

申明：本文档是根据当前公司获取的试题资源进行的知识点分类汇总。试题中的答案非官方标准答案，在练习时请参照阿里官方资料进行配合复习

数据同步

1. 24、在 Data IDE 中，当从云数据库 RDS(MySQL)抽取数据时，可配置切分键进行数据分片，提高数据同步效率，以下关于切分键的说法中正确的有（ ）。（正确答案的数量：2 个）B C
a. 可以选择任意字段作为切分键
b. 从云数据库 RDS 导出数据到 MaxCompute 配置切分键可以实现并发，提高数据同步速度
c. 推荐使用表的主键作为切分键
d. 配置切分键容易出现数据热点
2. Data IDE 中的数据同步任务只支持全量数据同步。（F）
3. 在 Data IDE 中使用数据同步任务时，可以将某个自定义变量的值写入目标表中，自定义变量采用“变量名=参数”的格式进行赋值。（T）
4. 在 Data IDE 中配置数据同步任务将云数据库 RDS(MySQL)同步至大数据计算服务（MaxCompute，原 ODPS）中，需要确认在云数据库 RDS 实例的 IP 白名单中是否包含了 Data IDE 的 IP 段，如未包含，则需要添加。（T）
5. 60、通过阿里云的数据集成服务（Data Integration，原云道 CDP），可以将云数据库 RDS、分布式关系型数据库 DRDS、分析型数据库（AnalyticDB，原 ADS）中存储的结构化数据同步到大数据计算服务（MaxCompute，原 ODPS）中去。（T）（本题在多套题中出现，但选择错，可是在在线文档中支持这些数据类型的导入）
6. 6、Data IDE 的数据通知任务 Task1，将数据从大数据计算服务（MaxCompute，原 ODPS）同步到云数据库 RDS（MySQL）中，源数据表中存在数据质量问题，加载的时候会有部分出错纪录，Task1 中已经配置了容错纪录条数，为了保证正确纪录能够保留，在数据加载与控制中，还需要对 Task1 配置（B）。
a. 源表切分主键
b. 视为脏数据，保留原有数据（insert into）
c. 作业速率上限
d. 数据过滤条件
7. 7、在 Data IDE 中，某数据同步任务执行成功，但是发现目标表数据为空。以下说法正确的有_____。（正确答案的数量：2 个）（AD）
a. 查看具体运行日志，看日志里具体源表信息，经检查满足抽取条件的数据纪录数为 0
b. 运行任务帐号只有目标表的读权限却没有写权限
c. 查看具体运行日志，看日志里具体源表信息，造成目标表为空的原因是源表不存在
d. 当配置了最大容错条数，即“当出错的纪录数超过”的值，如果源表中满足抽取条数的记录数小于当前配置的最大容错条数的值，并且所有的纪录全部同步出错，则整个同步任务的状态为执行成功，

批注 [s1]: 重点!!!!

这两个错误会抛出错误

目标表数据为空

8. 15、Data IDE 中，配置数据同步任务时，需要先在对应项目空间中配置数据源，即配置源和目标对应的数据源的信息等。当数据源（B）数据库时，肯定不属于云数据库 RDS 类型。

a. mysql
b. oracle
c. sqlserver
d. postgresql

批注 [s2]: RDS 支持数据库类型！不包括 oracle

9. 19、在使用 Data IDE 进行数据同步时，会因为数据质量问题，导致部分记录无法从源表同步至目标表中去。数据同步任务 Task1 的目标表为云数据库 RDS 中的表，假设数据同步过程中只会因主键冲突产生错误记录。以下关于主键冲突配置中说法正确的是（A）

a. 当主键冲突配置为“视为脏数据，保留原有数据（insert into）”，则后导入的数据被丢弃，并被视作脏数据
b. 当主键冲突配置为“替换原有数据（replace into）”时，会使用新数据直接替换原有纪录，并计入脏数据纪录
c. 无需配置主键冲突处理规则，所有的主键冲突都不会被视为脏数据
d. 无需配置主键冲突处理规则，所有的主键冲突都会被自动视为脏数据

10. 21、在 Data IDE 数据同步任务（云数据库 RDS 数据同步至 MaxCompute）中，需要根据源表中的时间戳字段实现数据的增量抽取，用户应该配置（C）选项。

a. 作业速率上限
b. 容错纪录条数
c. 数据过滤条件
d. 源表切分主键

11. 在 Data IDE 中可以使用数据同步任务，将云数据库 RDS(MySQL)数据同步至大数据计算服务（MaxCompute，原 ODPS）的分区表中，通常会把日期作为目标表的分区键，以方便实现增量数据加载。现在想把 RDS 中的表 A 中的数据，按照不同的日期导入到 MaxCompute 中表 B 对应的不同分区中（分区字段的格式为 yyyyymmdd），如将 20160318 的数据导入到 MaxCompute 表的分区'2016-03-18'中，MaxCompute 的分区可以通过（B）实现。

a. 在 MaxCompute 侧的分区中写入常量值 20160318 即可
b. 在 MaxCompute 侧中使用系统内置参数\${bdp.system.bizdate}
c. 在 MaxCompute 侧中使用系统内置参数\$bizdate
d. 在 MaxCompute 侧中使用系统内置参数\${bdp.system.cyctime}

12. 23、在使用 Data IDE 进行数据同步时，会因为数据质量问题，导致部分纪录无法从源表同步至目标表中去。任务 Task1 是从云数据库 RDS(MySQL)表同步至大数据计算服务（MaxCompute，原 ODPS）表中，可以通过设置以下哪两项进行容错控制？（正确答案的数量：2 个）（AD）

a. 设置“当出错的记录数超过”的值，即最大容错纪录条数
b. 设置作业速率上限
c. 设置切分键
d. 设置“错误百分比达到”的值，即最大容错纪录条数占总条数的比例

批注 [s3]: 考到!!!!

13. 26、在 Data IDE 数据同步任务中，合理的选择和配置切分键可以提高云数据库 RDS(MySQL) 到大数据计算服务（MaxCompute，原 ODPS）数据同步效率。（T）
14. 在 Data IDE 中使用数据同步任务将云数据库 RDS（MySQL）的用户日志表同步至大数据计算服务（MaxCompute，原 ODPS）中，其中 RDS 表中的用户访问时间字段为 visit-time（datetime 类型），格式为 YYYY-MM-DD HH:MM:SS,需要将该数据转化并存储在 MaxCompute 对应 string 类型列中，存储的日期格式为 YYYYMMDD，以下做法正确的是（B）
- a. 在 RDS Mysql 数据同步至 MaxCompute 中，RDS Mysql 侧不支持使用函数
 - b. 在数据同步字段配置界面中增加字段‘DATE-FORMAT(visit_tim， %Y%m%d)’并与 MaxCompute 时间列对应
 - c. 在数据同步字段配置界面中增加字段‘DATE-FORMAT(visit_tim， %Y-%m-%d)’并与 MaxCompute 时间列对应
 - d. 在数据同步字段配置界面中增加字段 DATE-FORMAT(visit_tim， ‘%Y%m%d’) 并与 MaxCompute 时间列对应
15. Data IDE 中表 A 的数据是通过按天调度的数据同步任务产生的，表 A 中的分区字段 dt 为 string 类型，填充的是数据产生的日期（格式为“yyyy-mm-dd”），即任务自动调度日期的前一天，比如 2016 年 4 月 9 日运行调度任务的时候，dt 的值为“2016-04-08”.在配置周期性同步任务的时候，A 表的 dt 字段需要（）才使得任务每次调度运行时自动替换日期。（C）
- A. 直接使用系统时间变量\${bdp.system.cyctime}
 - B. 引入自定义变量 var，令 var=\${yyyy-mm-dd-1}，使用 var 对 dt 进行赋值
 - C. 直接使用系统时间变量\${bdp.system.bizdate}
 - D. 引入自定义变量 var，令 var=\${yyyy-mm-dd}，使用 var 对 dt 进行赋值
16. 在 Data IDE 中使用数据同步任务将云数据库 RDS 数据导入大数据计算服务（MaxCompute，原 ODPS）时，将 MaxCompute 侧数据清理规则配置为导入前保留已有数据，可以有效避免测试引入脏数据带来的主键冲突问题 （T）
17. 在 Data IDE 中配置数据同步任务，将主键冲突选项配置为“视为脏数据，保留原有数据”时，数据同步任务进行时，脏数据还是会写入目标表中。（F）

角色授权

1. Data IDE 中不同角色拥有不同的权限，可以通过角色授权来控制成员的操作权限。角色主要有组织管理员、项目管理员、开发者、访客等，其中开发角色不具备（）权限。（B）
- a. 新建任务
 - b. 添加项目成员
 - c. 上传资源
 - d. 新建函数

2. 18、Data IDE 中，项目空间 prj1 只能配置一个成员为项目管理员 (F)
3. 16、Data IDE 项目空间的管理员角色拥有项目空间的各种管理权限，以下操作中 (BD) 必须由拥有项目管理员角色的成员来完成。(正确答案的数量：2 个) (BD)
- a. 添加调度资源组
 - b. 把组织成员添加为项目成员并分配合理的角色
 - c. 创建 Data IDE 项目空间
 - d. 添加数据源
4. 17、组织是 Data IDE 的特有概念，组织的基本对象为项目空间，以下说法正确的有 (ABC)。(正确答案的数量：3 个) (ABC)
- a. 一个帐号可以加入同一个组织的多个项目空间
 - b. 一个帐号只能加入一个组织
 - c. 一个组织可以创建多个项目空间
 - d. 一个项目空间可以属于多个组织
5. 出于业务需要，要把云数据库 RDS 中的某张表的数据同步到大数据计算服务 (MaxCompute，原 ODPS) 的某张表中去，希望通过 Data IDE 的同步任务来实现，以下关于实现流程的说法正确的是 (C)
- A. 拥有项目管理员角色的成员先配置 RDS 数据源，然后拥有运维角色的成员创建 MaxCompute 表，再创建、配置数据同步任务
 - B. 拥有开发角色的成员先配置 RDS 数据源，然后创建 MaxCompute 表，再创建、配置数据同步任务
 - C. 拥有项目管理员角色的成员先配置 RDS 数据源，然后拥有开发角色的成员创建 MaxCompute 表，再创建、配置数据同步任务
 - D. 拥有开发角色的成员先配置 RDS 数据源，然后拥有运维角色的成员创建 MaxCompute 表，再创建、配置数据同步任务
6. Data IDE 中组织管理员是指组织的管理者，拥有组织层面的管理权限，以下操作中，(BCD) 只能由组织管理员才能完成 (正确答案 3 个) (BCD)
- A. 添加数据源
 - B. 添加调度资源组
 - C. 添加组织成员
 - D. 创建 Data IDE 项目空间

批注 [s4]: 项目管理员配置数据源，开发者创建表，配置数据同步任务，运维角色负责运维工作

任务调度

7. Data IDE 中，周期性调度任务的调度规则是：若任务有上游任务则上游任务必须执行成功，且当前时间点已经等于或超过任务的定时时间，两个条件都满足后当前任务才可以运行；若该任务没有上游任务，则定时时间满足即可运行。Task2 是按天调度的周期任务，配置定时时间 12:00，依赖属性的上游任务配置为任务 task1，task1 是按天调度的周期任务，定时时间配置为 0:00，关于 task2 每天调

度情况的说法中正确的是_____。(A)

- a. task1 执行成功后，如果时间没到 12 点，那么等到 12 点 task2 可以运行
- b. task1 执行成功后，task2 会立即运行
- c. task1 执行成功后，如果时间超过 12 点，那么 task2 任务已经无法运行
- d. task1 任务执行成功时，时间必须刚好是 12 点 task2 才能运行

8. 在 Data IDE 中配置数据同步任务时，将云数据库 RDS (MySQL) 中的数据同步至大数据计算服务 (MaxCompute, 原 ODPS) 分区表对应的日期分区中，采用数据同步任务周期性的调度向不同分区中写入数据，把 yyyymmdd 的数据写入对应的 yyyy-mm-dd 分区中，如 20160308 的数据需要写入 MaxCompute 表的 2016-03-08 分区中，MaxCompute 的分区可以通过 () 实现。(D)

- a. 在 MaxCompute 侧中使用系统内置参数\$bizdate
- b. 在 MaxCompute 侧中使用变量\${datetime}，并在节点参数出赋值 datetime=\${YYYY/MM/DD}
- c. 在 MaxCompute 侧的分区中写入常量值 2016-03-18 即可
- d. 在 MaxCompute 侧中使用变量\${datetime}，并在节点参数出赋值 datetime=\${YYYY-MM-DD}

9. 在 Data IDE 中，某数据同步任务执行成功，但是发现目标表数据为空。以下说法正确的有 ()。(正确答案的数量：2 个) (BD)

- a. 查看具体运行日志，看日志里具体源表信息，造成目标表为空的原因是源表不存在
- b. 查看具体运行日志，看日志里具体源表信息，经检查满足抽取条件的数据记录数为 0
- c. 运行任务帐号只有目标表的读权限却没有写权限
- d. 当配置了最大容错条数，即“当出错的纪录数超过”的值，如果源表中满足抽取条件的纪录数小于当前配置的最大容错条数的值，并且所有的纪录全部同步出错，则整个同步任务的状态为执行成功，目标表数据为空

10. Data IDE 中，事实表 A 需要按天增量更新 (配置成任务 task1)，事实表 B 需要按天增量更新 (配置成任务 task2)，每天事实表 A 中的部分数据信息来自事实表 B 前一天的数据。task1 和 task2 需要如何配置才能实现上述场景? (C)

- a. task1 的依赖属性中的上游任务配置为 task2
- b. 不需要配置依赖关系，在 task1 中直接使用表 B 生成表 A 即可
- c. task1 的跨周期依赖属性设置为“等待自定义任务的上一周期结束”，自定义任务选择 task2
- d. task2 的依赖属性中的上游任务配置为 task1

11. Data IDE 中，周期性调度任务是指按配置的时间周期进行自动调度的任务，以下场景中，() 不适合配置成周期性调度任务。(C)

- a. MaxCompute 中基于商品库存表每隔一小时生成一张库存盘点的汇总表
- b. 每隔 5 分钟对 MaxCompute 中的访问日志表进行一次数据清洗
- c. 今天市场部经理临时要求分析师基于 MaxCompute 中的用户交易表生成一份昨天有成交纪录的品牌列表
- d. 把云数据库 RDS 中的用户登录记录表每天 0 点 30 分同步到 MaxCompute 表中

12. 33、Data IDE 中，调度系统可以根据配置好的任务属性进行自动调度。属性中包含的生效日期是指任务在有效期内可以进行调度运行，任务 task1 的生效日期为 2016-01-01 至 2016-06-30，以下关于任务 task1 的说法中正确的有 ()。(正确答案的数量：2 个) (AD)

- a. task1 在 2016 年 7 月 1 日当天在调度上无法操作测试

- b. task1 在 2016 年 4 月 1 日那天操作补数据时无法补业务日期为 2016 年 1 月 1 日的数据
- c. task1 在 2016 年 6 月 30 日当天就不会再自动调度运行
- d. task1 如果在 2016 年 1 月 1 日前添加, 则 2016 年 1 月 1 日当天才会自动调度运行

13. 34、Data IDE 中, 周期性调度任务在配置依赖属性时, 上游任务只能是同个项目空间的任务。(F)

14. 35、Data IDE 中, task1 和 task2 都是按天调度的周期性调度任务, task1 设置定时时间为 02:00, task2 设置定时时间为 00:00, task2 依赖属性的上游任务配置为 task1, 每天 task2 何时可以运行?

