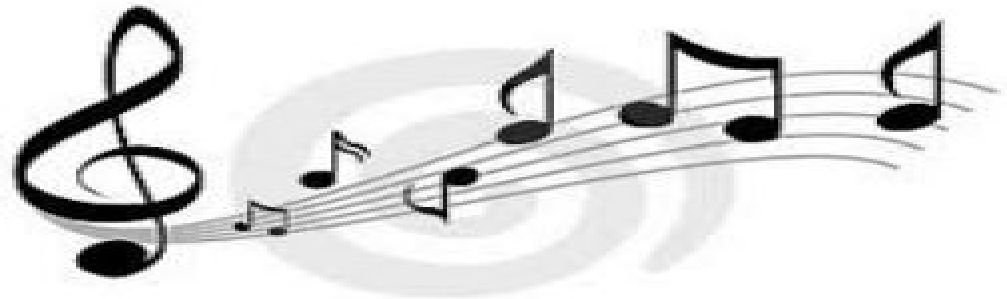


Music Counsel

Because We Know Your Choice!!!

Team Members:

Aditi Goel (800966155)

Nikita Kunchanwar (800962459)

Priya Kumari (800964015)

Objective

Motivation-

A **recommender system** or a **recommendation system** is a subclass of information filtering **system** that seeks to predict the "rating" or "preference" that a user would give to an item. (Source: Wikipedia)

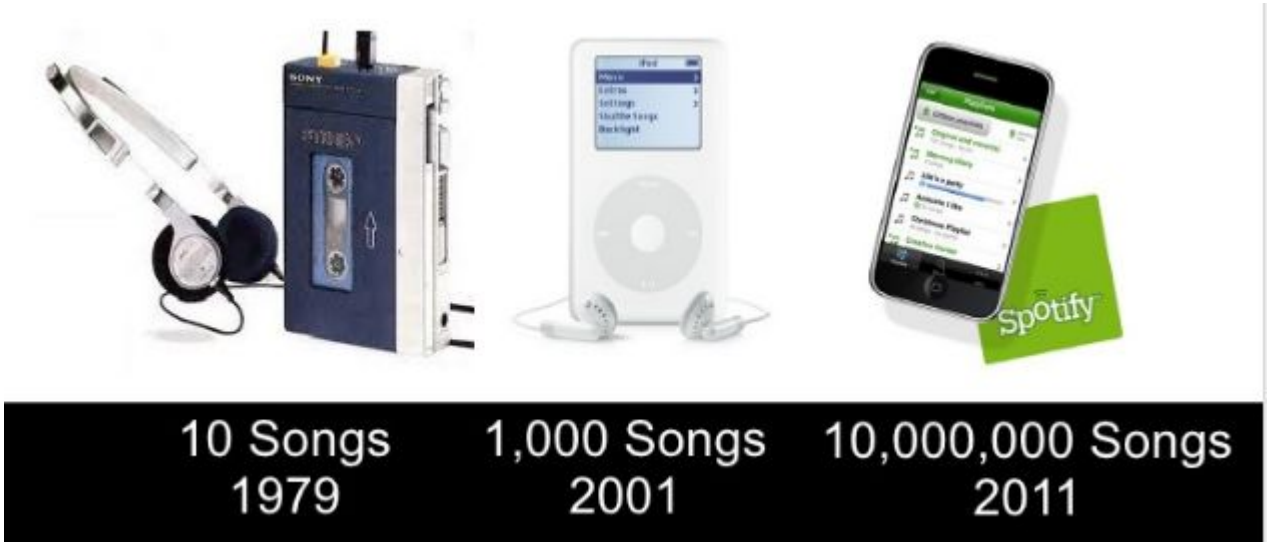
Offering online personalized recommendation services helps to improve customers' satisfaction and needs. Personalization of product information has become one of the most important factors that impact a customer's product selection and satisfaction in today's competitive and challenging market.

Music recommendation systems help in predicting the choice of songs for the users based on the interests and historical data and it is one of the most popular application of big data processing.

The goal of this project was to recommend songs to users based solely on similarity of users' ratings, with no information about the music.

Music Recommendation is important

How many songs fit in my pocket?



Description

Collaborative filtering (CF) is a technique used by us to recommend songs to the users. Applications of collaborative filtering typically involve very large data sets. It predicts user preferences in songs selection based on the known user ratings of songs. We used user based collaborative filtering which is based on the assumption that people who agreed in the past are likely to agree again.Similarity between users is decided by looking at their overlap in opinions for other items.

We have used star representation to indicate the ratings of songs which are getting recommended for customer. This star representation consists of 1 to 5 stars:

1 star indicates song recommended is rated high by less than 10% of users with similar interest.

2 stars indicates song recommended is rated high by greater than 10% and less than 30% of users with similar interest.

3 stars indicates song recommended is rated high by greater than 30% and less than 60% of users with similar interest.

4 stars indicates song recommended is rated high by greater than 60% and less than 80% of users with similar interest.

5 stars indicates song recommended is rated high by greater than 80% and less than 100% of users with similar interest.

