

CIS 192: Python Programming 🐍

# Fun Stuff: Cloud Computing

Arun Kirubarajan

([kiruba@seas](mailto:kiruba@seas))

@kirubarajan

[kirubarajan.com](http://kirubarajan.com)

# Game Plan

1. What is **the cloud**?
2. Cloud architectures
3. Virtualization techniques
4. Python on the cloud
5. How to use Dokku

# What is the cloud?

**Answer:** some collection of accessible computers connected to the internet.

# Why do we care now?

**Answer:** lots more users. **LOTS** more users.

# Economies of Scale

**Definition:** competitive advantage of operating at scale (e.g. Amazon giving compute resource “leftovers”).

# Quick Maths:

Amazon is estimated to maintain over  
2,000,000 servers.

# Philosophy Time:

Is Amazon a consumer goods company or a **cloud** company?

# Virtualization:

Emulating a little computer's operation system that runs on a bigger computer's operating system.

# Containerization:

Wrapping applications in self-contained boxes. Stack boxes on top of each other and swap boxes accordingly.

Assume we have developed  
Bibly, a disruptive startup that  
is changing the world. **How  
do we deploy a scalable  
app?**

# Plan 1:

Port forward your router, run the app on your laptop and start accepting traffic through your internet connection.

## **Plan 2:**

Rent a computer in the cloud, copy/paste the source code to the instance and accept.

## **Plan 3:**

Use a Git repository to manage source code. Create a script SSH into the instance and pull the source code.

## **Plan 4:**

Containerize your application. Copy the image to the instance and run the application.

## **Plan 5:**

Set up a Git server on your instance. Push to the instance and write a script to containerize and run the received source code.

## **Plan 6:**

Rent multiple instances. Denote one instance as the master node and distribute traffic between the nodes.

# Python on the Cloud:

1. Development
2. Deployment
3. Data Analysis

# Introducing: Colab

- Free Jupyter Notebook environment
- Tesla K80 GPU access and lots of RAM
- Keras, PyTorch, Numpy already installed
- Supports all Google Drive features

# Introducing: Spark

- Large-scale data processing
- Support in Python to use multiple machines
- Open Sourced by the Apache Foundation
- Industry standard for data processing

# Introducing: Dokku

- Container orchestration platform
- Full emulation of Heroku build system
- Docker container support
- Domain name (DNS) support

**See you next week!**

