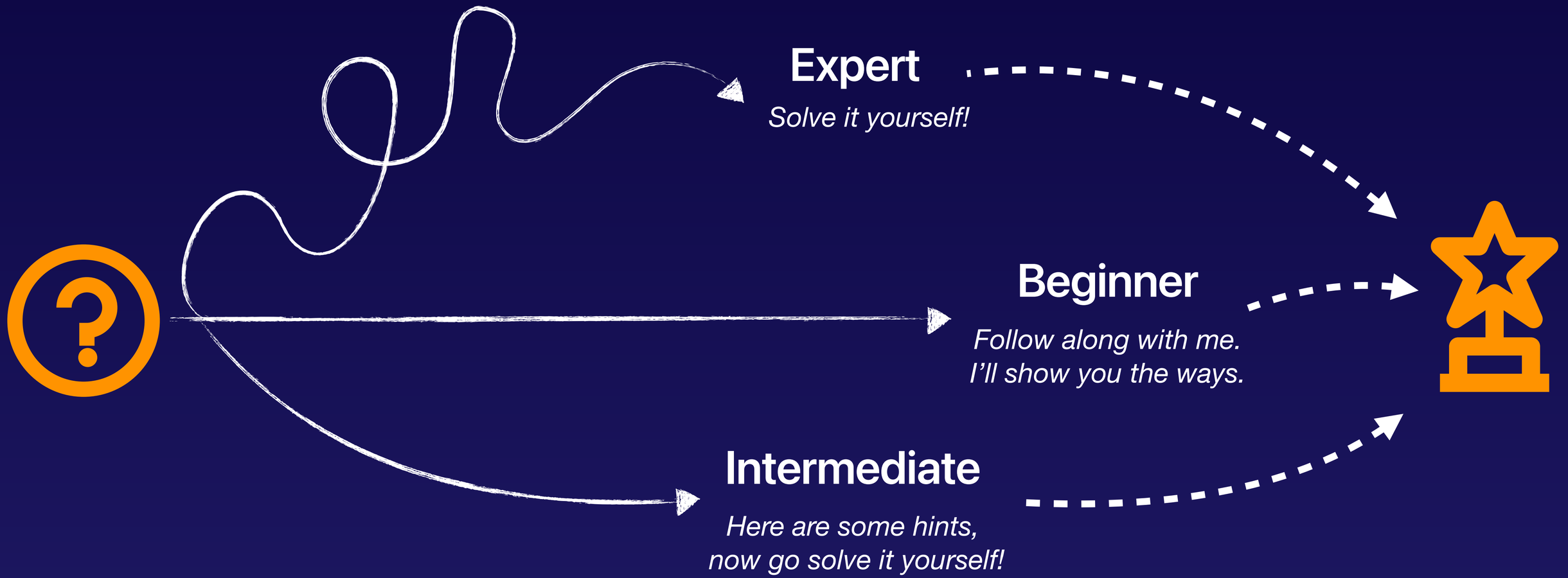
A CLOUD GURU

Streaming Data Collection Lab

Brock Tubre
INSTRUCTOR

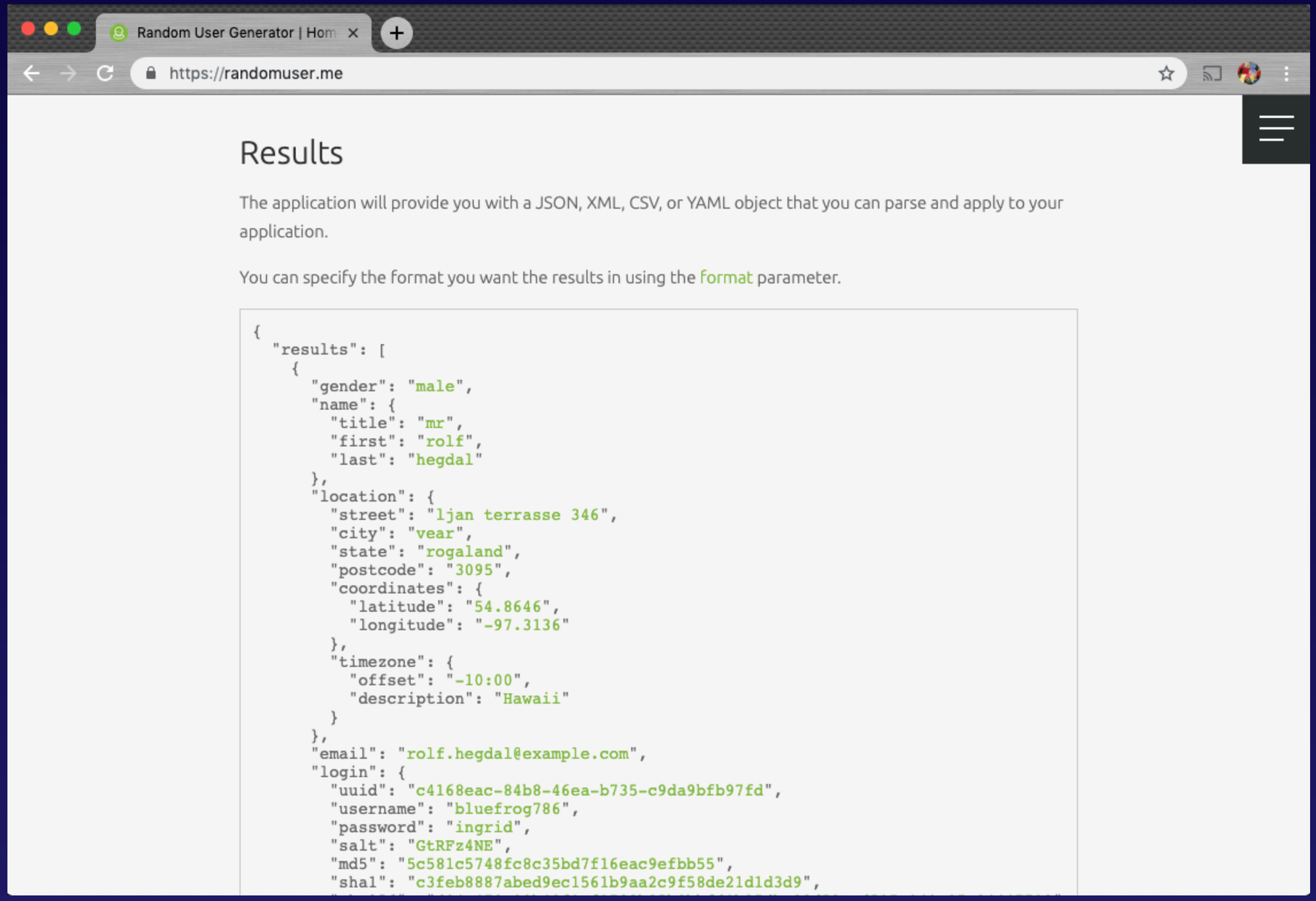
Choose Your Path



Use Case

You work for a company who has thousands of users interacting with the company's application. You have been tasked with capturing real-time data about the users for a marketing campaign. You need to capture information like name, age, gender, and location for users who are 21 and older.

Dummy User Data

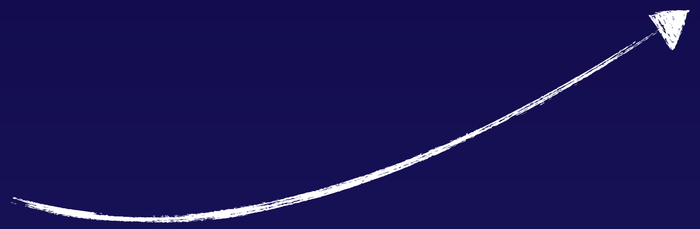


Final results?

- JSON files
- Stored in S3
- First Name
- Last Name
- Age (>= 21)
- Gender
- Latitude
- Longitude

