Cloud Image Processing Infrastructure

Picture Perfect LLC

Key Requirements

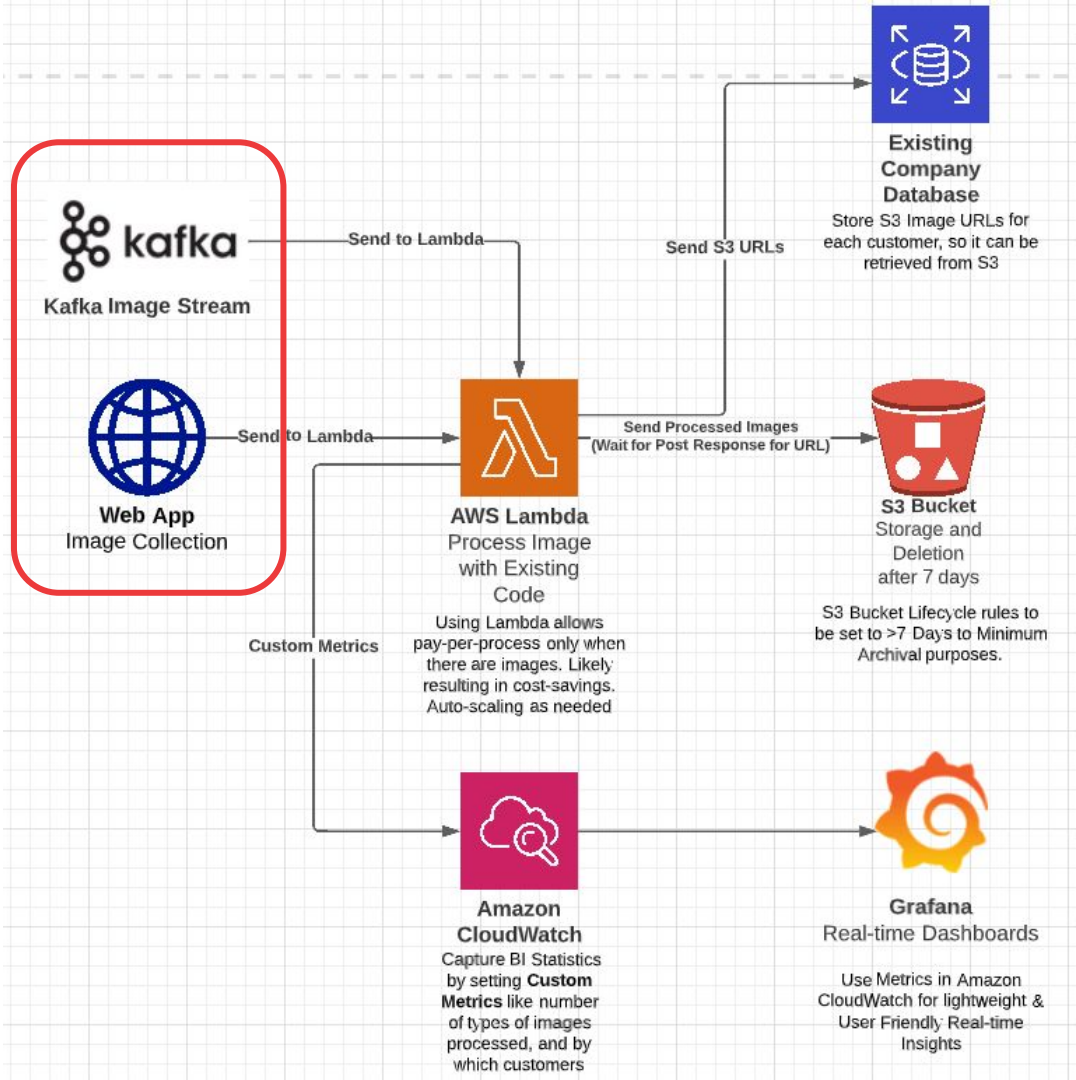
With these requirements in mind, it is only natural that we utilise **cloud infrastructures**.

AWS:

- Auto-scale resources to **meet large unstable resource demands**
- S3 has rule-based storage → **Auto deletion** → Save Spaces
- CloudWatch → Custom Metrics for BI

Data Producers

Both **Kafka Image Stream** and **Image Collection Web App** will send new images directly to **AWS Lambda**.



