

# 巨量資料探勘與應用

## HW4

### 一、組員名單與分工

資工碩一 112598017 楊淨雯

資工碩一 112598028 余珮綺

### 二、環境設定

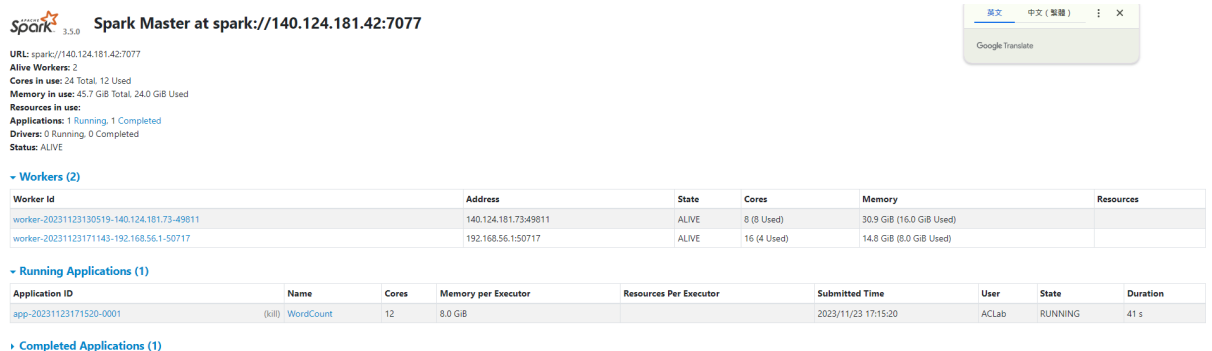
#### Multi-node on spark

master 和 worker 皆為 windows, 一樣都要先照之前的作業把 python、spark、hadoop 安裝好

master: 140.124.181.42

worker 1: 192.168.56.1, 4 cores, 8GB Used

worker 2: 140.124.181.73, 8 cores, 16GB Used



The screenshot shows the Spark Master web interface at `spark://140.124.181.42:7077`. It displays the following information:

- URL:** `spark://140.124.181.42:7077`
- Alive Workers:** 2
- Cores in use:** 24 Total, 12 Used
- Memory in use:** 45.7 GiB Total, 24.0 GiB Used
- Resources in use:**
- Applications:** 1 Running, 1 Completed
- Drivers:** 0 Running, 0 Completed
- Status:** ALIVE

**Workers (2)**

Worker Id	Address	State	Cores	Memory	Resources
worker-20231123130519-140.124.181.73-49811	140.124.181.73:49811	ALIVE	8 (8 Used)	30.9 GiB (16.0 GiB Used)	
worker-20231123171143-192.168.56.1-50717	192.168.56.1:50717	ALIVE	16 (4 Used)	14.8 GiB (8.0 GiB Used)	

**Running Applications (1)**

Application ID	Name	Cores	Memory per Executor	Resources Per Executor	Submitted Time	User	State	Duration
app-20231123171520-0001	(kill) WordCount	12	8.0 GiB		2023/11/23 17:15:20	ACLab	RUNNING	41 s

**Completed Applications (1)**

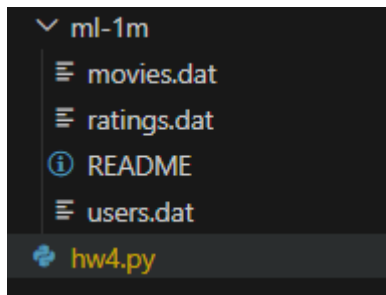
### 三、實驗執行說明

#### Step1. 連線

- Master:  
以 Admin 開啟 cmd, 到 `spark \ bin` 資料夾, 輸入指令:  
`spark-class org.apache.spark.deploy.master.Master`  
直到出現 Alive 即可
- worker:  
以 Admin 開啟 cmd, 到 `spark \ bin` 資料夾, 輸入指令:  
`spark-class org.apache.spark.deploy.worker.Worker spark://<master_ip>:7077`

