

下载相关创建表格和插入数据脚本

[练习题1](#)

[练习题2](#)

[练习题3](#)

[练习题4](#)

[练习题5](#)

[练习题6](#)

[练习题7](#)

[练习题8](#)

[练习题9](#)

[练习题10](#)

SQL训练营页面地址：<https://tianchi.aliyun.com/specials/promotion/aicampsql>

天池龙珠计划训练营地址：<https://tianchi.aliyun.com/specials/promotion/aicamps>

下载相关创建表格和插入数据脚本

这里为了方便大家，数据脚本已经给大家准备好啦。

点击下方链接直接下载

创建数据表脚本：http://tianchi-media.oss-cn-beijing.aliyuncs.com/dragonball/SQL/create_table.sql

插入数据脚本：<http://tianchi-media.oss-cn-beijing.aliyuncs.com/dragonball/SQL/data.zip>

大家下载好脚本后，先在MySQL环境中运行 `create_table.sql` 脚本，创建数据表，然后解压下载好的 `data.zip`，解压后目录如下：

```
8-10ccf_offline_stagel_train.sql
6-winequality-white.sql
5-8-10ccf_online_stagel_train.sql
4-macro industry.sql
3-ccf_offline_stagel_test_revised.sql
2-winequality-red.sql
1-9income statement.sql
1-9company operating.sql
1-7market data.sql
```

脚本文件名前面的序号表示用到该数据集的题目序号，例如 `1-7market data.sql` 表示第1题和第7题用到了该数据集。

同样的，这里给大家的也是sql脚本，里面是插入数据的语句，大家只需打开后在MySQL环境中运行即可将数据导入到数据表中。

练习题1

数据来源: <https://tianchi.aliyun.com/dataset/dataDetail?dataId=1074>

请使用A股上市公司季度营收预测数据集《Income Statement.xls》和《Company Operating.xlsx》和《Market Data.xlsx》，以Market Data为主表，将三张表中的TICKER_SYMBOL为600383和600048的信息合并在一起。只需要显示以下字段。

表名	字段名
Income Statement	TICKER_SYMBOL
Income Statement	END_DATE
Income Statement	T_REVENUE
Income Statement	T_COGS
Income Statement	N_INCOME
Market Data	TICKER_SYMBOL
Market Data	END_DATE_
Market Data	CLOSE_PRICE
Company Operating	TICKER_SYMBOL
Company Operating	INDIC_NAME_EN
Company Operating	END_DATE
Company Operating	VALUE

答案及思路

将数据表导入数据库，根据表结构和字段含义，Company Operating表使用sheet-EN(使用CN也可以)，Income Statement表使用sheet-General Business，Market Data使用sheet-Data。

```
SELECT MarketData.*,
       OperatingData.INDIC_NAME_EN,
       OperatingData.VALUE,
       IncomeStatement.N_INCOME,
       IncomeStatement.T_COGS,
       IncomeStatement.T_REVENUE
FROM (
  SELECT TICKER_SYMBOL,
         END_DATE,
         CLOSE_PRICE
  FROM `market data`
  WHERE TICKER_SYMBOL IN ('600383','600048') ) MarketData
LEFT JOIN -- operating data
  (SELECT TICKER_SYMBOL,
         INDIC_NAME_EN,
```

```

        END_DATE,
        VALUE
    FROM `company operating`
    WHERE TICKER_SYMBOL IN ('600383','600048') ) OperatingData
    ON MarketData.TICKER_SYMBOL = OperatingData.TICKER_SYMBOL
    AND MarketData.END_DATE = OperatingData.END_DATE
LEFT JOIN -- income statement
    (SELECT DISTINCT TICKER_SYMBOL,
        END_DATE,
        T_REVENUE,
        T_COGS,
        N_INCOME
    FROM `income statement`
    WHERE TICKER_SYMBOL IN ('600383','600048') ) IncomeStatement
    ON MarketData.TICKER_SYMBOL = IncomeStatement.TICKER_SYMBOL
    AND MarketData.END_DATE = IncomeStatement.END_DATE
ORDER BY MarketData.TICKER_SYMBOL, MarketData.END_DATE