Similarity Computation:

1. For each user u, generate song pairs and corresponding rating pairs. <User_id Track_id Track_name rating>
2. For a given pair of songs, collect rating pairs from all users.
3. For each song pair, form a vector of rating values for each song in the pair.
4. Apply similarity functions on the two vectors and derive a similarity score for each song pair.
5. The output of this stage will be of the form: (userid##tracknameTTTTtrackid_rating;score)
6. Once similarity score is generated for all pairs, use these values to put together a list of songs similar to a given song.
7. The output of the final stage will be of the form: (Track_name ***)

Framework : Hadoop Map Reduce

Dataset

Dataset Gathering:

We have used the dataset from http://www.dtic.upf.edu/~ocelma/MusicRecommendationDataset/lastfm-1K.html

This dataset contains <user, timestamp, artist, song> tuples collected from Last.fm API, using the user.getRecentTracks() method.

Data Statistics:

File userid-timestamp-artid-artname-traid-traname.tsv

Total Lines: 19,150,868

Unique Users: 992

Data Format:

The data is formatted one entry per line as follows (tab separated, "\t"):

Data Analysis and Preprocessing:

Our dataset contains over 100 million songs with 1.2 million users of lastFm Music services. Tab is used as a delimiter for all data files. Each song in the data set is accompanied by artist name-id, songs name-id, timestamp and users attributes. There was no rating associated. We pre-processed this whole dataset and generated random rating for the users:

```
from random import randint
import csv
import sys
csv.field_size_limit(sys.maxsize)
with open('Dataset.tsv','r') as f_in:
    with open('finaldataset.tsv','w') as f_out:
            writer=csv.writer(f_out,delimiter='\t', lineterminator='\n')
            reader=csv.reader(f_in)
            all=[]

            for row in reader:
              row.append(randint(0,5))
              all.append(row)

            writer.writerows(all)
```

Algorithms Implemented

1. Jaccard Coefficient-

The Jaccard index, also known as Intersection over Union and the Jaccard similarity coefficient is a statistic used for comparing the similarity and diversity of sample sets. The Jaccard coefficient measures similarity between finite sample sets, and is defined as the size of the intersection divided by the size of the union of the sample sets.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

In this implementation, we have considered only users with similar interest. This coefficient is determined by calculating two factors. First being total no. of tracks which are getting matched with customer . Other factor is finding no. of tracks which are getting rated equally by the customer and other users. While calculating jaccard coefficient first factor is taken as denominator and second factor as numerator. Rating comparison is done using below logic :

If both customer and other user's rating ranges between 3 to 5 or 1 to 3, , it is considered as equal.

2. Pearson Correlation-

It is a measure of the linear correlation between two variables *X* and *Y*. It has a value between +1 and −1, where 1 is total positive linear correlation, 0 is no linear correlation, and −1 is total negative linear correlation.

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

r -> Pearson Coefficient
{x1, x2, … xn} -> rating values for the user for whom recommendation is to be done
{y1, y2, … yn} -> rating values for the users whose songs matched with input user
$\bar{x}$-> average of all values in X
$\bar{y}$ -> average of all values in Y

The values of r range between [-1, +1]. The values nearing -1 are the most dissimilar and values around 0 have no similarity. Hence, in the Pearson Similarity computation, only values greater than 0 are taken for computing recommendation.

3. Cosine Similarity-

Cosine similarity is a measure of similarity between two non-zero vectors of an inner product space that measures the cosine of the angle between them.

The cosine of two non-zero vectors can be derived by using the Euclidean dot product formula:

Given two vectors of attributes, *A* and *B*, the cosine similarity, *cos(θ)*, is represented usi $\mathbf{a} \cdot \mathbf{b} = \|\mathbf{a}\|_2 \|\mathbf{b}\|_2 \cos\theta$ nitude as:

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|_2 \|\mathbf{B}\|_2} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}}$$, where A_i and B_i are components of vector A and B respectively.

The resulting similarity ranges from −1 meaning exactly opposite, to 1 meaning exactly the same, with 0 indicating orthogonality (decorrelation), and in-between values indicating intermediate similarity or dissimilarity.

Example

Table 1 gives an example of a user-song rating matrix. We assume that there are four songs and five users in the systems. The missing ratings of the rating matrix are represented by the symbol. Then we calculate the similarities of users in the table according to those similarity measures described the above. Fig. 1 shows the results of the user similarities in Table 1. Since user similarity matrix is symmetric, we only show partial values.

Table 1: An example of the user-song rating matrix. The missing ratings are represented by the symbol –

	Item1	Item2	Item3	Item4
User1	4	3	5	4
User2	5	3	–	–
User3	4	3	3	4
User4	2	1	–	–
User5	4	2	–	–

Example contd..

