# Current Proposal

**Attn: Michael Strachan**

## Architecture:

This project utilizes a modular workflow approach to manage the entire data pipeline. Each workflow represents a distinct stage of the process, ensuring a clear and organized structure.

- BigQuery Export: Google Analytics
- BigQuery Data Transfer Service: Google Ads, Meta Ads
- Storage Transfer Service: Bing Ads, TikTok Ads

## Analytics Engineering:

1. Data Ingestion: Data from various sources like BingAds and TikTokAds is ingested into Cloud Storage. Google Analytics data is exported to BigQuery ML.
2. Data Processing: BigQuery ML processes data from Cloud Storage and other sources. DataForm integrates with GitHub for version control and uses Cloud Functions for automation.
3. SQLX is used within DataForm for data transformations.
4. Data Management: Looker is used for data visualization and reporting. LookML integrates with Looker for data modeling. Spectacles is used for LookML validation.

## Outputs:

- Dashboards: Generated by Looker for data visualization.
- Looker Actions: Custom actions integrated into Looker.
- Looker API: Used for programmatically accessing Looker data.
- Looker SDK: For integrating Looker with other applications.
- Cloud Marketplace: For publishing and accessing Looker applications.

## Cloud Environment:

- Docker: Containerization for applications.
- Resource Manager: For managing cloud resources.
- Cloud APIs: Various APIs provided by Google Cloud.
- Billing: Managing billing and costs.
- VPC: Virtual Private Cloud for networking.
- Firewall: Security and access control.
- Workflows: Orchestrating complex workflows.
- Cloud SDK: Tools and libraries for cloud management.
- Cloud Build: CI/CD for building and deploying applications.
- Vertex AI: For machine learning and AI applications.
- Cloud Logging: Centralized logging for monitoring and troubleshooting.
- Cloud Composer: Manage the dailty data source ffeds and transformation.

## Proposal Amendment:

I have concluded that this environment is too complex to not have a fully dedicated engineer supporting it and that the heterogeneous mix of technologies would require almost all of your (the customer) time to learn, operate, and maintain.

There are too many moving parts and although I have made attempts to automate some it is not enough. Subsequently, I am proposing a significant change to the Analytics Pipeline.

**We remove Looker from the architecture.**

I assert Looker adds too much complexity, has too steep a learning curve, and can be replaced by Machine Learning models for a lower cost and less time to implement and manage; removing Looker also reduces the overall infrastructure complexity, cost, and maintenance.

## Assertions:

Most of the heavy lifting is performed by BigQuery and BQML. We still have a data model with Dataform (SQLX), the ML models were already planned for BigQuery not Looker, and we remove the risk of translating SQL and SQLX to LookML.

Looker, in the current scenario, is not contributing much more than visualizations and an API for delivery. Visualizations can be replaced by Looker Studio for $9 per user and the API delivery can be replaced by Cloud Functions, a low-cost microservice.

Even with Looker cloud Core being native in the Console, it still can't be managed with Cloud API or the Cloud SDK.

We increase the workload on Gemini Pro 1.5 to build our Data Product with Machine Learning models that are not available in Looker and, frankly, easier to build.

Additionally, this will be easier to support. For example, at DoiT there are 15 qualified engineers to handle BigQuery customers. For Looker, there are only 4 and that's because we thought Google was not sufficiently supporting their Looker customers. Outside of Looker this ratio is 100 to 4 not 15 to 4.

And, down the road, we can always reintroduce Looker after some operational stability without having to refactor the entire pipeline.

# Updated Proposal

The major changes are here:

## Analytics Engineering:

1. Data Ingestion: Data from various sources like BingAds and TikTokAds is ingested into Cloud Storage. Google Analytics data is exported to BigQuery ML.
2. Data Processing: BigQuery ML processes data from Cloud Storage and other sources. DataForm integrates with GitHub for version control and uses Cloud Functions for automation.

SQLX is used within DataForm for data transformations.
3. Data Management: ~~Looker is used for data visualization and reporting. LookML integrates with Looker for data modeling. Spectacles is used for LookML validation.~~
   Everything stays in BigQuery and Cloud Storage.

## Outputs:

An AI/ML augmented Data Prodcut and set of reports based on daily feeds: BigQuery. Google Analytics, Google Ads, Meta Ads, Bing Ads, and TikTok Ads.
With visualizations build in Looker Studio and python libraries like Motplotlib and Seaborn.

## Cloud Environment:

Docker: Containerization for applications.

Resource Manager: For managing cloud resources.

Cloud APIs: Various APIs provided by Google Cloud.

Billing: Managing billing and costs.

VPC: Virtual Private Cloud for networking.

Firewall: Security and access control.

Workflows: Orchestrating complex workflows.

Cloud SDK: Tools and libraries for cloud management.

Cloud Build: CI/CD for building and deploying applications.

Vertex AI: For machine learning and AI applications.

Cloud Logging: Centralized logging and troubleshooting.

~~Cloud Monitoring:~~

~~Cloud Composer:~~

VR/dion