

End to End Testing

Data Pipeline

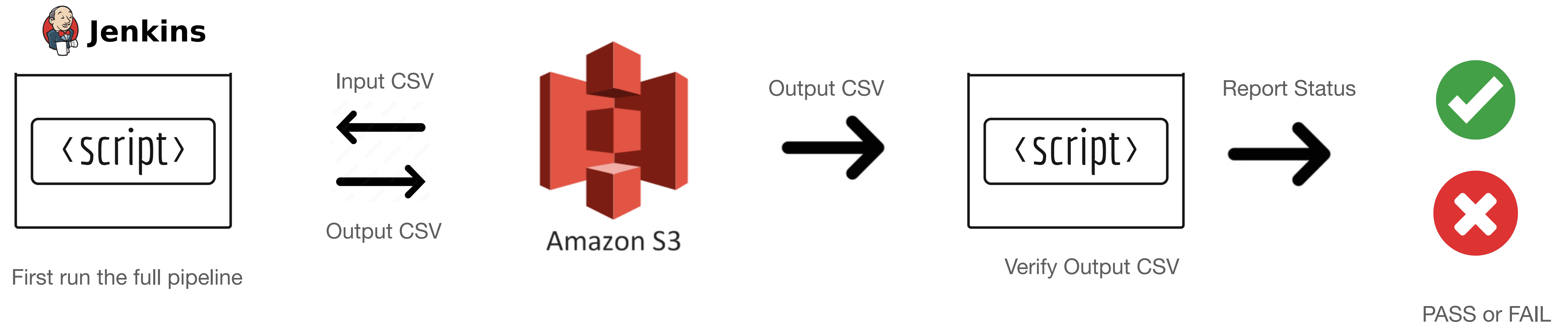
Shruti Priya

E2E Testing of the Pipeline

- We will run the whole pipeline with a sample dataset: sample csv
- A sample functionality of the same is implemented in `src/etl.py`. See following CLI:

```
PYTHONPATH='.' python src/etl.py --input-csv data/sample.csv --output-csv data/output.csv
```

- Here we can see that it takes input csv path as an input and saves the output csv.
- After we run the pipeline completely, we can verify the schema and values for the output csv to verify that the end2end test PASSED successfully.
- Also note that the sample dataset would be on cloud object store during e2e testing, hence we will supply s3 paths instead of local path



An example of the E2E Setup of the same on a CI Server