



# Playing the Long Game

Transforming Ricardo's Data Infrastructure  
with Apache Beam

By Tobias Kaymak



BEAM  
SUMMIT

Austin, 2022





Search for products, sellers or product number



All categories



Antiques & artwork



Vehicles



Manual work & gardening



Household & home furnishings



Clothing & accessories



Hobbies & model building



Sports



Watches & jewellery

### Currently on Ricardo

Nearby

Ending soon

From CHF 1.-

Popular

New

### Our picks for you



Keentec Carbon Gabel

17 May 2022, 20:22

Starting bid **150.00**



Zeus FCM Carbon Rennvelo Gabel Neu

15 May 2022, 23:01

Starting bid **149.00**



Specialized S-Works Diverge Carbon

15 May 2022, 17:54

Buy now **3'990.00**



CARBON-RENNVELO-GABEL 1" PORTOFREI!

14 May 2022, 23:44

Buy now **149.95**



Carbon Gabel

15 May 2022, 17:46

Starting bid **420.00**

# Ricardo

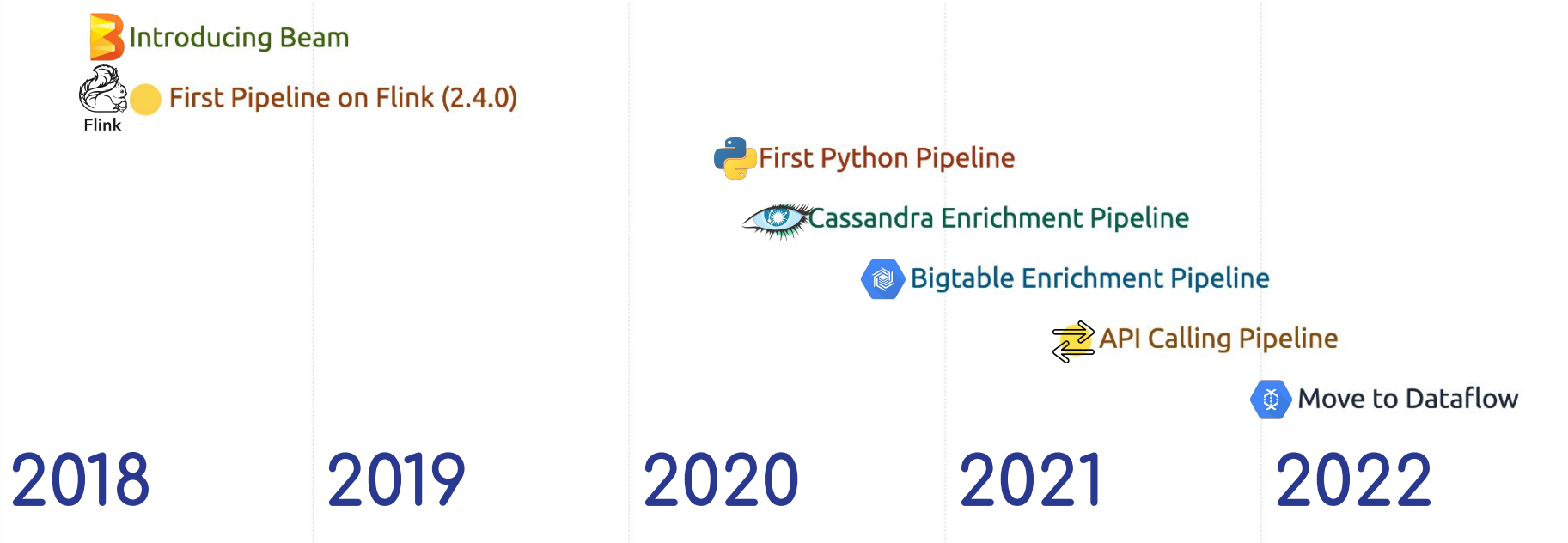
# Why Apache Beam?



Ricardo

# Will We Make It to the Promised Land?

# Timeline





# Why move?



# smg | swiss marketplace group





What is the most  
important ingredient  
of a successful  
migration?

