

## 阿里云大数据专业认证（ACP 级）考试样题

## 一、 单选题

1. 阿里云大数据计算服务（MaxCompute，原 ODPS）是阿里巴巴自主研发的海量数据处理平台，主要服务于批量结构化数据的存储和计算。以下哪个场景不适合使用大数据计算服务实现？
  - a) 在线交易系统
  - b) 数据仓库
  - c) 大数据的分析建模
  - d) 网站日志离线分析
2. 阿里云大数据计算服务（MaxCompute，原 ODPS）中的表 event 是分区表，分区键是 dt（类型为 string），每天生成一个分区，现在表中有 dt='20160101' 至 dt='20160531' 共 5 个月的数据，为了统计 3 月份 eventid 非空的数据量，开发人员运行了以下语句：select count(\*) from event where substr(dt,1,6)='201603' and eventid is not null; 对此任务的描述正确的是\_\_\_\_\_。
  - a) 此任务需要读 event 表中所有分区中的数据
  - b) 此任务只需要读 event 表中 dt='20160301' 至 dt='20160331' 共 31 个分区中的数据
  - c) 此任务会读 event 表中所有分区中 eventid 非空的数据
  - d) 此任务只需要读 event 表中 dt='20160301' 至 dt='20160331' 共 31 个分区中 eventid 非空的数据
3. 某信贷公司推出一款线上贷产品，采用阿里云的数据处理技术，通过对会员的历史数据进行分析，包括交易量、网上信用评价、企业自身经营状况等等，每天处理的数据量在 20P 左右，基于这些数据对用户信用进行聚类分析建模并产生预测数据，用户线上申请贷款时，能在 1 秒钟内根据用户模型预测结果得到该用户的资质评估、授信等信息。请根据您对阿里云产品的理解判断，下述方案中成本最低、性能最好的一个是\_\_\_\_\_。
  - a) 使用阿里云大数据计算服务（MaxCompute，原 ODPS）对数据进行建模，然后使用该模型对所有符合贷款条件的会员进行资质评估和授信评估，将处理结果保存在 MaxCompute 中
  - b) 使用阿里云大数据计算服务（MaxCompute，原 ODPS）对数据进行建模，然后使用该模型对所有符合贷款条件的会员进行资质评估和授信评估，将处理结果保存在阿里云表格存储（Table Store，原 OTS）中，以供实现快速查询
  - c) 使用阿里云分析型数据库（Analytic DB，原 ADS）对数据进行建模，然后使用该模型对所有符合贷款条件的会员进行资质评估和授信评估，将处理结果保存在 Analytic DB 中

- d) 使用阿里云分析型数据库（Analytic DB，原 ADS）对数据进行建模，然后使用该模型对所有符合贷款条件的会员进行资质评估和授信评估，将处理结果保存在阿里云表格存储（Table Store，原 OTS）中
4. 某企业选择了使用阿里云大数据计算服务（MaxCompute，原 ODPS）进行数据仓库建设，由一个多人的团队负责实施和运维。为了更好的进行角色分工、数据权限、多用户协同等管理工作，同时还能实现任务的自动调度、任务监控、错误告警及处理等运维工作，并且需要有一个简单易用的开发环境，以下方案中\_\_\_\_\_是最便捷、最高效的。
- a) 用 MaxCompute 的 console 客户端作为主要开发工具
  - b) 使用 Maxcompute 提供的 SDK 自己开发一个客户端工具
  - c) 使用 Data IDE 作为开发工具
  - d) 使用 MaxCompute 的管理控制台

## 二. 多选题

1. DataX 是阿里集团发起的开源项目，是一种离线数据同步的工具，支持多种异构数据源之间的高效的数据同步。DataX 支持的数据源包括\_\_\_\_\_。（正确答案的数量：3 个）
- a) 关系型数据库：Mysql、SQLserver、Oracle
  - b) 阿里云大数据分析处理 MaxCompute（原 ODPS）、Hbase、阿里云分析型数据库（Analytic DB，原 ADS）
  - c) 半结构化数据存储：阿里云表格存储（Table Store，原 OTS）
  - d) 非结构化数据存储：图片、视频
2. 阿里云大数据计算服务（MaxCompute，原 ODPS）中的 MapReduce 是一种编程模型，用于大规模数据集的并行运算，它极大地方便了编程人员在不会分布式并行编程的情况下，将自己的程序运行在分布式系统上。其中，两个主要阶段 Map 和 Reduce 相互配合，可以完成对海量数据的处理。关于这两个阶段的关系，说法正确的有\_\_\_\_\_。（正确答案的数量：3 个）
- a) 一个 map 的输出结果可能会被分配到多个 reduce 上去
  - b) 一个 reduce 的输入数据可能来自于多个 map 的输出
  - c) 一个 MR 处理可以不包括任何 map
  - d) 一个 MR 处理可以不包含任何 reduce
3. 阿里云大数据计算服务（MaxCompute，原 ODPS）中，在做 web 日志分析时如果发现日志内容（对应字段 log\_content）中有 'select' 字符串，则有可能是恶意的 sql 注入攻击，为了从日志表 log 中找出所有符合这种特征的访问，可以使用以下哪些方式？（正确答案的数量：3 个）
- a) `select * from log where instr(tolower(log_content),'select')>0;`
  - b) `select * from log where regexp_instr(tolower(log_content),'select',1)>0;`

- c) `select * from log where 'select' in tolower(log_content);`
- d) `select * from log where tolower(log_content) like '%select%';`

### 三. 判断题

1. 在阿里云大数据计算服务（MaxCompute，原 ODPS）中，当一个用户被移除出项目空间（project）后，与该用户有关的 ACL 授权仍然会被保留。一旦该用户以后被再添加到该项目空间时，该用户的历史的 ACL 授权访问权限将被重新激活。
2. 在为企业的 IT 系统做架构设计时，成本是一个很重要的考虑因素。对于分析型数据库（Analytic DB，原 ADS）和大数据计算服务（MaxCompute，原 ODPS），前者使用了 SSD 盘存储数据，价格比较贵，因此在两个产品都能实现的场景下，优先选择大数据计算服务，会降低用户的整体使用成本。
3. Data IDE 中，调度任务每次运行前都先将任务实例化，即生成实例，调度运行时实际上在执行相应的实例。按天调度的周期性任务每天生成一个自动调度实例；按 00:00 至 23:59 之间每隔一小时执行一次的周期性任务，每天生成 24 个自动调度实例。