

# CS 327E Class 11

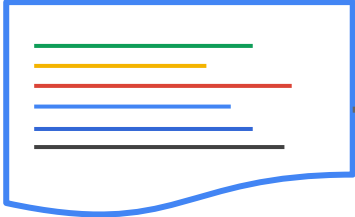
May 7, 2021

# Beam Concepts

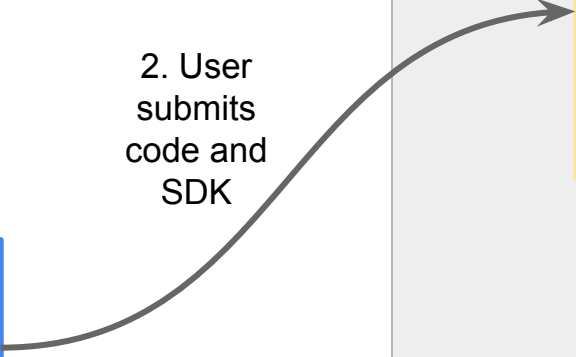
- SDK: the library that lets the user author a Beam pipeline
- Runner: the engine that executes the Beam pipeline

# Life of a Beam pipeline

1. Users writes pipeline code

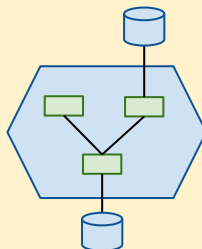


2. User submits code and SDK



## Dataflow Service

3. Runner validates and optimizes pipeline

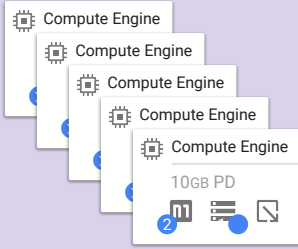


Execution graph

4. Runner deploys and schedules pipeline



5. Runner executes distributed pipeline on workers

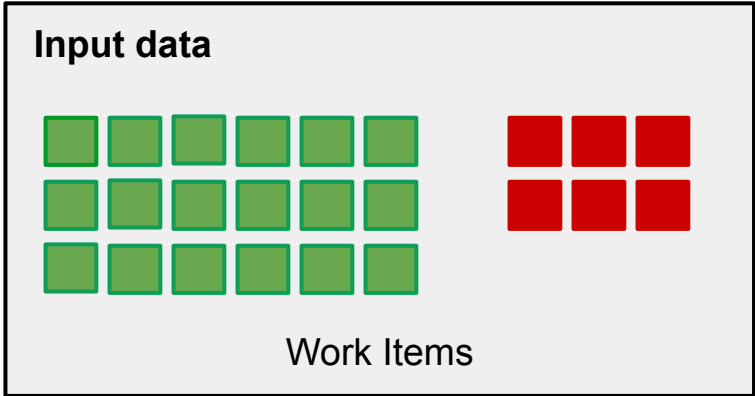
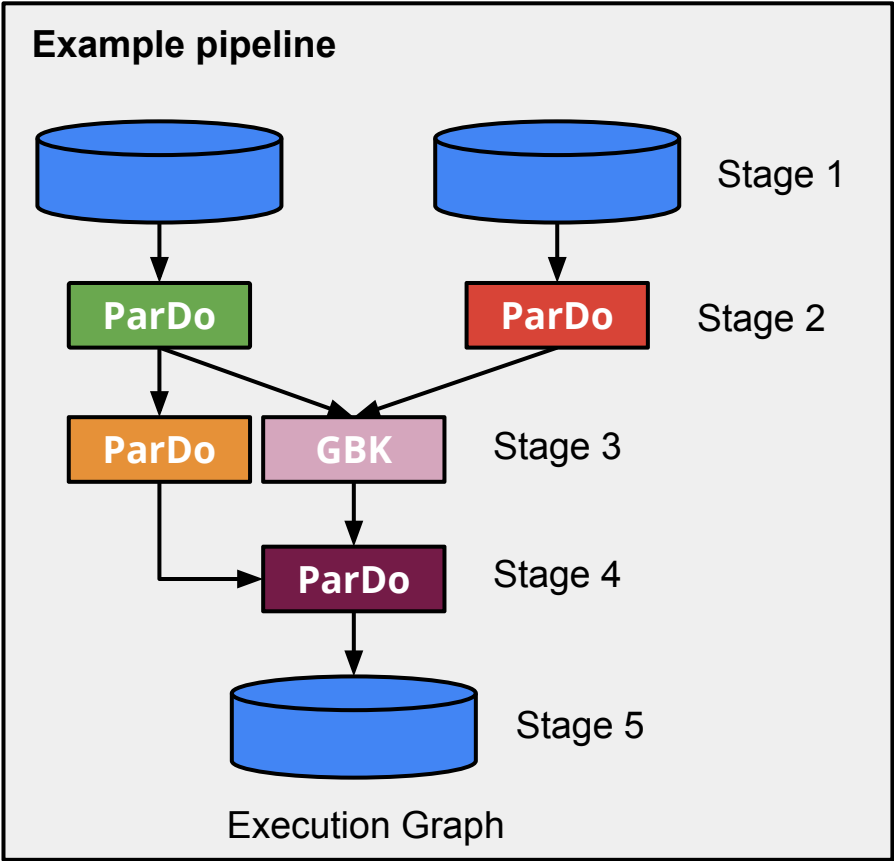


Workers

UI

logs

# Under the hood



# Code samples

- Clone [snippets](#) repo
- Step through [instructor\\_dataflow.py](#), [student\\_dataflow.py](#), and [class\\_dataflow.py](#)
- Run code from [beam\\_college.ipynb](#)

# Last but not least...

- [Milestone 4](#)
- [Final Presentation](#)

# Looking forward

- Find your focus: Data Engineering versus Data Science
- Continue your learning
- Work on side projects
- Get certified
- Don't let obstacles slow you down
- Say yes to things that come up
- Be collaborative
- Be open to ideas and arguments
- Don't be afraid to go your own way

Thank you! (it's up to you now)