# 美团面试要求

**1）技术部分**

（1）HDFS架构理解（基础）

（2）HDFS源码/工作原理（高级）

（3）MapReduce架构理解（基础）

（4）MapReduce源码/工作原理/Shuffle原理（高级）

（5）MapReduce二次排序（编程，必选）

（6）YARN架构理解（基础）

（7）YARN源码/工作原理（高级）

（8）HBase架构理解（基础）

（9）HBase源码/工作原理（高级）

（10）HBase性能优化（高级）

（11）Hive原理理解（基础）

（12）Hive性能优化（高级）

（13）Flume架构理解（可选）

（14）Kafka架构理解（可选）

（15）Spark RDD理解（基础）

（16）Spark reduceByKey与groupByKey区别（基础）

（17）Spark Broadcast与Accumulator功能（基础）

（18）Spark工作原理（高级）

（19）Spark Shuffle原理（高级）

（20）Spark源码理解/贡献（高级）

（21）Spark性能优化/数据倾斜（高级）

**2）项目部分**

**2.1）数仓理论**

（1）数据仓库分层架构（初级）

（2）事实表与维度表（初级）

（3）星型模型与雪花模型 （初级）

（4）日增量表与日全量表 （初级）

（5）拉链表 (初级)

（6）缓慢变化维 （初级）

（7）两种建模理论（初级）

（8）data-vault理论（高级）

**2.2）数仓周边系统**

（1）血缘关系

（2）调度系统

（3）展示，olap（kylin）

（4）指标管理

（5）数据质量管理 （高级）

（6）元数据管理，调度系统元数据，任务运行数据，表存储数据，数据字典等

**3）手写SQL部分**

（1）点：窗口函数，行列转换，聚合函数

加一点复杂的：collect\_set，split，case when，聚合函数，union等

（2）函数考察

eg: rank, dense\_rank, row\_number区别 （基础）

行列转换 （基础）

sort by，distribute by，order by，clustered by区别

（3）3个表关联，写下大概的mr过程 (中级)

（4）去重，2个写法的区别:

select count(distinct id) from t1;

select count(1) from ( select id from t1 group by id) a (高级)

（5）map join 原理 （基础）

（6）bloom filter去重，原理(高级)

（7）hive tunning

**4）场景部分**

（1）Job，Reduce长时间百分之99；

（2）Reduce的3个阶段，Reduce百分之90+，又降下去，反复几次

（3）Map一直百分之90多，之后报错

## 1

上海科锐国际外派美团点评

**1）面试前HR给的重点**

（1）SQL要强；

（2）分层是否有参与；

（3）熟悉哪些框架；

（4）项目的数据量有多少；

（5）调度工具用到哪些；

（6）数据源在哪里；

（7）数据可视化怎么做；

（8）跟Java怎么对接；

（9）面试官还会现场出一些题目，现场做

**2）一面：**

（1）问Spark，Flink优劣势，

（2）Flink如何处理乱序数据，

（3）Flink内存溢出怎么办，

（4）Flink有没有试过优化，

（5）Flink精准一次消费，

（6）建模理论，用过什么模型，雪花和星型的分别解决什么问题。

（7）用过什么工具进行数据迁移，导入导出。

（8）数仓大体干了什么，自己做了哪些指标，

（9）SQL题：用Hive解析Json字符串 或者 七天连续登陆用户

**3）二面：**

基本情况：去了是帮美团点评做和银行对接的数据仓库，数据量大，面试是完全和美团点评面试，签合同是和科瑞国际签约。

（听内部消息说尚大好像已经进去了两个，目前HR说还缺两个数据开发岗位）

## 2

美团金融

第一轮：

输入: 表T(user\_id, timestamp), 用户在某一页面的所有曝光历史记录

输出: 2020年1月份1~31日，指标: 首次曝光用户数量，次1-7日留存用户数量

#

select t1.min\_visit\_date

,sum(if(t1.min\_visit\_date == t2.visit\_date, 1, 0)) as base\_count

,sum(if(datediff(t2.visit\_date, t1.min\_visit\_date) = 1, 1, 0)) as liucun\_day01

,sum(if(datediff(t2.visit\_date, t1.min\_visit\_date) = 2, 1, 0)) as liucun\_day02

,sum(if(datediff(t2.visit\_date, t1.min\_visit\_date) = 3, 1, 0)) as liucun\_day03

,sum(if(datediff(t2.visit\_date, t1.min\_visit\_date) = 4, 1, 0)) as liucun\_day04

,sum(if(datediff(t2.visit\_date, t1.min\_visit\_date) = 5, 1, 0)) as liucun\_day05

,sum(if(datediff(t2.visit\_date, t1.min\_visit\_date) = 6, 1, 0)) as liucun\_day06

,sum(if(datediff(t2.visit\_date, t1.min\_visit\_date) = 7, 1, 0)) as liucun\_day07

from (

select user\_id

,min(to\_date(timestamp)) as min\_visit\_date

from visit\_info

group by user\_id

having min\_visit\_date >= '2020-01-01' and min\_visit\_date <= '2020-01-31'

