

尚硅谷大数据技术之 Hive 面试题

(作者: 尚硅谷研究院)

版本 2.0

第 1 题 连续问题

如下数据为蚂蚁森林中用户领取的减少碳排放量

id	dt	lowcarbon
1001	2021-12-12	123
1002	2021-12-12	45
1001	2021-12-13	43
1001	2021-12-13	45
1001	2021-12-13	23
1002	2021-12-14	45
1001	2021-12-14	230
1002	2021-12-15	45
1001	2021-12-15	23
...

找出连续 3 天及以上减少碳排放量在 100 以上的用户

第 2 题 分组问题

如下为电商公司用户访问时间数据

id	ts (秒)
1001	17523641234
1001	17523641256
1002	17523641278
1001	17523641334
1002	17523641434
1001	17523641534
1001	17523641544
1002	17523641634
1001	17523641638
1001	17523641654

某个用户连续的访问记录如果时间间隔小于 60 秒，则分为同一个组，结果为：

id	ts (秒)	group
1001	17523641234	1
1001	17523641256	1
1001	17523641334	2
1001	17523641534	3
1001	17523641544	3
1001	17523641638	4
1001	17523641654	4
1002	17523641278	1
1002	17523641434	2
1002	17523641634	3

第 3 题 间隔连续问题

某游戏公司记录的用户每日登录数据

id	dt
1001	2021-12-12
1002	2021-12-12
1001	2021-12-13
1001	2021-12-14
1001	2021-12-16
1002	2021-12-16
1001	2021-12-19
1002	2021-12-17
1001	2021-12-20

计算每个用户最大的连续登录天数，可以间隔一天。解释：如果一个用户在 1,3,5,6 登录游戏，则视为连续 6 天登录。

第 4 题 打折日期交叉问题

如下为平台商品促销数据：字段为品牌，打折开始日期，打折结束日期

brand	stt	edt
oppo	2021-06-05	2021-06-09
oppo	2021-06-11	2021-06-21
vivo	2021-06-05	2021-06-15
vivo	2021-06-09	2021-06-21
redmi	2021-06-05	2021-06-21
redmi	2021-06-09	2021-06-15
redmi	2021-06-17	2021-06-26
huawei	2021-06-05	2021-06-26
huawei	2021-06-09	2021-06-15
huawei	2021-06-17	2021-06-21

计算每个品牌总的打折销售天数，注意其中的交叉日期，比如 vivo 品牌，第一次活动时间为 2021-06-05 到 2021-06-15，第二次活动时间为 2021-06-09 到 2021-06-21 其中 9 号到 15 号为重复天数，只统计一次，即 vivo 总打折天数为 2021-06-05 到 2021-06-21 共计 17 天。

第 5 题 同时在线问题

如下为某直播平台主播开播及关播时间，根据该数据计算出平台最高峰同时在线的主播人数。

id	stt	edt
1001	2021-06-14 12:12:12	2021-06-14 18:12:12
1003	2021-06-14 13:12:12	2021-06-14 16:12:12
1004	2021-06-14 13:15:12	2021-06-14 20:12:12
1002	2021-06-14 15:12:12	2021-06-14 16:12:12
1005	2021-06-14 15:18:12	2021-06-14 20:12:12
1001	2021-06-14 20:12:12	2021-06-14 23:12:12
1006	2021-06-14 21:12:12	2021-06-14 23:15:12
1007	2021-06-14 22:12:12	2021-06-14 23:10:12
...