

Viscosity黏度相关

1.线上MR黏度批处理

每天计算当前的前30天的总播放时长top100以及占比情况

- (1) 通过shell脚本控制产生今天前的30天的数据目录
- (2) 通过mr接参决定输入路径
- (3) 结果分为直播、点播|playlength、zb(pl/set(mac))
- (4) 结果reduce输出到hdfs的两个不同文件中，通过sqoop导入150的mysql
- (5) mysql中通过天数切表，每天访问不同表展示

热播节目top100				
直播&占比				
直播&播放时长				
点播&占比				
点播&播放时长				
查询日期 2016-12-04				
合作商 AH_CMCC				
频道 CCTV-1-HD				
查询				
排名	合作商	频道	节目名	占比
1	AH_CMCC	CCTV-1-HD	电视剧:绝命后卫师20	39772.13
2	AH_CMCC	CCTV-1-HD	电视剧:彭德怀元帅23	3582.11
3	AH_CMCC	CCTV-1-HD	动物世界	3463.01
4	AH_CMCC	CCTV-1-HD	好记者讲好故事-2016年中国记者节特别节目	3450.39
5	AH_CMCC	CCTV-1-HD	新闻30分	3027.27
6	AH_CMCC	CCTV-1-HD	电视剧:彭德怀元帅15	2952.33
7	AH_CMCC	CCTV-1-HD	2016挑战不可能第二季-13	2525.58
8	AH_CMCC	CCTV-1-HD	加油向未来-精编7	2259.44
9	AH_CMCC	CCTV-1-HD	2016挑战不可能-5	2241.06
10	AH_CMCC	CCTV-1-HD	电视剧:彭德怀元帅18	2239.9
11	AH_CMCC	CCTV-1-HD	2016挑战不可能第二季-10	2213.87
12	AH_CMCC	CCTV-1-HD	电视剧:彭德怀元帅20	2186.86
13	AH_CMCC	CCTV-1-HD	电视剧:我和我的传奇奶奶9	2155.5
14	AH_CMCC	CCTV-1-HD	电视剧:长征大会师11	2140.89
15	AH_CMCC	CCTV-1-HD	电视剧:长征大会师21	2132.87
16	AH_CMCC	CCTV-1-HD	寻宝-我有传家宝	2111.95
17	AH_CMCC	CCTV-1-HD	电视剧: 父亲的身份30	2085.67
18	AH_CMCC	CCTV-1-HD	戏苑百家-叮咚晓玲哈(第二季)-8	2068.82
19	AH_CMCC	CCTV-1-HD	电视剧:长征大会师19	2017.28

2.PingBack_log To Cassandra (Storm)

收集来自active_user发进kafka的数据,处理后数据入casandra的结构

如下图大红色和蓝色为主键(红色分区键、蓝色排序键)、玫红色为count类型

重复字段累加click_num、playlength

添加时间戳字段，只细分到天，为后面的读取cassandra到hdfs作为一天一同步的过滤条件

【liv_viscosity_program】												
partner	logdate	srcname	chncode	albumname	playorder	province	city	click_num	playlength	timelength		
ZJYD	2016-11-25 16:00:00.000000+0000	频道列表	AHMS	好运属于你	20161126	浙江省	none	3	7200	2400		
ZJYD	2016-11-25 16:00:00.000000+0000	频道列表	AHMS	每日新闻报	20161126	浙江省	none	1	1800	1800		
ZJYD	2016-11-25 16:00:00.000000+0000	频道列表	AHMS	海豚真情剧场:锦绣未央25	25	浙江省	none	1	3120	3120		
ZJYD	2016-11-25 16:00:00.000000+0000	频道列表	BTV-1	<锦绣未央>剧宣	0	浙江省	none	1	3360	3360		
ZJYD	2016-11-25 16:00:00.000000+0000	频道列表	BTV-1	跨界喜剧王	20161126	浙江省	none	2	12420	4200		
ZJYD	2016-11-25 16:00:00.000000+0000	频道列表	CCTV-1-HD	开讲啦	20161126	浙江省	none	1	3480	3480		
ZJYD	2016-11-25 16:00:00.000000+0000	频道列表	CCTV-1-HD	星光大道	20161126	浙江省	none	2	14640	7320		
ZJYD	2016-11-25 16:00:00.000000+0000	频道列表	CCTV-1-HD	第四批 中国梦"主题新创作歌曲-四有军人歌	20161126	浙江省	none	1	480	480		
ZJYD	2016-11-25 16:00:00.000000+0000	频道列表	CCTV-10	探索·发现	20161126	浙江省	none	1	2580	2580		
ZJYD	2016-11-25 16:00:00.000000+0000	频道列表	CCTV-10	欣赏	20161126	浙江省	none	1	180	180		