) t1

left join (

select user\_id

,to\_date(timestamp) as visit\_date

from visit\_info

where to\_date(timestamp) >= '2020-01-01'

group by user\_id, to\_date(timestamp) -- 去重

) t2 on t1.user\_id = t2.user\_id

group by t1.min\_visit\_date;

第二轮

求：每个uid上一笔成功订单id，没有的记录为null，

状态值：1-成功 0-失败

id time uid is\_suc 结果列[最终展示结果]

1 2020-01-01 1 1 Null

2 2020-01-02 1 0 1

3 2020-01-03 1 0 1

4 2020-01-04 1 1 1

5 2020-01-05 1 0 4

6 2020-01-06 2 0 Null

7 2020-01-07 2 1 Null

8 2020-01-08 2 0 7

#

我开始理解错题目了。我以为求上一个正确的状态值

实际上计算的是上一个状态是正确的id，还要面试给过了

drop table visit\_info;

create table visit\_info

as

select 1 as id, '2020-01-01' as visit\_date, '1' as user\_id, 1 as is\_suc

union all

select 2 as id, '2020-01-02' as visit\_date, '1' as user\_id, 0 as is\_suc

union all

select 3 as id, '2020-01-03' as visit\_date, '1' as user\_id, 0 as is\_suc

union all

select 4 as id, '2020-01-04' as visit\_date, '1' as user\_id, 1 as is\_suc

union all

select 5 as id, '2020-01-05' as visit\_date, '1' as user\_id, 0 as is\_suc

union all

select 6 as id, '2020-01-06' as visit\_date, '2' as user\_id, 0 as is\_suc

union all

select 7 as id, '2020-01-07' as visit\_date, '2' as user\_id, 1 as is\_suc

union all

select 8 as id, '2020-01-08' as visit\_date, '2' as user\_id, 0 as is\_suc

;

select t1.id

,t1.visit\_date

,t1.user\_id

,t1.is\_suc

,max(if(datediff(t1.visit\_date, t2.visit\_date) <= 0, Null, t2.id)) as result\_flag

from visit\_info t1

left join (

select id

,visit\_date

,user\_id

from visit\_info

where is\_suc = 1

) t2 on t1.user\_id = t2.user\_id

group by t1.id ,t1.visit\_date,t1.user\_id,t1.is\_suc

;

t1.id t1.visit\_date t1.user\_id t1.is\_suc result\_flag

1 2020-01-01 1 1 NULL

2 2020-01-02 1 0 1

3 2020-01-03 1 0 1

4 2020-01-04 1 1 1

5 2020-01-05 1 0 4

6 2020-01-06 2 0 NULL

7 2020-01-07 2 1 NULL

8 2020-01-08 2 0 7

第三轮:

确定是 技术面

面试官有点奇怪，开头就问为啥没考研，大四实习了吗？成绩很差吗？

什么专业？为什么学习java? 为什么又做大数据开发？为什么在某个公司要离职？

基本没问什么项目, 我尽量回答，最后问了HR 说沟通下，后续有其他部门的offer，就不管这个部门了, 应该是三面挂了

## 3

美团bp数据中心

一轮说自己写个top10

写完了说有啥问题吗？

我说挺好的？

row\_number 如果order by 多个字段一样有问题吗？

然后就解释了一波 row\_number, rank, dense\_rank的区别

第2轮没有写sql, 问的都是项目内容

## 4

美团财务集市的面试题

第一轮

面试官不怎么问sql, 就说咱们随便来一个吧

dptid, user\_id, age

求每个部门年龄的top10

第二轮，第三轮

也没写sql，面试官问的原理大部分都是基础的东西，觉得我准备的很充足，表达的欲望就是想要你，虽然方向不偏向数仓

第四轮交叉面

应该是个大佬，据说交叉面是定级的

面试偏向于管理方向，比如给你一个团队，你怎么带？怎么安排，我坦诚没带过，然后就回复 假如让你带，你构思下？

最后面试过了，因为选择了其他部门，这个就没进行HR面

## 5

美团外卖

第一轮：

create table if not exists tmpA(

wm\_order\_id bigint comment '订单id',

wm\_poi\_id bigint comment '商家id',

userid string comment '用户id',

month string comment '月份，格式：yyyyMM',

actual\_amt double comment '实付金额',

dt string comment '日期，格式：yyyyMMdd'

)

row format delimited fields terminated by ",";

1.统计每月支付额排名前3的用户清单

2.展示每个用户每月的最大实付金额和截止到该月的总实付金额

-- 1.统计每月支付额排名前3的用户清单

select month

,useid

,actual\_amt

,rn

from (

select month

,useid

,actual\_amt

,row\_number()over(partition by month order by actual\_amt desc) as rn

from (

select month

,useid

,sum(actual\_amt) as actual\_amt

from tmpA

group by month, userid

) t

) t

where rn <= 3

;