(A)

- a. 只要 task1 执行成功, task2 就马上运行
- b. task1 在 2:00 之后才执行成功, 则 task2 不再运行
- c. task1 执行成功时, 时间若已到 2:00, 则 task2 可以运行
- d. task1 执行成功时, 时间没到 2:00, 则 task2 不再运行

task2处于等待状态, 因为时间是在task1之前

批注 [s5]: 考到!!!

15. 36、Data IDE 中, 调度任务每次运行前都先将任务实例化, 即生成一个实例, 调度运行时实际上在执行相应的实例。任务配置不同的调度周期, 自动调度生成的实例个数也就不一样, 以下说法正确的是?

(正确答案的数量: 3 个) (ABD)

- a. 调度周期配置成按天调度的任务, 每天生成一个自动调度实例
- b. 调度周期配置成在 00:00 至 23:59 之间按分钟调度, 每隔 30 分钟执行一次的任务, 每天生成 48 个自动调度实例
- c. 调度周期配置为按周调度, 每周调度一次的任务, 只有周一才会生成一个自动调度实例, 周二到周日都不会生成实例
- d. 调度周期配置成在 00:00 至 23:59 之间按小时调度, 每天生成 24 个自行调度实例

16. 37、Data IDE 中, 虚拟节点类型是空跑节点, 即调度时直接返回成功状态不会真正执行。两个周期调度任务 task1、task2 之间没有依赖关系, 业务上也没有共同上游任务, 但是希望可以有一个共同的父任务进行统筹管理, 那么可以创建一个虚拟节点作为 task1 和 task2 的上游任务。(T)

批注 [s6]: 考到!!!!

17. 38、Data IDE 中, 调度任务每次运行前都先将任务实例化, 即生成一个实例, 调度运行时实际上在执行相应的实例。可以通过运维中心的任务运维管理暂停任务自动调度的实例。如果按天调度的周期性任务 task1 的 2016 年 4 月 9 日自动调度的实例被操作暂停, 以下关于 task1 的说法中正确的是。

(A)

- a. 仅 2016 年 4 月 9 日的自动调度实例被暂停
- b. 2016 年 4 月 8 日的自动调度的实例也会被暂停
- c. 所有在 2016 年 4 月 9 日之后生成的自动调度的实例都会被暂停
- d. 2016 年 4 月 10 日自动调度的实例也会被暂停

批注 [s7]: 考到!!!!

18. 39、Data IDE 中, 运维中心的“任务管理视图”以 DAG 图(有向无环图)的方式显示调度任务之间的依赖关系和运行约束。通过任务管理视图可以查看任务的定义, 但不能查看任务的运行状态, 以下关于任务管理视图说法错误的是。

(C)

- a. 可以查看任务配置的调度周期粒度
- b. 可以查看任务 owner
- c. 可以查看任务影响的记录条数
- d. 可以查看任务描述

19. 40、Data IDE 中，调度任务每次运行前都先将任务实例化，即生成一个实例，调度运行时实际上在执行相应的实例。在调度的不同阶段，实例会处于不同的状态，包括未运行、运行中、运行成功、运行失败、等待时间、等待资源等。当实例处于_____状态时不能手工停止。(A)

a. 未运行
b. 运行中
c. 等待时间
d. 等待资源

批注 [s8]: 考到!!!!

20. 41、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。在调度的不同阶段，实例会处于不同的状态，包括_____。(正确答案的数量: 3 个) (BCD)

a. 未完成
b. 运行中
c. 未运行
d. 失败

批注 [K9]: 运行成功状态

未运行状态
运行失败状态
正在运行状态
终止运行手动
等待状态

21. 42、Data IDE 中，调度任务每次运行前都先将任务实例化，即生产实例，调度运行时实际上在执行相应的实例。运维中心的任务运维视图显示的实例是任务自动调度的实例，对这些实例可以进行_____。(正确答案的数量: 2 个) (BC)

a. 补数据
b. 重跑并恢复调度
c. 查看节点运行日志
d. 测试

补数据是要重新生成示例的

22. 43、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。(T)

23. 3、Data IDE 中，调度任务每次运行前都先将任务实例化，即生成一个实例，调度运行时实际上在执行相应的实例，每个实例带有具体的定时时间。对于按天调度的周期性任务 task1，以下说法下正确的是_____。(A)

a. 每天生成 1 个自动调度实例
b. 使用运维中心补数据时，一天补 10 天的数据，则会生成 10 个自动调度的实例
c. 使用运维中心测试时，会生成一个自动调度的实例
d. 使用运维中心补数据时，会生成一个自动调度的实例

24. 27、Data IDE 的数据同步任务 Task1，将云数据库 RDS(MySQL) 的表 Table1 中的数据增量同步至大数据计算服务 (MaxCompute，原 ODPS) 的表 Table2 对应的分区中。可以在 MaxCompute 侧配置变量分区表达式，将对应日期数据写入对应的分区中去。(T)

25. 28、Data IDE 工作流任务支持一次性调度和周期调度两种类型，对于周期性调度可以设置不同时间粒度的调度周期，以下对任务按月调度的调度周期配置，说法错误的是 (C)。
- a. 可以配置每月有固定几天需要运行

补数据生成的补数据实例将按照指定的业务日期运行！！

批注 [K10]: •查看节点运行日志: 节点正在运行、成功、失败等状态时查看任务运行的日志。

•查看节点属性: 查看实例属性，包括实例运行的各种时间信息、运行状态等。

•终止: 可以 kill 等待时间、re 等待资源、运行状态的任务，并将任务置为失败状态。

前置条件: 只能终止等待时间、等待资源、运行中状态的任务。

•重跑并恢复调度: 可以重跑某任务，任务执行成功后可以触发下游未运行状态任务的调度。常用于处理出错节点和漏跑节点。

前置条件: 只能重跑未运行、成功、失败状态的任务。

•置成功并恢复调度: 将当前节点状态改为成功，并运行下游未运行状态的任务。常用于处理出错节点。

前置条件: 只能失败状态的任务能被置成功。

•重跑下游并恢复调度: 可以重跑某任务及其下游任务，需要用户自定义勾选，勾选的任务将被重跑，任务执行成功后可以触发下游未运行状态任务的调度。常用于处理数据修复。

前置条件: 只能勾选未运行、完成、失败状态的任务，如果勾选了其他状态的任务，页面会提示“已选节点中包含不符合运行条件的节点”，并禁止提交运行。

•暂停: 暂停节点当前周期的调度。

批注 [s11]: 重点!!!

- b. 配置每月 30 日运行，则 2 月份该任务不会运行
c. 无法配置每月 1 日和 15 日运行
d. 无法配置每月最后一天运行
26. 31、Data IDE 中，通过数据开发面板的任务开发，可以创建节点任务，也可以创建工作流任务，而工作流任务内部又可以创建工作流节点，关于节点任务和工作流节点的说法中错误的是 (A)。
- a. 节点任务依赖属性的上游任务可以是工作流任务，也可以是某个工作流节点
b. 节点任务依赖属性的上游任务可以是工作流任务，工作流节点无法单独配置依赖属性
c. 同个工作流的工作流节点之间可以配置依赖关系
d. 节点任务可以配置调度周期，工作流节点不能单独配置调度周期
27. 32、Data IDE 中，周期性调度任务的依赖属性配置决定了任务运行的先后顺序，任务 task1、task2、task3 都是按天调度的周期性任务，下列关于 3 个任务说法正确的有？（正确答案的数量：2 个）AD
- a. 若 task1 执行时需要用到 task2 当天生成的最新数据，那么 task1 依赖属性的上游任务应该配置 task2
b. 若 task2 执行时需要用到 task1 当天生成的最新数据，task3 每天需要用到 task2 当天生成的最新数据，那么 task3 依赖属性的上游任务应该直接配置 task1
c. 若 task2 执行时需要用到 task1 当天生成的最新数据，task3 每天需要用到 task2 当天生成的最新数据，那么 task3 依赖属性的上游任务必须直接配置 task2 和 task1
d. 若 task2 执行时需要用到 task1 当天生成的最新数据，task3 每天需要用到 task2 当天生成的最新数据，那么 task3 依赖属性的上游任务直接配置 task2
28. 34、Data IDE 中，调度系统提供了一些常用的时间参数，使周期任务自动调度运行时参数会自动替换具体的时间值。其中 \${bdp.system.bizdate} 参数值是取任务运行时的定时时间所在的年月日（即业务日期），按天调度的周期任务配置定时时间为 20:00，2016-01-01 当天自动调度时该参数赋值结果是 20160101。(F)
29. 35、Data IDE 中，task1 和 task2 都是按天调度的周期性调度任务，task1 生成分区表 table1，task2 生成分区表 table2，每天 task2 取 table1 最新分区的数据进行统计，而 task1 有部分数据来源是 table2 前一天分区的数据。task1 和 task2 的依赖关系需要任何配置？(D)
- a. 只需要 task2 依赖属性的上游任务配置为 task1
b. task1 和 task2 都不需要配置依赖属性
c. 只需要 task1 依赖属性的上游任务配置为 task2
d. task1 跨周期依赖选择“等待自定义任务的上一周期结束，才能继续运行”，自定义任务填写 task2，task2 依赖属性的上游任务配置为 task1
30. 36、Data IDE 中，任务 task1 是按天调度的周期任务，task1 生成分区表 table1，该表每次执行生成的分区（分区名 ds）取前一天日期的年月日（格式为 yyyyymmdd），要使用调度系统提供的时间参数定义 table1 的分区，使得 task1 每次调度运行时自动替换时间值，task1 应该如何配置？（正确答案的数量：2 个）(AC)
- a. task1 的代码中 table1 的分区 ds=\${bdp.system.bizdate}
b. task1 的代码中 table1 的分区 ds=\${var}，task1 的参数配置：var=\${yyyyymmdd}
c. task1 的代码中 table1 的分区 ds=\${var}，task1 的参数配置：var=\${yyyyymmdd-1}
d. task1 的代码中 table1 的分区 ds=\${bizdate}

工作流内部节点不能被依赖！！！！

批注 [s12]: 考到!!!!

biz是前一天！！！！！！

互相引用，需要使用周期配置

31. 37、Data IDE 中，虚拟节点类型是空跑节点，即调度时直接返回成功状态不会真正执行。两个周期调度任务 task1、task2 之间没有依赖关系，业务上也没有共同上游任务，但是希望可以有一个共同的父任务进行统筹管理，那么可以创建一个虚拟节点作为 task1 和 task2 的上游任务。(T)

32. 38、Data IDE 中，调度任务每次运行前都将任务实例化，即生成一个实例，调度运行时实际上在执行相应的实例。在调度的不同阶段，实例会处于不同的状态，包括未运行、运行中、运行成功、运行失败、等待时间、等待资源等。当实例处于_____状态时可以将任务量为成功并恢复调度。(D)

- a. 运行中
- b. 未运行
- c. 等待时间
- d. 失败

批注 [s13]: 重点!!!

33. 40、Data IDE 中，通过运维中心的任务运维视图可以对调度任务操作重跑并恢复调度，即重跑该任务同时把下游未运行的任务恢复调度。任务 task3 依赖任务 task2，任务 task2 依赖任务 task1，对 task2 操作重跑并恢复调度，以下哪个场景会使 task2 和 task3 执行?(D)

- a. task1 为失败状态，task2、task3 都是未运行状态
- b. task1、task2、task3 都是运行成功状态
- c. task1、task2、task3 都是未运行状态
- d. task1 为成功状态，task2 为失败状态，task3 为未运行状态

34. 41、Data IDE 中，调度任务每次运行前都将任务实例化，即生成实例，调度运行时实际上在执行相应的实例。在调度的不同阶段，实例会处于不同的状态，包括_____。(正确答案的数量：3 个)

- (ABC)
- a. 运行中
- b. 失败
- c. 未运行
- d. 未完成

35. Data IDE 上某个数据同步任务执行失败，运行日志中提示：“Code:[OdpReader-33],Description:[分区配置错误.]-分区配置错误，根据您所配置的分区没有匹配到源头表中的分区。源头表所有分区是:[ds=' 2010328' ds=' 20160328'.done' ds=' 20160329' ds=' 20160330' ds=' 20160330.done']，您配置的分区是:[ds='\${daytime}]，请您根据实际情况作出修改。”(B) 会造成类似错误。(注：本题答案不明确，原试题是 B)

- a. 改任务在运行的时候，分区 ds='\${daytime} 被删除了
- b. 配置任务的源表时，制定分区值时使用了自定义变量 \${daytime}，但在数据同步节点的参数配置中未给变量 daytime 赋值
- c. 该任务配置的源表分区还没有生成，该任务就已经开始运行了。可能的原因是该任务没有配置合适的依赖关系
- d. 配置任务的源表时，指定分区值时应该写成：ds=' 20160328' ds=' 20160328.done' ds=' 20160329' ds=' 20160329.done' ds=' 20160330' ds=' 20160330.done'

36. Data IDE 中, 事实表 A 需要按天增量更新 (配置成任务 task1), 事实表 B 需要按天增量更新 (配置成任务 task2), 每天事实表 A 中的部分数据信息来自事实表 B 前一天的数据。Task1 和 task2 需要如何配置才能实现上述场景。(D)
- A. task1 的依赖属性中的上游任务配置为 task2
 - B. 不需要配置以来关系, 在 task1 直接使用表 B 生成表 A 即可
 - C. task2 的依赖属性中的上游任务配置为 task1
 - D. task1 的跨周期依赖属性设置为“等待自定义任务的上一周期结束”, 自定义任务选择 task2
37. Data IDE 中, 通过数据开发面板的脚本开发, 可以创建 SQL 脚本文件并直接编辑、执行。关于脚本文件, 以下说法正确的是 (B)
- A. 脚本文件可以自动调度
 - B. 脚本文件不能发布到目标项目空间
 - C. 脚本文件可以设置依赖属性
 - D. 脚本文件不能保存
38. Data IDE 中, 项目空间 prj1 的 SQL 任务 task1 中, 部分数据来源于项目空间 prj2 的任务 task2 生成的表 table2, task1 和 task2 都是按天调度的周期性调度任务, task1 每次执行要使用 table2 同周期生成的数据, 为了保证 task1 能正常执行, 即每次运行时所需的 table2 数据已经生成, 以下 task1 和 task2 配置说法正确的是? (B)
- A. task1 代码中表 tabke2 的写法为“table2”, task1 依赖属性的上游任务配置为 prj2 中的 task2
 - B. task1 代码中表 tabke2 的写法为“prj2.table2”, task1 依赖属性的上游任务配置为 prj2 中的 task2
 - C. task1 代码中表 tabke2 的写法为“prj2.table2”, 但是 task1 和 task2 不在同一个项目无法配置成依赖关系
 - D. task1 代码中表 tabke2 的写法为“table2”, 但是 task1 和 task2 不在同一个项目无法配置成依赖关系
39. Data IDE 中, task1 和 task2 都是按天调度的周期性调度任务, task1 生成分区表 table1, task2 生成分区表 table2, 每天 task2 取 table1 最新分区的数据进行统计, 而 task1 有部分数据来源是 table2 前一天分区的数据。Task1 和 task2 的依赖关系需要如何配置? (C)
- A. 只需要 task1 依赖属性的上游任务配置为 task2
 - B. task1 和 task2 都不需要配置依赖属性
 - C. task1 跨周期依赖选择“等待自定义任务的上一周期结束, 才能继续运行”, 自定义任务填写 task2, task2 依赖属性的上游任务配置为 task1
 - D. 只需要 task2 依赖属性的上游任务配置为 task1
40. Data IDE 中, SQL 任务 task1 是按小时调度的周期性调度任务, 任务配置为每 1 小时执行一次, 结果会插入到表 table 中去, table1 为一个有两个分区键 (ds 和 hh) 的分区表, 其中 ds 取前一个小时的年月日 (格式为 yyyyymmdd), hh 取前一小时的小时 (格式为 hh24), 要使用调度系统提供的时间参数定义分区, 使得 task1 每次调度运行中自动替换时间值, task1 应该如何配置? (B)
- 注: 本题倾向 D**
- A. task1 代码里 table1 的分区 ds=\${daytime}, hh=\${hh}, 任务参数配置: daytime=\${yyyyymmdd}, hh=\${hh24-1/24}
 - B. task1 代码里 table1 的分区 ds=\${bdp.system.bizdate}, hh=\${hh}, 任务参数配置: hh=\${hh24-1/24}
 - C. task1 代码里 table1 的分区 ds=\${bdp.system.bizdate}, hh=\${hh}, 任务参数配置: hh=\${hh24}
 - D. task1 代码里 table1 的分区 ds=\${daytime}, hh=\${hh}, 任务参数配置: daytime=\${yyyyymmdd}-

