

Projects

Projects will be performed in groups - of **at most 4 students per group**. Please organize yourselves into groups as appropriate. You need to pick a topic by **March 26**. There will be an interim review, and the final review **in finals week**. Please see the main page for deliverables on the project.

Suggested Topics for Projects

- 1. Implement all algorithms from Section 10 of the book as re-usable stream analytics.
- 2. Implement all algorithms from Section 11 of the book as re-usable stream analytics.
- 3. Implement the rule-based marketing platform from Section 4.5, Exercise 1.
- 4. Implement the query-based network monitoring application from Section 4.5, Exercise 2.
- 5. Implement a news fusion application (that finds correlations between two data sources)
- 6. Compare Streams, Spark and Storm (performance, ease of use, debugging, distributed processing, fault tolerance)
- 7. Build an application that processes real-world datasets (examples below)

Datasets

- [Physionet](http://www.physionet.org/): <http://www.physionet.org/>
- [NOAA](http://w1.weather.gov/xml/current_obs/): http://w1.weather.gov/xml/current_obs/
- [Stock Ticker Data Stream](#) or [Yahoo](#) or Google
- [Youtube](#) API
- [Twitter](#) API
- [UCI Machine Learning Datasets](#)
- RSS News Feeds (e.g. [BBC](#))
- [MTA Live Feeds](#)
- Other Public [Streaming Datasets](#)

Final Project Reviews: April 23 and April 30

Location: TBD

Slot	Group	Topic
3:00-3:30		
3:30-4:10		
4:10-4:40		
4:40-5:10		
5:10-5:40		
5:40-6:10		
6:10-6:40		
6:40-7:10		
7:10-7:40		
7:40-8:10		
8:10-8:40		

Final Project Report and Source Code (due May 4)

Please submit a 5 page report that summarizes your project, its goals, results, and a description of how it could be further improved.

Please also package up your source code, including visualization interfaces, makefiles, etc