

GitHub Project link: <https://github.com/SolsticeDante/50043-Project>

Application URL: kensim.online

Description:

We created a webpage to allow users to view ebooks and their reviews. Users are also able to submit their own reviews and mark others as helpful.

The application was created with Vue.js as the frontend framework, and Flask as the backend framework.

We webscrapped as many titles as we could as we realised that most books from the original database did not have any title.

The readme for the hadoop portion is in a separate file titled 'ANALYTICS_README.md'

Things to note:

The page will turn dark when clicking a book. Just wait for it to load, don't click away.

Automation is done for frontend/backend, mongo db and mysql deployment only.

By default the key-name is **key1** and the file name is **key1.pem**. If you wish to change it, modify lines **2, 18, 20, 22** in **misc/build.sh**.

Instructions how to run automated deployment:

1. Run `aws configure`
(<https://docs.aws.amazon.com/cli/latest/userguide/cli-chap-configure.html>) and ensure that **output format is 'json'**
2. Rename .pem file as 'key1.pem' and put it in misc folder
3. Run `./build.sh` in the misc folder

Endpoint specifications

GET main_top_row_books

Gets a list of preselected recommended and best selling books

Sample return:

```
{
  "recommended": [
    {
      "title": "Faux Leather Kindle Sleeve Case for Kindle (Fits 9.7" Display, Latest Generation Kindle DX) - Yellow (Smooth finish)",
      "asin": "B002HWRIIQ",
      "price": 1.99,
      "imageUrl": "http://ecx.images-amazon.com/images/I/513mtcKv9HL._B02,204,203,200_PIsitb-sticker-v3-big,TopRight,0,-55_SX278_SY278_PIKin4,BottomRight,1,22_AA300_SH20_OU01_.jpg",
    },
    {
      "asin": "B002HWRKES",
      "price": 0.99,
      "imageUrl": "http://ecx.images-amazon.com/images/I/51KA6T01xdL._B02,204,203,200_PIsitb-sticker-v3-big,TopRight,0,-55_SX278_SY278_PIKin4,BottomRight,1,22_AA300_SH20_OU01_.jpg",
    },
    {
      "asin": "B002HWRKES",
      "price": 1.99,
      "imageUrl": "http://ecx.images-amazon.com/images/I/51KA6T01xdL._B02,204,203,200_PIsitb-sticker-v3-big,TopRight,0,-55_SX278_SY278_PIKin4,BottomRight,1,22_AA300_SH20_OU01_.jpg",
    }
  ],
  "best": [
    {
      "title": "Faux Leather Kindle Sleeve Case for Kindle (Fits 9.7" Display, Latest Generation Kindle DX) - Yellow (Smooth finish)",
      "asin": "B002HWRIIQ",
      "price": 1.99,
      "imageUrl": "http://ecx.images-amazon.com/images/I/513mtcKv9HL._B02,204,203,200_PIsitb-sticker-v3-big,TopRight,0,-55_SX278_SY278_PIKin4,BottomRight,1,22_AA300_SH20_OU01_.jpg",
    },
    {
      "asin": "B002HWRKES",
      "price": 0.99,
      "imageUrl": "http://ecx.images-amazon.com/images/I/51KA6T01xdL._B02,204,203,200_PIsitb-sticker-v3-big,TopRight,0,-55_SX278_SY278_PIKin4,BottomRight,1,22_AA300_SH20_OU01_.jpg",
    },
    {
      "asin": "B002HWRKES",
      "price": 1.99,
      "imageUrl": "http://ecx.images-amazon.com/images/I/51KA6T01xdL._B02,204,203,200_PIsitb-sticker-v3-big,TopRight,0,-55_SX278_SY278_PIKin4,BottomRight,1,22_AA300_SH20_OU01_.jpg",
    }
  ]
}
```

GET main_bot_row_books?start_list=0&seed=123

Gets a randomized list of books of length 18.

Query Params	Description
start_list	An integer, used to get a new list each page of the website
seed	An integer, used to randomly generate 18 books to show

Sample return:

```
{
  "collection": [
    {
      "title": "Faux Leather Kindle Sleeve Case for Kindle (Fits 9.7" Display, Latest Generation Kindle DX) - Yellow (Smooth finish)",
      "asin": "B002HWRIIQ",
      "price": 1.99,
      "imUrl": "http://ecx.images-amazon.com/images/I/513mtcKv9HL._B02,204,203,200_PIsitb-sticker-v3-big,TopRight,0,-55_SX278_SY278_PIKin4,BottomRight,1,22_AA300_SH20_OU01_.jpg",
    },
    {
      "asin": "B002HWRKES",
      "price": 0.99,
      "imUrl": "http://ecx.images-amazon.com/images/I/51KA6T01xdL._B02,204,203,200_PIsitb-sticker-v3-big,TopRight,0,-55_SX278_SY278_PIKin4,BottomRight,1,22_AA300_SH20_OU01_.jpg",
    },
    {
      "asin": "B002HWRKES",
      "price": 1.99,
      "imUrl": "http://ecx.images-amazon.com/images/I/51KA6T01xdL._B02,204,203,200_PIsitb-sticker-v3-big,TopRight,0,-55_SX278_SY278_PIKin4,BottomRight,1,22_AA300_SH20_OU01_.jpg",
    }
  ]
}
```