1/24],hh=\${hh24-1/24}]

41. Data IDE 中,通过运维中心的任务运维视图可以对调度任务操作重跑并恢复调度,即重跑该任务同时把下游未运行的任务恢复调度,任务 task3 依赖任务 task2,任务 task2 依赖任务 task1,对 task2 操作重跑并恢复调度,以下哪个场景会使 task2 和 task3 执行? (A)

A. task1 为成功状态,task2 为失败状态,task3 为未运行状态
B. task1、task2、task3 都是未运行状态
C. task1、task2、task3 都是运行成功状态
D. task1 为失败状态,task2、task3 都是未运行状态

42. Data IDE 中,调度任务每次运行前都先将任务实例化,即生成一个实例,调度运行时实际上在执行相应的实例。运维中心中的任务运维视图有运维、测试、补数据 3 种类型实例,运维实例是任务日常自动调度生成的实例,测试实例和补数据实例是通过任务管理视图对任务进行测试和补数据生成的实例。以下哪种类型任务不会生成自动调度实例? (C)

A. 按天调度的周期调度 workflow 任务
B. 按月调度的周期调度 workflow 任务
C. 一次性 workflow 任务
D. 按周调度的周期调度 workflow 任务

43. Data IDE 中,调度任务每次运行前都先将任务实例化,即生成一个实例,调度运行时实际上在执行相应的实例。通过运维中心的任务运维视图可以查看任务自动调度的实例。任务 task1 是按天调度的周期性任务,2016 年 4 月 9 日 11 点 30 分成功提交到调度系统。2016 年 4 月 10 日到任务运维视图查看 task1 的实例时可以查到____。(C)

A. task1 业务日期为 2016-04-08 的实例
B. task1 业务日期为 2016-04-10 的实例
C. task1 业务日期为 2016-04-09 的实例
D. 无相关实例

批注 [s14]: 考到!!!

44. Data IDE 中,项目空间 prj1 随着调度任务量的不断增多,经常出现任务调度时长时间处于等待资源状态。为了对自己负责的周期性任务的调度时间进行规划调整,prj1 的开放角色成员 A 想统计 prj1 每天执行时间最长且任务负责人为自己的 10 个任务,通过__可以快速查到。(B)

注:原答案 A,修改后为 B

A. 运维中心的任务管理列表
B. 运维中心的任务运维列表
C. 数据开发面板中的节点属性
D. 项目管理中的血缘分布

45. 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。周期性调度任务每天自动调度的实例都是在当天 23:30 开始生成的,以下说法正确的是? (3 个) (ABD)

A. 任务 task1 是按天调度的周期性调度任务,2016 年 4 月 9 日 23 点 00 分,把 task1 的定时时间从原来零点整改成 8 点 30 分,4 月 10 日,task1 的实例定时时间为 8 点 30 分
B. 任务 task1 是任务 task2 的上游任务,2016 年 4 月 9 日 11 点 30 分,把 task2 的上游任务换成 task3,当天 task2 的自动调度实例依赖的还是 task1 的实例

- C. 2016 年 4 月 9 日 11 点 30 分新增的按天调度的周期性调度任务，当天可以自动调度运行
- D. 2016 年 4 月 9 日 11 点 30 分新增的按天调度的周期性调度任务，4 月 10 日可以自动调度运行
46. Data IDE 中，一次性 workflow 任务不会被自动调度运行，需要运行时，可以到运维中心对一次性任务操作测试的方式手法运行，以下关于一次行 workflow 说法正确的是（）（2 个）（AC）
- A. 一次性 workflow 创建好后，无法修改成周期调度 workflow
- B. 一次行 workflow 测试的时候，业务日期只能选择前一天的日期
- C. 周期调度任务的依赖属性的上游任务不能配置为一次性 workflow 不然拿不到结果！！
- D. 一次行 workflow 的依赖属性的上游任务可以配置为周期调度任务
47. Data IDE 中，任务 task1 是按天调度的周期任务，task1 生成分区表 table1，该表每次执行生成的分区（分区名 ds）取前一天日期的年月日（格式为 yyyymmdd），要使用调度系统提供的时间参数定义 table1 的分区，使得 task1 每次调度运行时自动替换时间值，task1 应该如何配置？（2 个）（AB）
- A. task1 的代码中 table1 的分区 ds=\${bdp.system.bizdate}
- B. task1 的代码中 table1 的分区 ds=\${var}，task1 的参数配置：var=\${yyyymmdd-1}
- C. task1 代码中 table1 的分区 ds=\${bizdate}
- D. task1 的代码中 table1 的分区 ds=\${var}，task1 的参数配置：var=\${yyyymmdd}
48. Data IDE 中，可以通过脚本开发创建 SQL 脚本文件，用于执行一些临时性的代码。以下哪些情况下可以使用脚本文件？（3 个）（BCD）
- A. 数据科学家在脚本文件里调度机器学习任务实现复杂分析
- B. 数据分析师使用 SQL 探查表数据
- C. 数据开发人员使用建表语句建表
- D. 数据开发人员发调度任务调试成功后，使用 SQL 检查表中的数据是否正常
49. Data IDE 中，某个数据同步任务的源表为云数据库 RDS（MySQL）中的表，目标表为大数据计算服务中的表，运行的时候任务报错，执行日志中提示“ODPS-0130131:Table not found - 'xxx_prj.xxx_table'”，以下说法正确的是（ACD）（3 个）
- A. 表不存在有可能是在配置完任务之后被删除了，所以执行就报错了
- B. 该日志说明目标表不存在
- C. 该表不存在有可能是配置的时候选错了表名
- D. 该表不存在有可能是当前所属项目为生产项目，任务在开发项目配置测试好了，发布到生产项目后，没有到生产项目创建该表
50. 在 Data IDE 中使用数据同步任务将云数据库 RDS（MySQL）加载到大数据计算服务时，为了提高单个数据同步任务的速度及效率，可以采取（BD）方式？（2 个）
- A. 数据过滤条件
- B. 将云数据库 RDS 表的主键配置为切分键
- C. 容错记录条数
- D. 提高作业速率上限的配置
51. Data IDE 中的数据同步支持离线批量数据同步模式，离线数据同步是指数据周期性、成批量地从源端系统传输到目标端系统，关于数据同步周期说法正确的是（）（正确答案 4 个）（ABDE）

- A. 支持天级调度
- B. 支持周级调度
- C. 支持实时调度
- D. 支持小时级调度
- E. 支持月级调度

批注 [s15]: 考到!!!!

52. Data IDE 中，调度系统提供一些常用的时间参数，使周期任务自动调度运行时参数会自动替换具体的时间值。其中`bdp.system.cycctime`参数值是取任务运行时的定时时间（年月日时分秒），若按天调度的周期任务配置定时时间为 20:00，则 2016-01-01 当天自动调度时该参数赋值结果是 20160101200000。（T） **注意是到时分秒的**
53. 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。（T）

部署、运维、发布

1. 1、Data IDE 可以实现开发环境和生产环境隔离，使用两个项目空间分别为 prj1 和 prj2，把项目空间 prj1 基本属性的发布目标设置为项目空间 prj2，以下说法错误的是（C）。（C）
- a. 项目空间 prj1 和 prj2 必须属于同一个组织
 - b. 项目空间 prj1 的中表中的数据不能直接通过发布流程发布到项目空间 prj2
 - c. 项目空间 prj1 的中的表结构变更可以直接通过发布流程发布到项目空间 prj2
 - d. 项目空间 prj1 的调度任务、资源文件以及自定义函数都可以通过发布流程发布到项目空间 prj2
2. Data IDE 中，通过运维中心的任务管理列表可以对任务进行一些批量修改，从而提高项目管理员管理调度任务的效率，其中批量操作功能包括？（3 个）（ABD）
- A. 添加报警
 - B. 修改资源组
 - C. 修改调度周期
 - D. 修改负责人

批注 [s16]: 考到!!!!

概念与参数配置

1. 32、Data IDE 中，调度系统提供了一些常用的时间参数，周期性调度任务使用这些参数后，在调度运行时参数会自动赋值为具体的时间，以下哪些参数是调度系统内置时间参数？（正确答案的数量：2 个）（BD）
- a. `${yyyymmdd}`
 - b. `bdp.system.bizdate`
 - c. `${yyyy-mm-dd}`

- d. `${bdp.system.cyctime}`
2. 2、Data IDE 中，SQL 任务 task1 是按月调度的周期性调度任务，生成分区表 table1，每次运行生成一个分区，分区取值为上个月最后一天日期的年月日（分区名 ds，格式 yyyymmdd），要使用调度系统提供的时间参数定义 table1 的分区，使得 task1 每次调度运行时自动替换时间值，task1 应该如何配置？（A）
- a. task1 的调度周期设置为每月 1 日，代码里 table1 分区 `ds=${bdp.system.bizdate}`
- b. task1 的调度周期设置为每月 31 日，代码里 table1 分区 `ds='${var}31'`，参数配置：var=`$(yyyymm)`
- c. 任务调度周期设置每月 31 日。代码分区表达式 `ds='${var}'`，参数配置：var=`$(yyyymm)31`
- d. task1 的调度周期设置为每月 31 日，代码里 table1 分区 `ds=${bdp.system.cyctime}`
3. 14、通过 Data IDE 的数据管理模块，将项目空间 prj1 的表 table1 设置成“隐藏”后，除了表 owner 或项目空间 prj1 的项目管理员之外的成员都不能通过数据管理模块查看 table1 的表结构，以下说法正确的是（A）。
- a. table1 的 owner 可以撤销 table1 的隐藏状态
- b. 只有 prj1 的项目空间管理员可以撤销 table1 的隐藏状态
- c. 所有项目空间的成员都无法申请表 table1 的查询权限
- d. 所有项目空间的成员都无法通过执行 `desc prj1.table1` 查看表信息

设置为31号二月份就gg了，要用表达式-1天

数据管理

1. Data IDE 的数据管理模块主要是对项目空间中的表的元数据、权限申请等进行管理操作，以下说法错误的是（D）
- a. 可以通过数据管理模块查看项目空间中的表的生命周期
- b. 可以通过数据管理模块查看项目空间中的表的结构
- c. 可以通过数据管理模块申请项目空间中的表的查询权限
- d. 可以通过数据管理模块查看项目空间中的表的数据
2. 在数据仓库中，随着表数量越来越多，对于表的管理会越来越复杂，希望可以按一定的规则划分类目，将不同的表归到不同类目中去，方便管理，即在业务层面上定义类目，对表进行类目划分，以下关于表类目说法正确的是（D）
- a. Data IDE 的组织管理模块实现了类目管理
- b. 可以通过 Data IDE 的项目管理模块实现
- c. 大数据计算服务（MaxCompute，原 ODPS）中提供了类目管理功能
- d. Data IDE 的数据管理模块实现了类目管理
3. Data IDE 中，通过数据管理可以查看表的血缘关系以及表元数据信息，以下（ ）信息变更后通过数据管理可以实时刷新查看（正确答案 3 个）（ABC）
- A. 表新增的字段结构信息
- B. 表修改的描述信息
- C. 表修改了某个字段信息

D. 表中新增的数据内容

批注 [s17]: 考到!!!

