**NetFlix Recommender System**

The Netflix recommender system is a famous system that Netflix provided $1M in 2009 to support the challenge to improve the accuracy of prediction about how much someone is going to enjoy a movie based on their movie preferences.

The movie rating files contain over 100 million ratings from 480 thousand randomly-chosen, anonymous Netflix customers over 17 thousand movie titles. The data were collected between October, 1998 and December, 2005 and reflect the distribution of all ratings received during this period. The ratings are on a scale from 1 to 5 (integral) stars. To protect customer privacy, each customer id has been replaced with a randomly-assigned id. The date of each rating and the title and year of release for each movie id are also provided

For details information, visit below links:

<http://netflixprize.com/>

<https://gist.github.com/janisozaur/3192952>

<http://academictorrents.com/details/9b13183dc4d60676b773c9e2cd6de5e5542cee9a>

**What we will do:**

1. Setup a Hadoop system

2. Upload those training data to Hadoop HDFS

3. Apply some specified data mining algorithm to classify those data and analysis the similarity

4. After training, for any test input, will give a recommend move list or predict the customer rating.

**What we will deliver:**

1. A running Hadoop system

2. A training result set that applied specified algorithms

3. A way that can accept user’s input and give a full recommend list

**How we design:**

<TBD>

1,101,5.0

1,102,3.0

1,103,2.5

2,101,2.0

2,102,2.5

2,103,5.0

2,104,2.0

3,101,2.0

3,104,4.0

3,105,4.5

3,107,5.0

4,101,5.0

4,103,3.0

4,104,4.5

4,106,4.0

5,101,4.0

5,102,3.0

5,103,2.0

5,104,4.0

5,105,3.5

5,106,4.0

（101,101）（101,102）（101,103）

（102,101）（102,102）（102,103）

（103,101）（103,102）（103,103）

[101][102][103][104][105][106] [107]

[101] 5 3 4 4 2 2 1

[102] 3 3 3 2 1 1 0

[103] 4 3 4 3 1 2 0

[104] 4 2 3 4 2 2 1

[105] 2 1 1 2 2 1 1

[106] 2 1 2 2 1 2 0

[107] 1 0 0 1 1 0 1