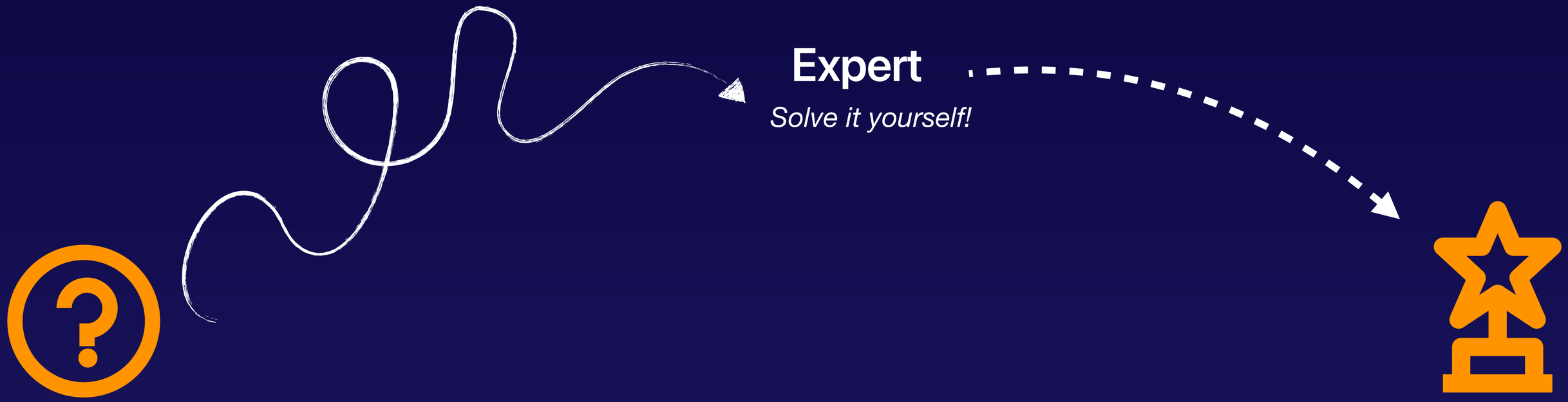
Final results?

- JSON files
- Stored in S3
- First Name
- Last Name
- Age (>= 21)
- Gender
- Latitude
- Longitude




```
{
   “FIRST”:"gladimir",
   "LAST":"silveira",
   "AGE":40,
   "GENDER":"male",
   "LATITUDE":-72.6634,
   "LONGITUDE":54.8243
}
{
   "FIRST":"aquino",
   "LAST":"peixoto",
   "AGE":44,
   "GENDER":"male",
   "LATITUDE":-71.4789,
   "LONGITUDE":-124.9399
}
{
   "FIRST":"elisete",
   "LAST":"ferreira",
   "AGE":42,
   "GENDER":"male",
   "LATITUDE":-72.8232,
   "LONGITUDE":131.9257
}
…
```

Good Luck!



Using PutRecord

Create a simple python program that simulates streaming data. Call the `PutRecord` to send data to Kinesis Streams for ingestion.

Using PutRecord


```
import requests
import boto3
import uuid
import time
import random
import json

client = boto3.client('kinesis', region_name='<INSERT_YOUR_REGION>')
partition_key = str(uuid.uuid4())

while True:
        r = requests.get('https://randomuser.me/api/?exc=login')
        data = json.dumps(r.json())
        client.put_record(
                StreamName='<INSERT_YOUR_STREAM_NAME>',
                Data=data,
                PartitionKey=partition_key)
        time.sleep(random.uniform(0, 1))
```

Using PutRecord


```
import requests
import boto3
import uuid
import time
import random
import json

client = boto3.client('kinesis', region_name='<INSERT_YOUR_REGION>')
partition_key = str(uuid.uuid4())

while True:
        r = requests.get('https://randomuser.me/api/?exc=login')
        data = json.dumps(r.json())
        client.put_record(
                StreamName='<INSERT_YOUR_STREAM_NAME>',
                Data=data,
                PartitionKey=partition_key)
        time.sleep(random.uniform(0, 1))
```

Using PutRecord

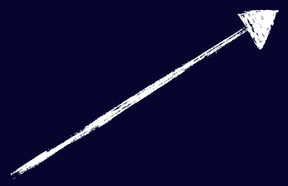

```
import requests
import boto3
import uuid
import time
import random
import json

client = boto3.client('kinesis', region_name='<INSERT_YOUR_REGION>')
partition_key = str(uuid.uuid4())

while True:
        r = requests.get('https://randomuser.me/api/?exc=login')
        data = json.dumps(r.json())
        client.put_record(
                StreamName='<INSERT_YOUR_STREAM_NAME>',
                Data=data,
                PartitionKey=partition_key)
        time.sleep(random.uniform(0, 1))
```

