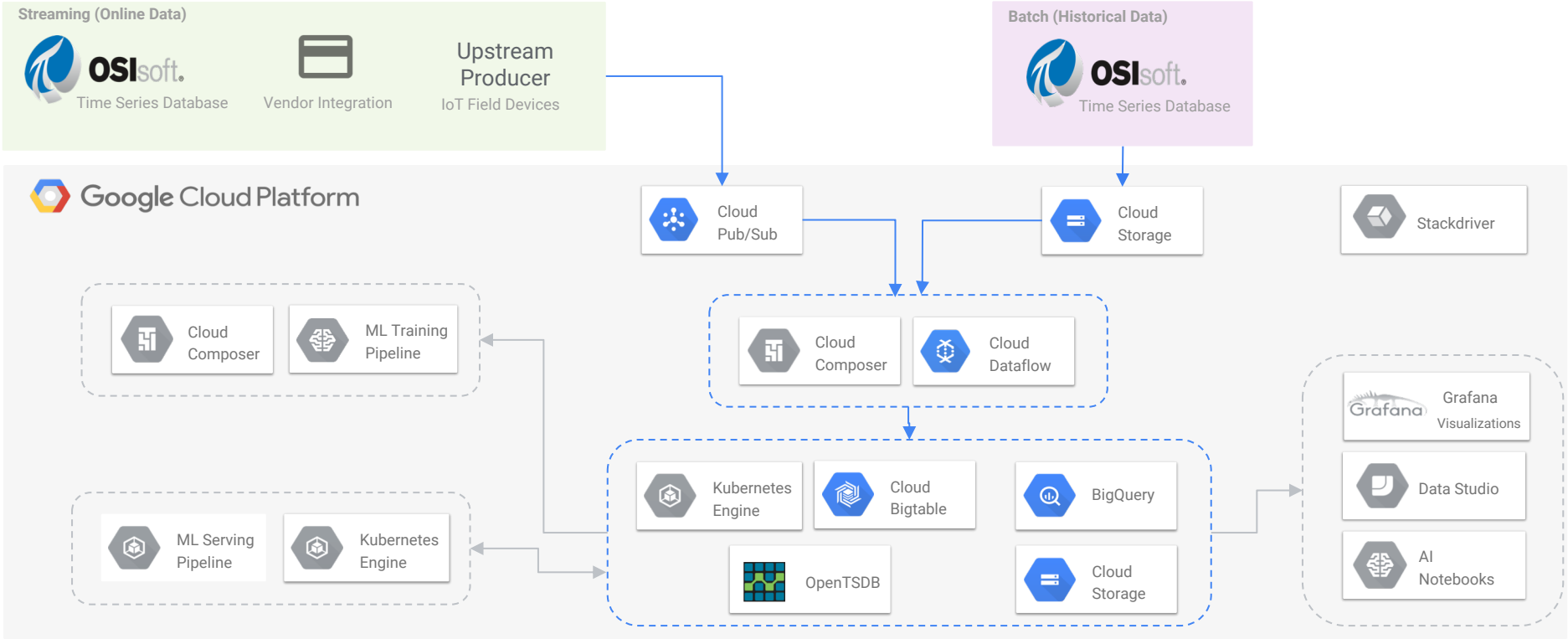
IoT Blueprint: Time Series Data Processing & ML Pipeline