Pearson Correlation Coefficient

	u_2	u_3	u_4	u_5
u_1	0.707	0.0	0.707	0.707
u_2		1.0	1.0	1.0
u_3			1.0	1.0
u_4				1.0

Jaccard Coefficient

	u_2	u_3	u_4	u_5
u_1	0.5	1.0	0.5	0.5
u_2		0.5	1.0	1.0
u_3			0.5	0.5
u_4				1.0

Cosine Similarity

	u_2	u_3	u_4	u_5
u_1	0.612	0.975	0.606	0.605
u_2		0.703	0.997	0.997
u_3			0.696	0.695
u_4				1.0

Intermediate output

Job1: Read each row from the input file. The data read is in form: <User_id Track_id Track_name rating>. The first mapper gives output as: <user_id as key, tracknameTTTTtrack_id::rating>


```
user_000010     ReasonTTTT40ac4aa3-14c3-4462-b95a-5fe6f700d146::3;Seniorita (Gorąca Krew)TTTTb4f6e1d6-f28b-4afe-a389-0778e809d066::0;Seniorita (Gorąca Krew)TTTTb4f6e1d6-f28b-4afe-a389-0778e809d066::4;Remo
ver (Club Mix)TTTTe4ae3620-e141-498f-962e-f573c5141097::0;Waiting 4 (Thomas Gold Remix)TTTT6ca91be3-1f75-419a-960d-bd89a0c17e11::5;SexybackTTTTe2218d47-55fe-4c82-8aed-a12f3c203e2e::2;SwitchbackTTTT3d7bbb8
3-1a6b-4425-ada4-52e1e638d061::2;FrozenTTTTd477a5a4-800f-49ff-9c86-0ae7019e18fe::3;SymbiontTTTTc229228b-a616-491e-9b07-750eaec86d5b::3;GoodbyeTTTTb33f48e3-eea5-4e33-9781-35091bc91504::3;GoodbyeTTTTb33f48e
3-eea5-4e33-9781-35091bc91504::4;SwitchbackTTTT3d7bbb83-1a6b-4425-ada4-52e1e638d061::0;SwitchbackTTTT3d7bbb83-1a6b-4425-ada4-52e1e638d061::0;Stay With Me (Unlikely)TTTTe08fed2d-c717-4fed-bf59-3901f9c8f4b2
::4;The Last FirstbornTTTTea0530ec-97bb-4426-a934-e2316492f5e5::3;Under My FeetTTTT2862133f-170c-4296-a0bb-effa779a5fa0::3;I Believe YouTTTTb7a65689-6801-4948-aaf3-d58055724160::4;FrozenTTTTd477a5a4-800f-
49ff-9c86-0ae7019e18fe::4;SymbiontTTTTc229228b-a616-491e-9b07-750eaec86d5b::1;Afraid This TimeTTTT7abc37ef-5b81-4ea5-be3e-046ba846bd79::3;FadeawayTTTT05c07b56-b7c3-4e3a-9777-64907d6a6779::0;So Sorry To Sa
yTTTTada256eb-7df6-419a-aa57-764a6bdb64b7::2;Own Little WorldTTTT2d533a2b-d7f0-4b4f-b1f2-2751d3c62894::5;Unlikely (Stay With Me)TTTTf48c2eaf-86f4-4ebb-a6b6-a4a4abd22f63::1;One Good ReasonTTTT5c19292f-7691
-4466-9d4c-1b040ffd8aa0::2;The Stars Of OrionTTTT4f305446-6ff1-4b2e-8ecb-0453dbf0352e::2;Welcome To The EndTTTTadb65cd7-1306-4f58-88a4-abc4805613fb::0;This Is Why I'M Hot (Radio)TTTT6707a1cd-9093-4828-93d
0-73d26bc7427f::5;Party Like A RockstarTTTTb4dd27b4-95ca-4cb5-9a29-a4cf937dc478::4;De Janeiro (S & H Project Remix)TTTTe3d65042-bb80-4cf0-9e67-a55bea871767::2;Children 2007TTTTf35b7eb7-d778-4fb0-b423-3e0f
285a5eed::0;Seniorita (Gorąca Krew)TTTTb4f6e1d6-f28b-4afe-a389-0778e809d066::2;Children 2007TTTTf35b7eb7-d778-4fb0-b423-3e0f285a5eed::2;Seniorita (Gorąca Krew)TTTTb4f6e1d6-f28b-4afe-a389-0778e809d066::2;R
easonTTTT40ac4aa3-14c3-4462-b95a-5fe6f700d146::4;Seniorita (Gorąca Krew)TTTTb4f6e1d6-f28b-4afe-a389-0778e809d066::2;Seniorita (Gorąca Krew)TTTTb4f6e1d6-f28b-4afe-a389-0778e809d066::5;SexybackTTTTe2218d47-
55fe-4c82-8aed-a12f3c203e2e::5;De Janeiro (S & H Project Remix)TTTTe3d65042-bb80-4cf0-9e67-a55bea871767::2;SexybackTTTTe2218d47-55fe-4c82-8aed-a12f3c203e2e::5;Seniorita (Gorąca Krew)TTTTb4f6e1d6-f28b-4afe
-a389-0778e809d066::1;Summer: Coney Island DreamingTTTT23ba2246-35e2-4fe6-879f-00150bb65ec4::0;Winter: Lux AeternaTTTTc44a6709-91f1-4b89-9696-7114df89320c::4;Winter: Lux AeternaTTTTc44a6709-91f1-4b89-9696
-7114df89320c::1;SexybackTTTTe2218d47-55fe-4c82-8aed-a12f3c203e2e::5;Winter: Lux AeternaTTTTc44a6709-91f1-4b89-9696-7114df89320c::1;Winter: Lux AeternaTTTTc44a6709-91f1-4b89-9696-7114df89320c::3;Winter: L