Using PutRecord


```
import requests
import boto3
import uuid
import time
import random
import json

client = boto3.client('kinesis', region_name='<INSERT_YOUR_REGION>')
partition_key = str(uuid.uuid4())

while True:
        r = requests.get('https://randomuser.me/api/?exc=login')
        data = json.dumps(r.json())
        client.put_record(
                StreamName='<INSERT_YOUR_STREAM_NAME>',
                Data=data,
                PartitionKey=partition_key)
        time.sleep(random.uniform(0, 1))
```

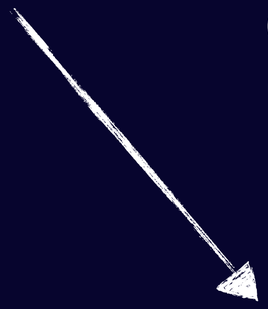
Using PutRecord


```
import requests
import boto3
import uuid
import time
import random
import json

client = boto3.client('kinesis', region_name='<INSERT_YOUR_REGION>')
partition_key = str(uuid.uuid4())

while True:
        r = requests.get('https://randomuser.me/api/?exc=login')
        data = json.dumps(r.json())
        client.put_record(
                StreamName='<INSERT_YOUR_STREAM_NAME>',
                Data=data,
                PartitionKey=partition_key)
        time.sleep(random.uniform(0, 1))
```



Using PutRecord


```
import requests
import boto3
import uuid
import time
import random
import json

client = boto3.client('kinesis', region_name='<INSERT_YOUR_REGION>')
partition_key = str(uuid.uuid4())

while True:
        r = requests.get('https://randomuser.me/api/?exc=login')
        data = json.dumps(r.json())
        client.put_record(
                StreamName='<INSERT_YOUR_STREAM_NAME>',
                Data=data,
                PartitionKey=partition_key)
        time.sleep(random.uniform(0, 1))
```

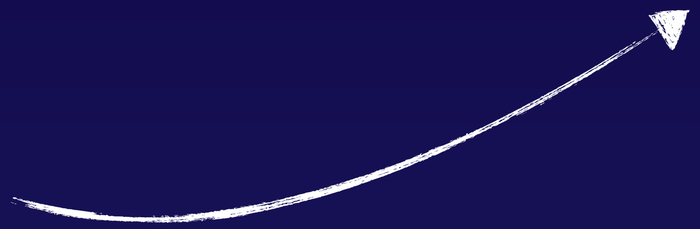


Transform and Load into S3

Find a way to transform the streaming data and output the results onto S3.

Final results?

- JSON files
- Stored in S3
- First Name
- Last Name
- Age (> 21)
- Gender
- Latitude
- Longitude




```
{
   “FIRST”:"gladimir",
   "LAST":"silveira",
   "AGE":40,
   "GENDER":"male",
   "LATITUDE":-72.6634,
   "LONGITUDE":54.8243
}
{
   "FIRST":"aquino",
   "LAST":"peixoto",
   "AGE":44,
   "GENDER":"male",
   "LATITUDE":-71.4789,
   "LONGITUDE":-124.9399
}
{
   "FIRST":"elisete",
   "LAST":"ferreira",
   "AGE":42,
   "GENDER":"male",
   "LATITUDE":-72.8232,
   "LONGITUDE":131.9257
}
…
```

Good Luck!