Compute

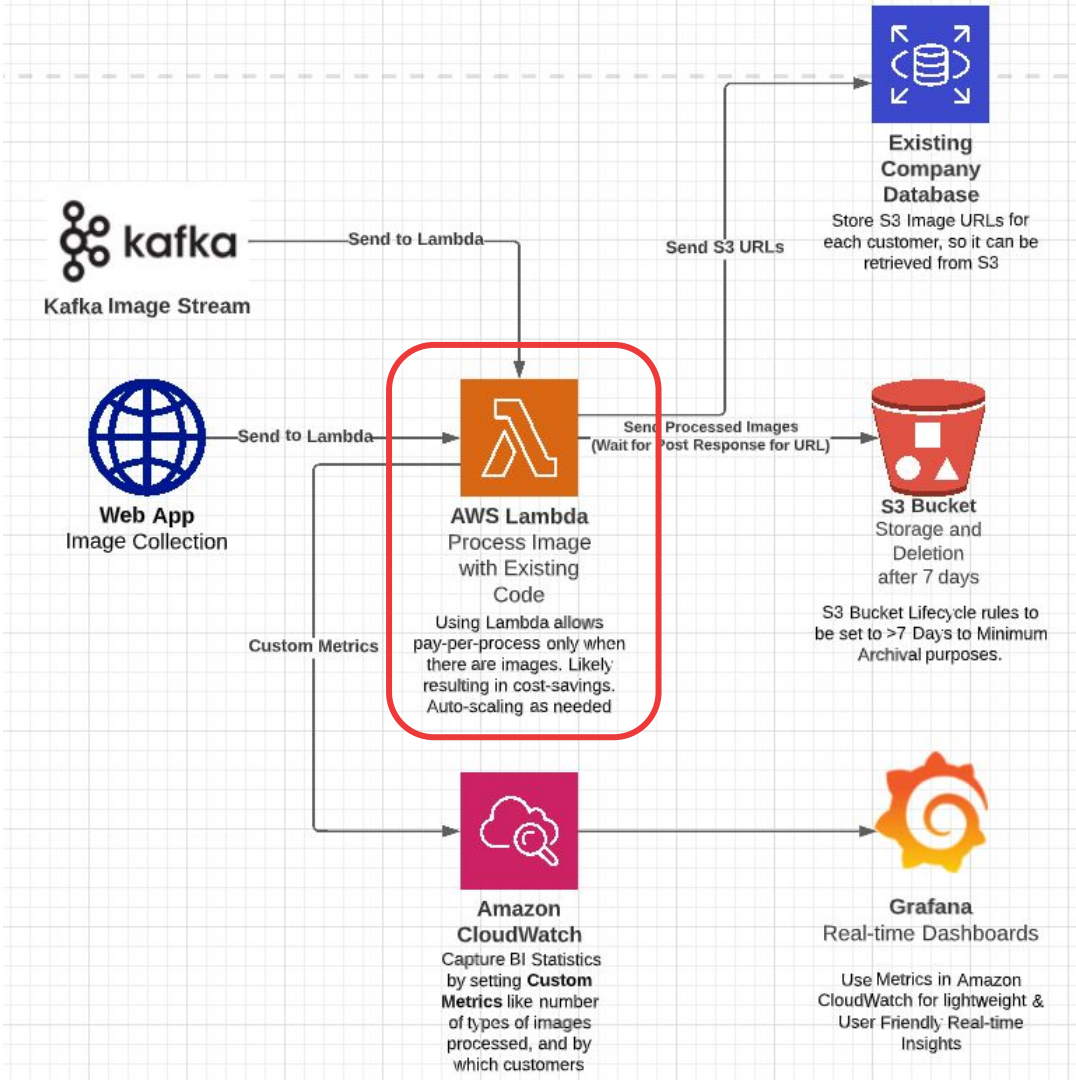
Triggered by Data Producers, AWS Lambda will **process the images using the existing code.**

Advantages:

- Lambda being **pay-per-use** will provide down to the minute **cost-savings.**
- **Zero idle time** compared to a normal server.