4. Data IDE 中, 通过数据管理可以查看表的元数据, 但无法查看某个表使用哪些表 (F)

申明: 本文档是根据当前公司获取的试题资源进行的知识点分类汇总。试题中的答案非官方标准答案, 在练习时请参照阿里官方资料进行配合复习

概念与参数配置

18. 32、Data IDE 中, 调度系统提供了一些常用的时间参数, 周期性调度任务使用这些参数后, 在调度运行时参数会自动赋值为具体的时间, 以下哪些参数是调度系统内置时间参数? (正确答案的数量: 2 个) (BD)

- a. \${yyyymmdd}
- b. \${bdp.system.bizdate}
- c. \${yyyy-mm-dd}
- d. \${bdp.system.cycltime}

19. 98、大数据计算服务 (MaxCompute, 原 ODPS) 有别于我们常见的关系型数据库, 以下关系型数据库中支持的功能 (feature), 在大数据计算服务中也支持的是哪一项? (C)

- a. 索引
- b. 事务
- c. 视图
- d. 主键

20. 100、大数据计算服务 (MaxCompute, 原 ODPS) 的分区 (Partition) 是指一张表下, 根据分区字段 (一个或多个字段的组合) 对数据存储进行划分。也就是说, 如果表没有分区, 数据是直接放在表所在的目录下; 而如果表有分区, 每个分区对应表下的一个目录, 数据是分别存储在不同的分区目录下。 (T)

21. 大数据计算服务存储到文件系统上时, 使用了_____的文件格式, 降低了无效的磁盘读取操作。 (A)

- A. 列存储
- B. 文档存储
- C. 行存储
- D. 键值对存储

22. 大数据计算服务的整体架构包括了四个主要的部分, 其中第一部分为客户端, 即客户发起请求的方式。客户端包括_____。 (AB)

- A. Java SDK
- B. ODPS CLT(Command line tools)
- C. Nodejs SDK
- D. RESTful API

23. 正则表达式可以用非常灵活的方式来处理字符串，在大数据计算服务 SQL 中的正则函数支持 POSIX 格式，以下对正则表达式描述正确的有 _____（4 个）（ABCD） **本题见文档正则章节**
- A. `[[:blank:]]`匹配空格和 TAB 制表符
 - B. `$`匹配行尾
 - C. `[[:alnum:]]`匹配字母字符和数字字符
 - D. `[[:digit:]]`匹配数字字符
 - E. `$^`匹配空行

匹配行位和行首

24. 大数据常见的计算方式有三种：离线计算、流计算以及在线计算，关于这几种计算方式的正确的说法有？（3 个）（BCD）
- A. 流计算是消息驱动的，数据更新一般是定时更新
 - B. 离线计算的数据更新以定期的批量更新为主
 - C. 在线计算一般在用户发起请求时（比如查询）发生
 - D. 流计算可以应用于实时数值统计、实时事件预警等领域
25. 表是大数据计算服务的数据存储单元，它在逻辑上是由行和列组成的二维结构。所有的数据都被存储在表中，各种不同类型的计算任务的操作对象都是表，同时，用户可以创建表、删除表、上传数据到表中，同时也可以更新表中符合条件的数据。（T）

这他妈是实时计算

组织授权

1. 5、大数据计算服务（MaxCompute，原 ODPS）项目 prj1 中，用户 owner 创建表 T1，通过 ACL 方式将读权限给用户 U1（此前用户 U1 无项目空间 prj1 的任何权限），现在使用 ACL 授权 U1 可以通过 `select* from prj1.T1` 来访问 T1 中的数据。接下来 owner 将表 T1 删除，重新创建一张新表，名字仍然叫 T1，以下说法中正确的是 _____。（C）
- a. 如果新建后的 T1 和原来的表 T1 表结构一致，则 U1 仍可以访问 T1
 - b. 用户 U1 仍然可以访问表 T1
 - c. 用户 U1 无法访问当前表 T1
 - d. 如果新建后的 T1 和原来的表 T1 表结构、数据内容均一致，则 U1 仍可以访问 T1
2. 13、项目空间（project）是大数据计算服务（MaxCompute，原 ODPS）的基本组织单元，它类似于传统数据库的 Database。以下关于项目空间的正确的说法有 _____。（正确答案的数量：3 个）（ABD）
- a. 所有的对象都会属于某个项目空间
 - b. 一个帐号可以拥有多个项目空间的权限
 - c. 一个帐号可以创建多个项目空间，最多不能超过 10 个
 - d. 通过安全授权，可以在一个项目空间中访问另一个项目空间中的数据
3. 88、大数据计算服务（MaxCompute，原 ODPS）的开发人员原来已经获得了视图 V 的读权限，但是这个视图被管理员使用 `create or replace view` 语句修改了视图的定义，则这位开发人员必须重新申请权限，否则无法读取这个视图。（F）

批注 [K18]: •创建视图时，必须有对视图所引用表的读权限。

- 视图只能包含一个有效的 `select` 语句。
- 视图可以引用其它视图，但不能引用自己，也不能循环引用。
- 不允许向视图写入数据，例如使用 `insert into` 或者 `insert overwrite` 操作视图。
- 当视图建好以后，如果视图的引用表发生了变更，有可能导致视图无法访问，例如：删除被引用表。用户需要自己维护引用表及视图之间的对应关系。
- 如果没有指定 `if not exists`，在视图已经存在时用 `create view` 会导致异常。这种情况可以用 `create or replace view` 来重建视图，重建后视图本身的权限保持不变。

4. 93、大数据计算服务（MaxCompute，原 ODPS）中的角色（Role）是一组访问权限的集合。当需要读一组用户赋予相同的权限时，可以使用角色来授权。基于角色的授权可以大大简化授权流程，降低授权管理成本，当需要对用户授权时，应当优先考虑是否应该使用角色来完成。关于角色的说法中正确的是_____。
- a. 可以将 Owner 视为一个角色，赋给其他用户
 - b. Admin 是项目空间自动创建的角色，可以为该角色添加权限
 - c. 删除一个角色时，会先检查该角色内是否还存在其他用户。若存在，则删除该角色失败
 - d. 创建完角色后，需要给角色添加权限，仅支持以 ACL 授权的方式添加
5. 96、使用大数据计算服务（MaxCompute，原 ODPS）中的 ACL 授权方式进行授权时，可以使用通配符，但是，目前仅支持星号（*）（F）
6. 94、大数据计算服务（MaxCompute，原 ODPS）中的 ACL 授权时类似 SQL92 定义的 GRANT/REVOKE 语法来实现，更接近于传统关系型数据库的授权模式，Policy 授权主要解决 ACL 授权机制无法解决的一些复杂授权场景。以下场景中，哪些既能使用 ACL 授权又能使用 Policy 授权实现？（BC）
- a. 一次操作对一组对象进行授权
 - b. 只允许用户使用 SQL（而不允许其他类型的 Task）来访问某张表
 - c. 将一个对象的多个权限一次性授予一个用户
 - d. 带限制条件的授权
7. 95、大数据计算服务（MaxCompute，原 ODPS）中的授权是指把某些对象的某些权限赋给一些指定的用户，必须包括以下哪三个要素？（正确答案的数量：3 个）（ACD）
- a. 操作（Action，比如读、写等）
 - b. 效果（Effect，比如同意、拒绝等）
 - c. 主体（Subject，可以是用户、角色等）
 - d. 客体（Object，可以是表、资源等）
8. 5、大数据计算服务（MaxCompute，原 ODPS）中，通常情况下，prj1 项目空间的成员若要访问 prj2 项目空间的数据，则需要 prj2 项目空间管理员执行繁琐的授权操作：首先需要将 prj1 项目空间中的用户添加到 prj2 项目空间中，再分别对这些新加入的用户进行普通授权。实际上，prj2 项目空间管理员并不期望对 prj1 项目空间中的每个用户都进行授权管理，而期望有另一种更高效的机制能使得 prj1 项目空间管理员能对许可的对象进行自主授权控制。以下哪种授权方式可以帮助 prj2 项目空间管理员实现他的预期目标？（B）
- a. ProjectProtection
 - b. Package
 - c. Role
 - d. LabelSecurity
9. 13、项目空间（project）是大数据计算服务（MaxCompute，原 ODPS）的基本组织单元，它类似于传统数据库的 Database 或 Scheme 的概念。以下关于项目空间的说法中正确的是_____。（正确答案的数量：3 个）（ABC）
- a. 通过安全授权，可以在一个项目空间中访问另一个项目空间中的数据
 - b. 一个帐号可以拥有多个项目空间的权限
 - c. 所有的对象都会属于某个项目空间

批注 [K19]: https://help.aliyun.com/document_detail/27934.html?spm=5176.doc27807.2.5.55WUHF

删除一个角色时，ODPS 会检查该角色内是否还存在其他用户。若存在，则删除该角色失败。只有在该角色的所有用户都被撤销时，删除角色才会成功。

- d. 一个帐号可以创建多个项目空间，最多不能超过 10 个
10. 大数据计算服务中同步数据到分析型数据库中时，特别要注意用户权限问题。账户 ALIYUN\$Bob@aliyun.com 要从大数据计算服务中将表 T_ODPS 导入到分析型数据库的表 T_ADS 中去，在大数据服务中，必须将表 T_ODPS 的 select 和 describe 权限赋给以下哪个账户？ (A)
- a. ALIYUN\$garuda_build@aliyun.com
 - b. ALIYUN\$root@aliyun.com
 - c. ALIYUN\$Bob@aliyun.com
 - d. ALIYUN\$garuda_data@aliyun.com
11. 大数据计算服务支持基于标签的安全，它是项目空间级别的一种强制访问控制策略，它的引入是为了让项目空间管理员能更加灵活地控制用户对敏感数据的访问。LabelSecurity 可以对敏感数据的哪个粒度进行控制？ (C)
- A. 分区
 - B. 表
 - C. 列
 - D. 行
12. 大数据计算服务项目空间 prjl 中存在表 t1，其中包含一列名字为 id，owner 打开 LabelSecurity 设置后，执行如下命令：SET LABEL 1 TO TABLE t1; SET LABEL 2 TO TABLE t1 (id); SET LABEL 3 TO TABLE t1; 此时，t1 表中的 id 列的敏感等级标签是什么？ (C)
- A. 1 级
 - B. 3 级
 - C. 2 级
 - D. 0 级
13. 关于大数据计算服务中的 Policy 授权，说法正确的有？ (3 个) (BCD)
- A. 授权或撤销授权时，对 Subject (如 User) 没有要求，即授权的 User 存在或者不存在都可以
 - B. 删除一个对象时，与该对象关联的 Policy 授权不会被删除
 - C. 支持带限制条件的授权
 - D. 授权或撤销授权时，要求 Object (如 Table) 必须已经存在
14. 当大数据计算服务项目空间的 Owner Alice 决定对另一个用户授权时，Alice 虚线将该用户添加到自己的项目空间中，只有添加到项目空间中的用户才能够被授权。 (F) 可以通过资源授权的方式

工具与 SDK

1. 10、大数据计算服务 (MaxCompute, 原 ODPS) 提供了完整的安全策略和授权策略，来保护我们的数据，在想下载数据到本地的时候，除了要有相应的权限之外，还要选择合适的下载方式。现在大数据计算服务的某表中有 1000 万行数据，需要把这些数据全部导出到本地。下面哪一种是可行的方法？ (C)

- a. 在客户端工具 odpscmd 中用 select 命令，把结果重定向到文件中
b. 管理控制台中读出该表，下载数据
c. 在命令行工具 odpscmd 中使用 tunnel 命令下载
d. 找到该表在大数据计算服务上的保存路径，然后把该路径下的文件全部下载
2. 11、大数据计算服务（MaxCompute，原 ODPS）中关于 tunnel 命令行工具的说法，正确的是？（正确答案的数量：2 个）(AC)
a. tunnel 上传到分区表时，会把数据直接上传到指定的分区中去，不保证业务逻辑
b. tunnel 命令上传文件时，不能进行压缩（注：通过 -cp 参数进行压缩）
c. tunnel 上传的数据是先写到临时目录，最后确定成功后才写到结果目录
d. tunnel 上传过程中不加密，数据以明文方式在网络中传输（注：通过 https 连接就是加密方式）
3. 19、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数据同步。DataX 任务配置文件使用（）文件格式。(D)
a. java
b. bson
c. xml
d. json
4. 89、大数据计算服务（MaxCompute，原 ODPS）tunnel 命令提供了丰富的参数，可以通过参数设置，对上传或者下载操作进行控制，现在有一个大文件要上传至 odps 某表中，想先检查一下该文件内容是否和表的定义匹配，并不实际上传，以免文件中有错误记录会导致上传出错。可以使用哪个参数实现这个功能？(C)
a. 错误记录丢弃：-dbr
b. 字段分隔符：-fd
c. 扫描选项：-s
d. 空字符串显示：-ni
5. 90、大数据计算服务（MaxCompute，原 ODPS）的 tunnel 命令行工具对上传的数据有格式上的要求，它支持以下哪种格式的数据？(D)
a. JSON 文件
b. xml 文件
c. xls 文件
d. csv 文件
6. 91、以下哪些场景可以使用大数据计算服务（MaxCompute，原 ODPS）的 Tunnel SDK 方式实现？（正确答案的数量：3）(ABC)
a. 定时批量上传数据
b. 下载某个分区表的一个分区的数据
c. 一次性上传本地格式化文本文件
d. 随时变化的数据，通过事件（消息）驱动触发
7. 92、使用大数据计算服务（MaxCompute，原 ODPS）的 Tunnel SDK 上传下载数据时，任务会被提交至 Tunnel 服务器（tunnel endpoint），由其进行响应和处理。(F)

批注 [s20]: 考到!!!!

批注 [s21]: 考到!!!!

批注 [K22]: •-bs, -block-size: 每次上传至 Tunnel 的数据块大小，默认值：100MiB（MiB=1024*1024B）。

•-c, -charset: 指定本地数据文件编码，默认为 'UTF-8'；不设定，默认下载源数据。

•-cp, -compress: 指定是否在本地压缩后再上传，减少网络流量，默认开启。

•-dfp: DateTime 类型数据格式，默认为 'yyyy-MM-dd HH:mm:ss'；

•-dbr: 是否忽略脏数据（多列，少列，列数据类型不匹配等情况）。值为 'true' 时，将全部不符合表定义的数据忽略。值为 'false'，若遇到脏数据，则给出错误提示信息，目标表内的原始数据不会被污染。

•-e: 指定 odps 的 endpoint。

•-fd: 本地数据文件的列分割符，默认为逗号 ','；

•-h: 数据文件是否包括表头，如果为 true，则 dship 会跳过表头从第二行开始上传数据。

•-mbr, -max-bad-records: 默认情况下，当上传的脏数据超过 1000 条时，上传动作终止。通过此参数，可以调整可容忍的脏数据量；

•-ni: NULL 数据标志符，默认为 ""（空字符串）；

•-rd: 本地数据文件的行分割符，默认为 '\r\n'；

•-s: 是否扫描本地数据文件，默认值为 'false'。值为 'true' 时，先扫描数据，若数据格式正确，再导入数据。值为 'false'，不扫描数据，直接进行数据导入。值为 'only' 时，仅进行扫描本地数据，扫描结束后不继续导入数据。

•-te: 指定 tunnel 的 Endpoint。

•-tz: 指定时区。默认为本地时区：Asia/Shanghai。

8. 10、大数据计算服务（MaxCompute，原 ODPS）提供了完整的安全策略和授权策略，来保护我们的数据，在想下载数据到本地的时候，除了要有相应的权限之外，还要选择合适的下载方式。现在大数据计算服务的某表中有 1000 万行数据，需要把这些数据全部导出到本地。下面哪一种是可行的方法？（C）

a. 管理控制台中选出该表，下载数据
b. 找到该表在大数据计算服务上的保存路径，然后把该路径下的文件全部下载
c. 在命令行工具 odpscmd 中使用 tunnel 命令下载
d. 在客户端工具 odpscmd 中用 select 命令，把结果重定向到文件中

9. 12、大数据计算服务（MaxCompute，原 ODPS）提供了客户端 odpscmd，支持 Windows、Linux 或者 OS X 等操作系统。在执行前需要对 odpscmd 的配置文件 odps_config.ini 中的某些参数进行设定，包括用于鉴权的 access_id/access_key，指定服务地址的 end_point，以及 tunnel 的服务地址 tunnel_endpoint 等。其中有一项参数为：project_name，关于这个参数的说法，正确的是？（B）

a. 指定的项目必须存在，否则连接过程中会报错
b. 如果给该值指定了恰当的 project 名字，连接成功后，会自动进入该 project
c. 如果不指定该值，连接成功后会自动进入上次退出时所在的 project
d. 如果不指定该值，且帐户下只有一个 project 的话，连接成功后会自动进入该 project

批注 [s23]: 考到!!!

10. 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数据同步，DataX 任务配置文件使用（D）

a. xml
b. java
c. bson
d. json

批注 [s24]: 考到!!!