ux AeternaTTTTc44a6709-91f1-4b89-9696-7114df89320c::2;This Is Why I'M Hot (Radio)TTTT6707a1cd-9093-4828-93d0-73d26bc7427f::1;This Is Why I'M Hot (Radio)TTTT6707a1cd-9093-4828-93d0-73d26bc7427f::5;This Is
Why I'M Hot (Radio)TTTT6707a1cd-9093-4828-93d0-73d26bc7427f::5;Party Like A RockstarTTTTb4dd27b4-95ca-4cb5-9a29-a4cf937dc478::3;This Is Why I'M Hot (Radio)TTTT6707a1cd-9093-4828-93d0-73d26bc7427f::5;This
Is Why I'M Hot (Radio)TTTT6707a1cd-9093-4828-93d0-73d26bc7427f::3;Party Like A RockstarTTTTb4dd27b4-95ca-4cb5-9a29-a4cf937dc478::4;This Is Why I'M Hot (Radio)TTTT6707a1cd-9093-4828-93d0-73d26bc7427f::4;Th
is Is Why I'M Hot (Radio)TTTT6707a1cd-9093-4828-93d0-73d26bc7427f::5;This Is Why I'M Hot (Radio)TTTT6707a1cd-9093-4828-93d0-73d26bc7427f::4;Hold Me Till The End (Ronski Speed Radio Mix)TTTTf4428610-2b0a-4
3c3-82aa-763cdd340b7f::1;Always A FoolTTTT657bb6e2-9b10-4d27-a3e7-bd16fea523bc::5;Dare Me (Jump Radio Mix)TTTT5f6a88f4-8544-4ed9-bcd0-f7da1a1a9cd5::0;Hold Me Till The End (Ronski Speed Radio Mix)TTTTf4428
610-2b0a-43c3-82aa-763cdd340b7f::0;Always A FoolTTTT657bb6e2-9b10-4d27-a3e7-bd16fea523bc::2;Dare Me (Jump Radio Mix)TTTT5f6a88f4-8544-4ed9-bcd0-f7da1a1a9cd5::3;Hold Me Till The End (Ronski Speed Radio Mix
)TTTTf4428610-2b0a-43c3-82aa-763cdd340b7f::0;Always A FoolTTTT657bb6e2-9b10-4d27-a3e7-bd16fea523bc::5;Dare Me (Jump Radio Mix)TTTT5f6a88f4-8544-4ed9-bcd0-f7da1a1a9cd5::3;This Is Why I'M Hot (Radio)TTTT670
7a1cd-9093-4828-93d0-73d26bc7427f::2;Hold Me Till The End (Ronski Speed Radio Mix)TTTTf4428610-2b0a-43c3-82aa-763cdd340b7f::2;Always A FoolTTTT657bb6e2-9b10-4d27-a3e7-bd16fea523bc::2;Dare Me (Jump Radio M
ix)TTTT5f6a88f4-8544-4ed9-bcd0-f7da1a1a9cd5::4;This Is Why I'M Hot (Radio)TTTT6707a1cd-9093-4828-93d0-73d26bc7427f::1;Hold Me Till The End (Ronski Speed Radio Mix)TTTTf4428610-2b0a-43c3-82aa-763cdd340b7f:
:0;Always A FoolTTTT657bb6e2-9b10-4d27-a3e7-bd16fea523bc::0;Hold Me Till The End (Ronski Speed Radio Mix)TTTTf4428610-2b0a-43c3-82aa-763cdd340b7f::2;Always A FoolTTTT657bb6e2-9b10-4d27-a3e7-bd16fea523bc::
1;This Is Why I'M Hot (Radio)TTTT6707a1cd-9093-4828-93d0-73d26bc7427f::2;Hold Me Till The End (Ronski Speed Radio Mix)TTTTf4428610-2b0a-43c3-82aa-763cdd340b7f::4;Always A FoolTTTT657bb6e2-9b10-4d27-a3e7-b
d16fea523bc::5;Hold Me Till The End (Ronski Speed Radio Mix)TTTTf4428610-2b0a-43c3-82aa-763cdd340b7f::0;Always A FoolTTTT657bb6e2-9b10-4d27-a3e7-bd16fea523bc::1;This Is Why I'M Hot (Radio)TTTT6707a1cd-909
3-4828-93d0-73d26bc7427f::2;Hold Me Till The End (Ronski Speed Radio Mix)TTTTf4428610-2b0a-43c3-82aa-763cdd340b7f::0;Hold Me Till The End (Ronski Speed Radio Mix)TTTTf4428610-2b0a-43c3-82aa-763cdd340b7f::
2;Always A FoolTTTT657bb6e2-9b10-4d27-a3e7-bd16fea523bc::4;Hold Me Till The End (Ronski Speed Radio Mix)TTTTf4428610-2b0a-43c3-82aa-763cdd340b7f::1;Always A FoolTTTT657bb6e2-9b10-4d27-a3e7-bd16fea523bc::4
;Hold Me Till The End (Ronski Speed Radio Mix)TTTTf4428610-2b0a-43c3-82aa-763cdd340b7f::1;This Is Why I'M Hot (Radio)TTTT6707a1cd-9093-4828-93d0-73d26bc7427f::4;This Is Why I'M Hot (Radio)TTTT6707a1cd-909
3-4828-93d0-73d26bc7427f::1;This Is Why I'M Hot (Radio)TTTT6707a1cd-9093-4828-93d0-73d26bc7427f::4;Hold Me Till The End (Ronski Speed Radio Mix)TTTTf4428610-2b0a-43c3-82aa-763cdd340b7f::4;Always A FoolTTT
T657bb6e2-9b10-4d27-a3e7-bd16fea523bc::0;Hold Me Till The End (Ronski Speed Radio Mix)TTTTf4428610-2b0a-43c3-82aa-763cdd340b7f::5;Always A FoolTTTT657bb6e2-9b10-4d27-a3e7-bd16fea523bc::0;Hold Me Till The
```