## Step2. 執行

- 確保指定路徑中的資料夾裡有以下這些檔案



- 打開 cmd，進入spark \ bin 資料夾
- 輸入 spark-submit --master spark://http://<master\_ip>:7077 --executor-cores 4 --executor-memory 8g <code\_file\_name>

附註：

在我們的電腦上

<master\_ip> 為 140.124.181.42

<code\_file\_name> 為 hw4.py

畫底線部分為分配worker資源

#### 四、實驗結果

##### 1. The top-rated movies by all users

MovieID	avg(Rating)
3280	5.0
989	5.0
787	5.0
3881	5.0
1830	5.0
3656	5.0
3607	5.0
3172	5.0
3233	5.0
3382	5.0
3245	4.8
53	4.75
2503	4.666666666666667
2905	4.608695652173913
2019	4.560509554140127
318	4.554557700942973
858	4.524966261808367
745	4.52054794520548
50	4.517106001121705
527	4.510416666666667

only showing top 20 rows

2.

The top-rated movies grouped by		
gender	age group	occupation
<pre> +-----+-----+-----+  MovieID Gender avg(Rating)  +-----+-----+-----+   3292     F       5.0    3641     F       5.0     989     M       5.0     854     F       5.0     787     M       5.0    3280     M       5.0    2675     F       5.0    3817     F       5.0    2129     F       5.0    1780     F       5.0    3233     M       5.0     787     F       5.0    3656     M       5.0     681     F       5.0    3517     M       5.0    3881     F       5.0    2332     F       5.0    3012     F       5.0    3245     F       5.0    1830     M       5.0  +-----+-----+-----+ only showing top 20 rows </pre>	<pre> +-----+---+-----+  MovieID Age avg(Rating)  +-----+---+-----+    696  25       5.0    1859  45       5.0     117   1       5.0     613   1       5.0    3338  45       5.0    1421  45       5.0    2624   1       5.0    1286   1       5.0     659  18       5.0    3881  18       5.0     358  50       5.0    1780  25       5.0    1354   1       5.0    2512  50       5.0    2166   1       5.0    1471  56       5.0    2215  18       5.0     687  50       5.0     299   1       5.0    1585  50       5.0  +-----+---+-----+ only showing top 20 rows </pre>	<pre> +-----+-----+-----+  MovieID Occupation avg(Rating)  +-----+-----+-----+    669       4       5.0    1423      13       5.0    2935      11       5.0    1827      10       5.0     334      13       5.0    3742      18       5.0    3894      20       5.0    2731      18       5.0    1795      12       5.0    3570       3       5.0    1817       6       5.0     853      17       5.0    3307      19       5.0    3093       8       5.0     423       8       5.0     119      13       5.0    1257       8       5.0    3853       3       5.0    1212       9       5.0    2727      11       5.0  +-----+-----+-----+ only showing top 20 rows </pre>

3.

The average rating score of each user for all movies.	The average rating score grouped by genre.
<pre> +-----+-----+  UserID       avg(Rating)  +-----+-----+    283  4.962962962962963    2339  4.956521739130435    3324  4.904761904761905    3902  4.890909090909091     446  4.8431372549019605     447  4.837837837837838    4649  4.818181818181818    4634  4.813725490196078    1131  4.796116504854369    4925  4.761904761904762    5069            4.76    4755            4.76    4801  4.73469387755102     682  4.733333333333333    1670  4.714285714285714      91  4.704545454545454    5768  4.702702702702703    3461  4.702702702702703    5984            4.7    1021  4.694656488549619  +-----+-----+ only showing top 20 rows </pre>	<pre> +-----+-----+-----+-----+  UserID       Genres avg(Rating)  +-----+-----+-----+-----+   1860  Film-Noir Sci-Fi            5.0     438  Film-Noir Sci-Fi            5.0    2736  Action Drama War            5.0    1138    Crime Thriller            5.0    1063  Film-Noir Sci-Fi            5.0    1186  Action Adventure ...            5.0     967   Comedy Romance War            5.0    1765  Action Comedy Sci...            5.0    2928  Action Adventure ...            5.0    1905            Comedy Crime            5.0    2993            Crime Drama            5.0    1374  Adventure Childre...            5.0    2261  Children's Comedy...            5.0    2933  Children's Comedy...            5.0    2762  Action Crime Dram...            5.0    2062            Comedy Crime            5.0    1389  Drama Mystery Sci...            5.0    2232  Children's Comedy...            5.0    2480  Animation Comedy            5.0    1035  Adventure Childre...            5.0  +-----+-----+-----+-----+ only showing top 20 rows </pre>