11. 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数据同步。使用 DataX 将存储在对象存储 OSS 中的数据同步至阿里云端大数据计算服务（MaxCompute，原 ODPS）时，以（B）为粒度进行切分实现并发处理

a. 没有任何切分
b. OSS 的 object
c. OSS 的 endpoint
d. OSS 的 Bucket

解释：<https://github.com/alibaba/DataX/blob/master/ossreader/doc/ossreader.md>

大数据计算服务提供了 Tunnel SDK，支持高速上传下载数据。实际使用的时候有两种方式，第一种是通过 odpscmd 客户端，调用封装好的 tunnel 命令，第二种是直接使用 Tunnel SDK 进行编程，通过调用程序进行个性化的上传下载。目前 SDK 支持以下哪种语言？（A）

A. java
B. C++
C. Perl
D. C#

12. 使用大数据计算服务 tunnel 命令上传数据文件 data.csv 到表 t_data 成功后，不小心又之行了一次该上传命令，以下说法中正确的是？（A）

- A. 最后一次上传的数据会追加插入到表 t_data 中去
- B. 最后一次上传的数据会覆盖插入到表 t_data 中去
- C. 记录重复会报错退出
- D. 最后一次上传的数据会把 t_data 中已存在的相关记录更新

批注 [s25]: 考到!!!!

13. 使用大数据计算服务 tunnel 命令上传 log.txt 文件到表 t_log 中去, t_log 为分区表, 分区列为 (p1 string, p2 string)。下列命令中正确的是? (C)
- A. tunnel upload log.txt t_log/p1="b1"/p2="b2"
 - B. tunnel upload log.txt t_log/(p1="b1",p2="b2")
 - C. tunnel upload log.txt t_log/p1="b1",p2="b2"
 - D. tunnel upload log.txt t_log(p1="b1",p2="b2")
14. tunnel upload 命令可以上传数据到大数据计算服务中去, 请判断以下哪些数据对象可以作为上传的目标? (2个) (BD)
- A. 资源
 - B. 非分区表
 - C. 视图
 - D. 分区表
15. 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数据同步。在使用 DataX 进行数据同步时, 源端数据的读取支持断点续传。 (F)
16. 大数据计算服务中, 在命令行工具 odpscmd 中使用 select 语句查看表中的记录时, 可以结合 limit 子句来限制输出结果的行数。 (T)
17. 使用大数据计算服务的 Tunnel SDK 上传下载数据时, 任务会被提交至 Tunnel 服务器(tunnel endpoint), 由其进行响应和处理。 (T) 此题请查询相关章节确定

UDF

1. 63、小王正在使用大数据计算服务 (MaxCompute, 原 ODPS) 进行自然语言分析, 他从门户网站上搜集了 1000 篇新闻报道, 然后每隔存成一条记录, 然后基于这些记录去做分词, 即将每条记录中的内容拆成很多独立的词语, 保存到另一张表中, 每个词语一条记录。小王对自定义函数比较熟悉, 他想通过编写自定义函数实现分词功能。请帮小王做一下选择, 从功能角度来看, 哪种自定义函数更适合处理这种场景? (D)
- a. User Defined Aggregation Function (用户自定义聚组函数)
 - b. User Defined Scalar Function (用户自定义标量函数)
 - c. User Defined Splitted Function (用户自定义分组函数)
 - d. User Defined Table Valued Function (用户自定义表值函数)
2. 83、在使用 java 创建大数据计算服务 (MaxCompute, 原 ODPS) 的 UDF 时, 可以在函数定义中引用资源, 并且在函数中读取这些资源。 (T)

3. 30、资源(Resource)是大数据计算服务(MaxCompute,原 ODPS)的特有概念。用户如果想使用MaxCompute的MapReduce 或自定义函数(UDF)功能,需要依赖资源来完成。Data IDE 中,可以通过数据开发面板的资源管理上传资源,目前支持的资源类型有 (A)。
- a. jar、file、archive
 - b. jar、file、python
 - c. jar、archive、python
 - d. file、archive、python
4. 某企业需要一个能够从身份证号中分析得出所属省份的自定义函数,哪种自定义函数适用于这种场景? (B)
- A. User Defined Splitted Function (用户自定义分组函数)
 - B. User Defined Scalar Function (用户自定义标量函数)
 - C. User Defined Table Valued Function (用户自定义表值函数)
 - D. User Defined Aggregation Function (用户自定义聚组函数)
5. 当大数据计算服务的内置函数无法满足业务需求时,可以使用 Java 开发自定义函数,对于普通的自定义标量函数的描述正确的有_____。(3个) (BCD)
- A. 可以实现多个参数类型不同的 evaluate 方法,系统调用函数时会根据参数的类型自动匹配到正确的方法
 - B. 在函数类中,函数的逻辑通过 evaluate 方法实现
 - C. 每个用户只能使用自己创建的 UDF
 - D. 在 UDF 中可以使用 MaxCompute 的资源(Resource),比如通过 add file 上传的文件资源
6. 大数据计算服务的函数分为标量函数、聚合函数及表值函数几种类型,各种函数都可以嵌套使用。(T)

批注 [s26]: 考到

图计算

1. 44、在大数据计算服务(MaxCompute,原 ODPS)的表中存有一批网站及其相互之间链接关系,开发人员希望通过 pagerank 算法来分析这批网站的排名,即把网站看做节点,把网站之间的连接关系看做权重,从而实现网站的评分。_____计算框架更适合实现这个场景。(C)
- a. SQL
 - b. MapReduce
 - c. Graph
 - d. Tunnel
2. 45、大数据计算服务(MaxCompute,原 ODPS)的图计算引擎是分布式的,在加载数据时,系统默认如何将数据分配到不同的 Worker 中? _____
- a. 根据边的权值进行 hash 并对 Worker 数取模
 - b. 完全随即分配
 - c. 根据顶点 ID 的 hash 值对 Worker 数取模

批注 [K27]: 图加载: 框架调用用户自定义的 GraphLoader 将输入表的记录解析为点或边;
分布式化: 框架调用用户自定义的 Partitioner 对点进行分片(默认分片逻辑: 点 ID 哈希值然后对 Worker 数取模), 分配到相应的 Worker;

d. 按照点 ID 值排序，再平均分配给 Worker

批注 [s28]: 考到!!!

3. 46、大数据计算服务（MaxCompute，原 ODPS）图计算框架再计算过程中通过对图进行编辑、演化最终求解出结果，在计算过程中用户可以对图进行哪些编辑操作？（正确答案的数量：4）（BCDE）
 - a. 读取本地数据修改点的权值
 - b. 增加边
 - c. 删除点
 - d. 修改边的权值
 - e. 增加点
4. 47、MaxCompute Graph 是 MaxCompute 提供的编程框架之一，主要是用来进行图像处理。（F）
5. 图是一种数据结构，由点和边组成，点和边包括权值，用点表示要研究的对象，边可以用来反映之间的各种关系。（T）
6. 开发人员在使用大数据计算服务的图计算时，主要的输入数据在表中，另外还有一些非结构化的配置信息存在一个本地文件中，则可以采用哪种方式使得这些配置信息在图计算过程中可用？（A）
 - A. 将这个配置信息做为 resource 上传到大数据计算服务中，在图计算过程中可以读取 resource 中的数据
 - B. 输入输出只能是 Table，必须要把这部分数据放到表中
 - C. 在 group 程序中直接读取本地的文件信息
 - D. 输入只能是一张表，因此无法满足这个场景
7. 大数据计算服务中，任何有创建函数权限开发者都可以创建与系统内置函数同名的函数，系统在调用时会优先调用用户自定义的函数。（F）

MapReduce

1. 4、在大数据计算服务（MaxCompute，原 ODPS）中，使用 MapReduce 开发一个分词程序，即输入为保存在表中的多篇文章（每篇文章一行纪录），参考一个字典文件（该文件会经常更新）进行分词。简而言之，在执行该程序的时候，既需要输入待分词的文章（已经保存在表中了），又需要输入字典（为经常变动的本地文件），以下最合理的解决方案为_____。（B）
 - a. 将这个字典与源代码一起进行打包，直接执行即可
 - b. 每次都字典文件作为资源上传到大数据计算服务上，然后 MR 中通过资源的方式来访问
 - c. 将字典打包 jar 包，使用 jar 命令执行 MR 程序时，通过 resources 参数指定该 jar 包
 - d. 将字典文件保存在本地，在 MR 中通过访问本地最新的字典文件实现
2. 48、大数据计算服务（MaxCompute，原 ODPS）的 MapReduce 提供了分布式的编程框架，以下哪类大数据计算服务可以作为 MR 的输入和输出？（A）
 - a. 表
 - b. 视图

- c. 资源
d. 项目空间
3. 49、大数据计算服务（MaxCompute，原 ODPS）提供了 MapReduce 编程接口，用户可以使用 MapReduce 提供的 SDK 编写程序处理大数据计算服务中的数据。目前提供了_____语言的 SDK。 **(D)**
- a. Perl
b. C++
c. Python
d. Java
4. 50、使用大数据计算服务（MaxCompute，原 ODPS）的 odpscmd 客户端中的 jar 命令调用 MapReduce 时，需要指定两个参数：classpath 和 resources。比如以下命令执行了一个 MapReduce 程序，输入为表 t_in，输出为表 t_out：jar-resources test.jar-classpath test.jar test t_in t_out；其中，test.jar 出现了两次，请判断以下说法正确的是_____。 **(D)**
- a. 两次出现的意义相同
b. 两个引用的 test.jar 都是保存在大数据计算服务上的资源
c. 两个引用的 test.jar 都是保存在本地
d. classpath 引用的是 main 函数，运行于本地，指定的 jar 包路径，即本地文件路径
5. 51、大数据计算服务（MaxCompute，原 ODPS）提出了安全沙箱的概念，为了保证整个计算环境的安全，某些特殊操作会被禁止。以下哪些操作会受到沙箱的影响？（正确答案的数量：2 个） **(BD)**
- a. 使用内置函数
b. 使用 MR
c. 使用 DDL 创建表
d. 使用 UDF
6. 52、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） **(ACD)**
- a. 一个 reduce 的输入数据可能来自于多个 map 的输出
b. 一个 MR 处理可以不包括任何 map
c. 一个 MR 处理可以不包括任何 reduce
d. 一个 map 的输出结果可能会被分配到多个 reduce 上去
7. 53、大数据计算服务（MaxCompute，原 ODPS）执行 MapReduce 时，每个 Map Worker 在读入各自的数据后，需要为每一条数据指定一个 Key，这个 Key 的值决定了记录会被发送到哪一个 Reduce Worker。单个 Reduce Worker 有可能接收到多个 Key 值的记录，一个 Key 值对应的记录也有可能发送到多个 Reduce Work。 **(F)**
8. 54、大数据计算服务（MaxCompute，原 ODPS）的 MapReduce 由多个步骤组成，以下哪个步骤不是必须的？ **(A)**
- a. combine
b. shuffle

批注 [s29]: 考到

批注 [K30]: 分片完毕后，多个 Map Worker 就可以同时工作了。每个 Map Worker 在读入各自的数据后，进行计算处理，最终输出给 Reduce。Map Worker 在输出数据时，需要为每一条输出数据指定一个 Key。这个 Key 值决定了这条数据将会被发送给哪一个 Reduce Worker。Key 值和 Reduce Worker 是多对一的关系，具有相同 Key 的数据会被发送给同一个 Reduce Worker。单个 Reduce Worker 有可能会接收到多个 Key 值的数据。

- c. map
- d. reduce

9. 55、在开发大数据计算服务（MaxCompute，原 ODPS ）的 MapReduce 时，可以使用 JobClient 类中的接口提交 MapReduce 作业。目前提交方式支持_____。（正确答案的数量：2 个）(AC)

- a. 非阻塞方式
- b. 远程提交
- c. 阻塞方式
- d. 本地提交

批注 [s31]: 考到!!

10. 56、大数据计算服务（MaxCompute，原 ODPS ）的 MapReduce 的输入可以为空，也可以为视图。(F)

11. 传统的 MapReduce 模型要求每一轮 MapReduce 操作之后，数据必须落地到分布式文件系统上。而一般的 MapReduce 应用通常由多个 MapReduce 作业组成，每个作业结束之后需要写入磁盘，接下去的 Map 任务很多情况下只是读一遍数据，为后续的 Shuffle 阶段做准备，这样其实造成了冗余的 IO 操作。为了解决这一问题，提供更优的性能，大数据计算服务提供了扩展的 MapReduce 模型，该模型区别于普通 MapReduce 模型的主要特点是？ (A)

- A. 支持 Map 后连接任意多个 Reduce 操作，如 Map-Reduce-Reduce
- B. 支持 Map 后不连接 Reduce，而是连接另一个 map，如 Map-Map-Reduce
- C. 支持 Chain Mapper / Reducer，即支持 Map-Reduce-Map-Reduce
- D. 支持没有 Map，直接进入 Reduce

12. 大数据计算服务的 MapReduce 中，归并函数是 MapReduce 计算框架的一种优化，即框架会在 map 端对相同 key 值的数据进行本地的合并操作。关于 Combiner（归并），以下说法正确的是？ (A)

- E. Combiner 的合理使用，可以减少传输带宽，提升整体性能
- A. Combiner 是必须的
- B. Combiner 的合理使用，可以减少传输带宽，提升整体性能
- C. Combiner 需要和 partitioner 配合使用

13. 从优化角度来看，大数据计算服务的 MapReduce 对于哪个节点处理哪个分片这个问题，采用了_____策略。(A)

- A. 本地性原理，即数据在哪台机器上，就在哪台机器上启动 map
- B. 能者多劳原理，将数据移动到计算能力最强的节点上，启动 map
- C. 基于历史处理信息的优化原理，即将数据移到从历史执行信息看处理效果最好的机器上，启动 map
- D. 负载均衡原理，即将数据放在当前负载最小的节点上启动 map

14. Bob 使用 java 开发了一个大数据计算服务的 MapReduce，用来对海量交易数据做时序分析，结果该 MR 性能较差，某些 map 存在明显的长尾效应，_____方法可以有效的消除长尾效应。(B)

- A. 只关心处理的列，忽略其余的多余列
- E. 合理使用 combiner，适当的减少数据交互量
- B. 通过 set 命令或者 Jobconf 修改 MR 参数，增大 map / reduce 的内存
- C. 通过 set 命令或者 Jobconf 修改 MR 参数，影响分片的大小和个数

批注 [s32]: 考到!!!