Job 2: This job gives output as: <user_id, tracknameTTTTtrack_id::rating> for the customer

user_000001 South Eastern DreamTTTTc51c4795-788b-45dd-bb61-28b08b14fbd3::0;Point Of No ReturnTTTT2abe3918-951c-4bb6-85fb-7150f15f4028::5;Attack Ships On FireTTTTddc1605e-b324-4fc1-bb3e-33397261bf32::1
;South Eastern DreamTTTTc51c4795-788b-45dd-bb61-28b08b14fbd3::3;All Hell Is Breaking LooseTTTT6d38ca02-3983-41f3-9fe5-3c0a3f1ef343::1;This Dark MatterTTTT135c2276-a69f-46b4-837d-c51645bc0863::3;Bare Relig
ionTTTTc946ad0c-0080-49a2-9a88-80dfbf173834::5;Waiting For The RaptureTTTT78bdc5ab-d1f5-4225-a2c1-4db767bb288d::1;The TurningTTTTb4f97522-88ea-4334-87f5-d5fe2ebbdf79::1;Bag It UpTTTTe5686ad6-536e-4aa3-ab9
4-dcca6f46483d::1;Cascades Of ColourTTTTeb74e745-4bdd-46e2-a5ee-4ef4377cc71b::3;Cowgirl / RezTTTTdac236ce-4037-4162-b9eb-b3191a003eaa::0;LoudTTTT4ea16024-911e-4546-ab3c-8a4e0e56e0a2::2;Hitchhiker'S Choice
TTTT0b5c0212-0736-447d-951d-99038ff4e526::5;Talk (Album Version)TTTT0cf3a110-b1bf-4274-9cbe-42f70df4d59d::1;Viva La VidaTTTT6cbbb97d-2180-4ba9-97a8-6158b6332bb4::2;Dayvan CowboyTTTTef756a43-0905-4e59-8f05
-a6b2bd977aa8::5;Liquid CherriesTTTT3a1a3db3-8387-4338-8c24-08a47d36eb9d::4;Into The Rainbow VeinTTTT37b0589c-2f06-4015-89a9-c02ff2a9af6d::4;Macquarie RidgeTTTT100efdc8-5387-4526-9cf7-383a7aba19dd::4;Fare
well FireTTTTf9dc952e-ded2-4c48-be73-bc17127aafa8::2;Tears From The Compound EyeTTTTf0d18ade-74b5-4a3d-83ce-615b64cdb919::0;Slow This Bird DownTTTTf817c371-6029-45ae-8f9f-08eeff26b0f2::4;Constants Are Cha
ngingTTTT3ca05a3e-6cac-4d57-b16f-a99a2861b2c0::3;Hey Saturday SunTTTT90da05fc-9d90-4244-8ac3-6ddeb0cf1fb5::4;AtaronchrononTTTT32184ab4-9ff9-4beb-8720-7200673b0353::5;Oscar See Through Red EyeTTTTa3c49027-
eb58-4e16-8848-28a857f936cd::0;Sherbet HeadTTTT9b0d3374-7968-4844-9bac-8b9397eccd42::3;'84 Pontiac DreamTTTTeb78521a-b57b-4e0c-9b94-83e1b437b82a::3;A Moment Of ClarityTTTTd8a1f18b-3bd3-49ed-bdaf-db271c189
010::0;Dayvan CowboyTTTTef756a43-0905-4e59-8f05-a6b2bd977aa8::0;Peacock TailTTTT93fa1eb2-dd7c-4e0e-8174-360d32176586::2;Satellite Anthem IcarusTTTT80ff268c-7ba4-45bd-ba3e-b633a2b24a62::0;HopeTTTT6c16d9b8-
4039-4d2a-8c16-41532be10287::2;Liquid CherriesTTTT3a1a3db3-8387-4338-8c24-08a47d36eb9d::2;Liquid CherriesTTTT3a1a3db3-8387-4338-8c24-08a47d36eb9d::2;Hang On To Your LoveTTTTa19f1bd1-2567-4078-8ef6-a34cb66
d33c6::0;HopeTTTT6c16d9b8-4039-4d2a-8c16-41532be10287::3;Liquid CherriesTTTT3a1a3db3-8387-4338-8c24-08a47d36eb9d::2;New FlatTTTT119c7eed-8cdb-43ab-9c4e-33a03882b931::2;Pia FliesTTTTb68e5849-b98d-45f6-9960
-09a268f3106c::2;HopeTTTT6c16d9b8-4039-4d2a-8c16-41532be10287::5;HopeTTTT6c16d9b8-4039-4d2a-8c16-41532be10287::3;Idiot FunkTTTTdabcd184-1cf7-4b94-8fce-6f1aa5e6ef57::5;WeekendTTTTb8a73773-f8fe-48dc-8428-f3
