

Final Project M4

CS 327E

November 20, 2017

Announcements:

- Today: Start on M4
- Next Week: Complete M4
- M5 Presentation Schedule: <https://tinyurl.com/yabzeyrm>

Monday 12/04		Tuesday 12/05		Wednesday 12/06		Thursday 12/07		Monday 12/11	
CLA 1.104		MEZ 1.216		CLA 1.104		MEZ 1.216		CLA 1.104	
Time	Group	Time	Group	Time	Group	Time	Group	Time	Group
6:30:00 PM	bug-free	6:30:00 PM	tejas	6:30:00 PM	327DB	6:30:00 PM	fallcs327lab	6:30:00 PM	xntj
6:40:00 PM	sudo-databases	6:40:00 PM	preql	6:40:00 PM	llamaschatila	6:40:00 PM	derry-andy-cs327e	6:40:00 PM	Databases
6:50:00 PM	Astros	6:50:00 PM	jetson	6:50:00 PM	codebusters	6:50:00 PM	snow-databases	6:50:00 PM	i-noe-you-heard-it
7:00:00 PM	estrerazhang	7:00:00 PM	327db	7:00:00 PM	trahim	7:00:00 PM	elmdatabase	7:00:00 PM	CS327E_Databases_BrandonDerek
7:10:00 PM	Koin	7:10:00 PM	econseniors	7:10:00 PM	alvinreagan327e	7:10:00 PM	csdb	7:10:00 PM	team-mangin_levine_cs327e
7:20:00 PM	dj	7:20:00 PM	AviMart	7:20:00 PM	512	7:20:00 PM	peacocks	7:20:00 PM	databases_natascha_kartik
7:30:00 PM	krish-vai	7:30:00 PM	lavendar	7:30:00 PM	327e-leo-and-nina	7:30:00 PM	schmidtgruber	7:30:00 PM	cs327e-databases
7:40:00 PM	jacogar	7:40:00 PM	dadabass	7:40:00 PM	snowfox	7:40:00 PM		7:40:00 PM	
7:50:00 PM		7:50:00 PM		7:50:00 PM		7:50:00 PM		7:50:00 PM	

Recap ETL Techniques:

- Postgres Copy command
- Postgres Copy + Create Table As Select ...
- Postgres Copy + Insert Into Table Select ...
- Spark + psycopg2

Recap ETL Techniques

- Select statements in SQL script
- Select statements + Postgres functions / UDFs in SQL script
- Spark RDDs and DataFrames

Common ETL Techniques

- FTP files
- Web API (REST or Streaming)
- Message queue
- Web scrape

Milestone 4

<http://www.cs.utexas.edu/~scohen/projects/m4-assignment.pdf>