#### 4. The cosine similarity of user

UserPair	CosineSimilarity
[2339, 283]	0.9983683
[283, 2339]	0.9983683
[3324, 283]	0.9852565
[3902, 283]	0.9817473
[446, 283]	0.96964574
[447, 283]	0.9683033
[4649, 283]	0.96332407
[4634, 283]	0.96219516
[1131, 283]	0.95773447
[4925, 283]	0.94906795
[4755, 283]	0.94858545
[5069, 283]	0.94858545
[4801, 283]	0.9421749
[682, 283]	0.9418303
[1670, 283]	0.9370051
[91, 283]	0.9345377
[3461, 283]	0.9340709
[5768, 283]	0.9340709
[5984, 283]	0.93338627
[1021, 283]	0.93203264

only showing top 20 rows

#### 5. The cosine similarity of movies

MoviePair	CosineSimilarity
[1830, 3280]	1.0
[3172, 3280]	1.0
[3233, 3280]	1.0
[3280, 989]	1.0
[3382, 3280]	1.0
[3607, 3280]	1.0
[3656, 3280]	1.0
[3881, 3280]	1.0
[787, 3280]	1.0
[989, 3280]	1.0
[3245, 3280]	0.95
[53, 3280]	0.9375
[2503, 3280]	0.9166667
[2905, 3280]	0.90217394
[2019, 3280]	0.89012736
[318, 3280]	0.88863945
[858, 3280]	0.88124156
[745, 3280]	0.88013697
[50, 3280]	0.8792765
[527, 3280]	0.8776042

only showing top 20 rows

6.

