Fundamentals of Data Engineering

Week 04 - sync session

While we're getting started

- Review your Project Questions
- Get ready to share

Overview

- Go over Assignment 3 results
- PRs from the command line
- Docker compose, beginning

Assignment 3: Our Project Questions

- What is a trip?
- What are the most common trips?
- How does this differ based on trip type (commuter vs all)?
- What is a commuter trip?

Due Friday (PR)

Housekeeping

- Class flow

Between Class 3 & Class 4

Between Class 3 & Class 4

- async material in Week 3 syllabus

Between Class 3 & Class 4

- async material in Week 3 syllabus
- Readings in Week 3 syllabus

Between Class 3 & Class 4

- async material in Week 3 syllabus
- Readings in Week 3 syllabus
- Assignment 03

Class 4 (i.e., today)

Class 4 (i.e., today)

- Review Assignment 03, questions, where did you hit a wall?

Class 4 (i.e., today)

- Review Assignment 03, questions, where did you hit a wall?
- Some lectur-ish stuff/activities on Assignment 04/Week 4 async topic

Class 4 (i.e., today)

- Review Assignment 03, questions, where did you hit a wall?
- Some lectur-ish stuff/activities on Assignment 04/Week 4 async topic
- Review Query Project (spans Assignments 2-5)

Class 4 (i.e., today)

- Review Assignment 03, questions, where did you hit a wall?
- Some lectur-ish stuff/activities on Assignment 04/Week 4 async topic
- Review Query Project (spans Assignments 2-5)
- Preview, discussion, do contentprep activities for Assignment 04 (refining and answering your project questions)

Class 4 (i.e., today)

- Review Assignment 03, questions, where did you hit a wall?
- Some lectur-ish stuff/activities on Assignment 04/Week 4 async topic
- Review Query Project (spans Assignments 2-5)
- Preview, discussion, do contentprep activities for Assignment 04 (refining and answering your project questions)
- Final Assignment 03 due on Friday

Between Class 4 & Class 5

Between Class 4 & Class 5

- async material in Week 4 syllabus

Between Class 4 & Class 5

- async material in Week 4 syllabus
- Readings in Week 4 syllabus

Between Class 4 & Class 5

- async material in Week 4 syllabus
- Readings in Week 4 syllabus
- Assignment 04

Where are we in the Query Project?

- Assignment 4: Answer your Project Questions
- Assignment 5: Use Jupyter Notebook to do some visualizations and write up your reasoning for the recommendations you choose to make.

Creating a GitHub Pull-Request(PR)

- From the command-line

Clone a repo from GitHub

```
git clone https://github.com/mids-w205-martin-mims/assignment-02-htm
cd assignment-02-htmartin
```

Create a branch to work from

```
git branch my-cool-feature
```

Switch to that branch

```
git checkout my-cool-feature
```

Make changes to code

```
vi README.md
```

Commit those changes

```
git commit -m'updated README' README.md
```

Push those up to GitHub

```
git push origin my-cool-feature
```

Note

- If this is the first time you've pushed to the remote `my-cool-feature` branch, then this command will automatically *create* that branch in your github repo and then push your changes to it.

Pull Request from the GitHub Web-UI

- Select "New Pull-Request"

- Select branches so that you are "Requesting to merge changes from `my-cool-feature` branch *into* `master`."

- Select your instructor(s) to review.

- Submit

Docker: Where am I?

- We've worked with Docker
- We've explore cli ways to find out what's up with data
- How do we find our way around with Docker?

Run the regular container

```
docker run -it --rm -v ~/w205:/w205 midsw205/base bash
```

What conatiners are running right now?

- New terminal window
- `docker ps`

What conatiners exist?

- `docker ps -a`

Container name

- `fervent_austin` is my running `midsw205/base:latest` container

What images do I have?

- Images vs. containers
- `docker images`

Image name

- Need both repository & tag

- e.g., `midsw205/base:latest`

Clean up containers

docker rm -f

Idiomatic docker

- start the container -pwd in midsw205/base:latest

- exit container

- from prompt

```
docker run -it --rm -v ~/w205:/w205 midsw205/base pwd
```

Docker compose

- What is docker compose?

Update your course content repo in w205

`cd w205` `cd course-content` `git pull --all`

Docker compose .yml file

- `cd w205`
- `mkdir kafka`
- save `docker-compose.yml` from recently pulled `~/w205/course-content` to recently created `~/w205/kafka` directory

Docker compose spin things up

- `cd ~/w205/kafka`
- `docker-compose up -d`
- `docker-compose ps`

Summary

Extras