Architecture: Internet of Things > Time Series Data Processing & ML Pipeline

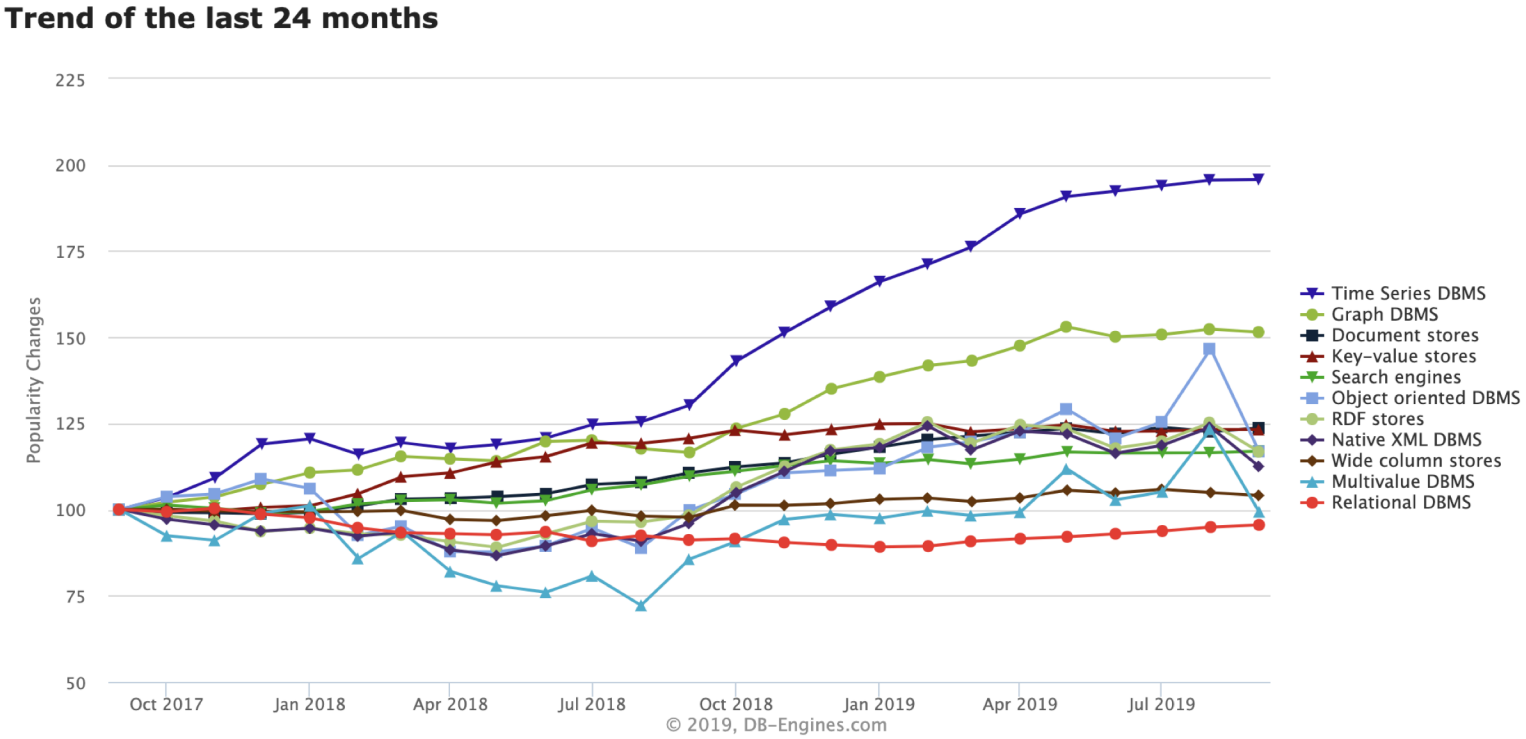


IoT Blueprint: Data Processing Pipeline

Architecture: Internet of Things > Time Series Data Processing & ML Pipeline



Why Data in both BigQuery and OpenTSDB/Bigtable?



Google Cloud

IoT Blueprint: Data Processing Pipeline Key Takeaways

IoT Analytics benefit from strengths of BigQuery as well as Time Series Interfaces