### (a) For user-based collaborative filtering

Top 10 movie recommendations for each user:	
userId	recommendations
1	[[{2562, 5.9895396}, {1741, 5.951258}, {1652, 5.8736134}, {309, 5.7896323}, {3282, 5.780738}, {572, 5.759351}, {2569, 5.754052}, {2388, 5.75286}, {3913, 5.71182}, {2636, 5.6748824}]]
3	[[{2998, 9.52328}, {37, 7.386846}, {3540, 7.173025}, {2831, 6.8575826}, {59, 6.5901423}, {1038, 6.5838003}, {2129, 6.404386}, {324, 6.395917}, {1546, 6.3517885}, {2843, 6.3202667}]]
5	[[{2962, 7.9037905}, {3490, 6.9524198}, {563, 6.9081903}, {1002, 6.767595}, {1570, 6.599475}, {2512, 6.4051123}, {3817, 5.943491}, {3749, 5.8941536}, {841, 5.867899}, {652, 5.855594}]]
6	[[{3092, 8.244707}, {2964, 7.6508636}, {649, 7.5498037}, {309, 7.325001}, {1725, 6.9006248}, {1780, 6.8097982}, {128, 6.697167}, {84, 6.625237}, {3670, 6.558164}, {831, 6.540138}]]
9	[[{2998, 7.5951614}, {1530, 6.7288346}, {2129, 6.489753}, {3540, 6.473192}, {1038, 6.1323185}, {572, 5.7809787}, {324, 5.728103}, {37, 5.6469364}, {853, 5.6444993}, {1651, 5.6336923}]]
12	[[{563, 10.616097}, {1685, 9.66391}, {2911, 9.4671135}, {3902, 9.234115}, {2388, 9.186652}, {1471, 9.180959}, {1864, 8.822442}, {974, 8.583023}, {1421, 8.560574}, {3116, 8.484338}]]
13	[[{1471, 6.887922}, {2964, 5.6233892}, {2911, 5.4988723}, {769, 5.3444147}, {572, 5.3410974}, {1725, 5.3048415}, {1652, 5.277052}, {2776, 5.1980643}, {1421, 5.077596}, {2175, 4.9496574}]]
15	[[{2964, 6.275356}, {2998, 6.1744804}, {2834, 5.8034167}, {572, 5.6424155}, {1749, 5.5237136}, {2911, 5.4756413}, {3288, 5.4678483}, {3474, 5.4521055}, {1471, 5.321652}, {324, 5.272166}]]
16	[[{411, 9.647766}, {2626, 6.8232584}, {956, 6.581363}, {2129, 6.5704226}, {831, 6.497583}, {572, 6.470781}, {131, 6.3306017}, {2345, 6.321882}, {3382, 6.280855}, {324, 6.257065}]]
17	[[{1369, 5.971708}, {2129, 5.8389363}, {3517, 5.660597}, {572, 5.6511197}, {751, 5.5909934}, {3817, 5.572334}, {2673, 5.526473}, {1859, 5.521916}, {649, 5.510001}, {3092, 5.503738}]]
19	[[{2767, 7.4834557}, {974, 6.953071}, {2305, 6.1446674}, {2675, 6.1134286}, {3282, 5.731082}, {3288, 5.650475}, {1685, 5.6129503}, {2897, 5.5871077}, {864, 5.5864935}, {2175, 5.5423927}]]
20	[[{2962, 9.565894}, {49, 8.146181}, {3749, 8.128284}, {3073, 7.88283}, {2482, 7.5537405}, {1000, 7.3272}, {2893, 7.246355}, {1312, 7.2048388}, {2562, 7.1327734}, {2127, 7.1262197}]]
22	[[{2962, 5.538701}, {2538, 5.3305464}, {2129, 5.3264723}, {563, 5.2791376}, {1539, 5.247116}, {2593, 5.1268477}, {72, 5.114027}, {3415, 5.094519}, {2493, 5.0886374}, {406, 5.085902}]]
26	[[{2998, 6.783158}, {2332, 6.3079543}, {2964, 6.0518327}, {119, 5.8874464}, {1846, 5.839117}, {128, 5.588673}, {681, 5.554248}, {618, 5.4396653}, {3539, 5.424505}, {240, 5.28087}]]
27	[[{1859, 9.965247}, {1369, 9.550068}, {2304, 9.13704}, {2211, 9.134435}, {2850, 8.579463}, {2931, 8.505317}, {864, 8.076642}, {3670, 8.00401}, {1423, 7.9315057}, {682, 7.917115}]]
28	[[{2962, 5.4255004}, {2512, 5.344499}, {3490, 5.286643}, {1000, 4.9596405}, {567, 4.934673}, {824, 4.9144707}, {3025, 4.895481}, {1002, 4.893924}, {3749, 4.887681}, {1117, 4.826899}]]
31	[[{1471, 6.6868196}, {2984, 6.6850696}, {572, 6.5806323}, {958, 6.472582}, {864, 6.467574}, {563, 6.410024}, {3924, 6.180409}, {1000, 6.0294595}, {600, 6.0125284}, {3832, 5.7518363}]]
34	[[{2998, 7.782036}, {1546, 7.2434897}, {572, 6.933036}, {2911, 6.7967863}, {3474, 6.7803135}, {2175, 6.7393584}, {324, 6.603598}, {2843, 6.2805953}, {877, 6.205108}, {1471, 6.208177}]]
35	[[{649, 8.728915}, {3092, 8.493096}, {2063, 7.8045044}, {1572, 7.361656}, {1859, 7.250364}, {3670, 6.960365}, {3862, 6.934491}, {617, 6.902969}, {1809, 6.8458133}, {2304, 6.5922084}]]
37	[[{3092, 10.024157}, {2897, 9.538182}, {1809, 8.6397295}, {411, 8.638151}, {649, 8.206267}, {2197, 8.044944}, {2767, 7.819145}, {2305, 7.7028623}, {1780, 7.70095}, {3711, 7.4903855}]]
only showing top 20 rows	