15. 使用大数据计算服务中的MR时,MR中的哪些操作会引发Java安全沙箱的限制而报错?(3个)(ABD)
- A. 起子进程执行Linux命令
 - B. 获取本地IP
 - C. 本地文件以资源方式上传至大数据计算服务,在MR中以资源方式访问该文件
 - D. Java反射限制
16. 在开发大数据计算服务的MapReduce时,可以使用JobClient类中的接口提交MapReduce作业。目前提交方式支持_____(2个)(CD)
- A. 远程提交
 - B. 本地提交
 - C. 阻塞方式
 - D. 非阻塞方式
17. MapReduce最早是由Google提出的分布式数据处理模型,随后受到了业内的广泛关注,并被大量应用到各种商业场景中。以下场景中,_____适合用MapReduce来实现。(3个)(BCD)
- A. 交互式查询:基于海量数据的透视分析,用户可以通过上卷、下钻、切片等交互操作,了解数据集细节
 - B. 机器学习:监督学习、无监督学习、分类算法如决策树、SVM等
 - C. 文本统计分析:比如词频TFIDF分析;学术论文、专利文献的引用分析和统计;维基百科数据分析等
 - D. Web访问日志分析:分析和挖掘用户在web上的访问、购物行为特征,分析用户访问行为
18. 大数据计算服务在正式执行Map前,需要将输入数据进行“分片”。所谓分片,就是将输入数据切分为大小差不多相等的数据块,每一块作为单个Map Worker的输入被处理,以便于多个Map Worker同时工作。(T)
19. 大数据计算服务的MapReduce的输入可以为空,也可以为视图。(F)

机器学习数据挖掘

1. 57、某销售企业基于自己顾客的若干个指标(价值、成本、使用的产品等)对已有的用户群进行划分:特征相似的用户做为一类,特征不同的用户分属于不同的分类。然后会针对不同类型的客户进行不同的营销策略。这个案例属于大数据分析的哪种类型的问题。(A)
- a. 分类
 - b. 关联分析
 - c. 聚类
 - d. 时序分析
2. 58、某超市通过分析顾客的购买记录后发现,购买纸尿裤的人由较大的概率也会购买啤酒,这个案例数据大数据分析的哪种类型的问题?(D)
- a. 聚类
 - b. 时序分析

- c. 分类
 - d. 关联分析
3. 4、在大数据计算服务（MaxCompute，原 ODPS）的 MapReduce 中可以使用资源（Resource，包括 jar 包、压缩文件、表等）。在使用 jar 命令执行 MR 程序时，需要在参数 resources 中指定资源，以下说法正确的是_____。（A）
- a. 包含已上传到大数据计算服务的 MR 程序本身所在的 jar 包，以及 MR 程序用到的大数据计算服务上的资源
 - b. 仅包含 MR 程序中用到的大数据计算服务上的资源
 - c. 仅包含保存在本地的 MR 程序本身所在的 jar 包
 - d. 仅包含已上传到大数据计算服务的 MR 程序本身所在的 jar 包

应用场景

1. 59、数据仓库是一个面向主题的（Subject Oriented）、集成的（Integrate）、相对稳定的（Non-Volatile）、反映历史变化（Time Variant）的数据集合，主要用于支持决策、面向分析型数据处理。大数据计算服务（MaxCompute，原 ODPS）出了能很好的处理大数据语境下的分析功能外，也是传统数据仓库上云的解决方案之一。下列哪些传统数据仓库的功能可以通过大数据计算服务实现？（正确答案的数量：3 个）（ABC）
- a. 定时的 ETL 任务（数据抽取、转换和加载）
 - b. 固定报表分析
 - c. 数据挖掘
 - d. 即席查询
2. 97、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？（C）
- a. 大数据的分析建模
 - b. 海量数据仓库
 - c. OLTP
 - d. 网站日志离线分析
3. 8、某企业准备上线一个项目，去搜集国内主要门户、论坛中和本公司相关的消息，一旦有符合预定义类型的消息出现，将第一时间反馈给公司的公共关系部，该场景属于典型的？（A）
- a. 流计算
 - b. 在线计算
 - c. OLTP
 - d. 离线计算
4. 9、大数据常见的计算方式有三种：离线计算（比如大数据计算服务提供的 MR、SQL 等）、流计算（阿里云的流计算 Stream Compute）以及在线计算（分析型数据库 AnalyticDB），关于这几种计算方式的说法正确的有？（正确答案的数量：3 个）（BCD）
- a. 流计算是消息驱动的，数据更新一般是定时更新
 - b. 流计算可以应用于实时数值统计、实时事件预警等领域

- c. 在线计算一般在用户发起请求时（比如查询）发生
d. 离线计算的数据更新以定期的批量更新为主
5. 大数据计算服务提供了大数据的存储和计算服务，非常适合应用于大数据分析的领域。以下说法中错误的是哪一项？（C）
A. 可以实现大型互联网企业的数据仓库和 BI 分析
B. 提供了便捷的分析处理海量数据的手段，用户可以不关心分布式计算细节，从而达到分析大数据的目的
C. 可以支持实时 OLAP 分析
D. 可以基于历史数据，进行用户特征和兴趣挖掘
6. 随着社会数据收集手段的不断丰富及完善，越来越多的行业数据被积累下来，数据规模已经增长到了传统软件行业无法承载的海量数据级别。大数据计算服务可以很多的利用存储和计算能力去处理、分析这些海量数据，发现价值。以下说法中正确的有_____（3 个）（BCD）
A. 大数据计算服务提供了高并发低延迟（毫秒级）的处理能力，对海量数据进行灵活分析
B. 大数据计算服务提供了高并发吞吐量上传下载工具，可以将外部数据源中的数据高效的同步到大数据计算服务中去
C. 大数据计算服务支持使用 SQL 或者 MapReduce 编程对海量数据进行处理
D. 大数据计算服务可以对海量数据进行基于复杂算法的分析，如聚类、分类、回归、主成分分析等。
7. 数据仓库是一个面向主题的、集成的、相对稳定的、反映历史变化的数据集合，主要用于支持决策、面向分析型数据处理。大数据计算服务除了能很好的处理大数据语境下的分析功能外，也是传统数据仓库上云的解决方案之一。（F）
8. 下列哪些传统数据仓库的功能可以通过大数据计算服务实现？（3 个）（BCD）
A. 即席查询
B. 定时的 ETL 任务（数据抽取、转换和加载）
C. 固定报表分析
D. 数据挖掘

DDL/ DML/内置函数

1. 61、大数据计算服务（MaxCompute，原 ODPS ）中的日志表 log 是一张分区表，分区表是 dt，每天产生一个分区用于存储当天新增的数据，数据是通过 tunnel 接口上传，运维人员需要在每天上传数据之前生成好当天的分区，假设当前日期是 20160301，语句_____可以创建合适的分区。（D）
a. add partition log (dt=' 20160301')
b. insert overwrite table log partition(dt=' 20160301')
c. insert into table log partition(dt=' 20160301')
d. alter table log add partition(dt=' 20160301')
2. 62、大数据计算服务（MaxCompute，原 ODPS ）中的订单表 fact_order 是一张分区表，分区键是 dt，每天新增的订单存储在当天的分区中，如果要更新 dt=' 20160310' 这个分区中的数据，对应的源表是 ods_order，则正确的 SQL 语句结构是_____。（C）

批注 [K33]: ALTER TABLE TABLE_NAME ADD [IF NOT EXISTS] PARTITION partition_spec

- a. `update fact_order select_from ods_order where dt=' 20160310'`
b. `insert overwrite table fact_order ds=' 20160310' select_from ods_order`
c. `insert overwrite table fact_order partition(dt=' 20160310') select_from ods_order`
d. `insert overwrite table fact_order select_from ods_order where dt=' 20160310'`
3. 65、大数据计算服务（MaxCompute，原 ODPS ）的运维人员想要知道一张日志表 log 已经占用了多少存储空间，可以在命令行工具 odpscmd 中使用_____命令查看。(B)
- a. `ls log;`
b. `desc log;`
c. `select* from log;`
d. `size log;`
4. 66、运营人员为了搞促销，需要找出目标客户名单，这批客户的信息存储在大数据计算服务（MaxCompute，原 ODPS ）的 user 表中，并且用户名字段 username 中包含“vip”字样，通过执行 SQL 语句_____可以找出这批客户。(D)
- a. `select* from user where username contains('vip')`
b. `select* from user where username=' vip'`
c. `select* from user where username like '*vip*'`
d. `select* from user where username like ' %vip%'`
5. 67、大数据计算服务（MaxCompute，原 ODPS ）中的日志表 log 是一张分区表，分区键是 dt，每天产生一个分区用于存储当天新增的数据，现在共有 dt=' 20160101' 至 dt=' 20160131' 共 31 个分区的数据，为了删除 20160101 当天新增的数据，可以通过_____方式实现。(D)
- a. `delete from log where dt=' 20160101'`
b. `truncate table where dt=' 20160101'`
c. `drop partition log (dt=' 20160101')`
d. `alter table log drop partition (dt=' 20160101')`
6. 68、大数据计算服务（MaxCompute，原 ODPS ）中的用户表 dim_user 是一张非分区表，另外有一部分新增用户的数据存在 user_delta 表中，两张表的结构相同，为了在 dim_user 表中得到全量的用户数据，可以采用_____方式。(D)
- a. `select* from user_delta into dim_user;`
b. `insert into dim_user select* from user_delta union all select*from dim_user`
c. `insert overwrite table dim_user select*from user_delta`
d. `insert into table dim_user select* from user_delta`
7. 69、大数据计算服务（MaxCompute，原 ODPS ）中的用户基本信息表 dim_user 和扩展信息表 dim_user_ext 都有一个整型字段 user_id(业务上该字段的每个取值都唯一标识一条用户记录)。表 dim_user 中共有 10000 条记录，其中 user_id 从 2000 到 11999，dim_user_ext 表中共有 5000 条记录，其中 user_id 从 0 到 4999。开发人员在建立用户模型时需要综合用到两张表中所有用户的数据，如果有扩展信息则加上，如果没有则只用基本信息，_____可以实现该需求。(C)
- a. 使用全关联: `select t1.*,t2.* from dim_user t1 full outer join dim_user_ext t2 on t1.user_id=t2.user_id;`
b. 使用右关联: `select t1.*,t2.* from dim_user t1 right join dim_user_ext t2 on`

批注 [s34]: 重点!!!

t1.user_id=t2.user_id;

c. 使用左关联: select t1.*,t2.* from dim_user t1 left join dim_user_ext t2 on t1.user_id=t2.user_id;

d. 使用内关联: select t1.*,t2.* from dim_user t1 join dim_user_ext t2 on t1.user_id=t2.user_id;

8. 70、对于大数据计算服务（MaxCompute，原 ODPS）内置绝对值函数 abs 描述正确的有_____。

（正确答案的数量：2 个）(AD)

a. 当输入参数是 bigint 时，返回值是 bigint 类型

b. 当输入参数是 string 类型时，一定会导致异常

c. 输入类型是 boolean 则返回值是 True

d. 当输入参数是 double 时，返回值是 double 类型

9. 71、大数据计算服务（MaxCompute，原 ODPS）中，表 ods_sale_detail 存放的是销售明细记录，建表语句如下：create table ods_sale_detail(sale_date datetime, region string, amt decimal); 其中 sale_date 字段表示销售的日期，amt 表示销售金额，想要找出所有汇总销售额大于 10000 的日期以及当日的销售额，以下哪几个 SQL 可以达到这个目的？（正确答案的数量：2 个）(BC)

a. select sale_date, sum(amt) from ods_sale_detail group by sale_date where amt>10000;

b. select* from(select sale_date,sum(amt) s_amt from sale_date group by sale_date)sub where s_amt>10000;

c. select sale_date, sum(amt) from ods_sale_detail group by sale_date having sum(amt)>10000;

d. select sale_date, sum(amt) from ods_sale_detail group by sale_date where sum(amt)>10000;

10. 72、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，可以使用_____从日志表 log 中找出所有符合这种特征的访问。（正确答案的数量：3 个）(ABD)

a. select* from log where tolower(log_content) like '%select%';

b. select* from log where instr(tolower(log_content),'select')>0;

c. select* from log where 'select' in tolower(log_content);

d. select* from log where regexp_instr(tolower(log_content),'select',1)>0;

11. 73、大数据计算服务（MaxCompute，原 ODPS）中，使用 CREATE TABLE LIKE 语句创建新表时，原表的哪几种属性会被复制到新表上？（正确答案的数量：3 个）(ABD)

a. 列注释

b. 分区

c. 表的生命周期属性

d. 表注释

12. 74、在一个大数据计算服务（MaxCompute，原 ODPS）的 SQL 语句中，可以读多张表中的数据，但是只能更新一张目标表中的数据。(F)

13. 75、大数据计算服务（MaxCompute，原 ODPS）的内置函数 substr 只能操作 string 类型的输入值，当输入的值是其他任何类型时，会导致异常。(F)

批注 [K35]: number: Double 或 bigint 类型，输入为 bigint 时返回 bigint，输入为 double 时返回 double 类型。若输入为 string 类型会隐式转换到 double 类型后参与运算，其它类型抛异常

批注 [K36]: • 表名与列名均对大小写不敏感。

- 在创建表时，如果不指定 if not exists 选项而存在同名表，则返回出错；若指定此选项，则无论是否存在同名表，即使原表结构与要创建的目标表结构不一致，均返回成功。已存在的同名表的元信息不会被改动。

- 数据类型只能是: bigint, double, boolean, datetime 及 string。

- 表名，列名中不能有特殊字符，只能用英文的 a-z, A-Z 及数字和下划线_，且以字母开头，名称的长度不超过 128 字节。

- Partitioned by 指定表的分区字段，目前仅支持 string 类型。分区值不允许有双字节字符(如中文)，必须是以英文字母 a-z, A-Z 开始后跟字母数字，名称的长度不超过 128 字节。允许的字符包括: 空格 ' ', 冒号 ':', 下划线 '_', 美元符 '\$', 井号 '#', 点 '.', 感叹号 '!' 和 '@', 出现其他字符行为未定义，例如: "\t", "\n", "/" 等。当利用分区字段对表进行分区时，新增分区、更新分区内数据和读取分区数据均不需要做全表扫描，可以提高处理效率。

- 注释内容是长度不超过 1024 字节的有效字符串。

- lifecycle 指明此表的生命周期，create table like 语句不会复制源表的生命周期属性。

- 目前，在表中建的分层层次不能超过 6 级。

批注 [K37]: •str: String 类型，若输入为 bigint, decimal, double 或者 datetime 类型会隐式转换为 string 后参与运算，其它类型报异常。

- start_position: Bigint 类型，起始位置为 1。当 start_position 为负数时表示开始位置是从字符串的结尾往前倒数，最后一个字符是 -1，往前数依次就是 -2, -3..., 其它类型抛异常。

- length: Bigint 类型，大于 0，其它类型或小于等于 0 抛异常。子串的长度。

14. 76、大数据计算服务（MaxCompute，原 ODPS ）的 insert overwrite 操作中，select 中的表达式类型与对应的目标中的数据类型必须一致，否则会导致异常。(F)注：会有隐式转换

15. 77、大数据计算服务（MaxCompute，原 ODPS ）中的客户表 customer 是一张非分区表，现在想要清空表的内容然后重新上传数据，可以通过 将数据清空。(C)

- a. delete from customer;
- b. insert overwrite table customer select* from customer;
- c. truncate table customer;
- d. drop table customer;

16. 78、大数据计算服务（MaxCompute，原 ODPS ）中客户信息表的建表语句如下：create table customer(cust_id bigint, province string, addr string, full_addr string); 其中 full_addr 字段需要根据源表中的 province 和 addr 拼成一个完整的地址，源表中的一条记录的 province 值为空 (NULL)，addr 为 '文一西路 969 号'，则 concat (province, addr) 的值为 。(D)

- a. 运行出错
- b. "<NULL>文一西路 969 号"
- c. '文一西路 969 号'
- d. 空值 (NULL)

17. 79、大数据计算服务（MaxCompute，原 ODPS ）中的 log 表保存了日志内容，都被存到了 log_content 字段中，以访问者的 IP 地址开头，示例格式如下：127.0.0.1—[22/Mar/2016:03:47:43-0700]" GET/favicon.ico HTTP/1.1" 404 209 http://127.0.0.1/"Mozilla/5.0(X11;Linux x86_64) AppleWebKit/537.36(KHTML, like Gecko)Chrome/49.0.2623.87 Safari/537.36" 为了提取出访问者的 IP 地址，可以用哪些内置的函数？（正确答案的数量：3 个）(ABC)