03d23bbc34::4;HonnojiTTTT37e0822b-cd4a-49d9-9fe1-11bffbdb8643::4;AsobiTTTTa5f274cc-a542-4b50-ab78-ac2c471b44e7::5;Look What You're Doin' To Me (Feat. Phonte)TTTT7167699e-2612-4f38-8917-e654ed034ad9::3;Loo
k What You're Doin' To Me (Feat. Phonte)TTTT7167699e-2612-4f38-8917-e654ed034ad9::4;Look What You're Doin' To Me (Feat. Phonte)TTTT7167699e-2612-4f38-8917-e654ed034ad9::3;Look What You're Doin' To Me (Fea
t. Phonte)TTTT7167699e-2612-4f38-8917-e654ed034ad9::5;Look What You're Doin' To Me (Feat. Phonte)TTTT7167699e-2612-4f38-8917-e654ed034ad9::2;Look What You're Doin' To Me (Feat. Phonte)TTTT7167699e-2612-4f
38-8917-e654ed034ad9::4;Look What You're Doin' To Me (Feat. Phonte)TTTT7167699e-2612-4f38-8917-e654ed034ad9::0;Look What You're Doin' To Me (Feat. Phonte)TTTT7167699e-2612-4f38-8917-e654ed034ad9::4;Look W
hat You're Doin' To Me (Feat. Phonte)TTTT7167699e-2612-4f38-8917-e654ed034ad9::4;Look What You're Doin' To Me (Feat. Phonte)TTTT7167699e-2612-4f38-8917-e654ed034ad9::0;Look What You're Doin' To Me (Feat.
Phonte)TTTT7167699e-2612-4f38-8917-e654ed034ad9::1;Look What You're Doin' To Me (Feat. Phonte)TTTT7167699e-2612-4f38-8917-e654ed034ad9::5;Look What You're Doin' To Me (Feat. Phonte)TTTT7167699e-2612-4f38-
8917-e654ed034ad9::2;Look What You're Doin' To Me (Feat. Phonte)TTTT7167699e-2612-4f38-8917-e654ed034ad9::5;Look What You're Doin' To Me (Feat. Phonte)TTTT7167699e-2612-4f38-8917-e654ed034ad9::5;NightowlT
TTT408ebcc1-4f07-4274-89cd-d925c47e308a::1;Lone CatTTTT468fa57f-a727-463f-95a6-3dc1d4f6e2ca::5;CyclicTTTT9178701f-e239-49bd-a64e-b1d15aa6d454::3;Fade OutTTTT6c6c954a-4a08-4db7-bdf3-4be36b695a79::3;Departu
reTTTTf0e111d5-c11c-4541-9945-34bb24fe637d::4;AsobiTTTTa5f274cc-a542-4b50-ab78-ac2c471b44e7::2;Idiot FunkTTTTdabcd184-1cf7-4b94-8fce-6f1aa5e6ef57::2;WeekendTTTTb8a73773-f8fe-48dc-8428-f303d23bbc34::4;Honn
ojiTTTT37e0822b-cd4a-49d9-9fe1-11bffbdb8643::4;AsobiTTTTa5f274cc-a542-4b50-ab78-ac2c471b44e7::0;Poppoya (Piano Version)TTTT8e463e7b-5f87-45e6-a430-b0f0ee246d85::0;Love Is The DevilTTTTfee83435-97de-4556-8
6b1-d6a293eed171::1;Snake Eyes [Short Version/Theme]TTTT8a77f52f-2de7-4ec9-9f11-c5655bd2864b::4;Acceptance (End Credit) - Little Buddha -TTTTb108f2a8-dd5f-4e53-98f6-525a97da1f79::4;Main Theme / End Titles
TTTTcf6fb14c-b9eb-4b0a-84a9-8a1816a20990::0;Main Theme (Piano Version)TTTTbca584d0-a601-421c-ad7e-ec2c47b1a06b::2;The Sheltering SkyTTTT989b783b-054d-46eb-8954-3e9fff7bf62d::0;Theme From The Sheltering Sk
yTTTTed4c84ff-0b94-4e91-a6f0-8d691b17c31a::3;The Sheltering Sky Theme (Piano Version)TTTT07288df2-c3c0-4c0b-b893-f8c6d42d4c5b::0;Wrong (Todd Terry Remix)TTTT93725fea-5bb9-482b-aac0-a18bcc8e4361::4;Walking