Top 10 movie recommendations for specified users:	
userId	recommendations
2453	[[{2964, 7.302004}, {1749, 6.0950556}, {2998, 6.047509}, {572, 5.918757}, {649, 5.864108}, {128, 5.7825217}, {1471, 5.444547}, {1546, 5.4345274}, {682, 5.3841143}, {1038, 5.366028}]]
26	[[{2998, 6.783158}, {2332, 6.3079543}, {2964, 6.0518327}, {119, 5.8874464}, {1846, 5.839117}, {128, 5.588673}, {681, 5.554248}, {618, 5.4396653}, {3539, 5.424505}, {240, 5.28087}]]
1697	[[{572, 6.8836513}, {1741, 6.591153}, {1652, 6.292038}, {831, 6.1373634}, {309, 6.1024213}, {958, 5.982988}, {324, 5.916804}, {2964, 5.847962}, {853, 5.811762}, {2562, 5.806903}]]

### (b) For item-based collaborative filtering

Top 10 user recommendations for each movie:	
movieId	recommendations
1	[[{2037, 6.821345}, {1371, 6.302769}, {41, 5.887873}, {4936, 5.882536}, {1573, 5.870336}, {1341, 5.813968}, {2151, 5.7188945}, {1444, 5.6459208}, {602, 5.643558}, {1751, 5.6116085}]]
3	[[{2441, 6.554672}, {1664, 6.468721}, {1346, 6.223833}, {5525, 6.0967875}, {1702, 5.787179}, {3915, 5.735427}, {3061, 5.714433}, {5810, 5.707639}, {4660, 5.7065606}, {430, 5.62558}]]
5	[[{2441, 6.1445622}, {2697, 5.4959936}, {1282, 5.404556}, {2364, 5.3857307}, {5052, 5.351468}, {1574, 5.3459406}, {4068, 5.286964}, {775, 5.266263}, {4719, 5.2603784}, {1070, 5.213218}]]
6	[[{3076, 5.7834663}, {3149, 5.612901}, {2534, 5.5497193}, {2819, 5.442894}, {2530, 5.4209614}, {794, 5.377537}, {1664, 5.3613334}, {1715, 5.347394}, {5003, 5.3445616}, {5760, 5.300594}]]
9	[[{89, 7.918287}, {1460, 7.168865}, {4419, 6.948296}, {14, 6.802034}, {1102, 6.7448463}, {5052, 6.7115717}, {1070, 6.626362}, {4504, 6.502143}, {5756, 6.490975}, {3911, 6.2652836}]]
12	[[{4504, 7.152516}, {1596, 6.4984164}, {5051, 6.3884263}, {3021, 6.360471}, {4392, 6.0491896}, {1039, 5.7294545}, {5991, 5.4460626}, {1076, 5.4410977}, {5218, 5.424595}, {1659, 5.4219}]]
13	[[{14, 8.198478}, {642, 8.141263}, {1310, 7.67798}, {43, 6.968624}, {2697, 6.7882013}, {4230, 6.7643}, {4504, 6.751701}, {4776, 6.6498485}, {2358, 6.576644}, {1282, 6.386712}]]
15	[[{4244, 8.495516}, {5901, 8.3357115}, {5051, 7.9994307}, {1596, 7.6904893}, {4504, 7.5534525}, {5072, 7.1778035}, {5199, 6.96901}, {1070, 6.8449244}, {2138, 6.8346796}, {5863, 6.6524}]]
16	[[{3076, 6.3786383}, {5756, 5.9701347}, {2534, 5.8269315}, {4698, 5.71746}, {2735, 5.6501126}, {4703, 5.64637}, {5670, 5.635728}, {1664, 5.624678}, {4142, 5.6157327}, {2530, 5.6055264}]]
17	[[{1213, 7.863852}, {41, 6.410016}, {807, 6.0389833}, {2640, 6.0281434}, {5944, 6.018391}, {4880, 5.996509}, {5810, 5.9012885}, {784, 5.8966966}, {3856, 5.8965826}, {5590, 5.8513203}]]
