综合实验指导手册：实现推荐系统

软件版本：

- Java：java-1.8.0
- Python：python-3.8.5
- Scala：scala-2.11.8
- Flume：apach-flume-1.9.0
- Kafka：kafka_2.13-2.6.1
- Hadoop：hadoop-2.10.1
- HBase：hbase-2.3.4
- Spark：spark-2.4.7
- Redis：redis-6.0.6

实验指导：

本次实验提供一套示例代码，包括了整个推荐系统中的基本组件。具体说明如下：

代码说明：

（1）启动 HBase，并配置 HBase Thrift 连接，以便于步骤二中使用 python 的 happybase 库连接 HBase：关于 thrift API 的配置可以在官网中查阅到：

http://hbase.apache.org/apidocs/org/apache/hadoop/hbase/thrift/package-summary.html#package_description

（2）load_train_ratings_hbase.py：将 json_train_ratings.json 作为历史数据直接存储到 HBase 中

（3）load_movie_redis.py：将 movies.csv 中存储的 movieId 与 movie title 的映射关系写入 Redis 中，用于后续推荐时从 movieId 转换成 movie title

（4）generatorRecord.py：将 json_test_ratings.json 作为实时数据发送到 Flume 中

（5）recommend_server.py 与 recommend_client.py：recommend_server 等待 recommend_client 连接，连接后 recommend_client 可以向 recommend_server 发送一个 userId，recommend_server 根据推荐算法从 Redis 中获取召回列表、特征参数与模型参数后，向 recommend_client 返回推荐列表。

（6）hbase2spark.scala：批式计算程序，计算后将结果写入 Redis

（7）kafkaStreaming.scala：流式计算程序，计算后将结果写入 Redis

（8）recommend.scala：推荐模型程序，计算后将结果写入 Redis

注：示例代码仅供参考，需要更改其中与配置相关的内容，如 IP、HBase 表名、Kafka 的 Topic 名、程序定期启动时间等；

运行流程：

（1）启动 load_train_ratings_hbase.py

python3 load_train_ratings_hbase.py bd 9090 "test_movie_records" "./data/json_train_ratings.json"

（2）启动 load_movie_redis.py

python3 load_movie_redis.py bd 6379 "./data/movies.csv"