【liv_viscosity_user】												
partner	logdate	mac	srcname	chncode	albumname	playorder	click_num	playlength	timelength			
ZJYD	2016-11-25 16:00:00.000000+0000	00:21:26:0c:e9:e2	首屏推荐	G_ZHEJIANG	转播中央电视台新闻联播	20161126	1	2100	2100			
ZJYD	2016-11-25 16:00:00.000000+0000	00:21:26:0c:f1:59	首屏推荐	G_JILIN	下午剧场:大刀记9	9	1	2220	2220			
ZJYD	2016-11-25 16:00:00.000000+0000	00:21:26:17:61:49	首屏推荐	G_CCTV-8	电视剧:妻子的谎言33	33	1	3240	3240			
ZJYD	2016-11-25 16:00:00.000000+0000	00:21:26:17:64:5b	频道列表	BTV-1	<锦绣未央>剧宣	0	1	3360	3360			
ZJYD	2016-11-25 16:00:00.000000+0000	00:21:26:17:6f:c2	首屏推荐	G_SHENZHEN	热播剧场:宝贝	0	1	2520	2520			
ZJYD	2016-11-25 16:00:00.000000+0000	00:21:26:17:70:a9	频道列表	G_CCTV-9	自然:冰冻星球6	20161126	1	3540	3540			
ZJYD	2016-11-25 16:00:00.000000+0000	00:21:26:1e:86:57	首屏推荐	CCTV-5	体育世界	20161125	1	2700	2700			
ZJYD	2016-11-25 16:00:00.000000+0000	00:21:26:1e:9a:39	频道列表	CCTV-12	普法栏目剧	20161126	1	3000	3000			
ZJYD	2016-11-25 16:00:00.000000+0000	00:21:26:20:90:e2	频道列表	G_CCTV-9	真相:小海鲜2	20161126	1	1980	1980			
ZJYD	2016-11-25 16:00:00.000000+0000	00:21:26:20:93:de	频道列表	CCTV-13	新闻联播	20161126	1	1800	1800			

【vod_viscosity_user】												
partner	logdate	mac	srcname	chnname	chnid	albumname	albumid	playorder	videoid	click_num	playlength	timelength
ZJYD	2016-11-25 16:00:00.000000+0000	00:24:68:c7:2c:2c	视频直播	电视剧	1000002	小别离	700543	10	525401400	1	2737	2737
ZJYD	2016-11-25 16:00:00.000000+0000	00:24:68:c8:24:61	视频直播	电视剧	1000002	女伯嫁错郎	757809	8	538463600	1	2700	2700
ZJYD	2016-11-25 16:00:00.000000+0000	00:24:68:c9:49:5e	频道列表	电影	1000001	X战警：天启	696588	1	502679100	1	8339	8339
ZJYD	2016-11-25 16:00:00.000000+0000	00:24:68:c9:8b:32	视频直播	电视剧	1000002	麻雀	739218	25	540444900	1	2720	2720
ZJYD	2016-11-25 16:00:00.000000+0000	00:24:68:e4:a1:2d	视频直播	电视剧	1000002	我本英雄	657413	25	509947400	1	2707	2707
ZJYD	2016-11-25 16:00:00.000000+0000	54:93:59:bb:b8:0f	频道列表	电影	1000001	超能敢死队	789013	1	543270900	1	7693	7693
ZJYD	2016-11-25 16:00:00.000000+0000	9c:62:ab:07:ed:72	标签筛选	电影	1000001	冲霄战警（杜比）	603412	1	2000727815	1	6001	6001
ZJYD	2016-11-26 16:00:00.000000+0000	00:21:26:17:6d:12	搜索列表	游戏	1000014	BOXBOX锐变-摄影集锦	61528	1	333467300	1	26	226
ZJYD	2016-11-26 16:00:00.000000+0000	00:21:26:17:6d:12	搜索列表	游戏	1000014	M7双排记-23-来自野兽灵魂的绝命拳法	39722	1	254961700	1	70	2918
ZJYD	2016-11-26 16:00:00.000000+0000	00:21:26:17:6d:12	搜索列表	游戏	1000014	TheShy锐雯精彩时刻2级塔下送你回温泉	51088	1	344074500	1	40	138

