Dataflow CI/CD

Surjit Singh
Cloud Data Engineer, Google

BEAM
SUMMIT

September 4-5, 2024

Sunnyvale, CA. USA

Agenda

- Cloud Source Repositories & Cloud Build
- Classic vs Flex Dataflow Template
- Demo Flex Dataflow Template
- Patterns for Deploying Dataflow
- Additional Links
- Q&A

Cloud Source Repositories & Cloud Build

Cloud Source Repositories

Design, develop, and securely manage your code. Collaborate easily on a fully featured, scalable, and private Git repository. Alternatively, Gitlab, Bitbucket etc.. can be used.

Cloud Build

Cloud Build enables you to build software quickly across all languages. Get complete control over defining custom workflows for building, testing, and deploying your applications. Alternatively, Gitlab pipelines can be used.
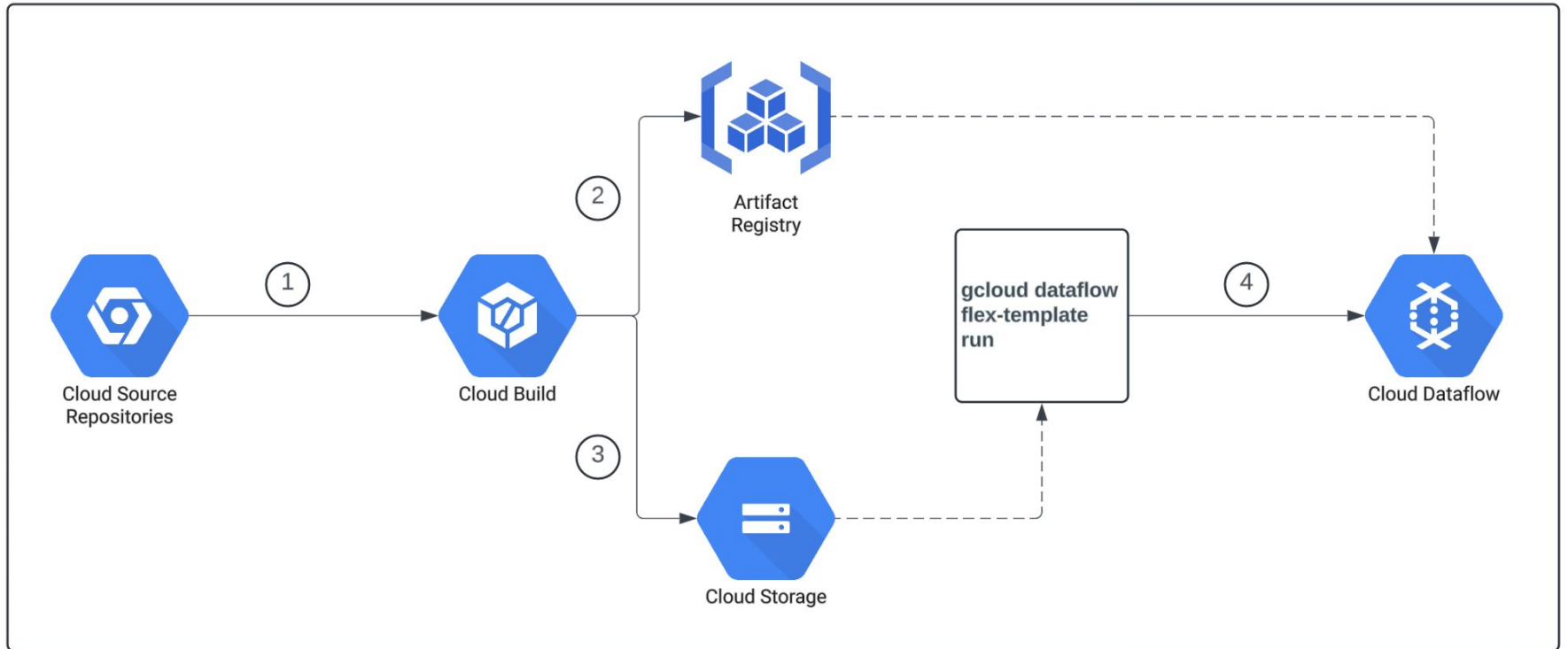
BEAM SUMMIT

Classic vs Flex Dataflow Template

Dataflow Templates: Dataflow templates allow you to package a Dataflow pipeline for deployment. Dataflow supports two types of template: Flex templates and Classic templates

Flex Templates: With a Flex template, the pipeline is packaged as a Docker image in Artifact Registry, along with a template specification file in Cloud Storage. The template specification contains a pointer to the Docker image. When you run the template, the Dataflow service starts a launcher VM, pulls the Docker image, and runs the pipeline. The execution graph is dynamically built based on runtime parameters provided by the user.