```

练习题2

数据来源: <https://tianchi.aliyun.com/dataset/dataDetail?dataId=44>

请使用 Wine Quality Data 数据集《winequality-red.csv》，找出 pH=3.03的所有红葡萄酒，然后，对其 citric acid 进行中式排名（相同排名的下一个名次应该是下一个连续的整数值。换句话说，名次之间不应该有“间隔”）

```

SELECT pH,
    `citric acid`,
    DENSE_RANK() OVER (ORDER BY `citric acid`) AS rankn
FROM `winequality-red`
WHERE pH= 3.03;

```

练习题3

数据来源: <https://tianchi.aliyun.com/competition/entrance/231593/information>

使用Coupon Usage Data for O2O中的数据《ccf_offline_stage1_test_revised.csv》，试分别找出在2016年7月期间，发放优惠券总金额最多和发放优惠券张数最多的商家。

这里只考虑满减的金额，不考虑打几折的优惠券。

```

-- 发放优惠券总金额最多的商家
SELECT Merchant_id,
    -- SUM(SUBSTRING_INDEX(`Discount_rate`,':', 1)) AS sale_amount,
    SUM(SUBSTRING_INDEX(`Discount_rate`,':',-1)) AS discount_amount
    FROM ccf_offline_stage1_test_revised
    WHERE Date_received BETWEEN '2016-07-01' AND '2016-07-31'
GROUP BY Merchant_id

```

```
ORDER BY discount_amount DESC
LIMIT 1;
-- 发放优惠券张数最多的商家
SELECT Merchant_id,COUNT(1) AS cnt
FROM ccf_offline_stagel_test_revised
WHERE Date_received BETWEEN '2016-07-01' AND '2016-07-31'
GROUP BY Merchant_id
ORDER BY cnt DESC
LIMIT 1;
```

练习题4

数据来源: <https://tianchi.aliyun.com/dataset/dataDetail?dataId=1074>

请使用A股上市公司季度营收预测中的数据集《Macro&Industry.xlsx》中的sheet-INDIC_DATA, 请计算全社会用电量:第一产业:当月值在2015年用电最高峰是发生在哪月? 并且相比去年同期增长/减少了多少个百分比?

```
-- 2015年用电最高峰是发生在哪月
SELECT PERIOD_DATE,
       MAX(DATA_VALUE) FianlValue
FROM `macro industry`
WHERE INDIC_ID = '2020101522'
      AND YEAR(PERIOD_DATE) = 2015
GROUP BY PERIOD_DATE
ORDER BY FianlValue DESC
LIMIT 1;
-- 并且相比去年同期增长/减少了多少个百分比?
SELECT BaseData.*,
       (BaseData.FianlValue - YoY.FianlValue) / YoY.FianlValue YoY
FROM (SELECT PERIOD_DATE,
       MAX(DATA_VALUE) FianlValue
FROM `macro industry`
WHERE INDIC_ID = '2020101522'
      AND YEAR(PERIOD_DATE) = 2015
GROUP BY PERIOD_DATE
ORDER BY FianlValue DESC
LIMIT 1) BaseData
LEFT JOIN -- YOY
      (SELECT PERIOD_DATE,
       MAX(DATA_VALUE) FianlValue
FROM `macro industry`
WHERE INDIC_ID = '2020101522'
      AND YEAR(PERIOD_DATE) = 2014
GROUP BY PERIOD_DATE ) YoY
ON YEAR(BaseData.PERIOD_DATE) = YEAR(YoY.PERIOD_DATE) + 1
AND MONTH(BaseData.PERIOD_DATE) = MONTH(YoY.PERIOD_DATE);
```

练习题5

数据来源: <https://tianchi.aliyun.com/competition/entrance/231593/information>

使用Coupon Usage Data for O2O中的数据《ccf_online_stage1_train.csv》，试统计在2016年6月期间，线上总体优惠券弃用率为多少？并找出优惠券弃用率最高的商家。