【vod_viscosity_program】

partner	logdate	srcname	chnname	chnid	albumname	albumid	playorder	videoid	province	city	click_num	playlength	timelength
AH_CMCC	2016-11-25 16:00:00.000000+0000	专题列表	娱乐	1000010	大眼：我的世界神奇宝贝EP4：坐骑铁甲暴龙兽	650174	1	510851900	安徽省	none	1	1634	1634
AH_CMCC	2016-11-25 16:00:00.000000+0000	主创相关	动漫	1000003	一零零一夜英文版	901018	1	569864700	安徽省	宣城	1	1298	1298
AH_CMCC	2016-11-25 16:00:00.000000+0000	主创相关	动漫	1000003	乐高幻影忍者魔咒幽灵的进击	914605	1	571653700	安徽省	none	2	5296	2648
AH_CMCC	2016-11-25 16:00:00.000000+0000	主创相关	动漫	1000003	喜羊羊与灰太狼之嘻哈闯世界	273982	3	572965000	安徽省	合肥	1	900	900
AH_CMCC	2016-11-25 16:00:00.000000+0000	主创相关	动漫	1000003	艾克斯奥特曼	82744	22	427522900	安徽省	none	1	1504	1504
AH_CMCC	2016-11-25 16:00:00.000000+0000	主创相关	电影	1000001	万物理论	271299	1	328558000	安徽省	none	1	7428	7428
AH_CMCC	2016-11-25 16:00:00.000000+0000	主创相关	电影	1000001	三个未结妈妈	342839	1	481492100	安徽省	none	1	5367	5367
AH_CMCC	2016-11-25 16:00:00.000000+0000	主创相关	电影	1000001	仙班校园2	709181	1	521028400	安徽省	none	1	4995	4995
AH_CMCC	2016-11-25 16:00:00.000000+0000	主创相关	电影	1000001	佐州自救兄弟2	275092	1	445139500	安徽省	none	1	6102	6102
AH_CMCC	2016-11-25 16:00:00.000000+0000	主创相关	电影	1000001	冰河追凶	298369	1	479299300	安徽省	none	1	6141	6141
AH_CMCC	2016-11-25 16:00:00.000000+0000	主创相关	电影	1000001	心归何处	70450	1	288497300	安徽省	滁州	1	6837	6837

结合上游数据量，本环节处理：

Window	Emitted	Transferred	Complete latency (ms)	Acked	Failed
10m 0s	1432853	1432654	7.676	246971	0

topo基本数据qps：

Id	Executors	Tasks	Emitted	Transferred	Capacity (last 10m)	Execute latency (ms)	Executed	Process latency (ms)	Acked	Failed
liv_distribute	3	3	30072980	30072980	0.000	0.009	23679600	0.009	23679620	0
liv_program_viscosity	3	3	6293880	6293880	0.000	0.083	6294180	24.589	6294200	0
liv_to_program_insert	3	3	0	0	0.000	0.316	124500	22.924	124500	0
liv_to_program_update	3	3	0	0	0.000	0.077	6157180	22.903	6157180	0
liv_to_user_insert	3	3	0	0	0.000	0.094	4807960	23.358	4807940	0
liv_to_user_update	3	3	0	0	0.000	0.109	1644020	22.610	1643980	0
liv_user_viscosity	3	3	6615560	6615560	0.000	0.085	6615760	25.879	6615760	0
nosrcName	1	1	0	0	0.001	0.016	16317860	0.580	16317860	0
top_filter	3	3	128254320	128254320	0.001	0.010	128285140	0.010	128285140	0
vod_distribute	3	3	207426200	207426200	0.001	0.007	103713060	0.007	103713040	0
vod_program_viscosity	3	3	97953240	97953240	0.004	0.032	97951160	19.884	97951160	0
vod_to_program_insert	3	3	0	0	0.001	0.131	6525460	24.972	6525460	0
vod_to_program_update	3	3	0	0	0.003	0.034	91342420	19.180	91342440	0
vod_to_user_insert	3	3	0	0	0.003	0.044	87540400	35.762	87540420	0
vod_to_user_update	3	3	0	0	0.000	0.061	15244600	13.230	15244600	0
vod_user_viscosity	3	3	104958820	104958820	0.003	0.041	104958780	37.088	104958800	0

