

Higher level Abstractions (Scalding)

1. What is scalding
2. installation
3. get to work part 1 (Movies)!
4. get to work part 2 (Books)!
- 5 .. if you are really awesome!

What is Scalding

- Scala , Cascading ---> Scalding
- Twitter build and maintained
- Higher level view, not aware of the inner workings
(*this to the annoyance of others*)
- Search for 'cascading for the impatient and Driven'



Installation

- * first use: `cd $HOME !`
- Download from here:
`git clone https://github.com/BrianWolfkamp/scalding_intro.git`
- Execute (and get coffee / tea / other drinks or do something..):
`sh installation.sh`
- ...

Get to work part 1

- Run the movie part!
- `Scald.rd ..jar -input ratings.csv -out rated.tsv`
(*output already provided, but you might want to play around*)
- Extract this: only movies starting with an 'A'
- And this : movies with more than 5 other movies in a range of '0.6'
(*range of 'correlation (..) ' ')*)

Get to work part 2

- - now use the 'BX-CSV-DUMP' from the data map.
- Run with new input, look at output
- Extract this: Books that compare to other books only with an 'A'



If you are really awesome!

- Setup a flask webservice
(*hint google for 'flask'*)
- Setup a interface, if a user searches for this.. he also wants this..
- Example url:

