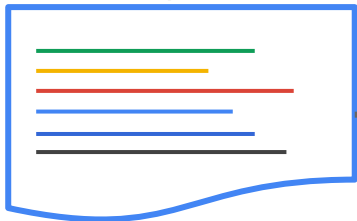


# CS 327E Class 11

Dec 4, 2020

# Life of a Beam pipeline

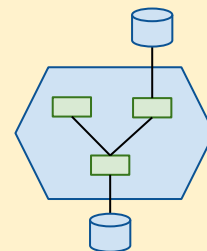
1. Write Pipeline Code



2. Submit code and SDK

## Dataflow Service

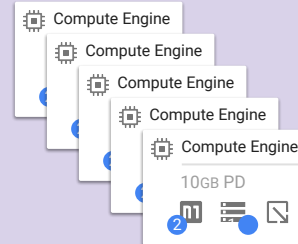
3. Validate and Optimize Pipeline



Execution graph

4. Deploy and Schedule

5. Dataflow Runner executes distributed pipeline on workers

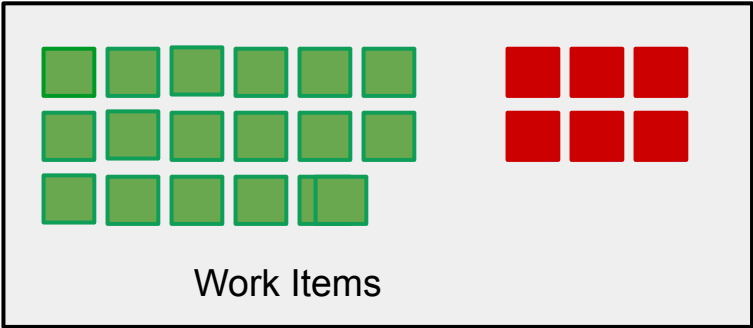
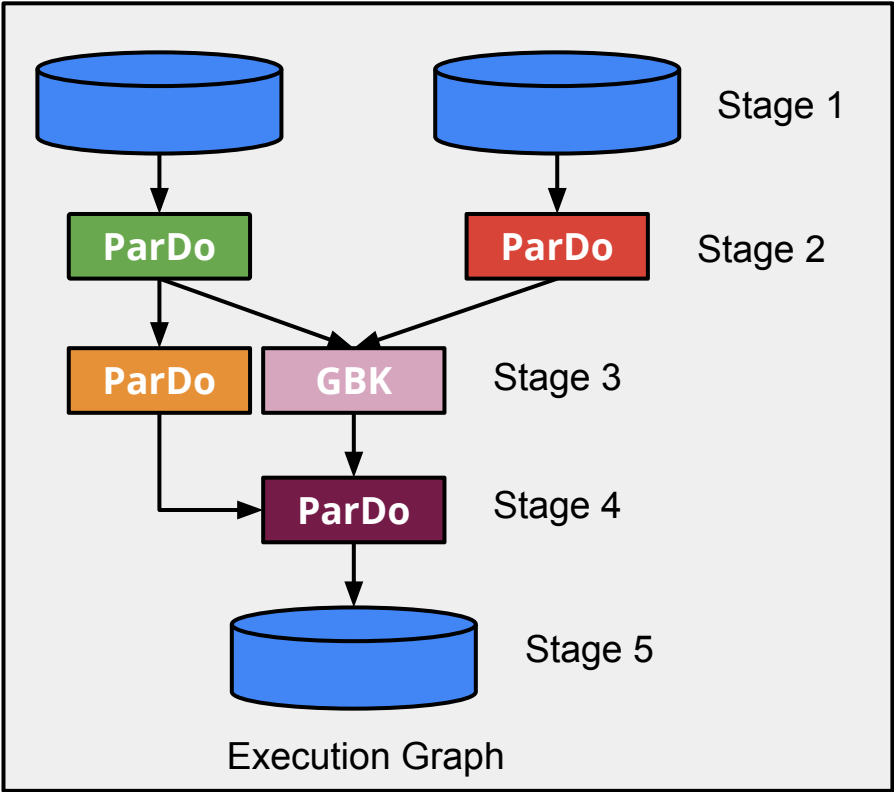


Workers

UI

logs

# Under the hood



# Code samples

[https://github.com/cs327e-fall2020/snippets/blob/master/teacher\\_dataflow.py](https://github.com/cs327e-fall2020/snippets/blob/master/teacher_dataflow.py)

[https://github.com/cs327e-fall2020/snippets/blob/master/student\\_dataflow.py](https://github.com/cs327e-fall2020/snippets/blob/master/student_dataflow.py)

[https://github.com/cs327e-fall2020/snippets/blob/master/class\\_dataflow.py](https://github.com/cs327e-fall2020/snippets/blob/master/class_dataflow.py)

# Last but not least...

<http://www.cs.utexas.edu/~scohen/projects/Milestone4.pdf>

<http://www.cs.utexas.edu/~scohen/projects/FinalPresentation.pdf>

# 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 is up to you now)