Showing 1 to 16 of 16 entries

3.Cassandra to HDFS

10.10.121.148:crontab -l

45 00 * * * sh /data/opt/script/viscosity/tohdfs.sh > /data/opt/script/viscosity/sh_tohdfs.log 2>&1 &

```
#!/bin/bash
. /etc/bashrc
today='date +%Y-%m-%d'
yesterday='date -d '-1 day' +%Y-%m-%d'
qt='date -d '-2 day' +%Y-%m-%d'
yyyy='echo ${yesterday}|cut -d "-" -f1'
mm='echo ${yesterday}|cut -d "-" -f2'
dd='echo ${yesterday}|cut -d "-" -f3'
nohup java -jar /data/opt/script/viscosity/vtoh.jar ${qt} ${today} 'vod' 'vod_viscosity_user' ${yesterday} > nohup1.out 2>&1 &
nohup java -jar /data/opt/script/viscosity/vtoh.jar ${qt} ${today} 'vod' 'vod_viscosity_program' ${yesterday} > nohup2.out 2>&1 &
nohup java -jar /data/opt/script/viscosity/vtoh.jar ${qt} ${today} 'liv' 'liv_viscosity_user' ${yesterday} > nohup3.out 2>&1 &
nohup java -jar /data/opt/script/viscosity/vtoh.jar ${qt} ${today} 'liv' 'liv_viscosity_program' ${yesterday} > nohup4.out 2>&1 &
#注意！无论分区location的路径是否存在，都可以制定这个分区
hive <<EOF
use likang;
ALTER TABLE vod_viscosity_user ADD IF NOT EXISTS PARTITION(yyyy='${yyyy}',mm='${mm}',dd='${dd}') location '/likang/viscosity/vod/vod_viscosity_user/${yyyy}/${mm}/${dd}';
ALTER TABLE vod_viscosity_program ADD IF NOT EXISTS PARTITION(yyyy='${yyyy}',mm='${mm}',dd='${dd}') location '/likang/viscosity/vod/vod_viscosity_program/${yyyy}/${mm}/${dd}';
ALTER TABLE liv_viscosity_user ADD IF NOT EXISTS PARTITION(yyyy='${yyyy}',mm='${mm}',dd='${dd}') location '/likang/viscosity/liv/liv_viscosity_user/${yyyy}/${mm}/${dd}';
ALTER TABLE liv_viscosity_program ADD IF NOT EXISTS PARTITION(yyyy='${yyyy}',mm='${mm}',dd='${dd}') location '/likang/viscosity/liv/liv_viscosity_program/${yyyy}/${mm}/${dd}';
EOF
```

启动四个java Application执行，导入hdfs后：

```
[root@10.10.121.148 viscosity]# hadoop fs -ls /likang/viscosity/*/*
Found 1 items
drwxr-xr-x - root supergroup 0 2016-12-02 00:45 /likang/viscosity/liv/liv_viscosity_program/2016
Found 1 items
drwxr-xr-x - root supergroup 0 2016-12-02 00:45 /likang/viscosity/liv/liv_viscosity_user/2016
Found 1 items
drwxr-xr-x - root supergroup 0 2016-12-02 00:45 /likang/viscosity/vod/vod_viscosity_program/2016
Found 1 items
drwxr-xr-x - root supergroup 0 2016-12-02 00:45 /likang/viscosity/vod/vod_viscosity_user/2016
```

执行结果会录入zk（成功或失败）：

```
[zk: localhost:2281(CONNECTED) 1] get /play_viscosity/hdfs_date/liv/liv_viscosity_program/2016-12-04
successed!
cZxid = 0x1036155d1

```