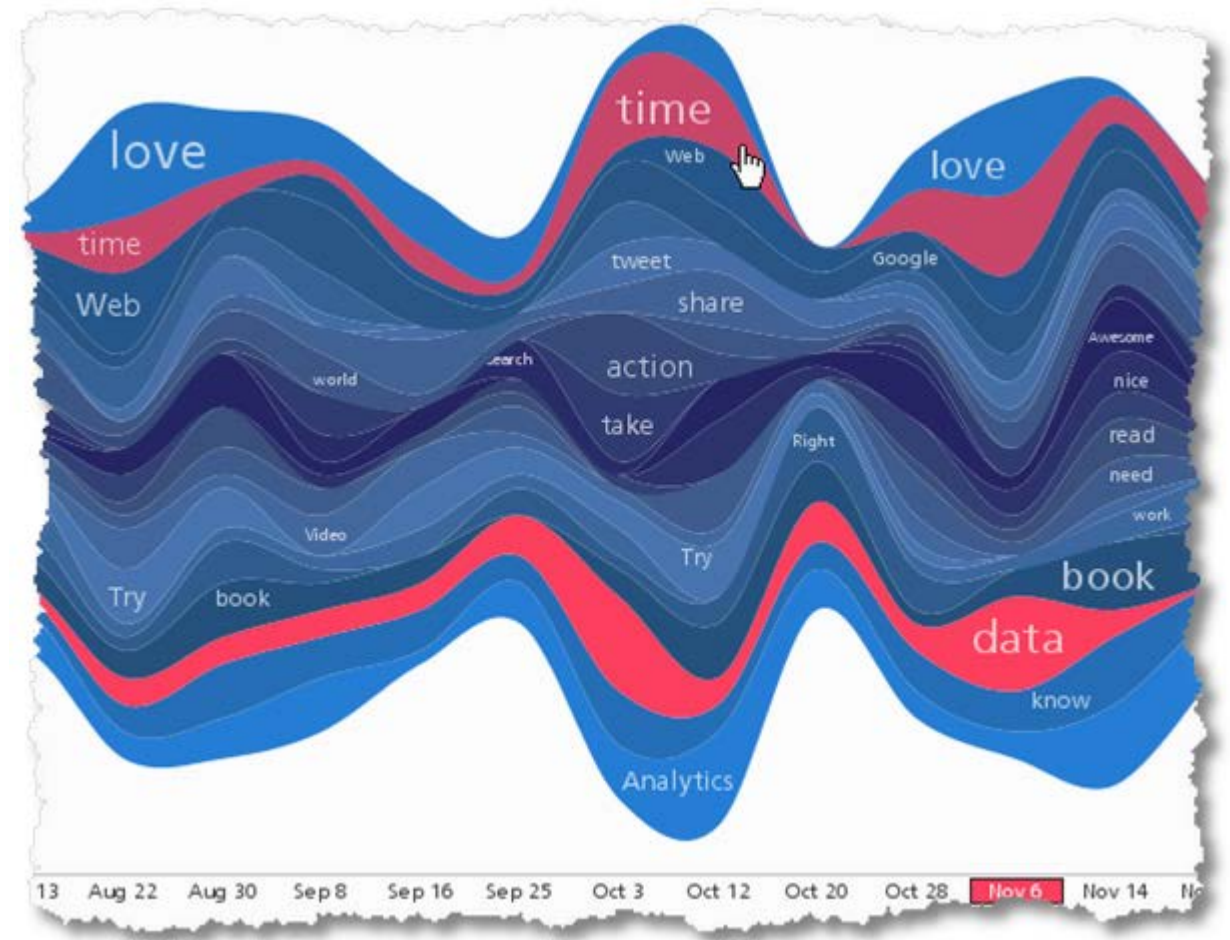# W205 Spring 2016 – Exercise #2
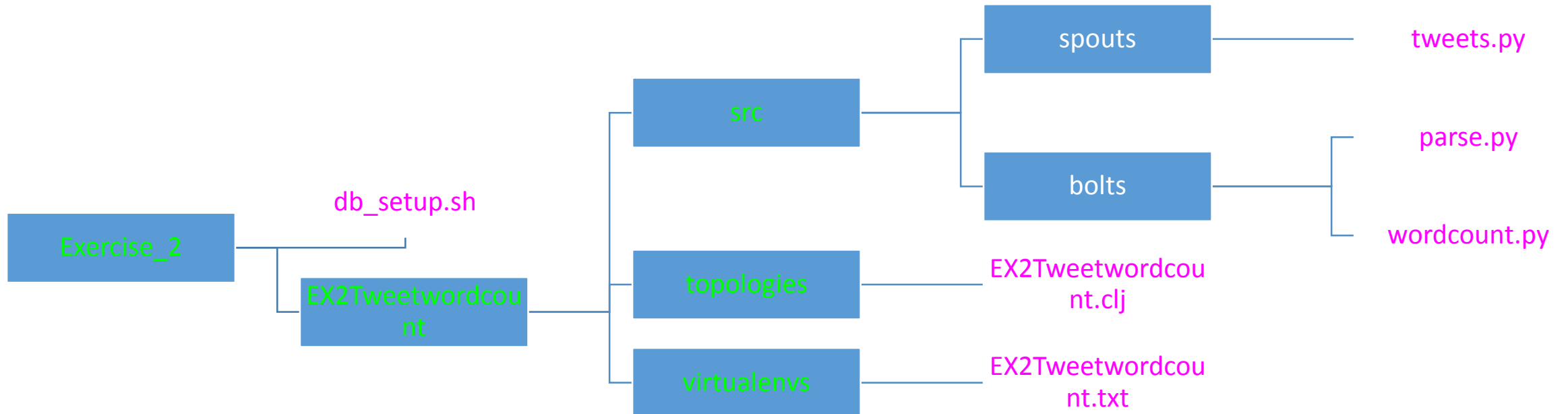
Beth Partridge

