Master Notes

Author: Unique Divine

Institute: Applied Technology Solutions, Inc. (ApTSi)

Date: April 12, 2020

ElegantLaTeX Program

Victory won't come to us unless we go to it.

Contents

Part I

Notes

Chapter 1 The Beginning (2020 Oct.)

1.1 Onboarding (Sep 30, Meeting # 1)

Largely, what we want to get started with: **devloping the kafka test case integration**

- Get access to github once NDA is in
- Our project management tool of choice is Trello: Kanban board

Kafka:

- tool used for streaming for high velocity flows
- used typically in datalake
- Integration will be from kafka to apache spark

Communication patterns:

There are multiple communication patterns that you can have. The patterns differ based on the **velocity** of interactions.

- Synchronous deals with rest services/APIs.

  > "Eventually, you will 100% be writing rest APIs on this project b/c our architecture's micro service based..."

  Rest services are synchronous by definition, which means there's a handshake. You send a request, somone sends back a response, but it's instantaneous. There's an immediate response that you need to send back.

- Another common pattern for communication is **queued communication**. Queued communication is an example of **asynchronous communication**.

  Q: Why is it called asynchronous?

  content

  callback: type of response

  - 2 types of asynchronous communication: fire & forget AND request-response driven (round track)

  message ID, correlation ID: how to know which message

  queued communication: Consumer, C, drops things off in a queue. There's a queue handler on the side of the consumer. When C sends something to the producer, P, this goes to the producer's queue handler.

  Cons: If one of the parties is slow, there can be blocking. The industry standard in response to this issue is "response codes".

- TCPIP based communication

end 10 minutes in

Part II

Log

Chapter 2 Onboarding

First Week 10/1- 10/7

Begin date: 10/2

- Refamiliarized myself with Shell scripting. Notes are described in my local computational science directory.
- Went over notes of objectives from yesterday's ApTSi meeting

Begin date: 10/3

The first thing I did was watch a few YouTube videos on 2x speed related to Apache Kafka. Through that, I found out that I'd need Java, Spark, and Zookeeper. I also found out that I'd have the easiest time using a Unix-like system such as WSL Ubuntu.

Install Java

Attempted to install Java development kit (JDK) from Oracle.

Installation was easier on Linux.

- At the shell, use `java --version` to see if it's already installed.
- It will prompt installation with `apt install openjdk-11-jre-headless` or something similar. After that installs, `java --version` should work.

Install Docker

From docs.docker.com/.../ubuntu/:

- To uninstall old versions:

```
sudo apt-get remove docker docker-engine docker.io containerd runc
```

- Before installing Docker Engine for the first time on a new host machine, you need to set up the Docker repository. Afterward, the Docker can be installed and updated form the repository.

  Update `apt` package index and install packaes to allow `apt` to use a repository over HTTPS:

```
sudo apt-get install \
apt-transport-https \
ca-certificates \
curl \
gnupg-agent \
software-properties-common
```

  Add Docker's official GPG key:

```
curl -fsSL https://download.docker.com/linux/ubuntu/gpg | sudo apt-key add
    -
```

To verify you have the key:

```
sudo apt-key fingerprint 0EBFCD88
```

Continue the rest of the installation guide until running the hello world script.

Attempted to install Kafka w/ quickstart and didn't have enough of the pre-requisite technologies.

Install Zookeeper

Begin date: 10/4

Begin date: 10/5

Bibliography

Goodfellow, I., Bengio, Y., Courville, A., and Bengio, Y. (2016). *Deep learning*, volume 1. MIT press Cambridge.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680.

Pinto, C., Gill, M., Heron, E. A., et al. (2019). Can artificial neural networks supplant the polygene risk score for risk prediction of complex disorders given very large sample sizes? *arXiv preprint arXiv:1911.08996*.

Raghu, M. and Schmidt, E. (2020). A survey of deep learning for scientific discovery. *arXiv preprint arXiv:2003.11755*. https://arxiv.org/pdf/2003.11755.pdf.

Singh, K., Sandhu, R. K., and Kumar, D. (2015). Comment volume prediction using neural networks and decision trees. In *IEEE UKSim-AMSS 17th International Conference on Computer Modelling and Simulation, UKSim2015 (UKSim2015)*, Cambridge, United Kingdom.

Wray, N. R. and Goddard, M. E. (2010). Multi-locus models of genetic risk of disease. *Genome Medicine*, 2(2):10.