Integrating Elasticsearch with Analytics Workflows
Advance Instructions

Stephanie Kirmer
ODSC West 2019

Before the session, users should:
1. **Download and install Docker**. This requires visiting hub.docker.com, creating a user account, and downloading the free personal use software that is correct for your operating system. You're welcome to do the accompanying tutorial but it is not necessary for this training.
2. **Clone the github repository**: https://github.com/skirmer/elastic_analytics
3. **If you use R**, download the libraries **elastic** and **uptasticsearch**. (CRAN)
4. **If you use Python**, download the libraries **elasticsearch-py** and **uptasticsearch**. (PYPI)
5. **Test that your system is correctly set up:**

To confirm that your installation is working, AND to check that Elasticsearch is downloaded to your system ahead of time, follow these steps (also found on the github readme).

> *After cloning this repo, if you have Docker installed and running, you can spin up your own Elasticsearch database on your local machine. Move into the top level of this repository after you clone it, and run the following at command line:*
> `./supporting_materials/setup_texas.sh 5.5`
>
> *You should see a screen similar to the following when this is operating correctly:*

```
skirmer:elastic_analytics skirmer$ ./supporting_materials/setup_texas.sh 5.5
collecting arguments...
major version: 5.5
Starting up Elasticsearch...
Unable to find image 'docker.elastic.co/elasticsearch/elasticsearch:5.5.3' locally
5.5.3: Pulling from elasticsearch/elasticsearch
364f9b7c969a: Pull complete
ddbb62cd8870: Pull complete
886fc8178f2e: Pull complete
c87ee85ce74a: Pull complete
04793f0e9f49: Pull complete
62df301e053d: Pull complete
2abe81925962: Pull complete
aec6453d8b64: Pull complete
ae99e04a1ae7: Pull complete
f0d9a3a098b9: Pull complete
Digest: sha256:0d449a58e8b98bf3fe49ec655c9a7ab1ec3ffe269a731abff15e88bd86423f1c
Status: Downloaded newer image for docker.elastic.co/elasticsearch/elasticsearch:5.5.3
98736e133e39e6bff224d8c52e44a197d0cdda600d0ad19271127a12ca718c5c
Elasticsearch v5.5 is now running on localhost:9200
Setting up local testing environment
```

```
{"acknowledged":true,"shards_acknowledged":true}{"_shards":{"total":14,"successful":7,"failed":0}}{"utexas":{"mappings"
{"_default_":{"properties":{"cellcount":{"type":"integer"},"ciptitle":{"type":"text","fields":{"raw":{"type":"keyword"}
},"degcip_4dig":{"type":"integer"},"deglevel":{"type":"text","fielddata":true},"deglevl_code":{"type":"integer"},"grad_
ohort":{"type":"double"},"grad_cohort_label":{"type":"text","fielddata":true},"institution_id":{"type":"integer"},"inst
tution_name":{"type":"text","fielddata":true},"p25_earnings":{"type":"integer"},"p50_earnings":{"type":"integer"},"p75_
arnings":{"type":"integer"},"system":{"type":"text","fielddata":true},"year_postgrad":{"type":"integer"}}}}}}
```

```
Your local environment is ready!
skirmer:elastic_analytics skirmer$
```

When the terminal indicates that the processes are complete, close down the database with this command:

`./supporting_materials/cleanup_local.sh`

If this has completed successfully, your screen should look similar to this.

```
skirmer:elastic_analytics skirmer$ ./supporting_materials/cleanup_local.sh
removing testing directory
killing running container
98736e133e39
done cleaning up test environment
```

There is NO need to run any other code ahead of time. Any questions before the session can be directed to me on Twitter @data_stephanie - I look forward to seeing you in October!