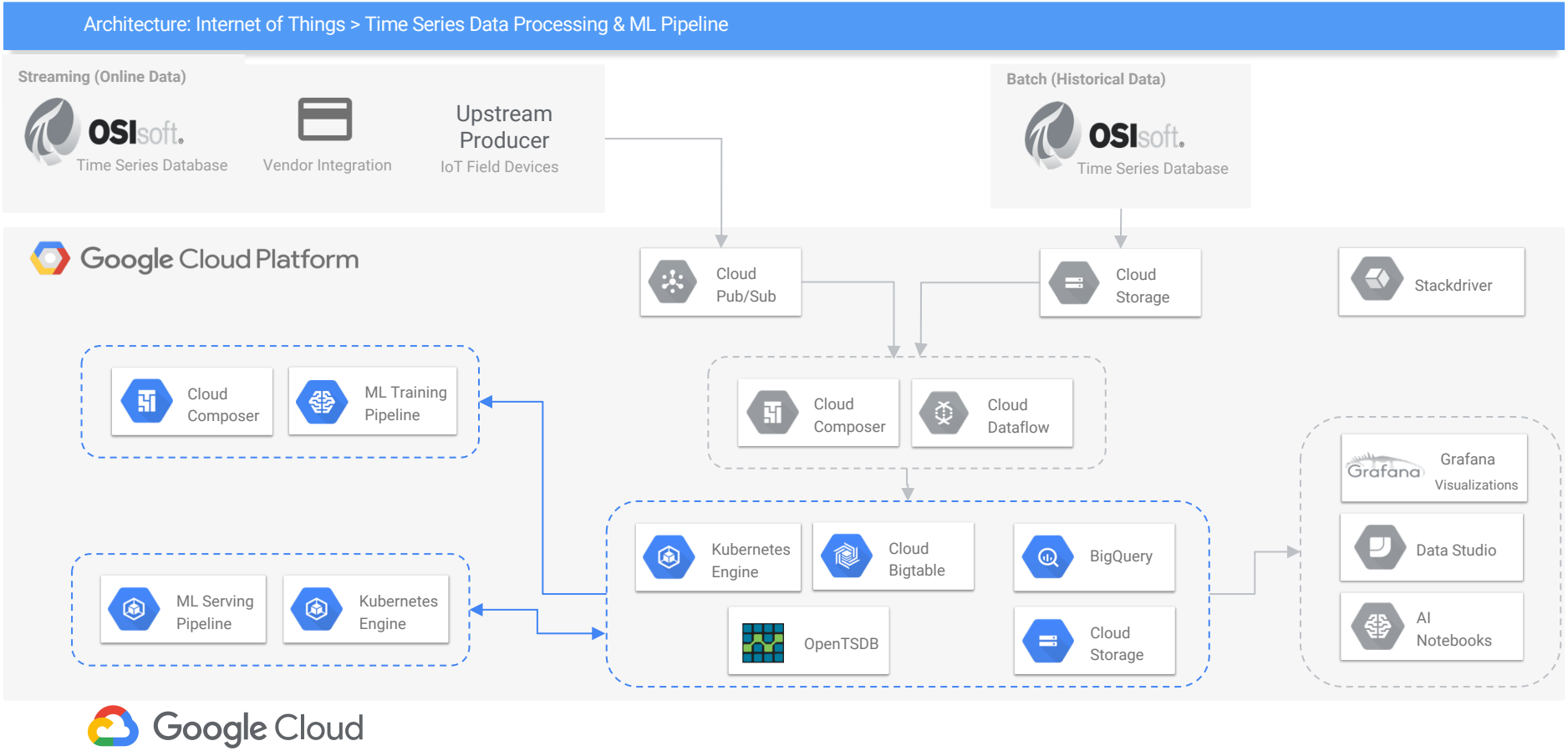
---

Focus on the Data, not the Infrastructure - lean on scalable managed services like Cloud Storage, Pub/Sub, Dataflow, BigQuery and Bigtable

---

Design batch and streaming pipelines using near identical Apache Beam pipeline code, running on Cloud Dataflow

IoT Blueprint: **Machine Learning**



IoT Blueprint: ML Training Pipeline - Driven via Cloud Composer

01 Read
Prepare and downsample telemetry data to BigQuery, ie one year across equip telemetry tags

02 Pre-Process
A chain of BigQuery SQL transforms are applied to the data, after which it is stored in Cloud Storage

03 Train
Using the processed data, AI Platform is used to submit a training job, and the weights are stored in Cloud Storage

04 Batch Predict
With trained weights and hyper parameters, AI Platform is used to submit a batch prediction job

05 Writeback
The results are stored in GCS, and are then transferred to BigQuery

06 Analyze
The BigQuery results are accessed via AI Notebooks for visualization and analysis

07 Online Predict
With trained weights and hyper parameters, deploys a Kubernetes container for live prediction

Google Cloud

IoT Blueprint: ML Online Prediction Pipeline - via Kubernetes

01 Read
The tag data is read in batch format from OpenTSDB

02 Pre-Process
The data is formatted and prepared for ML prediction input

03 Predict
The saved model weights generated by the ML training pipeline are used in local mode for prediction

04 Post-Process
The prediction values and scores are separated, and unpivoted

05 Writeback
The results are written to PubSub, from which they feed back into OpenTSDB and BigQuery

Google Cloud

IoT Blueprint: ML Pipeline Key Takeaways

Focus on the ML model & results, not the Infrastructure - lean on scalable managed services like Cloud Composer, AI Platform, and Google Kubernetes Engine

---

Consider build vs buy tradeoffs between a home grown versus off the shelf anomaly detection workflow

---

Take advantage of existing historian telemetry collection - shift towards Cloud IoT direct integration when ready

IoT Blueprint

Future Direction

01 **Further Cloud Automation**

02 **Automated Edge Model and App Deployment**

03 **Federated Learning**

Google Cloud


Thank you
Questions?
Google Cloud