GET single_book/{ASIN_NUMBER}/

e.g. /single_book/B0070YQGSO

Fetches all the details of a single book (minus reviews).

Sample return:

```
{
  "book": {
    "asin": "B002HWRR78",
    "price": 0.99,
    "averageRating": 4.2,
  }
}
```

```

        "imUrl":
        "http://ecx.images-amazon.com/images/I/51zBsXMfkFL._B02,204,203,200_PIsitb-sticker-v3-big,TopRight,0,-55_SX278_
        SY278_PIKin4,BottomRight,1,22_AA300_SH20_OU01_.jpg",
        "related": [{
            "asin": "B0070YQGS0",
            "imUrl":
            "http://ecx.images-amazon.com/images/I/51zBsXMfkFL._B02,204,203,200_PIsitb-sticker-v3-big,TopRight,0,-55_SX278_
            SY278_PIKin4,BottomRight,1,22_AA300_SH20_OU01_.jpg"
        }, {
            "asin": "B0070YQGS0",
            "imUrl":
            "http://ecx.images-amazon.com/images/I/51zBsXMfkFL._B02,204,203,200_PIsitb-sticker-v3-big,TopRight,0,-55_SX278_
            SY278_PIKin4,BottomRight,1,22_AA300_SH20_OU01_.jpg"
        }, {
            "asin": "B0070YQGS0",
            "imUrl":
            "http://ecx.images-amazon.com/images/I/51zBsXMfkFL._B02,204,203,200_PIsitb-sticker-v3-big,TopRight,0,-55_SX278_
            SY278_PIKin4,BottomRight,1,22_AA300_SH20_OU01_.jpg"
        }, {
            "asin": "B0070YQGS0",
            "imUrl":
            "http://ecx.images-amazon.com/images/I/51zBsXMfkFL._B02,204,203,200_PIsitb-sticker-v3-big,TopRight,0,-55_SX278_
            SY278_PIKin4,BottomRight,1,22_AA300_SH20_OU01_.jpg"
        }, {
            "asin": "B0070YQGS0",
            "imUrl":
            "http://ecx.images-amazon.com/images/I/51zBsXMfkFL._B02,204,203,200_PIsitb-sticker-v3-big,TopRight,0,-55_SX278_
            SY278_PIKin4,BottomRight,1,22_AA300_SH20_OU01_.jpg"
        }, {
            "asin": "B0070YQGS0",
            "imUrl":
            "http://ecx.images-amazon.com/images/I/51zBsXMfkFL._B02,204,203,200_PIsitb-sticker-v3-big,TopRight,0,-55_SX278_
            SY278_PIKin4,BottomRight,1,22_AA300_SH20_OU01_.jpg"
        }, {
            "asin": "B0070YQGS0",
            "imUrl":
            "http://ecx.images-amazon.com/images/I/51zBsXMfkFL._B02,204,203,200_PIsitb-sticker-v3-big,TopRight,0,-55_SX278_
            SY278_PIKin4,BottomRight,1,22_AA300_SH20_OU01_.jpg"
        }],
        "categories": [
            ["Books", "Christian Books & Bibles"],
            ["Books", "Literature & Fiction", "Classics"],
            ["Kindle Store", "Kindle Short Reads", "One hour (33-43 pages)", "Religion &
            Spirituality"],
            ["Kindle Store", "Kindle eBooks", "Religion & Spirituality", "Christian Books &
            Bibles"]
        ]
    }
}

```

GET review_list/{ASIN_NUMBER}

e.g. review_list/B002HWRKES

Returns a list of reviews for the specified book.