（3）开启 Flume

（4）启动 Kafka 并创建 Kafka Topic

（5）启动 Redis

（6）启动 generatorRecord.py

python generatorRecord.py -h bd -p 44444 -f ./data/json_test_ratings.json

{'headers': {'key': '1266504109rating'}, 'body': '{"userId": 390, "movieId": 1193, "rating": 1.0, "timestamp": 1266504109}'}
{'headers': {'key': '1266532488rating'}, 'body': '{"userId": 68, "movieId": 69069, "rating": 1.0, "timestamp": 1266532488}'}
{'headers': {'key': '1266609746rating'}, 'body': '{"userId": 68, "movieId": 71535, "rating": 1.0, "timestamp": 1266609746}'}
{'headers': {'key': '1266626448rating'}, 'body': '{"userId": 573, "movieId": 48385, "rating": 1.0, "timestamp": 1266626448}'}
{'headers': {'key': '1266789678rating'}, 'body': '{"userId": 253, "movieId": 2324, "rating": 1.0, "timestamp": 1266789678}'}
{'headers': {'key': '1266894558rating'}, 'body': '{"userId": 381, "movieId": 71057, "rating": 1.0, "timestamp": 1266894558}'}
{'headers': {'key': '1266929102rating'}, 'body': '{"userId": 434, "movieId": 1196, "rating": 1.0, "timestamp": 1266929102}'}
{'headers': {'key': '1267152883rating'}, 'body': '{"userId": 22, "movieId": 318, "rating": 1.0, "timestamp": 1267152883}'}
{'headers': {'key': '1267249501rating'}, 'body': '{"userId": 22, "movieId": 60950, "rating": 0.0, "timestamp": 1267249501}'}
{'headers': {'key': '1267450587rating'}, 'body': '{"userId": 22, "movieId": 49272, "rating": 1.0, "timestamp": 1267450587}'}
{'headers': {'key': '1267523616rating'}, 'body': '{"userId": 464, "movieId": 1210, "rating": 1.0, "timestamp": 1267523616}'}
{'headers': {'key': '1267656389rating'}, 'body': '{"userId": 434, "movieId": 541, "rating": 1.0, "timestamp": 1267656389}'}
{'headers': {'key': '1267678698rating'}, 'body': '{"userId": 483, "movieId": 68954, "rating": 1.0, "timestamp": 1267678698}'}
{'headers': {'key': '1267809852rating'}, 'body': '{"userId": 335, "movieId": 762, "rating": 0.0, "timestamp": 1267809852}'}
{'headers': {'key': '1267830936rating'}, 'body': '{"userId": 330, "movieId": 1732, "rating": 1.0, "timestamp": 1267830936}'}
{'headers': {'key': '1267844634rating'}, 'body': '{"userId": 590, "movieId": 1198, "rating": 1.0, "timestamp": 1267844634}'}
{'headers': {'key': '1267861691rating'}, 'body': '{"userId": 464, "movieId": 318, "rating": 1.0, "timestamp": 1267861691}'}
{'headers': {'key': '1267878801rating'}, 'body': '{"userId": 434, "movieId": 2571, "rating": 1.0, "timestamp": 1267878801}'}
{'headers': {'key': '1267914054rating'}, 'body': '{"userId": 590, "movieId": 904, "rating": 1.0, "timestamp": 1267914054}'}
{'headers': {'key': '1268184922rating'}, 'body': '{"userId": 454, "movieId": 7116, "rating": 1.0, "timestamp": 1268184922}'}
{'headers': {'key': '1268546602rating'}, 'body': '{"userId": 434, "movieId": 5378, "rating": 0.0, "timestamp": 1268546602}'}
{'headers': {'key': '1268561583rating'}, 'body': '{"userId": 232, "movieId": 69757, "rating": 1.0, "timestamp": 1268561583}'}
{'headers': {'key': '1268725898rating'}, 'body': '{"userId": 215, "movieId": 50, "rating": 1.0, "timestamp": 1268725898}'}
{'headers': {'key': '1268919860rating'}, 'body': '{"userId": 330, "movieId": 288, "rating": 1.0, "timestamp": 1268919860}'}
{'headers': {'key': '1268991646rating'}, 'body': '{"userId": 605, "movieId": 1210, "rating": 1.0, "timestamp": 1268991646}'}
{'headers': {'key': '1269685661rating'}, 'body': '{"userId": 288, "movieId": 57669, "rating": 1.0, "timestamp": 1269685661}'}
{'headers': {'key': '1269690495rating'}, 'body': '{"userId": 464, "movieId": 593, "rating": 1.0, "timestamp": 1269690495}'}
{'headers': {'key': '1269815207rating'}, 'body': '{"userId": 434, "movieId": 3863, "rating": 0.0, "timestamp": 1269815207}'}
{'headers': {'key': '1269877852rating'}, 'body': '{"userId": 434, "movieId": 741, "rating": 1.0, "timestamp": 1269877852}'}
{'headers': {'key': '1269966043rating'}, 'body': '{"userId": 253, "movieId": 4973, "rating": 1.0, "timestamp": 1269966043}'}
{'headers': {'key': '1270089033rating'}, 'body': '{"userId": 330, "movieId": 1923, "rating": 1.0, "timestamp": 1270089033}'}
{'headers': {'key': '1270156949rating'}, 'body': '{"userId": 590, "movieId": 2338, "rating": 0.0, "timestamp": 1270156949}'}
{'headers': {'key': '1270184383rating'}, 'body': '{"userId": 454, "movieId": 904, "rating": 1.0, "timestamp": 1270184383}'}
{'headers': {'key': '1270191619rating'}, 'body': '{"userId": 522, "movieId": 68205, "rating": 0.0, "timestamp": 1270191619}'}
{'headers': {'key': '1270355105rating'}, 'body': '{"userId": 78, "movieId": 6537, "rating": 0.0, "timestamp": 1270355105}'}
{'headers': {'key': '1270386617rating'}, 'body': '{"userId": 434, "movieId": 2858, "rating": 1.0, "timestamp": 1270386617}'}
{'headers': {'key': '1270482752rating'}, 'body': '{"userId": 414, "movieId": 71535, "rating": 1.0, "timestamp": 1270482752}'}
{'headers': {'key': '1270725570rating'}, 'body': '{"userId": 522, "movieId": 40148, "rating": 1.0, "timestamp": 1270725570}'}