- a. regexp_substr
- b. regexp_extract
- c. substr 和 instr
- d. repeat

18. 80、在大数据计算服务（MaxCompute，原 ODPS ）的客户端工具 odpscmd 中开发 SQL 时，用户对 project1 和 project2 都有足够的操作权限，当前所在的项目空间是 project1，想要查看 project2 中有哪些表，可以通过以下哪几种操作？（正确答案的数量：2 个）(AC)

- a. show tables in project2; (注：语法 show tables [in <projectname>])
- b. show project2.tables;
- c. use project2;show tables;
- d. show tables project2;

19. 81、以下哪几个大数据计算服务（MaxCompute，原 ODPS ）的 SQL 语句可以用用户表 user 中找出用户名包含 'hu' 的记录？（正确答案的数量：2 个）(AB)

- a. select* from user where username rlike '.*hu.*'
- b. select* from user where username like '%hu%'
- c. select* from user where username contains('hu')
- d. select* from user where username like '*hu*'

批注 [K38]: 需要注意的是，MaxCompute SQL 不支持事务、索引及 Update/Delete 等操作，同时 MaxCompute 的 SQL 语法与 Oracle，MySQL 有一定差别，用户无法将其他数据库中得 SQL 语句无缝迁移到 MaxCompute 上来。此外，在使用方式上，MaxCompute 无法满足实时查询，查询计算时间在分钟级别，无法在秒、乃至毫秒级别返回用户结果。
https://help.aliyun.com/document_detail/27810.html

批注 [K39]: a, b 等为 String 类型，若输入为 bigint, double 或 datetime 类型会隐式转换为 string 后参与运算，其它类型报异常
concat('ab', 'c') = 'abc'
concat() = NULL
concat('a', null, 'b') = NULL

批注 [K40]: REGEXP_SUBSTR 从 start_position 位置开始，source 中第 nth_occurrence 次匹配指定模式 pattern 的子串
REGEXP_EXTRACT 将字符串 source 按照 pattern 正则表达式的规则拆分，返回第 occurrence 个 group 的字符

批注 [K41]: show tables [in <projectname>]

批注 [K42]: •LIKE 和 RLIKE 的 source 和 pattern 参数均仅接受 string 类型；
•其他类型不允许参与运算，也不能进行到 string 类型的隐式类型转换；IN 的使用方式形如：

20. 82、大数据计算服务（MaxCompute，原 ODPS ）中，在命令行工具 odpscmd 中使用 select 语句查看表中的记录时，可以结合 limit 子句来限制输出结果的行数。(T)
21. 84、在大数据计算服务（MaxCompute，原 ODPS ）的数据仓库中的订单表 fact_order，建表语句如下：
create table fact_order(order_id string, order_amt double, order_dt string)partitioned by (dt string); 此表中的数据是从 ods_order 加工而来，ods_order 建表语句如下：create table ods_order(order_id string, order_amt bigint, order_dt string);ods_order 中有一条记录数据值是 order_id order_amt order_dt 0001 100 20160301 运行 SQL 语句将数据从 ods_order 加载到 fact_order 中：insert overwrite table fact_order partition(dt=' 20160301') select* from ods_order; 对此语句的执行结果描述正确的是_____。(D)（注：请进行实际测试）
- a. 语句可以执行，但是这条数据会被当作脏数据丢弃
 - b. 目标表与源表中的数据类型不一致，执行出错
 - c. 提交 SQL 时会报语法错误
 - d. 语句可以执行，order_amt 的值会被自动的转为 double 类型
22. 85、大数据计算服务（MaxCompute，原 ODPS ）的开发人员需要创建一张用户消费习惯表，其中一个字段想要命名为 like，则以下说法正确的是_____。(D)
- a. 可以将字段命名为 like，无需特殊处理
 - b. 可以将字段命名为 like，在创建表及引用该字段时，需要用括起来，即表示为' like'
 - c. 在创建表时无需特殊处理，但是在引用该字段时需要表示为' like'（注：没有测试成功）
 - d. 无法将字段命名为 like，因为这是关键字
23. 86、大数据计算服务（MaxCompute，原 ODPS ）SQL 中支持 join、left join、right join、full join 等，对于连接条件描述正确的有_____。（正确答案的数量：3 个）(BCD)
- a. 只能有一个连接条件
 - b. 没有 mapjoin hint 时，必须是等值连接
 - c. 没有 mapjoin hint 时，多个连接条件之间必须是 AND 关系
 - d. 使用 mapjoin hint 时，语法上可以允许没有连接条件
24. 87、在大数据计算服务（MaxCompute，原 ODPS ）SQL 中可以用逻辑运算连接多个条件，以下逻辑运算结果正确的有_____。（正确答案的数量：4 个）(ABCD)
- a. True or NULL = True
 - b. False OR False = False
 - c. True AND False = False
 - d. True AND True = True
 - e. NULL and False = NULL
25. 大数据计算服务中的日志表 log 是一张分区表，分区键是 dt，每天产生一个分区用于存储当天新增的数据，现在共有 dt= '20160101' 至 dt= '20160131' 共 31 个分区的数据，为了删除 20160101 当天新增的数据，可以通过_____方式实现。(B)
- A. delete from log where dt='20160101'
 - B. alter table log drop partition (dt='20160101')
 - C. truncate table where dt='20160101'

批注 [K43]: MaxCompute 将 SQL 语句的关键字作为保留字。在对表、列或是分区命名时请不要使用，否则会报错。保留字不区分大小写。

- D. drop partition log (dt='20160101')
26. 某企业的数据仓库运行在大数据计算服务上，开发人员在加工数据时发现用户表 user 中的用户标识字段 user_id 有部分脏数据，正确的格式应该是 8 位的字符串。可以使用 SQL 语句_____将脏数据过滤出来。(D) (本题请测试，主要测试 length 参数为 null 是返回是不是整型)
- A. select* from user where length(userid)<>8
- B. select* from user where length(userid)>8
- C. select* from user where length(userid)<8
- D. select* from user where length(userid)<>8 or userid is null
27. 大数据计算服务的数据仓库的管理员发现用户表 user 没有备注信息，为了方便其他人员的使用，可以通过_____方式给这张表加上备注。(A)
- A. alter table user set comment' my comment';
- B. 删除表 user，在重建时指定 comment 选项
- C. touch table user set commnt' my comment';
- D. change table user set comment' my comment;
28. 某 BI 系统构建在大数据计算服务上，某数据分析员在分析订单的地域分布时，需要把订单表 order（占用存储空间约 10G）与地域维表 region（占用存储空间 100M）关联起来，两张表的结构如下：
create table order(order_id string, region_id string, order_amt bigint); create table region(region_id string, region_name string)在运行时发现由于订单大量集中在上海地区，数据的倾斜导致整体运行时间较长，运行的 SQL 语句如下：select region_name, sum(order_amt)amt from order t1 join region t2 on t1.region_id=t2.region_id; 此时可以考虑通过_____优化 SQL 的运行效率。(A) 本题原标红，有可能有歧义，请查阅相关章节文档
- A. 使用 mapjoin hint: select/*+mapjoin(t2)*/region_name,sum(order_amt)amt from order t1 join region t2 on t1.region_id=t2.region_id;
- B. 使用 mapjoin hint: select/*+mapjoin(t1)*/region_name,sum(order_amt)amt from order t1 join region t2 on t1.region_id=t2.region_id;
- C. 调整两张表的位置：select region_name, sum(order_amt) amt from region t1 join t2 on t1.region_id=t2.region_id;
- D. 减小 SQL 的 splitsize，增加计算资源
29. 大数据计算服务中的客户表 customer 是一张非分区表，现在想要清空表的内容然后重新上传数据，可以通过_____将数据清空。(B)
- A. insert overwrite table customer select* from customer;
- B. truncate table customer;
- C. delete from customer;
- D. drop table customer;
30. 大数据计算服务中，临时数据表 tmp_item 是一张非分区表，开发人员在建表时指定了 lifecycle 属性为 30，且使用一次后未再进行任何操作和访问。30 天后这张表会_____。(B)
- A. tmp_item 会被自动重命名为 tmp_item.deleted
- B. tmp_item 表会被自动删除掉
- C. 不会任何变更

mapjoin综合512m，所以加入内容的应该是小标

批注 [s44]: 使用 mapjoin(hint)时，把小表放入内存中。

- D. tmp_item 中的数据会被清空，表结构仍存在
31. 大数据计算服务的开发人员创建了日志表，并且导入了历史数据后发现表的一个字段命名拼写有错误，可以采用__D__更正此问题成本最低。(D)
- A. 将原表重命名，然后建张正确的表，再导入历史数据
- B. 在表的末尾新添加一个字段
- C. 删除表后重建，然后重新导入历史数据
- D. 通过 alter 命令重新命名该字段
32. 在大数据计算服务的数据仓库中的订单表 fact_order，建表语句如下：create table fact_order(order_id string, order_amt double, order_dt string) partitioned by (dt string); 此表中的数据是从 ods_order 加工而来,ods_order 建表语句如下:create table ods_order(order_id string, order_amt bigint, order_dt string); ods_order 中有一条记录数据值是 order_id order_amt order_dt 0001 100 20160301 运行 SQL 语句将数据从 ods_order 加载到 fact_order 中：insert overwrite table fact_order partition(dt=' 20160301') select* from ods_order; 对此语句的执行结果描述正确的是____。(B)
- A. 提交 SQL 时会报语法错误
- B. 语句可以执行，order_amt 的值会被自动的转为 double 类型
- C. 语句可以执行，但是这条数据会被当作脏数据丢弃
- D. 目标表与源表中的数据类型不一致，执行出错
33. 大数据计算服务中的 fact_order 表的生命周期属性设置成了 30 天，开发人员通过以下命令创建了一张新的 fact_lottery_order 表：create table fact_lottery_order like fact_order; 对于新表的描述正确的是____。(A) 注:[create table t like]不复制生命周期，生命周期为默认值
- A. fact_lottery_order 的生命周期属性是无穷大
- B. fact_lottery_order 的生命周期属性不会被设置
- C. fact_lottery_order 的生命周期属性是 30
- D. fact_lottery_order 的生命周期属性是 0
34. 大数据计算服务中，日志表 ods_order 是一张非分区表，结构如下:create table ods_order(order_id string, order_date string, order_amt decimal); fact_order 是一张按日期分区的表，结构如下：create table fact_order(order_id string, order_amt decimal) partitioned by (order_dt string); 现在需要将数据从 ods_order 加载到 fact_order 表中，并且按照 order_date 进行分区，执行了以下 SQL: insert overwrite table fact_order partition(order_dt) select order_id, order_amt, order_date as order_dt from ods_order; ods_order 表中的 order_date 有部分记录取值为 NULL，对执行结果描述正确的是____。(B)
- A. 语句可以正常执行，其中 order_date 为 NULL 的记录会被放入 order_dt='BAD'的分区
- B. 语句可以正常执行，在提交时会报语句错误
- C. 语句可以正常执行，其中的 order_date 为 NULL 值的记录会被丢弃
- D. 语句无法正常执行，系统会抛异常
35. 在大数据计算服务 SQL 中使用 distribute by 和 sort by 可以对数据进行局部排序，以下对 distribute by、sort by 的用法描述正确的有____。(3 个) (ABD)
- A. 数据会按照 distribute by 指定的列进行分片，hash 值相同的数据会在一个分片中

- B. 使用 `sort by` 子句时，必须有 `distribute by` 子句
- C. `sort by` 和 `order by` 的用法没有区别
- D. 在一个分片中，数据会按照 `sort by` 指定的列进行局部排序
36. 大数据计算服务 SQL 中支持 `join`、`left join`、`right join`、`full outer join` 等，对于连接条件描述正确的有_____。（3 个）（ACD）
- A. 使用 `mapjoin hint` 时，语法上可以允许没有连接条件
- B. 只能有一个连接条件
- C. 没有 `mapjoin hint` 时，多个连接条件之间必须是 `AND` 关系
- D. 没有 `mapjoin hint` 时，必须是等值连接
37. 大数据计算服务中，检查数据何时被修改时需要参考 `lastmodifiedtime` 属性，对于分区表，在分区级别和表级别都有 `lastmodified` 属性。对此属性的影响，以下描述正确的有？（2 个）（AC）
注：原题有标红，请验证
- A. 更新一个分区中的数据时，只会影响该分区的 `lastmodifiedtime` 属性
- B. 更新一个分区中的数据时，表的 `lastmodifiedtime` 属性会被更新
- C. 更新一个分区中的数据时，分区的 `lastmodifiedtime` 属性会被更新
- D. 更新一个分区中的数据时，不会更新表和分区的 `lastmodifiedtime`
38. 大数据计算服务表 `fact_sale_detail` 的建表语句如下：`create table fact_sale_detail(sale_date datetime, region string, amt decimal) partitioned by (dt string);` 包括 `dt='20160229'`，`dt='20160301'`，`dt='20160302'`，`dt='20160303'` 共四天的数据，每个分区中只包含当天的销售数据。`Udf_substr` 为自定义的数据（UDF），功能和 `substr` 相同但是性能更好。现在想要统计 3 月份的按天销售汇总金额，以下哪种方式经过分区剪裁后只需要付 3 个分区的数据久可以完成计算？（4 个）（ABCD）
- A. `select dt, sum(amt) from fact_sale_detail where dt like '201603%' group by dt;`
- B. `select dt, sum(amt) from fact_sale_detail where substr (dt,1,6) = '201603' group by dt;`
- C. `select dt, sum(amt) from fact_sale_detail where dt in('20160301','20160302','20160303') group by dt;`
- D. `select dt, sum(amt) from fact_sale_detail where udf_substr(dt,1,6)='201603' group by dt;`
- E. `select dt, sum(amt) from fact_sale_detail where dt>='20160301' group by dt;`
39. 用大数据计算服务构建海量的数据仓库时，分区表是一种很常见的做法，对于分区表的描述正确的有_____。（3 个）（BCD）
- A. 支持 `hash` 分区、范围分区、列表分区及组合分区
- B. 在通过 SQL 读取分区表中的数据时，可以通过指定分区的方法只读取一部分分组，减少 IO 开销
- C. 可以单独处理指定分区中的数据，不会对其他分区数据产生影响
- D. 对于过期的数据可以将对应的分区 `drop` 掉，不会影响其他分区中的数据
40. 大数据计算服务中，表 `ods_sale_detail` 存放的是销售明细记录，建表语句如下：`create table ods_sale_detail(sale_date datetime, region string, amt decimal);` 其中 `sale_date` 字段表示销售的日期，`amt` 表示销售金额，想要找出所有汇总销售额大于 10000 的日期以及当日的销售额，以下哪几个 SQL 可以达到这个目的？（2 个）（CD）注：原题有标红，请验证
- A. `select sale_date, sum(amt) from ods_sale_detail group by sale_date where amt>10000;`
- B. `select sale_date, sum(amt) from ods_sale_detail group by sale_date where sum(amt)>10000;`

