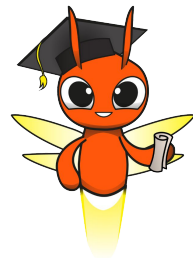




# Beam YAML: A new SDK for authoring pipelines

Svetak Sundhar  
Cloud Data Engineer, Google



# What is Beam YAML?

## 3 SDKs

- Java
- Python
- GoLang

But these are all verbose in the amount of code required.

# So...YAML?

- Yet another markup language
- Think of it as config files where you essentially set properties (ie Kubernetes)
- Much less verbose!

# What does it look like?

```
pipeline:
  transforms:
    - type: ReadFromCsv
      config:
        path: /path/to/input*.csv
    - type: WriteToJson
      config:
        path: /path/to/output.json
  input: ReadFromCsv
```



# Brief Demo on Beam SQL

---

# What about dependency management

- Beam YAML supports containerization
- Can still use a gcloud command to deploy!

# Complicated pipelines?

- Can write the complicated parts of your logic in Java and package it into a JAR
- Can call that JAR in your YAML code
- Reusability

# Future Directions?

- Testing frameworks for these
- More transforms supported natively!
- Performance benchmarking against other SDKs.



# Thank you!

Please reach out!

Linkedin: [www.linkedin.com/in/svetakvsundhar](https://www.linkedin.com/in/svetakvsundhar)