（7）启动 hbase2spark.scala、kafkaStreaming.scala、recommend.scala

（8）启动 recommend_server.py

python3 recommend_server.py "bd" 6379 23456

base > python3 recommend_server.py "bd" 6379 23456
[(318, 0.9811299999541332), (2959, 0.9639364843974809), (858, 0.9557722322844537), (858, 0.9557722322844537), (3114, 0.9301126058152407), (6377, 0.9051398924056727), (8580, 0.893587645317098), (59387, 0.8872557791056576), (38038, 0.8803205972101956), (2067, 0.8761623682959341), (59784, 0.8692124979641218), (370, 0.8563072042263599), (7669, 0.8527455465154442), (5989, 0.8514554213310371), (80831, 0.8494479097346284), (99114, 0.8433668606156731), (364, 0.8410304144837444), (4816, 0.8075374845613453), (1394, 0.802389254373134), (2739, 0.795540827353193), (27803, 0.7912756976076144), (1298, 0.7658867085477851), (589, 0.754380647866059), (296, 0.7377313414662906), (296, 0.7377313414662906), (61348, 0.728088920382693), (1201, 0.7097180823019783), (918, 0.6538864355025186), (65188, 0.6311937957890522), (3037, 0.587509634260148), (551, 0.5566520537744689)]
[INFO]userId:51, list:[318, 2959, 858, 858, 3114, 6377, 8580, 59387, 38038, 2067]
[(318, 0.9958856137219882), (858, 0.9901577305191877), (858, 0.9901577305191877), (2959, 0.9847650307842487), (3114, 0.9712031069332193), (6377, 0.9697113323241764), (2067, 0.9638783664490663), (5989, 0.9638783664490663), (370, 0.9610597503399089), (4816, 0.9610597503399089), (38038, 0.9610597503399089), (918, 0.960074148902 9206), (59387, 0.9600741489029206), (1394, 0.9598137665231422), (2739, 0.9590646663849441), (7669, 0.9590646663849441), (27803, 0.9580307707426491), (80831, 0.958030707426491), (99114, 0.9580307707426491), (8580, 0.9580307707426491), (59784, 0.9580307707426491), (589, 0.9580307707426491), (65188, 0.9580307707426491), (3037, 0.9521687529489351), (1298, 0.947198871129791), (364, 0.9417441903058404), (1201, 0.9417441903058404), (61348, 0.9403000694784972), (551, 0.9110868478294171), (296, 0.8026082695199378), (296, 0.8026082695199378)]
[INFO]userId:123, list:[318, 858, 858, 2959, 3114, 6377, 2067, 5989, 370, 4816]
[(318, 0.9958856137219882), (858, 0.9901577305191877), (858, 0.9901577305191877), (2959, 0.9847650307842487), (3114, 0.9712031069332193), (6377, 0.9697113323241764), (2067, 0.9638783664490663), (5989, 0.9638783664490663), (370, 0.9610597503399089), (38038, 0.9610597503399089), (918, 0.960074148902 9206), (59387, 0.9600741489029206), (1394, 0.9598137665231422), (2739, 0.9590646663849441), (7669, 0.9590646663849441), (27803, 0.9580307707426491), (80831, 0.958030707426491), (8580, 0.9580307707426491), (59784, 0.9580307707426491), (589, 0.9580307707426491), (65188, 0.9580307707426491), (3037, 0.9521687529489351), (1298, 0.947198871129791), (364, 0.9417441903058404), (1201, 0.9417441903058404), (61348, 0.9403000694784972), (551, 0.9110868478294171), (296, 0.8026082695199378), (296, 0.8026082695199378)]
[INFO]userId:2333, list:[318, 858, 858, 2959, 3114, 6377, 2067, 5989, 370, 4816]
[(2959, 0.985554283206574), (318, 0.985309144979858508), (3114, 0.9802879644325283), (6377, 0.9696605143488157), (8580, 0.9695447036141813), (858, 0.965369143136358 5), (858, 0.9653691431363585), (59784, 0.9622273394875036), (370, 0.9609950014132327), (38038, 0.9609950007030447), (589, 0.9480099591560471), (59387, 0.9420337441 975971), (1201, 0.9281264392580812), (99114, 0.9211089000372183), (4816, 0.9099212169503195), (1394, 0.9071969429622464), (364, 0.8973556229143482), (5989, 0.88086 64281715951), (2067, 0.8748933377148739), (80831, 0.869920505964574), (61348, 0.865708261626087), (7669, 0.8530971086341902), (2739, 0.8308868319299505), (296, 0.8275159933046099), (296, 0.8275159933046099), (27803, 0.8271988805754923), (1298, 0.7990973812909724), (918, 0.7923781322343861), (65188, 0.7740900136614378), (3037, 0.7492603745228592), (551, 0.7286611264191808)]
[INFO]userId:520, list:[2959, 318, 3114, 6377, 858, 858, 59784, 370, 38038]