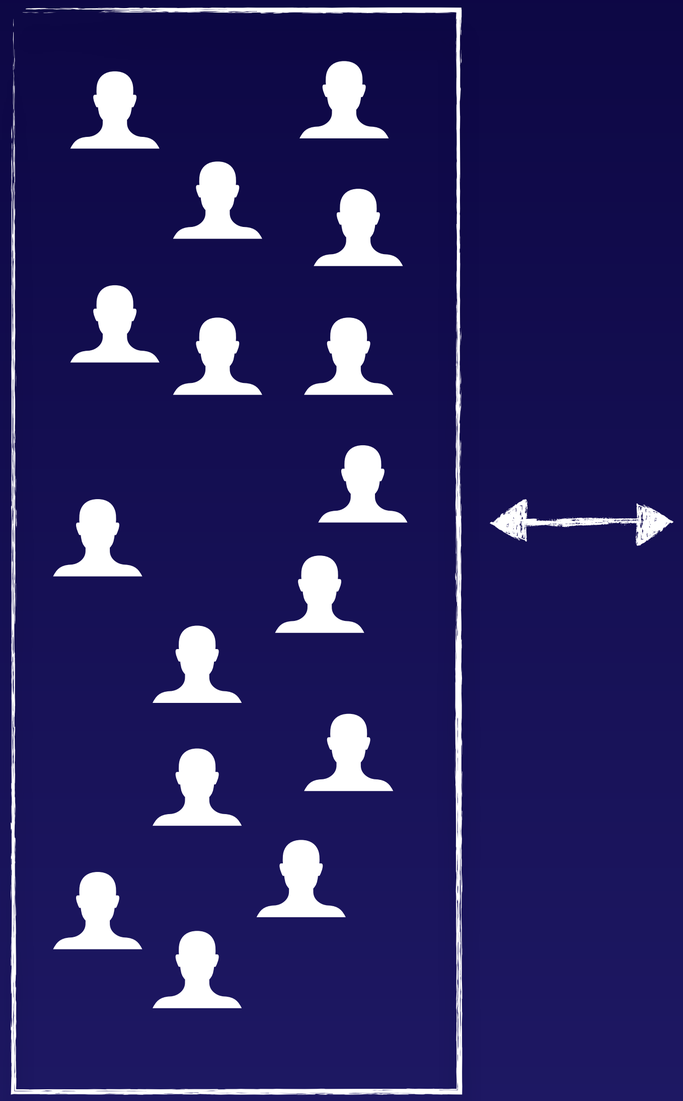
Intermediate

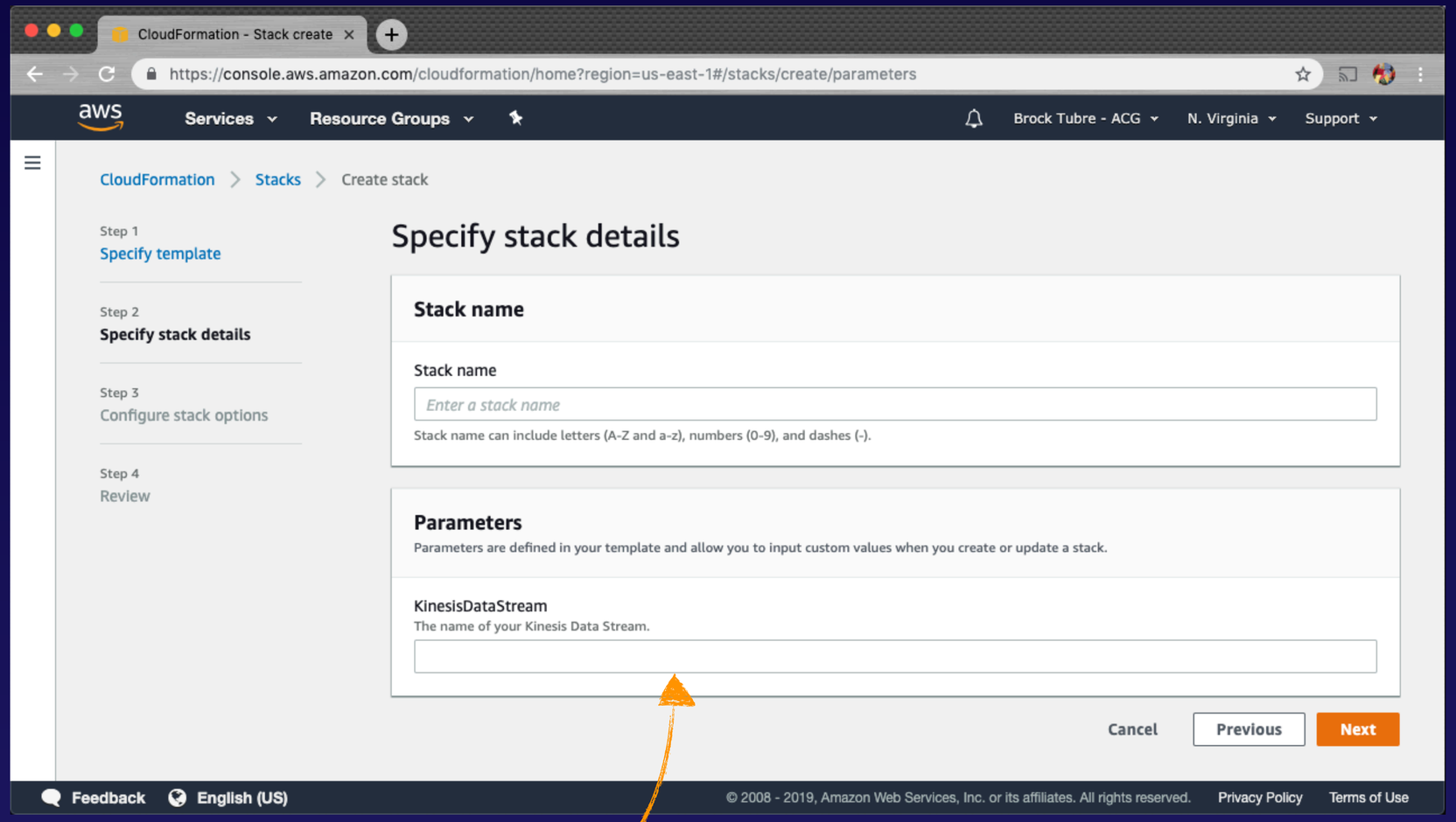
Here are some hints,
now go solve it yourself!