Job 3: This gives output <trackid##track_name key, useridDDDDsong>

```
b31f8a8e-32ee-4017-8723-aa0cbb7f80e8##Relic (Aghast View Intense Remix) user_000009DDDDsong
2c4c5f7c-6be1-49b9-887e-71f1a2278d23##Always You (Acoustic)     user_000009DDDDsong
edebead4-de22-4a2f-808c-3c57c1a100e2##Light Through Skin        user_000009DDDDsong
9c287d74-d998-47dd-a061-d96e68fac15d##What About Love   user_000009DDDDsong
bbc327b4-0d62-4237-8975-7708b37ab6d5##Another Void      user_000009DDDDsong
c32e2872-0658-4982-ab48-0e39db3be053##Sweetest Maleficia        user_000009DDDDsong
e93c7702-a68b-47d8-ab3d-ab17453bd601##Tearing The Veil From Grace       user_000009DDDDsong
95b47745-9dda-49ad-8131-90a573eaa2bd##Skin      user_000009DDDDsong
1a3f6617-140e-494c-9ba2-6e196b24524e##Living Hell       user_000009DDDDsong
323142b7-73fe-48f2-9922-8ebe13eb6c8e##Sympathetic       user_000009DDDDsong
14b5e624-5322-4a4e-a4f5-8ead6b15cf22##Indestructible    user_000009DDDDsong
212b5f22-2f64-4823-ae53-82024f25c99f##Kenta     user_000009DDDDsong
1b2feb99-fc9d-404b-93cd-aaec23d0eb43##Vegetarian Restaurant     user_000009DDDDsong
3389c6bd-ee3f-419b-bdbc-012888d76f65##No Rest For The Wicked    user_000009DDDDsong
7b1638ba-597a-48b6-8efc-8766e1522218##Resuscitation Of A Dead Man       user_000009DDDDsong
320f722f-3092-44e8-a894-ec47d217c5c8##Hey Lady  user_000009DDDDsong
9c8499e3-0914-4662-9afb-7f90c021d064##Broken    user_000009DDDDsong
7b0e50f8-6e49-49ea-a057-dd3ef96effb1##Steeds Of Thunder user_000009DDDDsong
9d72ed6b-5857-48a2-8c34-e52e605ac107##My Device user_000009DDDDsong
05fe1420-3238-49b1-bdf9-dfcb171eeead##Cocoon    user_000009DDDDsong
841f9fd7-746e-44fe-b911-9d67f22bb696##Give 'Em Hell Kid user_000009DDDDsong
09c4e7ef-e21c-4eea-b593-b1f5c135d3a0##For You   user_000009DDDDsong
09a4d0ba-fa3d-4741-8e8a-4c6c5386a615##Behind The Wheel (F.A. Preve Mix) user_000009DDDDsong
ab56df89-beb6-4521-9254-cbc79e29e492##This Is Halloween user_000009DDDDsong
2e25a2d0-f5e3-43ca-a8d5-a68c6f758f7b##I Wanna Love You  user_000009DDDDsong
f94d2b94-ca4a-4bde-907c-7fa2fb4fbe83##To Forever (Moonbeam Remix)       user_000009DDDDsong
37b1ad42-c47d-434a-ae58-b45596476668##Inside Out        user_000009DDDDsong
637efa16-9c9b-4035-a49f-12e1fe5827da##Frenetic Amnesic  user_000009DDDDsong
b66314ac-bdf8-4a1b-bebe-23c53ef15924##A Threnody For Modern Romance     user_000009DDDDsong
e732dc44-d36b-4e12-b09b-af96f4f1d92f##A user_000009DDDDsong
c9730197-f2c0-47c5-9d85-45b81a05bb7b##Baby Hold On      user_000009DDDDsong
8be6e698-ffe2-4897-8d4c-302925597167##Four Wall Blackmail       user_000009DDDDsong
28ad9682-a88a-4004-b09b-efcc4f019216##Runnin' Down A Dream      user_000009DDDDsong
259836d9-1d6b-4d6b-a05f-be6c9eed02e3##The Thin Ice      user_000009DDDDsong
35093e80-ed56-4194-a0b7-3ff473fe2bcb##Q (The Best One Of Our Lives)     user_000009DDDDsong
34d084ad-62ab-44f5-bcf8-7dc6f939b67d##Return To Sender  user_000009DDDDsong
915dbb48-7932-4280-87d8-21c8bcd86095##Because I Got High        user_000009DDDDsong
2a00d5b0-7f0f-467e-b581-675f0da9de9e##My Friend Of Misery       user_000009DDDDsong
10725e6b-2ac2-45c5-b189-b0ad95a0ad1e##The End Of Heartache      user_000009DDDDsong
6aafc6c9-d188-4dfe-b44d-efb4d2bf855f##Prologue (Feat. Cyber Axis)       user_000009DDDDsong
619225ce-cde0-4e7d-88a0-b9b217bb0068##As He Climbed The Dark Mountain   user_000009DDDDsong
a9db05f6-27d1-40d8-bbc7-f582e3baf556##Vigil     user_000009DDDDsong
```