Sample return:

```
{
```

```

"reviews":[
  {
    "reviewerID": "A1F6404F1VG29J",
    "asin": "B000F83SZQ",
    "reviewerName": "Avidreader",
    "helpful": [0, 0],
    "reviewText":
      "I enjoy vintage books and movies so I enjoyed reading this book. The plot was unusual. Don't think
      killing someone in self-defense but leaving the scene and the body without notifying the police or hitting
      someone in the jaw to knock them out would wash today.Still it was a good read for me.",
    "overall": 5.0,
    "summary": "Nice vintage story",
    "reviewTime": "June 8, 2019"
  },
  {
    "reviewerID": "AN0N05A9LIJEQ",
    "asin": "B000F83SZQ",
    "reviewerName": "critters",
    "helpful": [2, 2],
    "reviewText":
      "This book is a reissue of an old one; the author was born in 1910. It's of the era of, say, Nero
      Wolfe. The introduction was quite interesting, explaining who the author was and why he's been forgotten; I'd
      never heard of him.The language is a little dated at times, like calling a gun a &#34;heater.&#34; I also made
      good use of my Fire's dictionary to look up words like &#34;deshabille&#34; and &#34;Canarsie.&#34; Still, it
      was well worth a look-see.",
    "overall": 4.0,
    "summary": "Different...",
    "reviewTime": "May 18, 2019"
  },
  {
    "reviewerID": "A795DMNCJILA6",
    "asin": "B000F83SZQ",
    "reviewerName": "dot",
    "helpful": [2, 2],
    "reviewText":
      "This was a fairly interesting read. It had old- style terminology.I was glad to get to read a
      story that doesn't have coarse, crasslanguage. I read for fun and relaxation.....I like the free
      ebooksbecause I can check out a writer and decide if they are intriguing,innovative, and have enough of the
      command of Englishthat they can convey the story without crude language.",
    "overall": 4.0,
    "summary": "Oldie",
    "reviewTime": "September 20, 2014"
  }
]
}

```

POST /post_new_review

Posts a new review to the backend/database.

Frontend will send a post with body:

```

{
  "overall": 4.0,
  "summary": "Very good book",
  "reviewText": "Lorem ipsum dolor sit amet, consectetur adipiscing elit. Suspendisse rutrum libero mollis velit
  venenatis porta et in odio. In ante tortor, eleifend at fringilla id, semper et ipsum. Integer vitae nisl eu
  urna cursus tincidunt id aliquam turpis. Suspendisse id tempus nunc. Quisque pharetra massa nisl, eu
  consectetur risus lacinia sit amet. Donec cursus massa ac pellentesque vulputate.",
  "reviewerName": "mister_reviewer"
}

```

Backend will respond with a 200 if successful.

PUT /helpful_review

Increases the 'helpful review' count by one. This number is used by Amazon to indicate if a review is helpful. We replicated the same capability in our system.

Frontend will send a post with body:

```
{
  "reviewerID": "A3DE6XGZ2EPADS",
  "asin": "B000F83SZQ",
  "reviewerName": "WPY"
}
```

Backend will respond with a 200 if successful.

GET /filter_books?filtertext="Job Hunting %26 Careers"

Combs through the database to find a list of books that belongs in the category matching the filter text and return the first book to the 20th book if any.

Query Params	Description
filtertext	A string, used to filter out categories

Sample return (in this example the category is "Job Hunting & Careers", which only has one book so it returns just this sample json):

```
{
  "books": [{
    "title": "Faux Leather Kindle Sleeve Case for Kindle (Fits 9.7" Display, Latest Generation Kindle DX) - Yellow (Smooth finish)",
    "asin": "B002HWRIIQ",
    "averageRating": 4.6,
    "imUrl":
"http://ecx.images-amazon.com/images/I/513mtcKv9HL._B02,204,203,200_PIsitb-sticker-v3-big,TopRight,0,-55_SX278_SY278_PIKin4,BottomRight,1,22_AA300_SH20_OU01_.jpg",
    "categories": [
      ["Books", "Business & Money", "Job Hunting & Careers", "Guides"],
      ["Books", "Business & Money", "Job Hunting & Careers", "Job Hunting"],
      ["Books", "Business & Money", "Job Hunting & Careers", "Resumes"],
      ["Kindle Store", "Kindle eBooks", "Business & Money", "Job Hunting & Careers",
"Career Guides"],
      ["Kindle Store", "Kindle eBooks", "Business & Money", "Job Hunting & Careers", "Job
Hunting"],
      ["Kindle Store", "Kindle eBooks", "Business & Money", "Job Hunting & Careers",
"Resumes"]
    ]
  }]
}
```

GET /search_books/?searchtext="Faux Leather"&filtertext=""

Do a comb through the database and give books that matches the search text as well as a list of all categories that these books belong to. If filtertext is given, only return the books that belong to that category (but still return the list of categories).