Streaming Data Collection Lab

A CLOUD GURU

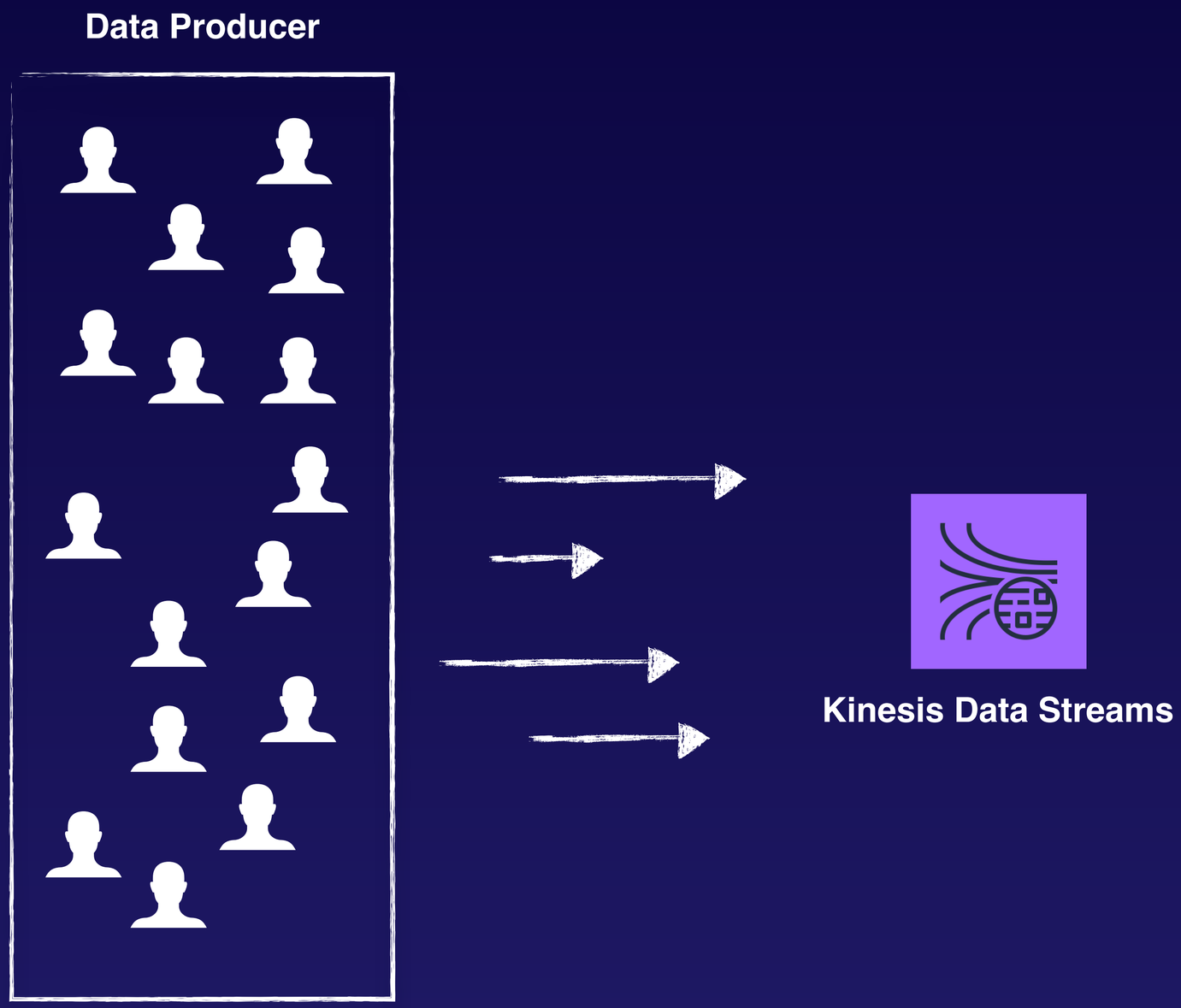
