

Pipeline Architecture

CSSE6400

Brae Webb

February 28, 2022

Question

Can you name a *pipeline architecture*?

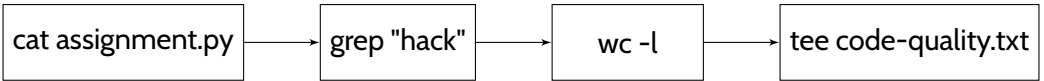
Question

Can you name a *pipeline architecture*?

Answer

How about *bash*?

```
1 >> cat assignment.py | grep "hack" | wc -l \  
2   | tee code-quality.txt
```







Filters

Modular software components



Filters

Modular software components

Pipes

The flow of data between filters

Producers

Source of data.

Producers

Source of data.

Transformers

Transform data.

Producers

Source of data.

Transformers

Transform data.

Testers

Filter data.

Producers

Source of data.

Testers

Filter data.

Transformers

Transform data.

Consumers

Target for results.

Exercise

Label the bash pipeline.



Producer



Producer

cat assignment.py

Tester

grep "hack"

wc -l

tee code-quality.txt

```
graph LR; A[cat assignment.py] --> B[grep "hack"]; B --> C[wc -l]; C --> D[tee code-quality.txt];
```






Question

Does this seem familiar?

Poll

Who has done *functional programming*?

```
1 let sum = reduce
2     (λ total value → total + value)
3     (map (λ seq → size seq) xs)
4     0
```

```
1 let sum = reduce + (map size xs) 0
```

Definition 1. *map*

1 $\text{map} : (\tau_1 \rightarrow \tau_2) \rightarrow \tau_1\text{Seq} \rightarrow \tau_2\text{Seq}$
2 $\text{map } f \text{ } xs$

Definition 2. *reduce*

1 **reduce** : $(\tau_1 \rightarrow \tau_1 \rightarrow \tau_1) \rightarrow \tau_1 \mathbf{Seq} \rightarrow \tau_1 \rightarrow \tau_1 \mathbf{Seq}$
2 **reduce** *f xs initial*

Question

What's the advantage of the map reduce pattern?

Question

What's the advantage of the map reduce pattern?

Answer

Parallelism [1]

Question

Using pipeline terminology, what filters do the *map* and *reduce* operators correspond to?

References

[1] Jeffrey Dean and Sanjay Ghemawat.

Mapreduce: Simplified data processing on large clusters.

In *OSDI'04: Sixth Symposium on Operating System Design and Implementation*, pages 137–150, San Francisco, CA, 2004.