Query Params	Description
searchtext	A string, used to search for a title or ASIN
filtertext	A string, used to filter out categories

Sample return:

```
{
  "books": [{
    "title": "Faux Leather Kindle Sleeve Case for Kindle (Fits 9.7" Display, Latest
Generation Kindle DX) - Yellow (Smooth finish)",
    "asin": "B002HWRIIQ",
    "averageRating": 4.6,
    "imUrl":
"http://ecx.images-amazon.com/images/I/513mtcKv9HL._B02,204,203,200_PIsitb-sticker-v3-big,TopRight,0,-55_SX278_
SY278_PIKin4,BottomRight,1,22_AA300_SH20_OU01_.jpg",
    "categories": [
      ["Books", "Business & Money", "Job Hunting & Careers", "Guides"],
      ["Books", "Business & Money", "Job Hunting & Careers", "Job Hunting"],
      ["Books", "Business & Money", "Job Hunting & Careers", "Resumes"],
      ["Kindle Store", "Kindle eBooks", "Business & Money", "Job Hunting & Careers",
"Career Guides"],
      ["Kindle Store", "Kindle eBooks", "Business & Money", "Job Hunting & Careers", "Job
Hunting"],
      ["Kindle Store", "Kindle eBooks", "Business & Money", "Job Hunting & Careers",
"Resumes"]
    ]
  }],
  "categories": [
    {
      "category": "Job Hunting & Careers"
    }
  ]
}
```

GET /categories

Returns all categories of all books in the database.

Sample return:

```
{
  "categories": [
    {
      "category": "Christian Books & Bibles"
    },
    {
      "category": "Job Hunting & Careers"
    },
    {

```

```
        "category": "Test 3"
    },
    {
        "category": "Test 4"
    }
]
}
```

Setup for PySpark and HDFS cluster

Instructions

For the setup of the PySpark and HDFS cluster, we'll be doing it through the framework Flintrock. This guide will apply on Ubuntu VMs.

1. At your VM, install flintrock through pip
pip3 install flintrock
2. Install aws cli through pip
pip3 install aws-cli
3. Configure your aws access key and aws secret access key of your aws-cli by running the following code:
aws configure
4. Export both your aws access key and aws secret access key:
export AWS_ACCESS_KEY_ID=access_key_here
export AWS_SECRET_ACCESS_KEY=secret_access_key_here
5. Configure flintrock by editing the config file directly:
vim /home/ubuntu/.config/flintrock/config.yaml

6. Ensure that the config file looks like this:

```
services:
  spark:
    version: 2.4.4
    # git-commit: latest # if not 'latest', provide a full commit SHA; e.g. d6dc12ef0146ae409834c78737c116050961f350
    # git-repository: # optional; defaults to https://github.com/apache/spark
    # optional; defaults to download from from the official Spark S3 bucket
    #   - must contain a {v} template corresponding to the version
    #   - Spark must be pre-built
    #   - must be a tar.gz file
    # download-source: "https://www.example.com/files/spark/{v}/spark-{v}.tar.gz"
    # executor-instances: 1
  hdfs:
    version: 2.8.5
    # optional; defaults to download from a dynamically selected Apache mirror
    #   - must contain a {v} template corresponding to the version
    #   - must be a .tar.gz file
    # download-source: "https://www.example.com/files/hadoop/{v}/hadoop-{v}.tar.gz"
    # download-source: "http://www-us.apache.org/dist/hadoop/common/hadoop-{v}/hadoop-{v}.tar.gz"

provider: ec2

providers:
  ec2:
    key-name: flint_cluster
    identity-file: /home/ubuntu/Desktop/flint_cluster.pem
    instance-type: t2.medium
    region: us-east-1
    # availability-zone: <name>
    ami: ami-0b8d0d6ac70e5750c # Amazon Linux 2, us-east-1
    user: ec2-user
    # ami: ami-61bbf104 # CentOS 7, us-east-1
    # user: centos
    # spot-price: <price>
    # vpc-id: <id>
    # subnet-id: <id>
    # placement-group: <name>
    # security-groups:
    #   - group-name1
    #   - group-name2
    # instance-profile-name:
    # tags:
    #   - key1,value1
    #   - key2, value2 # leading/trailing spaces are trimmed
    #   - key3, # value will be empty
    # min-root-ebs-size-gb: <size-gb>
    tenancy: default # default | dedicated
    ebs-optimized: no # yes | no
    instance-initiated-shutdown-behavior: terminate # terminate | stop
    # user-data: /path/to/userdata/script

launch:
  num-slaves: 8
  install-hdfs: True
  install-spark: True

debug: false
```

Under “launch”, the number you put next to “num-slaves” specifies the number of slaves for the cluster. We’ll require you to change that value to 2 and 5 later on in accordance to our project requirements.

7. And that’s it! Run “flintrock launch spark-hdfs-cluster” to launch your cluster.

Debugging

- Exporting the environment variable is required as in some cases aws-cli would fail to authenticate with the configuration in merely aws configure.

- Flintrock performs better on local VMs than on remote cloud instances. It is recommended that it is run on a local VM instead.