Data Producer





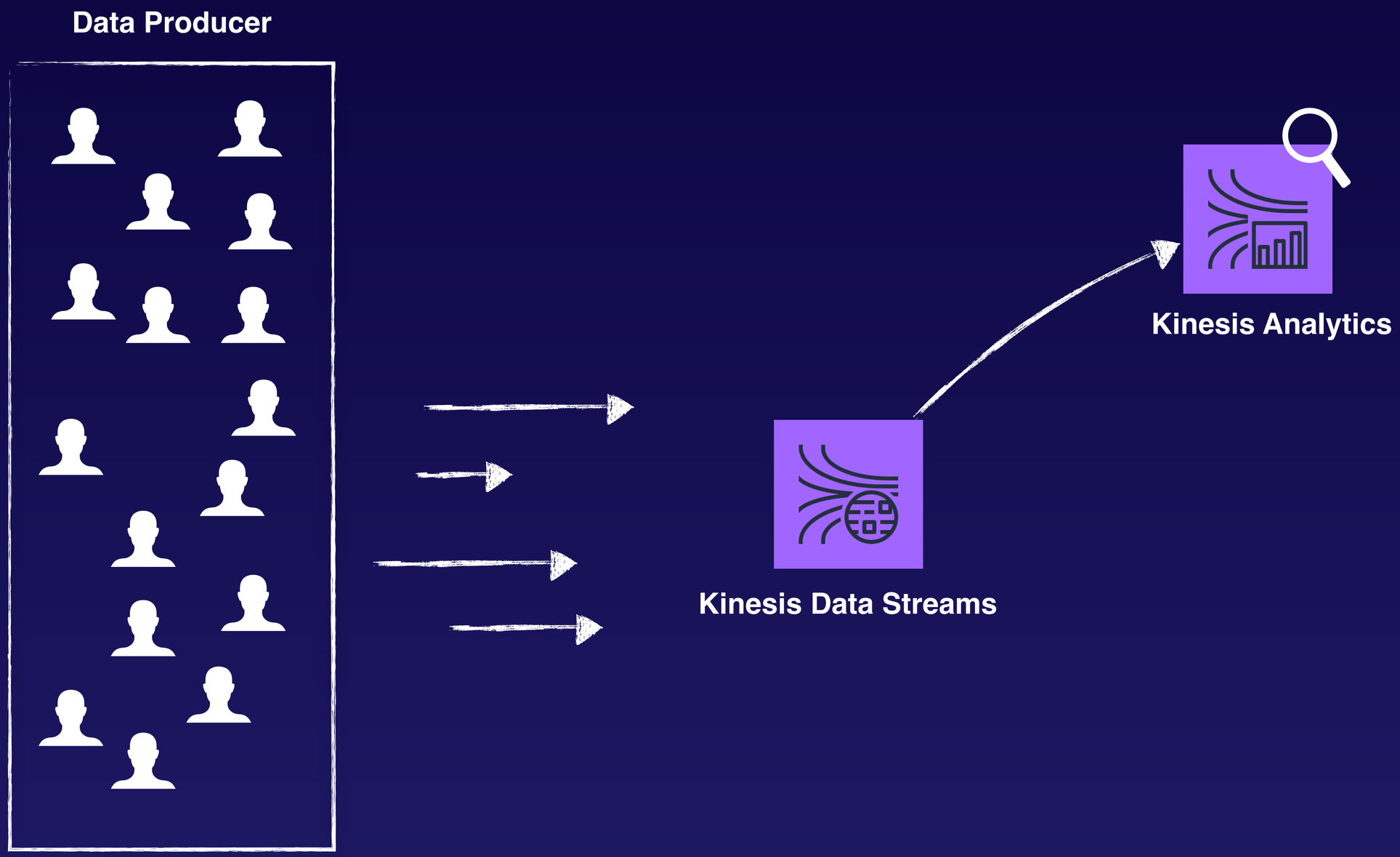
Insert your stream name here.