Classic Templates: A classic template contains the JSON serialization of a Dataflow job graph
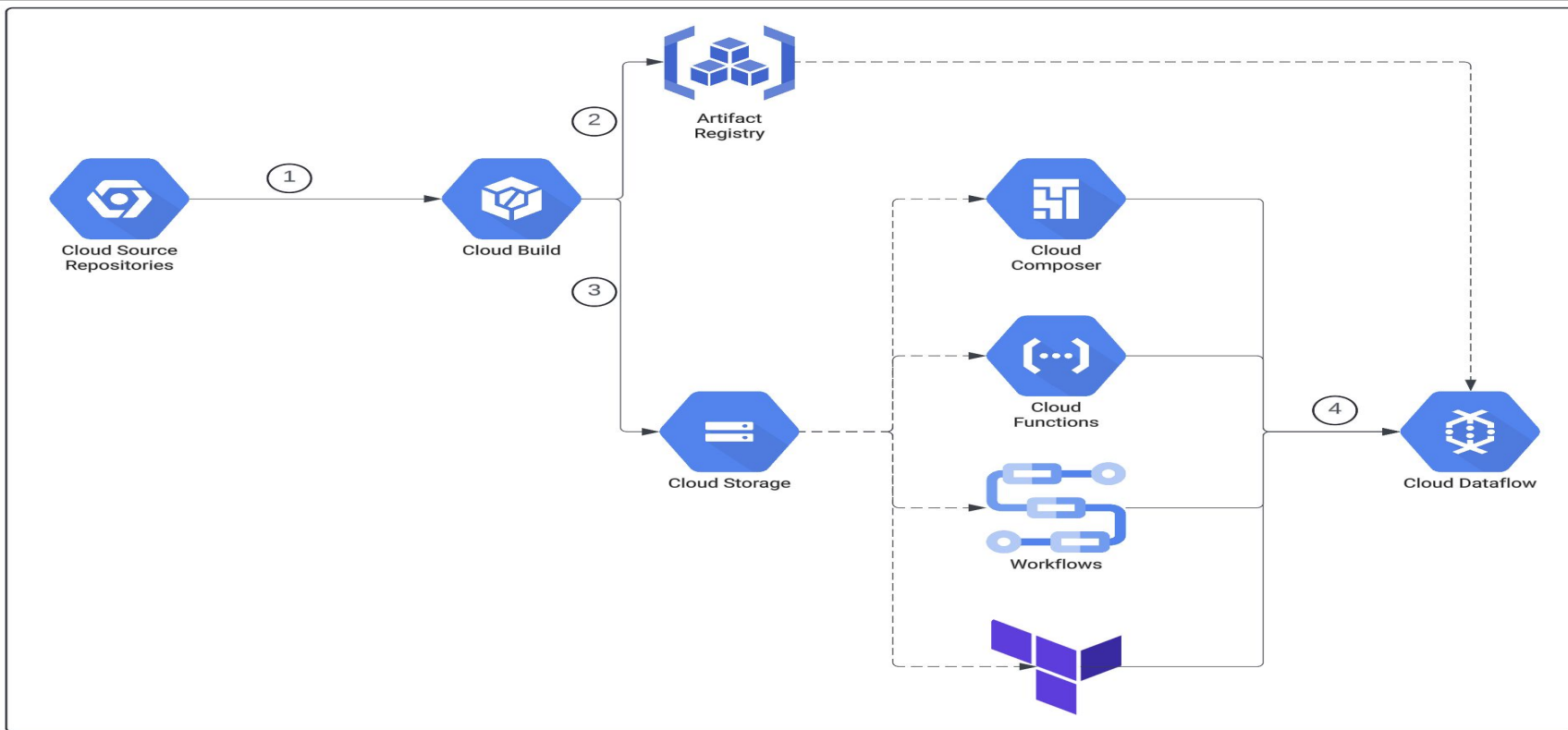
Demo Flex Dataflow Template

Demo

Patterns for Deploying Dataflow Template

Additional Links

- [Build and run a Flex Template](#)

- [Google-provided templates](#)

- [Why you should use Flex Templates](#)

Thank you!

Questions?

Connect on <u>LinkedIn</u>

https://www.linkedin.com/in/surjits254/