19	[[{2108, 6.4799757}, {5320, 6.2545753}, {2490, 5.910975}, {315, 5.719739}, {4583, 5.529673}, {4660, 5.493388}, {1573, 5.4889503}, {4936, 5.379226}, {3373, 5.3690577}, {2992, 5.3450704}]]
20	[[{1070, 7.2797747}, {491, 6.4836345}, {5297, 6.2587852}, {2141, 6.1143246}, {4209, 6.0742097}, {4533, 5.800133}, {1754, 5.7279086}, {653, 5.71869}, {189, 5.7065773}, {2618, 5.605971}]]
22	[[{1070, 5.971113}, {14, 5.8458166}, {484, 5.641631}, {89, 5.5652137}, {4419, 5.524737}, {2867, 5.4452486}, {775, 5.357733}, {2697, 5.3527865}, {283, 5.319575}, {2358, 5.307614}]]
26	[[{2640, 6.659348}, {14, 7.8097396}, {5901, 7.4724245}, {5590, 7.3059154}, {3176, 7.108586}, {1953, 7.049055}, {484, 7.0255804}, {5944, 6.97075}, {2532, 6.964947}, {5943, 6.951067}]]
27	[[{3061, 8.286673}, {217, 7.4849424}, {1319, 7.342832}, {540, 7.285132}, {5527, 7.231309}, {2618, 6.9662614}, {775, 6.7356157}, {2441, 6.721264}, {315, 6.625263}, {5274, 6.573148}]]
28	[[{41, 7.0384064}, {2640, 6.994604}, {2037, 6.663853}, {5901, 6.5695205}, {3856, 6.5052896}, {4880, 6.4782395}, {4504, 6.461916}, {917, 6.323841}, {5328, 6.214004}, {1111, 6.169469}]]
31	[[{4419, 6.8068824}, {484, 6.2592874}, {14, 6.2419543}, {814, 5.9135547}, {2560, 5.6714525}, {2884, 5.5992403}, {653, 5.574177}, {5756, 5.557227}, {775, 5.5166297}, {3448, 5.506696}]]
34	[[{2037, 7.1163692}, {1213, 6.4593277}, {41, 6.1734037}, {1371, 6.1732216}, {4936, 5.9181013}, {602, 5.8063493}, {4880, 5.776233}, {1039, 5.72553}, {1081, 5.699471}, {1573, 5.5845985}]]
35	[[{14, 8.17497}, {4996, 7.904059}, {932, 7.3004146}, {4504, 7.2573757}, {2640, 7.0974727}, {2133, 6.846909}, {700, 6.7669525}, {761, 6.686885}, {1566, 6.5985003}, {228, 6.515929}]]
37	[[{1070, 11.175665}, {4028, 10.752685}, {1808, 10.570334}, {3176, 10.225843}, {5204, 10.154271}, {3634, 9.930752}, {14, 9.769791}, {785, 9.764757}, {484, 9.7388735}, {3932, 9.681393}]]
only showing top 20 rows	

Top 10 user recommendations for specified movies:	
movieId	recommendations
3764	[[{4504, 6.482575}, {540, 5.930874}, {814, 5.676044}, {1070, 5.6243505}, {1460, 5.6211567}, {1145, 5.4800053}, {283, 5.3739357}, {4838, 5.349323}, {5973, 5.349081}, {4398, 5.3311133}]]
3506	[[{527, 6.712149}, {2160, 6.114269}, {5358, 6.001586}, {2151, 5.9838405}, {1083, 5.963216}, {98, 5.836952}, {3373, 5.7859063}, {4676, 5.7270174}, {1664, 5.660566}, {4102, 5.6206856}]]
29	[[{540, 7.4707375}, {5810, 7.2048573}, {1346, 6.8133264}, {4996, 6.418278}, {1452, 6.3660746}, {3856, 6.311527}, {378, 6.177482}, {5119, 6.172873}, {5313, 6.1708555}, {1703, 6.1617346}]]