弃用率 = 被领券但未使用的优惠券张数 / 总的被领取优惠券张数

-- 2016年6月期间，线上总体优惠券弃用率为多少？

```
SELECT SUM(CASE WHEN Date='0000-00-00' AND Coupon_id IS NOT NULL
                THEN 1
                ELSE 0
            END) /
        SUM(CASE WHEN Coupon_id IS NOT NULL
                THEN 1
                ELSE 0
            END) AS discard_rate
FROM ccf_online_stage1_train
WHERE Date_received BETWEEN '2016-06-01' AND '2016-06-30';
```

-- 2016年6月期间，优惠券弃用率最高的商家？

```
SELECT Merchant_id,
        SUM(CASE WHEN Date = '0000-00-00' AND Coupon_id IS NOT NULL
                THEN 1
                ELSE 0
            END) /
        SUM(CASE WHEN Coupon_id IS NOT NULL
                THEN 1
                ELSE 0
            END) AS discard_rate
FROM ccf_online_stage1_train
WHERE Date_received BETWEEN '2016-06-01' AND '2016-06-30'
GROUP BY Merchant_id
ORDER BY discard_rate DESC
LIMIT 1;
```

练习题6

数据来源: <https://tianchi.aliyun.com/dataset/dataDetail?dataId=44>

请使用 Wine Quality Data 数据集《winequality-white.csv》，找出 pH=3.63的所有白葡萄酒，然后，对其 residual sugar 量进行英式排名（非连续的排名）

```
SELECT pH,
        `residual sugar`,
        RANK() OVER (ORDER BY `residual sugar`) AS rankn
FROM `winequality-white`
WHERE pH= 3.63;
```

练习题7

数据来源: <https://tianchi.aliyun.com/dataset/dataDetail?dataId=1074>

请使用A股上市公司季度营收预测中的数据集《Market Data.xlsx》中的sheet-DATA,

计算截止到2018年底, 市值最大的三个行业是哪些? 以及这三个行业里市值最大的三个公司是哪些?
(每个行业找出前三大的公司, 即一共要找出9个)

-- 计算截止到2018年底, 市值最大的三个行业是哪些?

```
SELECT TYPE_NAME_CN,  
       SUM(MARKET_VALUE)  
FROM `market data`  
WHERE YEAR(END_DATE) = '2018-12-31'  
GROUP BY TYPE_NAME_CN  
ORDER BY SUM(MARKET_VALUE) DESC  
LIMIT 3
```

-- 这三个行业里市值最大的三个公司是哪些?

```
SELECT BaseData.TYPE_NAME_CN,  
       BaseData.TICKER_SYMBOL  
FROM (SELECT TYPE_NAME_CN,  
             TICKER_SYMBOL,  
             MARKET_VALUE,  
             ROW_NUMBER() OVER(PARTITION BY TYPE_NAME_CN ORDER BY MARKET_VALUE)  
      CompanyRanking  
      FROM `market data` ) BaseData  
LEFT JOIN  
( SELECT TYPE_NAME_CN,  
      SUM(MARKET_VALUE)  
FROM `market data`  
WHERE YEAR(END_DATE) = '2018-12-31'  
GROUP BY TYPE_NAME_CN  
ORDER BY SUM(MARKET_VALUE) DESC  
LIMIT 3 ) top3Type  
ON BaseData.TYPE_NAME_CN = top3Type.TYPE_NAME_CN  
WHERE CompanyRanking <= 3  
AND top3Type.TYPE_NAME_CN IS NOT NULL
```