Disadvantage:

- **15 Mins** max per usage (Large images may never get processed completely if it exceed 15 mins

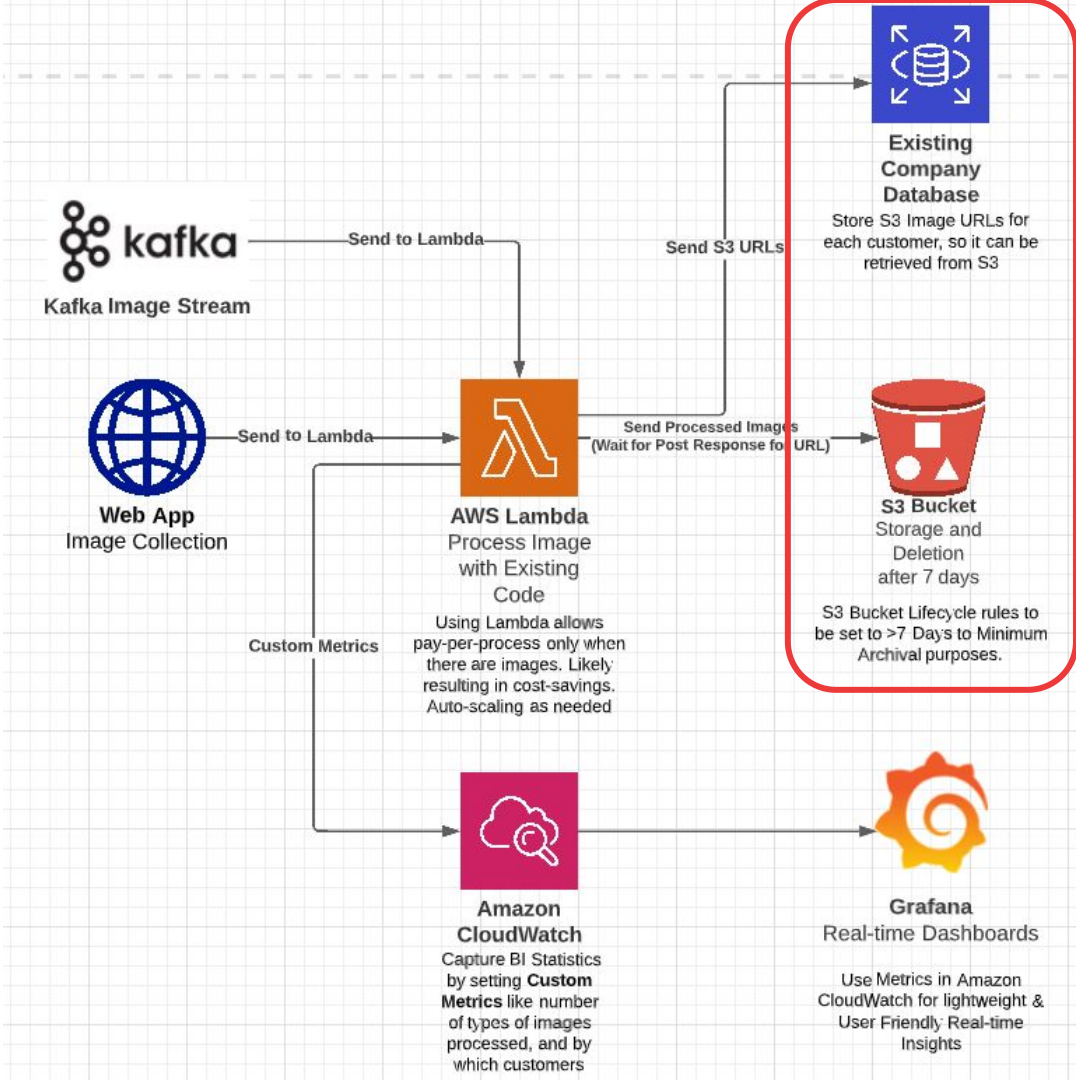


Storage

AWS Lambda will **send processed Images to S3**. S3 Image **URLs will be stored** in company database

Advantages:

- Large images that may cause unstable loads is storage in S3 which **auto-scales**, eliminating the problem.
- S3 Lifecycle Rules allows **auto-deletion**, can optimise space storage which results in cost-savings. (Min 7-days for archival purposes)