# People



# So let's migrate!



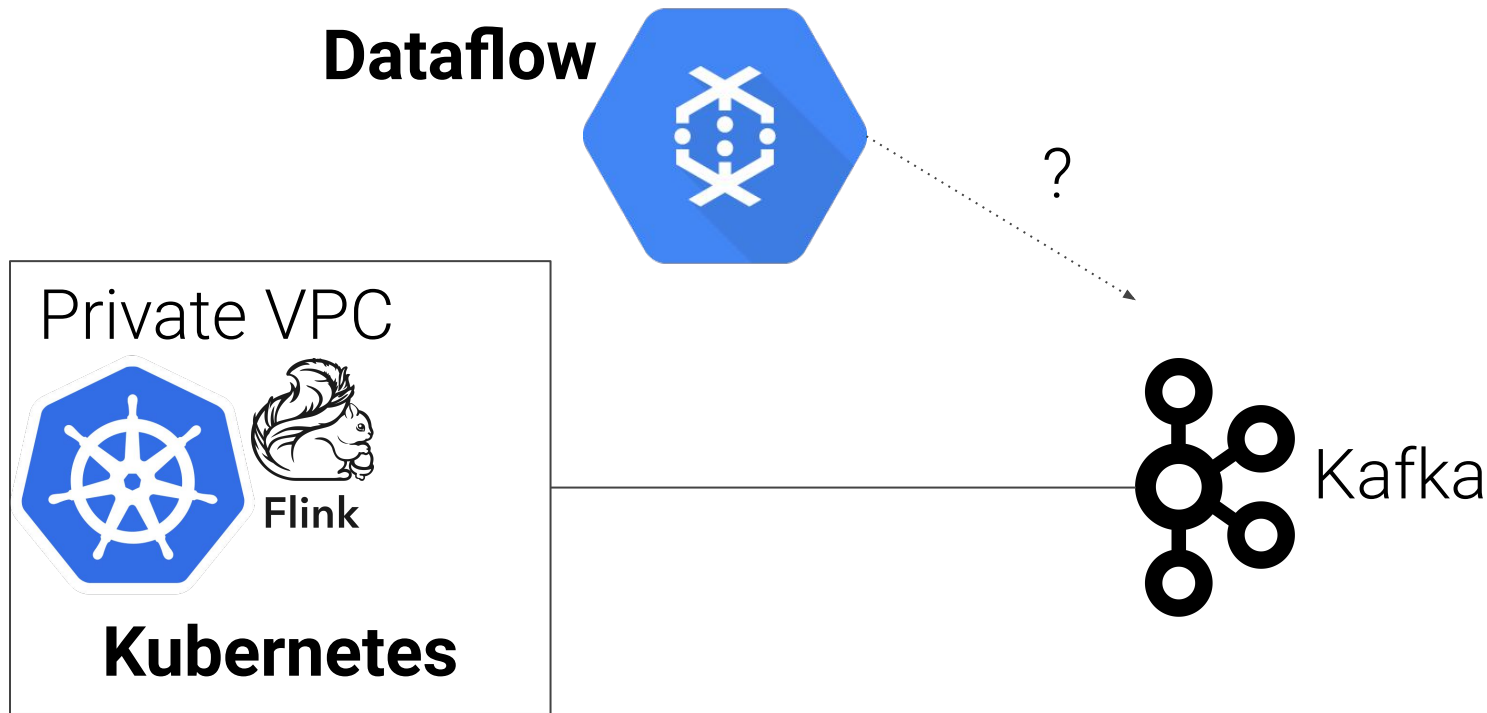
```
mvn package -Pflink-runner
```



```
mvn package -Pdataflow-runner
```

