

# Data 612 - Research Discussion 2

Paul Perez

7/14/2020

## Contents

<b>Music Recommendations at Scale with Spark</b>	<b>2</b>
Instructions . . . . .	2
Response . . . . .	2

# Music Recommendations at Scale with Spark

## Instructions

For this discussion item, please watch the following talk and summarize what you found to be the most important or interesting points. The first half will cover some of the mathematical techniques covered in this unit's reading and the second half some of the data management challenges in an industrial-scale recommendation system.

## Response

Christopher Johnson speaks about the importance of Recommender Systems for Spotify as well as the various ways the recommender systems can be created. Spotify uses a combination of collaborative filter and other content for music such as metadata and text analysis specifically because of the large user base and information available for their platform. They use 'Implicit' matrix factorization in order to minimize the RMSE. Alternating Least Squares was mentioned as one solve for implicit matrix factorizations. Looking back at it retroactively, since they were using Hadoop, and there were several issues with the process that slowed it down, having ALS built out in Spark would be the optimal choice. At the time, they started to experiment with Spark, and tried a few times to group, partition, and condense the data. This was a great watch in introducing and familiarizing some terms and concepts discussed in class, and provide me direction of further studying I need to do in order to become more well versed in this field.

## Sources

Spark Summit. (2014, July 17). Music Recommendations at Scale with Spark - Christopher Johnson (Spotify) [Video]. YouTube. [https://www.youtube.com/watch?v=3LBgiFch4\\_g](https://www.youtube.com/watch?v=3LBgiFch4_g)