Prog 1: Jaccard Coefficient

Output:

```
Blue Hour (Home Recording)      *****
Blue Orchid     *****
Blues In 3/4 Time       *****
Brace And Break *****
1349    *****
4Th Of July     *****
68 Guns *****
A Public Affair *****
Break Me Shake Me       *****
Breakdown       *****
Stuck In America        *****
Not Gonna Be Alone Tonight      *****
Button Up Your Overcoat *****
C'Etait Une Histoire D'Amour    *****
Full Of Stars   *****
Caide Sin Don Té Sin    *****
Du Bist Es      *****
Dust And Rocks  *****
Dying Breed     *****
Cardiff Afterlife       *****
Chainsawdomy    *****
Chelsea Dagger  *****
Cherry Bomb     *****
Love Letters    *****
Lucky Star (New Mix)    *****
Man In The Box  *****
Complicated     *****
Con Cassidy'S (Jig) / Dusty Millar (Slip Jig)   *****
Con Cossidy'S / Neil Gow'S Highland / Moll And Tiarna / Mcsweeney'S     *****
Cool 'N' Out    *****
I Built This City (Instrumental Mix)*****
Wishin' And Hopin'      ****
Won'T U Please B Nice   ****
Counting Down The Days  ****
County Jail Blues       ****
Crashdown       ****
Dan Abnormal    ****
Dan Y Dwr       ****
Danger Zone     ****
```

Prog2: Pearson

```
Bli-Blip        *****
Full Of Stars   *****
Caide Sin Don Té Sin    *****
Du Bist Es      *****
Dust And Rocks  *****
Dying Breed     *****
Cardiff Afterlife       *****
Blood Bleeds    *****
Blue Hour (Home Recording)      *****
Blue Orchid     *****
Blues In 3/4 Time       *****
Brace And Break *****
1349    *****
4Th Of July     *****
68 Guns *****
A Public Affair *****
Break Me Shake Me       *****
Breakdown       *****
Stuck In America        *****
Not Gonna Be Alone Tonight      *****
Button Up Your Overcoat *****
C'Etait Une Histoire D'Amour    *****
Chainsawdomy    *****
Cool 'N' Out    *****
I Built This City (Instrumental Mix)*****
Wishin' And Hopin'      ****
Won'T U Please B Nice   ****
Counting Down The Days  ****
County Jail Blues       ****
```

Prog3:Cosine

```
Blue Orchid     *****
Blues In 3/4 Time       *****
Brace And Break *****
1349    *****
4Th Of July     *****
Breakdown       *****
Stuck In America        *****
Not Gonna Be Alone Tonight      *****
Con Cassidy'S (Jig) / Dusty Millar (Slip Jig)   *****
Con Cossidy'S / Neil Gow'S Highland / Moll And Tiarna / Mcsweeney'S     *****
Button Up Your Overcoat *****
C'Etait Une Histoire D'Amour    *****
Chainsawdomy    *****
Cool 'N' Out    *****
I Built This City (Instrumental Mix)*****
68 Guns *****
A Public Affair *****
Break Me Shake Me       *****
Deadweight On Velveteen ****
Deep Down The Road (Outro)      ****
Deep Ocean Vast Sea     ****
Wishin' And Hopin'      ****
Won'T U Please B Nice   ****
Counting Down The Days  ****
County Jail Blues       ****
```

Performance Evaluation

Similarity Measure	Running Time (in minutes)
Pearson Correlation Coefficient	72.86
Jaccard Coefficient	88.93
Cosine Similarity	115.05

Challenges & Solutions

Challenges	**Solution**
Data gathering, analysis and processing of gigabytes of data was difficult to handle. Eg: Our dataset had (“ ”)	We handled it by removing (“ ”) using 010 editor to handle big data
Users-songs did not have ratings associated with it.	Wrote a python script to generate ratings randomly for all users-song pairs.
Sparsity problem – Songs which are not rated by users are hard to find similarity.	Tracks which are rated as 5 by other users are recommended.
Two inputs were needed in mapper class,	Introduced setup method to handle 2 inputs.
New user if entered who has no previous data	Tracks which are rated as 5 by other users are recommended.

Surprises & Cool Enhancements

Surprises:

- After analyzing outputs and code, we got to know that in the dataset the tracks had multiple ratings from the same user.

  What we did?- We used the latest rating given by the user for computation purposes.

Cool Enhancements:

- Representing song recommendations along with stars based on the calculation of number of users rated the particular track of all the users in the dataset.

Responsibilities Division

Aditi:

- Responsible for data gathering
- Data Cleaning and preprocessing.
- Implemented Cosine similarity algorithm.

Nikita:

- Responsible for implementing Jaccard Coefficient algorithm.
- Implemented the logic behind stars representation
- sorting the final output based on count of stars.

Priya:

- Responsible for implementing Pearson Correlation Coefficient algorithm.
- Look up to implement the songs names with star counts in the final output
- Implemented setup function to handle two inputs in mapper class.

Tasks Implemented

Implemented a music recommendation system using Hadoop MapReduce that recommends top songs for a user based on songs rated by other users of similar taste. Additionally, we have used star representation for recommended songs based on percentage of similar users' high rating(rating of 4,5 only) for the same song.

Future Enhancements

Instead of working with a static dataset, we would like to implement the same by implementing an additional job which would fetch songs from different sources through Music API.
Link: https://musicmachinery.com/music-apis/

References

- https://cseweb.ucsd.edu/~jmcauley/cse255/reports/wi15/Guanwen%20Yao_Lifeng_Cai.pdf
- http://www.iaeng.org/publication/IMECS2013/IMECS2013_pp380-384.pdf
- http://www.stat.wmich.edu/s216/book/node122.html