# KBhave_Discussion2.Rmd

*Kumudini Bhave*

*June 22, 2017*

---

## Overview

Christopher Johnson, (machine learning background with more experience in Hadoop and recent expereince in Spark), who works at Spotify which provides music recommendation, talks about the different algorithms and tradeoffs for providing better recommendations

---

### Points He Makes

1. Spotify is a music streaming service that offers on-demand music based on artist / album / song from the 40 million song catalogue. Personalized recommendations are offered at Spotify.
2. In general, personal recommendations can be based on artist/ song radio stations and play something similar to that. Recommendations can be summed up with manual approach , by tagging attributes, collaborative filtering , text analysis by checking the audio content, request logs etc
3. At Spotify , they are more in line with collborative filtering.
4. In comparison with Netfilx approach of Matrix Factorization, where they use *Explicit Data*, at Spotify they have *Implicit Data* and do not follow explicit ratings. They infer what user would like based on what users listen to, with binary data. They weight the RMSE based on the number of times a music piece is listened to.
5. How at Spotify they had Hadoop all along and how they were experimenting to use Spark for the music recommendations.
6. The I/O overhead with Hadoop, directed them to Spark, where ratings are loaded into memory and iterations are performed.
7. He explains the brodcast method , similar to Hadoop, the full gridify, half gridify method and the ALS running times for each of these.
8. PairRDDFunctions for keeping together the user / item related factors as vectors.

---

### Inference :

Spark offers itself as a step up from Hadoop w.r.t implicit data factorization. However, certain content-based, latent factor based analysis would need certain amount of customization at the implementation side to have tailor-made personalized recommendations.

---