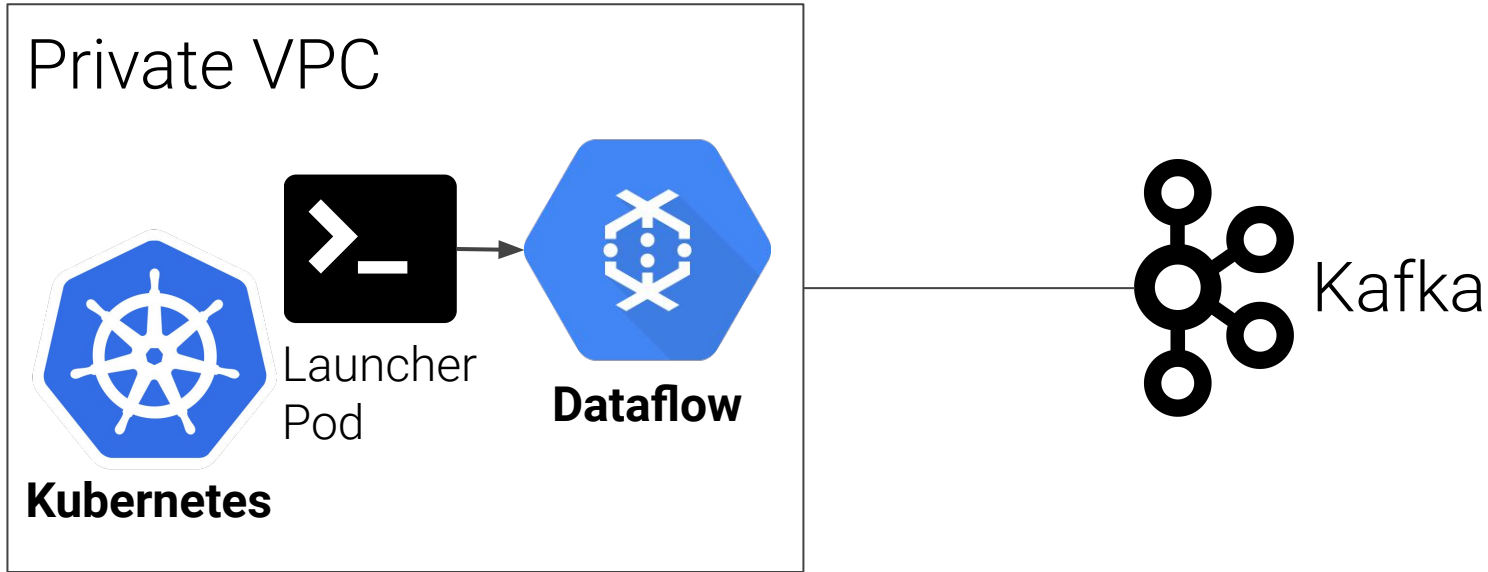


# Pitfall #1





# Solution #1



# Pitfall #2



UTF-8



**Kubernetes**



US-ASCII



**Dataflow**

# Solution #2



```
return JsonFormat.printer()  
    .omittingInsignificantWhitespace()  
    .preservingProtoFieldNames()  
    .print(message);
```

```
return new String(JsonFormat.printer()  
    .omittingInsignificantWhitespace()  
    .preservingProtoFieldNames()  
    .print(message).getBytes(StandardCharsets.UTF_8), StandardCharsets.UTF_8);
```

# Pitfall #3

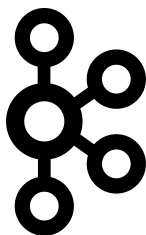


1:1

topic-pipeline



**Kubernetes**



Operations

many:1  
topic-pipeline



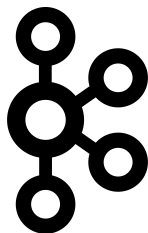
**Dataflow**



# Solution #3



Dataflow



Kafka

Operations

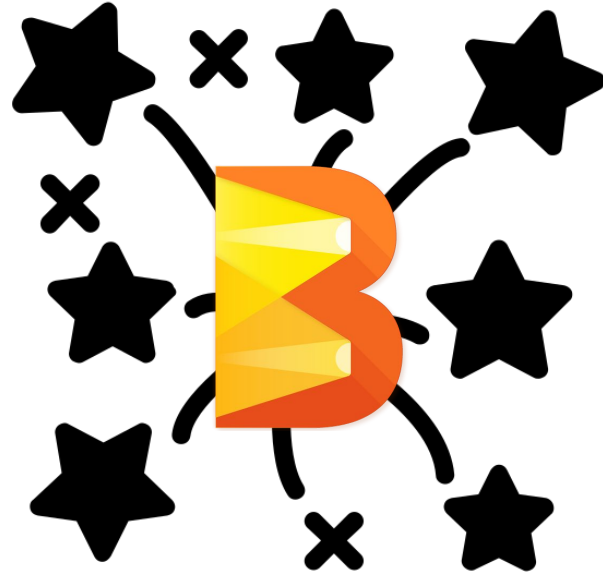
```
kafka-consumer-groups
  --group beam-consumers
  --reset-offsets
  --to-datetime 2022-07-19T00:00:00.000
  --all-topics
  --timeout 10000
  [--execute]
```

# Recap



**Dataflow**

# The future is bright





Thank you ❤️

# Questions?

Tobias Kaymak

tobias.kaymak@gmail.com  
tobtob.de



BEAM  
SUMMIT

Austin, 2022