CS 744: PYWREN

Shivaram Venkataraman
Fall 2021
Project checkins due Nov 30th
Poster presentation
  Dec 14th

Project grade breakdown
  Intro: 5%
  Mid-semester checkin: 5%
  Poster: 10%
  Final Report: 10%
NEW DATA, HARDWARE MODELS
Serverless Computing

Compute Accelerators

InfiniBand Networks

Non-Volatile Memory
SERVERLESS COMPUTING
MOTIVATION: USABILITY

What instance type?
What base image?
How many to spin up?
What price? Spot?
When to use the Cloud?

Data
- Large amounts of data. Can't store locally
- Shared data across users
- Long term storage

Compute
- Need lots of CPUs for short periods
- Varying compute needs (Yes)
- No admin overhead

Why is there no "cloud button"?
ABSTRACTION LEVEL?

Application

Compute Framework

Hardware

Logistic Regression

Spark

Amazon EC2
CloudLab
Private Cluster
...

Application

Compute Framework
STATELESS DATA PROCESSING

Function, Dependencies → Function Scheduler

Function Scheduler → Key Value Store (Low Latency)
Function Scheduler → Blob Store (High Bandwidth)

Key Value Store (Low Latency) → Containers
Blob Store (High Bandwidth) → Containers
“SERVERLESS” COMPUTING

300-900 seconds single-core

512 MB in /tmp

3-10GB RAM

Python, Java, node.js
```python
import pywren
import numpy as np

def addone(x):
    return x + 1

wrenexec = pywren.default_executor()
xlist = np.arange(10)
futures = wrenexec.map(addone, xlist)

print [f.result() for f in futures]
```

The output is as expected:

```
[1, 2, 3, 4, 5, 6, 7, 8, 9, 10]
```
PYWREN: HOW IT WORKS

future = runner.map(fn, data)

future.result()

your laptop  the cloud
future = runner.map(fn, data)

Serialize func and data
Put on S3
Invoke Lambda

pull job from s3
download anaconda runtime
python to run code
pickle result
stick in S3

result

poll S3
unpickle and return

your laptop

the cloud
What are the trade-offs?

<table>
<thead>
<tr>
<th>Storage Medium</th>
<th>Write Speed (MB/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSD on c3.8xlarge</td>
<td>208.73</td>
</tr>
<tr>
<td>SSD on i2.8xlarge</td>
<td>460.36</td>
</tr>
<tr>
<td>4 SSDs on i2.8xlarge</td>
<td>1768.04</td>
</tr>
<tr>
<td>S3</td>
<td>501.13</td>
</tr>
</tbody>
</table>
MAP AND REDUCE?

Input Data → Lambda 

Lambda → Reducer 

Reducer → Lambda 

Lambda → Output Data
PARAMETER SERVERS

Use lambdas to run “workers”

Parameter server as a service?
WHEN SHOULD WE USE SERVERLESS?

Yes!

Maybe not?
SUMMARY

Motivation: Usability of big data analytics
Approach: Language-integrated cloud computing

Features
- Breakdown computation into stateless functions
- Schedule on serverless containers
- Use external storage for state management

Open question on scheduling, overheads
DISCUSSION: MIDTERM 1
Happy Thanksgiving!

Post-Thanksgiving:

- Mid-semester project check-in, Nov 30th
- SplitFS, TPU papers
- Midterm 2, Dec 7th