

Fundamentals of Data Engineering

Week 04 - sync session

datascience@berkeley

While we're getting started

- Review your Project Questions
- Get ready to share

Overview

- Go over Assignment 3
- Intro Docker
- PRs from the command line

Assignment 3 Breakout

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?

Housekeeping

- Class flow

Where are we in the Query Project?

- Assignment 3: Answer your Project Questions
- Assignment 4: Use Jupyter Notebook to do some visualizations and write up your reasoning for the recommendations you choose to make.

Docker: Where am I?

- Working with Docker
- Use it to 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 containers are running right now?

- New terminal window
- `docker ps`

What containers 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., `midsww205/base:latest`

Clean up containers

```
docker rm -f <name-of-container>
```

Idiomatic docker

- start a `midsw205/base:latest` container
- run `pwd`
- exit container

-VS-

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


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-htmartin
```

```
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

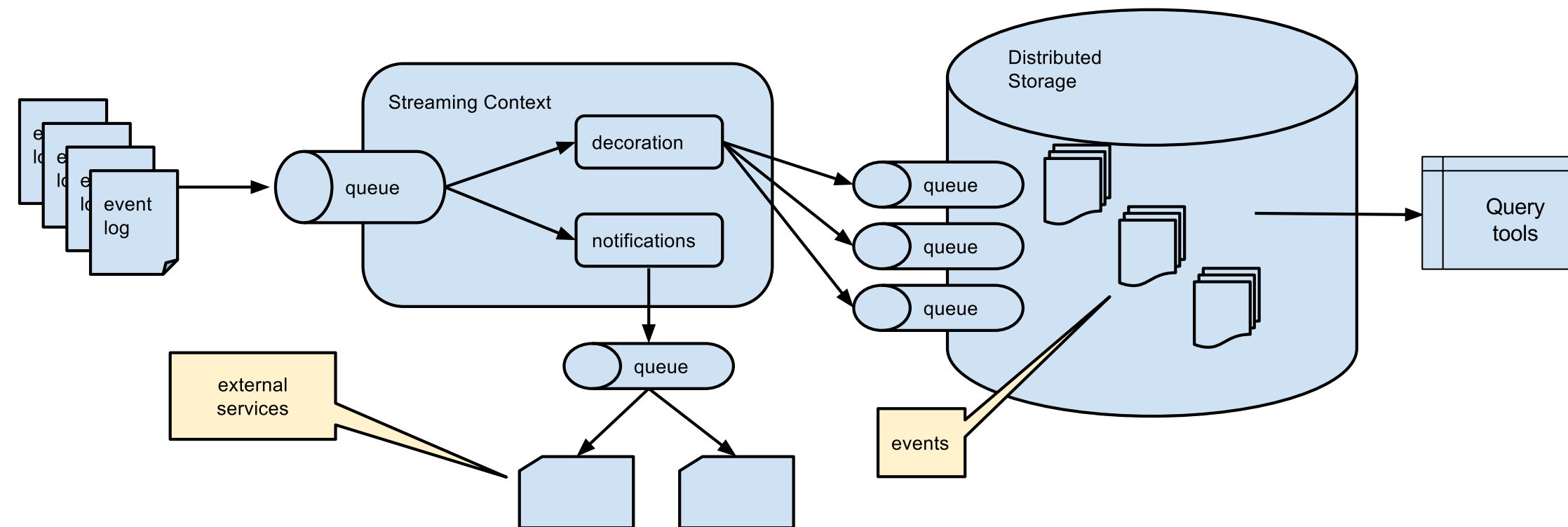
- Select “New Pull-Request”
- Select branches so that you are “Requesting to merge changes from `my-cool-feature` branch *into* `master`.”
- Select your instructor(s) to review.
- Submit

Summary

- git branching
- using docker

Week 4 videos

- Relational Data Stores
- NoSQL Data Stores
- Understand the RDB store we've been working with with big query
- Get ready to work with the nosql store redis next week as we start Project 2



Extras

Berkeley

SCHOOL OF
INFORMATION