CS 5463

**Final Project Report**

ASKMEMP: Cloud-based One-stop solution for Smart Shopping



# 

# Introduction

AskMeMP stand for Ask Me Market Place, which will gather the price of the item from multiple market pages as Walmart, Amazon and Ebay and display it in one page when the user query it. This project is powered by Spark since it is designed to deal with a lot of data to analyze and query. The works were distributed evenly among the team members:

Sharif: Writing python script to process user query.

Tuan: Writing python script to collect the data from Amazon.

Akash: Register Amazon, Walmart, Ebay account and find the Python wrapper for each API

Saeef: Testing script, keep track results, and suggest way to reduce load time and search time.

# Project Components

## Data Collecting:

To collect the data, we used the amazon python API [1] to achieve it. However, we were unable to obtain Walmart API key and Ebay API key to use the API. Amazon Advertising API is the only one we successfully set up and it has Python wrapper which can be downloaded through “pip install python-amazon-product-api”. In this project, we used the API to scrape 1800 books information from Amazon to get the UPC (serve as a key), title, author name, lowest new price, and URL. For each line of data, we make another version of Ebay and Walmart with different price by adding or subtracting a random value from the price. Therefore, we could have 1800 \* 3 = 5400 data in total. Those data is written in three different files and copied into HDFS after that.

## User Query:

Spark is used to process the data: load all the files into RDD and be ready to get the input from the user.

We also use the bash script to get input from command line, transfer it along with the code to submit the python script to Spark. When there is the keyword, the RDD will filter the data with the title that has the keyword, reduce the duplicated results, and display it to the user.

# Evaluation and Discussion

Here is the result we got when run ./query.sh Decorated (Decorated is the search keyword)

+++0847858723

\*\*\*"The Decorated Home: Living with Style and Joy" by "Meg Braff"

Amazon $28.46 <https://www.amazon.com/Decorated-Home-Living-Style-Joy/dp/0847858723?SubscriptionId=AKIAJZWFT54YVED6FYQA&tag=utsacloud-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=0847858723>

Walmart $28.51 <https://www.walmart.com/Decorated-Home-Living-Style-Joy/dp/0847858723?SubscriptionId=AKIAJZWFT54YVED6FYQA&tag=utsacloud-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=0847858723>

Ebay $28.93 <https://www.ebay.com/Decorated-Home-Living-Style-Joy/dp/0847858723?SubscriptionId=AKIAJZWFT54YVED6FYQA&tag=utsacloud-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=0847858723>

+++1542786401

\*\*\*"Interior Designs: An Adult Coloring Book with Beautifully Decorated Houses, Inspirational Room Designs, and Relaxing Modern Architecture" by "Jade Summer"

Amazon $6.29 <https://www.amazon.com/Interior-Designs-Beautifully-Inspirational-Architecture/dp/1542786401?SubscriptionId=AKIAJZWFT54YVED6FYQA&tag=utsacloud-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=1542786401>

Walmart $5.52 <https://www.walmart.com/Interior-Designs-Beautifully-Inspirational-Architecture/dp/1542786401?SubscriptionId=AKIAJZWFT54YVED6FYQA&tag=utsacloud-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=1542786401>

Ebay $6.01 <https://www.ebay.com/Interior-Designs-Beautifully-Inspirational-Architecture/dp/1542786401?SubscriptionId=AKIAJZWFT54YVED6FYQA&tag=utsacloud-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=1542786401>

+++0847847411

\*\*\*"How They Decorated: Inspiration from Great Women of the Twentieth Century" by "P. Gaye Tapp"

Amazon $31.53 <https://www.amazon.com/How-They-Decorated-Inspiration-Twentieth/dp/0847847411?SubscriptionId=AKIAJZWFT54YVED6FYQA&tag=utsacloud-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=0847847411>

Walmart $31.66 <https://www.walmart.com/How-They-Decorated-Inspiration-Twentieth/dp/0847847411?SubscriptionId=AKIAJZWFT54YVED6FYQA&tag=utsacloud-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=0847847411>

Ebay $32.11 <https://www.ebay.com/How-They-Decorated-Inspiration-Twentieth/dp/0847847411?SubscriptionId=AKIAJZWFT54YVED6FYQA&tag=utsacloud-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=0847847411>

Load time: 5.15189 seconds

Search time: 10.72773 seconds

The Amazon link is ready to use. Since we do not know how to load the data in Spark Shell and keep waiting for user input, the data has to be loaded for every time the job is submitted. Minus the load time, the search time is in the range of 8 - 12 second. It is not as good as we expected. Caching may help making the search time to be instant.

# Conclusion

This project is promising to bring convenient experience to the user to shop in less time, bring more traffic to other marketplace sites, get commission by referring customer to those sites as well. However, it still need a lot of work. As being mentioned above, caching may help reducing search time near to zero second. Web UI is also required to bring friendly interface to customer. Natural language processing is also needed to be considered so that AskMeMP can provide accurate results to real words questions such as “What is best present for 10 year old birthday boy”. With the help of cloud computing, we believe that live price monitoring will be easily implemented to aid customer shopping in the future.

# References

[1] Amazon python API: <https://pypi.python.org/pypi/python-amazon-simple-product-api>