Monitoring & Business Intelligence

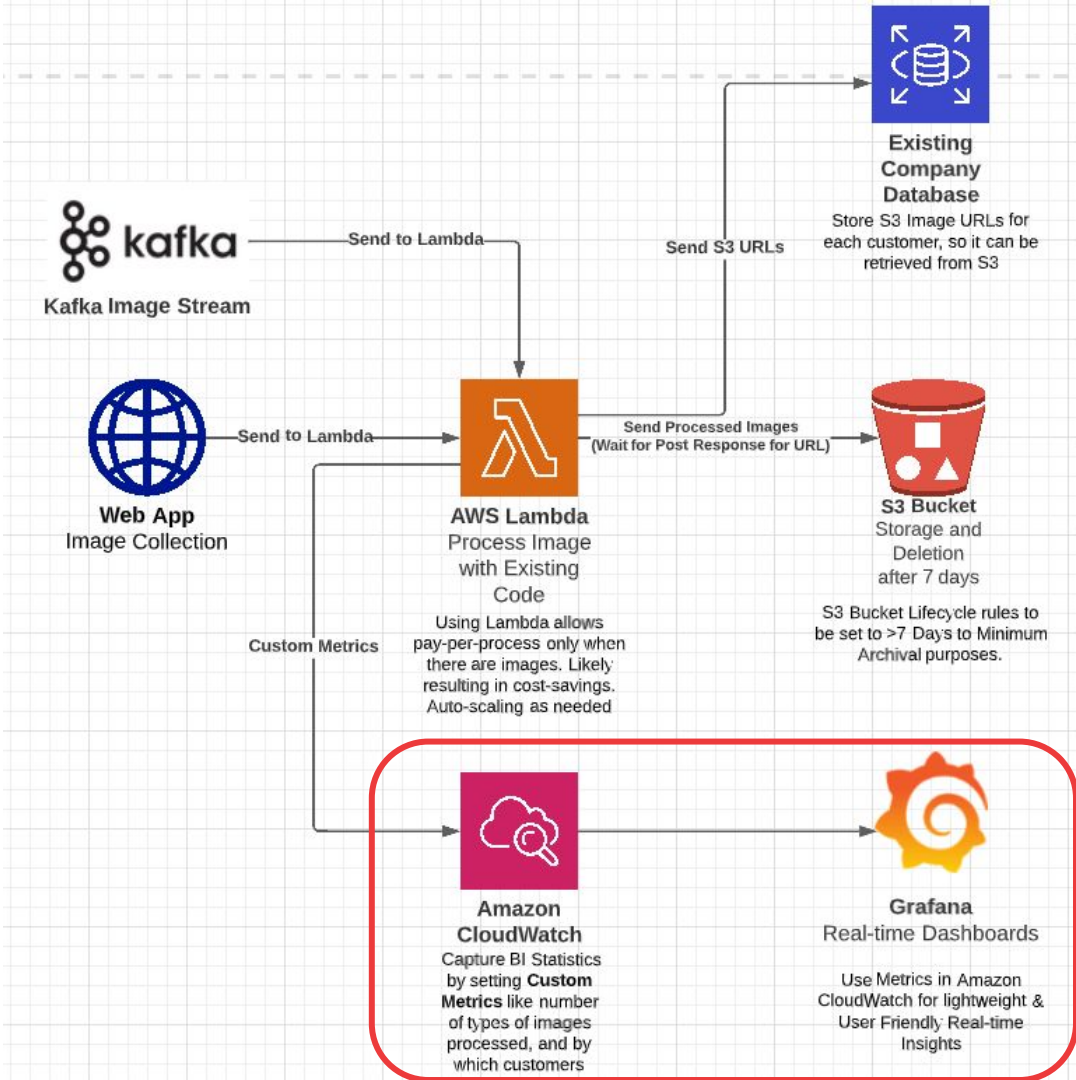
AWS Lambda will also send **custom metrics** to Amazon CloudWatch. Grafana then draws these metrics in real-time for more **liberal visualisations**.

Advantages:

- **No storage of metrics required**, no external monitoring system required.
- Grafana integration with CloudWatch provides seamless method of visualising metrics. (It's much better than what CloudWatch ships with and it's **free**)

Disadvantages:

- CloudWatch only retains metrics for a **limited time**. Metrics needs to be exported for historical purposes. (But this can be done easily by using Lambda to send metrics data to the database in addition to URLs.)



Thank You!