Application Idea

- Capture and analyze live Twitter data for a deeper understanding of social trends and demands using Apache Storm

- Use Tweepy library to read the live Twitter stream

- Parse the stream down to individual ascii words

- Store and update cumulative word counts in a Postgres database

Directory & File Structure

Architecture Block Diagram

Tweet-spout   Parse-tweet-bolt   Count-bolt
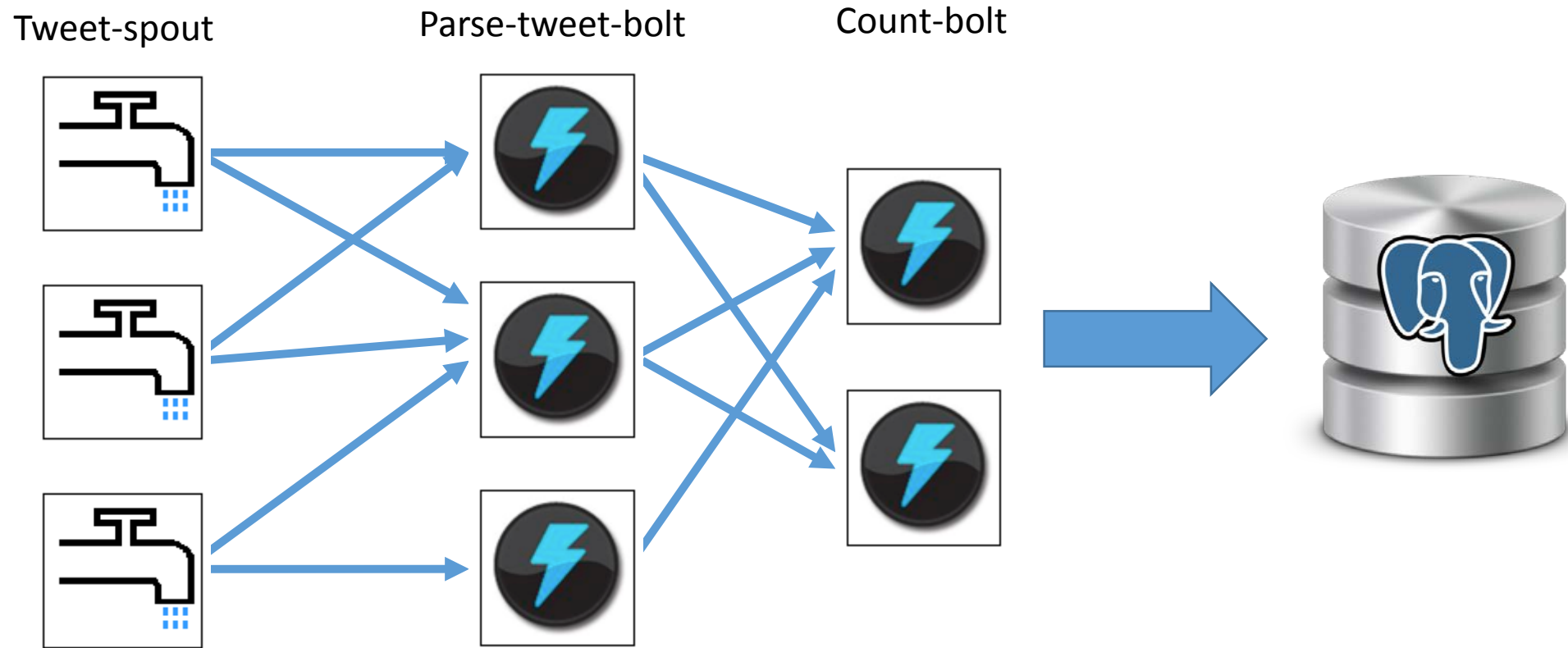
Assumptions & Dependencies

- Assumptions
  - Python 2.7
  - PostgreSQL 8.4.20

- Required packages:
  - Tweepy
  - Psycopg2
  - Streamparse
  - Matplotlib
  - Re

- Other
  - Twitter application and credentials

Instructions to Run Application

1. Launch AWS EC2 instance using UCB W205 Spring AMI

2. Attach a volume already prepared with assumptions listed above

3. Clone Exercise_2 repository from Github: git clone <repository URL>

4. Change to Exercise_2 directory: cd Exercise_2

5. Make the db_setup script executable: chmod a+x db_setup.sh

6. Execute the setup script: bash db_setup.sh

7. Change to the EX2Tweetwordcount directory: cd EX2Tweetwordcount

8. Run the application: sparse run

9. When you see a word count reach 10 stop the application with ^C

10. Run finalresults.py with no argument: python finalresults.py

11. Run finalresults.py with a common word argument: python finalresults.py the

12. Run histogram.py to see all the words with counts between 6-10: python histogram.py 6,10