练习题8

数据来源: <https://tianchi.aliyun.com/competition/entrance/231593/information>

使用Coupon Usage Data for O2O中的数据集《ccf_online_stage1_train.csv》和

《ccf_offline_stage1_train.csv》, 试找出在2016年6月期间, 线上线下累计优惠券使用次数最多的顾客。

```

SELECT User_id,
       SUM(couponCount) couponCount
FROM (SELECT User_id,
             count(*) couponCount
      FROM `ccf_online_stage1_train`
      WHERE (Date != 'null' AND Coupon_id != 'null')
            AND (LEFT(DATE,4)=2016 )
      GROUP BY User_id
      UNION ALL
      SELECT User_id,
             COUNT(*) couponCount
      FROM `ccf_offline_stage1_train`
      WHERE (Date != 'null' AND Coupon_id != 'null')
            AND (LEFT(DATE,4)=2016 )
      GROUP BY User_id ) BaseData
GROUP BY User_id
ORDER BY SUM(couponCount) DESC
LIMIT 1

```

练习题9

数据来源: <https://tianchi.aliyun.com/dataset/dataDetail?dataId=1074>

请使用A股上市公司季度营收预测数据集《Income Statement.xls》中的sheet-General Business和《Company Operating.xlsx》中的sheet-EN。

找出在数据集所有年份中, 按季度统计, 白云机场旅客吞吐量最高的那一季度对应的净利润是多少? (注意, 是单季度对应的净利润, 非累计净利润。)

-- 因为正好是第一季度, 所以不需要减。 如果是2季度, 单季度净利润需要用2季度的值减去1月份的

```

SELECT *
FROM (SELECT TICKER_SYMBOL,
             YEAR(END_DATE) Year,
             QUARTER(END_DATE) QUARTER,
             SUM(VALUE) Amount
      FROM `company operating`
      WHERE INDIC_NAME_EN = 'Baiyun Airport:Passenger throughput'
      GROUP BY TICKER_SYMBOL, YEAR(END_DATE), QUARTER(END_DATE)
      ORDER BY SUM(VALUE) DESC
      LIMIT 1 ) BaseData
LEFT JOIN -- income statement
      (SELECT TICKER_SYMBOL,
             YEAR(END_DATE) Year,
             QUARTER(END_DATE) QUARTER,
             SUM(N_INCOME) Amount
      FROM `income statement`
      GROUP BY TICKER_SYMBOL, YEAR(END_DATE), QUARTER(END_DATE) ) Income
ON BaseData.TICKER_SYMBOL = Income.TICKER_SYMBOL

```

```
AND BaseData.Year = Income.Year
AND BaseData.QUARTER = Income.QUARTER
```

练习题10

数据来源: <https://tianchi.aliyun.com/competition/entrance/231593/information>

使用Coupon Usage Data for O2O中的数据《ccf_online_stage1_train.csv》和《ccf_offline_stage1_train.csv》，试找出在2016年6月期间，线上线下累计被使用优惠券满减最多的前3名商家。

比如商家A，消费者A在其中使用了一张200减50的，消费者B使用了一张30减1的，那么商家A累计被使用优惠券满减51元。

```
SELECT Merchant_id,
       SUM(discount_amount) discount_amount
FROM (SELECT Merchant_id,
            SUM(SUBSTRING_INDEX(`Discount_rate`,':',-1)) AS discount_amount
      FROM `ccf_online_stage1_train`
     WHERE (Date != 'null' AND Coupon_id != 'null')
           AND (LEFT(DATE,4)=2016 )
           AND MID(DATE,5,2) = '06'
     GROUP BY Merchant_id
    UNION ALL
    SELECT Merchant_id,
           SUM(SUBSTRING_INDEX(`Discount_rate`,':',-1)) AS discount_amount
      FROM `ccf_offline_stage1_train`
     WHERE (Date != 'null' AND Coupon_id != 'null')
           AND (LEFT(DATE,4)=2016 )
           AND MID(DATE,5,2) = '06'
     GROUP BY Merchant_id ) BaseData
GROUP BY Merchant_id
ORDER BY SUM(discount_amount) DESC
LIMIT 1
```

作者: Datawhale成员: 王复振、杨煜、闫钟峰、杨梦迪、苏鹏

如果你对本次学习有任何问题，欢迎加入阿里云天池龙珠计划SQL训练营进行学习交流。

龙珠计划-SQL 训练营

48人



扫一扫群二维码，立刻加入该群。