Streaming Data Collection Lab



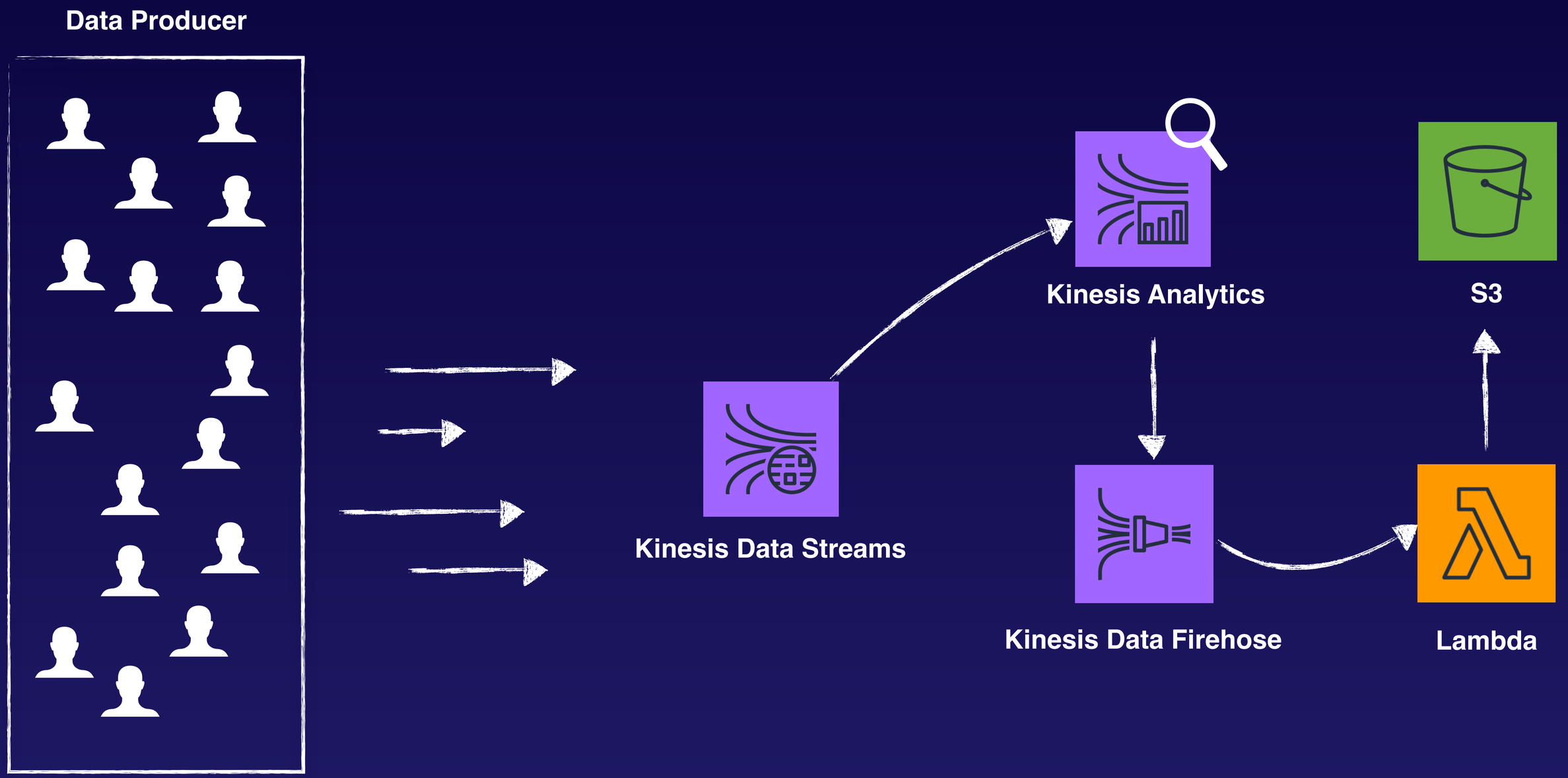
Streaming Data Collection Lab

A CLOUD GURU



Streaming Data Collection Lab

A CLOUD GURU



Good Luck!

Beginner

Follow along with me.
I'll show you the ways.