MSDS 5153 Worksheet—Week 1

What is the term for the doubling of microprocessors on a chip every 2 years?
- ☒ Moore's Law
- ☐ Fairchild's Theorem
- ☐ Newton's Second Law
- ☐ Lovelace's Law

Which of the following is NOT a cloud service provider?
- ☐ Amazon Web Service
- ☐ Kamatera Performance Cloud
- ☐ Rackspace
- ☒ Intelluron
- ☐ Microsoft Azure

What is AWS Sagemaker? Please describe.
AWS Sagemaker is a tool offered to programmers and data scientists to ease and accelerate the Machine Learning process. It is comprised of three independent components that can be used together or separately: Build, Train, Deploy. Sagemaker provides customized algorithms, pre-built frameworks, docker containers that can be used to bring in your own algorithms, and automatic model tuning. These can be managed from an integrated Jupyter notebook instance, and the final model can be deployed to a scalable and secure hosted environment. By utilizing all of these tools for a fee, you can decrease your ML training time and reduce overall project costs.

What is a cluster?
- ☒ Two or more servers grouped together to provide enhanced computational power
- ☐ A series of cloud technologies grouped together to create a pipeline
- ☐ Any technology that uses MapReduce

Activity: Complete the Cron Job tutorial at https://www.youtube.com/watch?v=QZJ1drMQz1A

What is the cost/hr of an m5.2xlarge EC2 instance from AWS? Add the link where you found the price.

The price of an m5.2xlarge EC2 instance from AWS for a linux machine is $0.384 per hour. The price for a windows machine is $0.752 per hour. Other OS prices are provided on the following website:
https://aws.amazon.com/ec2/pricing/on-demand/.

Using `virtualenv`, how do you activate a virtual environment named `venv` at the linux command line?
- ☐ `workon venv`
- ☐ `activate venv`
- ☒ `source venv/bin/activate`
- ☐ `sudo pip install venv`

What command lists all installed packages in a virtual environment?
- ☐ `list packages`
- ☒ `pip freeze`
- ☐ `find packages`
- ☐ `venv list packages`

==List 2 types of databases offered by _either_ AWS or Azure. What are the benefits and drawbacks of each?==

1. RDS – This relational database is offered by AWS, but has similar databases in Azure. There are many benefits to RDS but some include the ability to easily scale the data, robust encryption capabilities, and automatic software patching. It also is known as fast, secure, and inexpensive. Some cons include downtime required for patching and scaling operations and the prevention of root access to the node which would disallow the use of special database encryption software that may require root access.
2. Cosmos DB – This NoSQL database allows you to work with unstructured data very quickly and easily, allowing you to develop their relationships overtime and adapt the schema regularly. It offers guaranteed speed at any scale, fast app development, consistent availability and security, and a valuable fully managed system. NoSQL databases do have drawbacks such as query capabilities being limited and issues with portability; however, Cosmos DB does attempt to combat these using a query via an API in Mongo. Overall, Cosmos DB performs well for its specific workload with unstructured data, but relational databases are better for structured data.