（9）python3 recommend_client.py 127.0.0.1 23456

```
base ) python3 recommend_client.py 127.0.0.1 23456
51
Recommend_List:[userid=51]
[Movie-1     : Shawshank Redemption, The (1994)]
[Movie-2     : Fight Club (1999)]
[Movie-3     : Godfather, The (1972)]
[Movie-4     : Godfather, The (1972)]
[Movie-5     : Toy Story 2 (1999)]
[Movie-6     : Finding Nemo (2003)]
[Movie-7     : Into the Woods (1991)]
[Movie-8     : Fall, The (2006)]
[Movie-9     : Wallace & Gromit in The Curse of the Were-Rabbit (2005)]
[Movie-10    : Doctor Zhivago (1965)]
123
Recommend_List:[userid=123]
[Movie-1     : Shawshank Redemption, The (1994)]
[Movie-2     : Godfather, The (1972)]
[Movie-3     : Godfather, The (1972)]
[Movie-4     : Fight Club (1999)]
[Movie-5     : Toy Story 2 (1999)]
[Movie-6     : Finding Nemo (2003)]
[Movie-7     : Doctor Zhivago (1965)]
[Movie-8     : Catch Me If You Can (2002)]
[Movie-9     : Naked Gun 33 1/3: The Final Insult (1994)]
[Movie-10    : Zoolander (2001)]
2333
Recommend_List:[userid=2333]
[Movie-1     : Shawshank Redemption, The (1994)]
[Movie-2     : Godfather, The (1972)]
[Movie-3     : Godfather, The (1972)]
[Movie-4     : Fight Club (1999)]
[Movie-5     : Toy Story 2 (1999)]
[Movie-6     : Finding Nemo (2003)]
[Movie-7     : Doctor Zhivago (1965)]
[Movie-8     : Catch Me If You Can (2002)]
[Movie-9     : Naked Gun 33 1/3: The Final Insult (1994)]
[Movie-10    : Zoolander (2001)]
520
Recommend_List:[userid=520]
[Movie-1     : Fight Club (1999)]
[Movie-2     : Shawshank Redemption, The (1994)]
[Movie-3     : Toy Story 2 (1999)]
[Movie-4     : Finding Nemo (2003)]
[Movie-5     : Into the Woods (1991)]
[Movie-6     : Godfather, The (1972)]
[Movie-7     : Godfather, The (1972)]
[Movie-8     : Kung Fu Panda (2008)]
[Movie-9     : Naked Gun 33 1/3: The Final Insult (1994)]
[Movie-10    : Wallace & Gromit in The Curse of the Were-Rabbit (2005)]
```