-- 2. 展示每个用户每月的最大实付金额和截止到该月的总实付金额

select useid

,month

,max(month\_max\_actual\_amt) as month\_max\_actual\_amt

,max(total\_actual\_amt) as total\_actual\_amt

from (

select month

,useid

,max(actual\_amt)over(partition by month, useid) as month\_max\_actual\_amt

,sum(actual\_amt)over(partition by useid order by month) as total\_actual\_amt

from tmpA

) t

group by month, useid

;

第二轮:

有个table 有两个字段 dt(yyyyMMdd), jieqi(内容为汉字 比如 立春, 小寒 记录每年农历的24节气)

表中记录了5年的数据

求: 增加一个season列，比如 春，夏，秋，冬，可以根据 立春 和 立夏 之间划分 春 诸如此类的划分；

展示行效果如下

dt jieqi season

2020xxxx 立春 春

select t1.dt

,t1.jieqi

,t2.season

from info t1

left join (

select dt

,jieqi

,season

,lead(dt, 1, '99991231') over(order by dt) as next\_dt

from (

select dt

,jieqi

,case when jieqi = '立春' then '春'

when jieqi = '立夏' then '夏'

when jieqi = '立秋' then '秋'

else '冬' end as season

from info

where jieqi in ('立春','立夏','立秋','立冬')

) t

) t2

where t1.dt >= t2.dt and t1.dt < t2.next\_dt

二面偏向java,问了一些java的

比如list, set, map的理解

并发容器CopyWriteArrayList的原理，又延伸了下 mysql数据的读写锁是怎么回事？写的时候，如何读的？

sql注入怎么理解的？

能否写个3个线程交替打印的代码？

回答的很不好，还是给过了

第三轮:

tord表: 字段为 ord\_id, ord\_time, usr\_id, dt

增量存储的交易记录表，记录每个月新增用户首单在未来30天的复购交易额；

#

展示字段: 统计月, 复购交易金额

比如 求 7月份的时候，7月1号新增的，求 2~31号的，2号新增的求 3~8-1号的

insert into table order\_amt\_info partition (dt)

select substr(t1.min\_info.col1,1,7) as report\_month

,sum(t2.amt) as amt

,substr(t1.min\_info.col1,1,7) as dt

from (

select usr\_id

,min(struct(min\_ord\_time, ord\_id)) as min\_info

from tord

group by usr\_id

having min\_ord\_time >= '2020-07-01' and min\_ord\_time <= '2020-07-31'

) t1

left join (

select ord\_id

,to\_date(ord\_time) as ord\_time

,usr\_id

,amt

from tord

where dt >= '2020-07-01'

) t2 on t1.usr\_id = t2.usr\_id and t1.min\_info.col2 = t2.ord\_id

where datediff(t2.ord\_time, t1.min\_ord\_time) >= 1

and datediff(t2.ord\_time, t1.min\_ord\_time) <= 30

group by substr(t1.min\_ord\_time,1,7)

第三题的衍生发问

延伸题目

从上述的表中构建一个 增量表 记录

fusr: user\_id, f\_ord\_time, dt - 增量的分区，每个分区存当日首购用户

延伸题目

fusr: user\_id, f\_ord\_time,f\_ord\_amt, dt - 增量的分区，每个分区存当日首购用户以及首单的金额

初始化全量的 收购用户

insert into table order\_amt\_info partition (dt)

select user\_id

,to\_date(min\_info.col1) as f\_ord\_time

,min\_info.col2 as f\_ord\_amt

,to\_date(min\_info.col1) as dt

from (

select user\_id

,min(struct(ord\_time, amt)) as min\_info

from tord

where dt < '2020-07-01'

group by user\_id

) t

构建后续的每日增量数据

insert into table order\_amt\_info partition (dt='${}')

select t1.user\_id

,to\_date(t1.min\_info.col1) as f\_ord\_time

,t1.min\_info.col2 as f\_ord\_amt

from (

select user\_id

,min(struct(ord\_time, amt)) as min\_info

from tord

where dt = '${}'

group by user\_id

) t1

left join (

select user\_id

from order\_amt\_info

where dt < '${}'

) t2 on t1.user\_id = t2.user\_id

where t2.user\_id is null

;

继续延伸

用户im： 用户发起会话数，用户参与会话数，用户发起消息数，用户参与消息数，发起会话用户数，参与会话用户数

log：misid，sentid，reid，mis\_time，dt

dim：uid，idtype（poi，usr，dlvr）

模型设计 - 出一张公共表：

按照自己的思路展开就好，不是sql题目

HR面

美团面试感觉特别耽搁时间，某天下午直接面了1-3轮，然后第二天HR面，之后等了3天，说面试过了，简单说了下基本工资和期望薪资

然后提交了流水

之后又增了7天，微信问了HR, 反馈面试通过，晚上谈了具体薪资，然后隔了两天再沟通，反馈offer已经审批通过，然后现在还没发offer, 还在等