- C. `select sale_date, sum(amt) from ods_sale_detail group by sale_date having sum(amt)>10000;`
D. `select* from (select sale_date, sum(amt)s_amt from sale_date group by sale_date) sub where s_amt>10000;`

41. 在一个大数据计算服务的 SQL 语句中，可以读多张表中的数据，但是只能更新一张目标表中的数据。
(F)
42. 大数据计算服务中，SQL 的 order by 子句对结果数据进行排序，默认是升序，NULL 值会被认为是最小值。(T) 注：原题标红
43. 大数据计算服务的函数 getdate 可以用来获取系统当前时间，某 SQL 语句中使用了 getdate 函数，某次执行时，返回多条结果记录，这些返回记录中 getdate 的值是一样的。(T) 注：原题标红

1、试卷

单选题

1、云盾加密服务采用了什么机制来保证密钥的安全？ B

A 防篡改硬件密码机 前期

B 符合国家密码管理局要求的算法

C 符合银监会对密钥强度的要求

D 支持国际通用密码算法 加密算法支持全面支持国产算法以及部分国际通用密码算法

2、RDS 采用的备 M-M 的高可用架构，其备之间的数据同步依靠日志的方式实现，MySQL 数据和 SQL server 分别使用的日志是什么？ D

A 均为物理日志

B MySQL 为事务日志，SQL server 为 Binlog 日志

C 均为 Binlog 日

D MySQL 为 Binlog 日志，SQL server 为事务日志

3、ODPS 的屏显行数可以通过管理控制台实现，最大的显示条数是多少？ D

A 100 条

B 10000 条

C 1000 条

D 5000 条

注解：

limit number]的 number 是常数，限制输出行数。当使用无 limit 的 select 语句直接从屏幕输出查看结果时，最多只输出 5000 行

4、阿里云云监控产品里的监控数据可以采用哪种方式下载？ B

A 通过 openAPI 环境

B 支持管理控制台和 OpenAPI 两种方式下载

C 只能查看报表，无法下载

D 从管理控制台直接下载

5、在实际应用中，那种性能指标表示 RDS 实例磁盘空间的使用量？ A

A 磁盘空间

B 连接数

C CPU 利用率

D iops

6、ODPS 表 T 中某列 C 的数据类型为 bigint，需要修改为 double，以下哪种方式可以实现？

A

A 将表 T 删掉重建

B ALTER TABLE T COLUMN C RENAME TO C DOUBLE;

C ALTER TABLE T DROP COLUMN C;ALTER TABLE T ADD C DOUBLE;

D ALTER TABLE T CHANGE COLUMNS C BIGINT

7、以下哪个功能，是 ODPS 的 mapreduce 不支持的？ D

A 统计信息

B 任务进度汇报

C 推测执行

D 文件压缩

8、ODPS 的 ACL 授权操作一般涉及三个要素，不包括下列哪个？ C

A 操作

B 客体

C 限制条件

D 主体

9、以下哪个数据库是 RDS 支持的？ A

A MySQL

B db2

C MariaDB

D Sybase

10、下面哪个特性属于多可用区 RDS 的？ A

A 轻松实现同城容灾

B 将 RDS 的存储空间扩大 2 倍

C 有效控制云产品间的网络延迟

D 是为了防止 DDoS 攻击

11、用户在购买 RDS 实例时所选择的 __A_ 决定了该实例的最大连接数

A 内存大小

B CPU 利用率大小

C 网络流量

D 磁盘容量大小

12、ADS 中表组在创建时需要制定哪两个参数？ D

- A numhashPartitions & partiton Type
- B numhashPartitions & minRedundancy
- C numhashPartitions & executeTimeout
- D executeTimeout & minRedundancy

13、通过 DMS 管理后台或者数据库客户端，连接 RDS 数据库时，提示错误信息“max_user_connections”，代表什么含义？ D

- A IOPS 超出极限
- B RDS 空间满了
- C 网络中断
- D RDS 数据库的连接数满了

14、OTS API 中单行操作不包括以下哪个？ A

- A Postrow
- B putrow
- C updaterow
- D getrow

15、ADS 建表可指定的更新方式为 B

- A 导入更新和插入更新
- B 批量更新和实时更新
- C 自动更新和手工更新
- D 全量更新和增量更新

备注：分析型数据库的表根据 updateType 分为批量更新表（仅能够离线批量更新数据）和实时更新表（能够通过 insert/delete 实时更新数据）

批注 [s45]: 考到!!!

16、以下哪个不属于 ODPS 的访问策略的访问控制元素？ A

- A 效果
- B 操作
- C 权限
- D 主体

17、RDS for MySQL 不支持以下哪种引擎 B

- A 以上都不支持
- B memory
- C InnoDB
- D MyISAM

目前仅支持 MyISAM（新创建实例已不支持，只支持部分存量实例）、InnoDB、TokuDB 三种。

18、开发 ODPS 的用户自定义标量函数，主要是实现其中的哪方法

- A evaluate
- B main
- C iterate
- D process

!!! 19、ODPS 中，哪个命令可查看角色 role_test 的权限 C

- A list role role_test
- B show grants for role role_test
- C desc role role_test
- D show acl for role rol_test

20、下列关于 ADS 的说法中，哪一项是正确的？ B

- A 表组的属性可以修改，修改后马上生效；
- B 对于批量插入的表可以建立二级 list 分区，二级 list 分区是非动态分区，分区值由用户导入时指定
- C minredundancy 是表组的一个属性，可以指定不同的值没，比如 1、2、3、4 等，缺省值为 2 ；
- D 表组是一个逻辑概念，同一个表组中的表可以选择同时上线，再删除表组时，会把表组中的表一起删除

注解：

minRedundancy 表示该表组的副本数，默认为 2，可配置为 1、2、4、8。

仅允许删除没有任何表的空表组，维度表组不允许删除。

需要注意的是，minRedundancy 修改后需要下次装载数据时才会生效

21、ODPS 项目空间 prj1 中存在表 t1，owner 打开 LabelSecurity 设置后，执行如下命令：SET LABEL 1 TO TABLE t1；SET LABEL 2 TO TABLE t1(id)；SET LABEL 3 TO TABLE t1；此时，t1 表中的列 ID 的敏感等级标签是什么？ B

- A、0 级
- B、2 级
- C、1 级
- D、3 级

!! 22、对于 ODPS DDL 的操作，一般都是由那层完成？ A

- A worker
- B executor
- C Scheduler
- D controller

!! ??? 23、如果需要从 ODPS 同步一张表到 ADS，则该表的访问权限需要授予 ODPS 中的那个用户？

- A
- A aliyun\$gurada_build@aliyun.com
- B aliyun\$gurada_data@aliyun.com
- C aliyun\$gurada_ads@aliyun.com
- D aliyun\$ads_data@aliyun.com

24、使用 OTS Java SDK 进行表操作时需要创建 OTSClient 对象，以下哪项不属于构造此对象时必须提供的信息 A

- A AccountID
- B 实例名称

C accessKeySecret

D endpoint

注解：创建 OTSClient 对象，在构造函数中指定 EndPoint，AccessKeyId，AccessKeySecret，实例名字。

25、下列不属于 ODPS 项目空间的对象类型为？ C

A 表

B 实例

C jar 包

D 资源

解释：以操作该项目空间下的对象，例如：[表\(Table\)](#)，[资源\(Resource\)](#)，[函数\(Function\)](#)，[实例\(Instance\)](#)等，而不需要关心操作对象所在的项目空间

注解：DPS 提供了项目空间、表、资源、函数等多种对象类型。Jar 包属于资源的具体文件。

26、DRDS 默认每个实例创建几个数据库？ D

A 16 个

B 64 个

C 1 个

D 8 个

备注：默认每个实例创建 8 个数据库，也就是说 2 个实例就会有 16 个分库，3 个实例 24 个分库以此类推。

https://help.aliyun.com/document_detail/29664.html

27、关于 ODPS SQL，说法不正确的是？ D

A 在很多方面并不具备数据库的特征

B 适用于海量数据，实时性要求不高的场合

C 每个作业的准备，提交等阶段要花费较长时间

D 大部分标准 SQL 的功能都支持，但是不支持窗口函数、rownum 等

!!!!!! 28、网络安全专家服务是阿里云云盾什么服务的基础上，推出的安全代为托管服务？ C

A 阿里绿网

B DDoS 高防 IP

C 安骑士

D 基础 DDoS 防护

29、ODPS 中，当一个用户被移除后，与该用户有关的那些授权会保留？ C

A package 授权

B policy 授权

C ACL 授权

D 项目空间授权

30、在 ADS 中，什么是用户所关心的最大单元，也是用户和 ADS 系统管理员的管理职权的分界

点? B

A 列

B 数据库

C 表

D 表组

解释: 在分析型数据库中,数据库是用户所关心的最大单元,也是用户和分析型数据库系统管理员的管理职权的分界点。分析型数据库系统管理员最小可管理数据库粒度的参数,而无法未经用户授权来查看和管理数据库内部的结构和信息。

31、对于 ADS 的 list 分区,哪个说法是正确的? C

A list 分区可以通过制定最大保留的二级分区数来进行生命周期管理。一旦分区数大于这个值,将会根据导入时间进行排序,将最早导入的分区下线

B list 分区信息定义时可以指定一个现有数据中的列,但是对数据类型有要求

C list 分区为非动态分区,即分区值不由数据本身决定,而是每次导入时用户指定的

D list 分区列的类型目前仅支持数值型和字符串型两种

解释: 将经常需要进行 Join 的列(例如买家 ID)作为一级 Hash 分区列,而将日期列作为二级分区列。这样的表既可以进行大表 Join 的加速,又可以每天进行增量数据导入,并且指定保留若干天的数据在线上来进行生命周期管理。

二级 list 分区为非动态分区,暨分区值不是由数据本身决定的,而是由每次导入数据时用户指定的。所以在进行分区信息定义时需要指定一个和现有数据中的列不同的新列名,以及这个列的类型(目前仅支持 long)。二级分区有一个可选属性,available_partition_num,即为最大保留的二级分区数,当新的数据装载进来后,若线上存在的二级分区数大于这个值,那么会根据二级分区的值进行排序,下线最小的若干分区的数据。实时更新表不支持二级分区。

32、阿里绿网具有什么功能? B

A 绿网的目标是建立低碳高校的绿色 IDC,帮助用户降低服务器能耗

B 帮助用户检查网站内容的合规性,包括文字内容和图片

C 用来限制访问 ECS 云主机的源 IP,防止黑客窃取数据

D 用来限制服务器对外访问的资源,防止 ECS 云主机中毒

34、下列关于 RDS 的使用,标书正确的是? D

A 只读实例的配置必须与主实例完全相同

B 允许第三方工具进行物理备份

C MySQL 存储引擎仅支持 InnoDB

D 创建数据库时必须使用 open api 或 RDS 管理控制台进行操作

!!!33、针对 ODPS SQL 做逻辑分析的时候,会进行一些优化,不包括以下哪种优化? A

A 扫描方式

B 列映射

C 谓语动词下沉

D 常量表达式

备注: 来自 ODPS 权威指南

1、常量表达式的计算

- 2、列裁剪
- 3、谓词下推

!!! 34、OTS 中可以使用拼接的分片键，拼接就是把几个属性拼接成一个属性。拼接是为了解决什么问题？ C

- A 解决数据生命周期问题
- B 解决写压力过于集中的问题
- C 解决单个分片键下数据量过大的问题
- D 解决 OTS 预留读写量不足的问题

备注：Table Store 建议单个分片下的数据量大小不超过 1GB。如果您的表中单个分片键的所有行的总数据量大小可能超过 1GB，在设计表时可以将原来的多个主键列拼接成分片键。

35、ADS 中，对批量导入的表，那种分区模式是可取的？ D

- A hash+hash
- B list +list
- C list+hash
- D hash+list

36、OTS 中某秒内发生了 10 次读操作，其中 5 次每次操作读出当行数据大小为 2.78KB，另外 5 次操作读出当行数据大小为 5.36KB，则共消耗多少*** D

- A 11CU
- B 20CU
- C 10CU
- D 15CU

各操作消耗的写服务能力单元的计算规则：

PutRow - 本次消耗的写 CU 为修改的行主键数据大小与属性列数据大小之和除以 4KB 向上取整，若指定条件检查不为 IGNORE，还需消耗该行主键数据大小除以 4KB 向上取整的读 CU。如果操作不满足应用指定的行存在性检查条件，则操作失败并消耗 1 单位写服务能力单元和 1 单位读服务能力单元 **PutRow** 详解

UpdateRow - 本次消耗的写 CU 为修改的行主键数据大小与属性列数据大小之和除以 4KB 向上取整，**UpdateRow** 中包含的需要删除的属性列，只有其列名计入该属性列数据大小，若指定条件检查不为 IGNORE，还需消耗该行主键数据大小除以 4KB 向上取整的读 CU。如果操作不满足应用指定的行存在性检查条件，则操作失败并消耗 1 单位写服务能力单元和 1 单位读服务能力单元 **UpdateRow** 详解

DeleteRow - 被删除的行主键数据大小除以 4KB 向上取整。若指定条件检查不为 IGNORE，还需消耗该行主键数据大小除以 4KB 向上取整的读 CU。如果操作不满足应用程序指定的行存在性检查条件，则操作失败并消耗 1 单位写服务能力单元 **DeleteRow** 详解

37、ODPS 中哪种操作会使授权失效？ D

- A 对 package 授权过的表重建
- B 上述授权都会失效
- C 对 policy 授权过的表重建
- D 对 ACL 授权过的表重建

38、使用 ODPS SQL: create table t1 like t2; 建表时, 表 t1 不会具有表 t2 的哪种属性?

B

A 分区

B 生命周期

C 二级分区

D 字段的注释

注解:

lifecycle 指明此表的生命周期, create table like 语句不会复制源表的生命周期属性。除生命周期属性外, 列名、列注释以及表注释等均相同。

39、在 RDS 只读实例变更配置过程中, 并且只读实例的对应源数据库实例状态为”升降级中“。主实例可以进行下列哪种任务? C

A 重启实例

B 内外网切换

C 系统资源监控

D 创建实例

40、ODPS 表 t1、t2 都是分区表, 分区键是 pt, 对于 sql 语句 select t1.id from t1 left outer join t2 on t1.id=t2.id where t1.pt='20140101' and t2.pt='20140101',哪一个说法是正确的。答案: C

A、两张表都会进行分区裁剪

B、t2 表会进行分区裁剪

C、t1 表会进行分区裁剪

D、两张表都不会进行分区裁剪

备注 : 当两个表进行 join 操作的时候, 主表的 Where 限制可以写在最后, 从表要先完成分区筛选后再进行 Join 操作;从表分区限制条件不允许写在 Where 条件里, 要求写在 ON 条件或者子查询, 主表可以写在 WHERE 条件里(最好先用子查询过滤)。

41、DRDS 事务支持哪种一致性? A

A 最终一致性事务

B 会话一致性事务

C 强一致性事务

D 因果一致性事务


42、关于 ODPS MR 哪项说法是正确的? A

A mapworker 在输出数据时, 需要为每一条输出数据指定一个 key

B 进入 reduce 前, 数据需要进行合并操作, 然后按照 key 排序

C map 和 reduce 前都需要对数据进行分片

D reducer 的个数和 mapper 的个数一致

43、大型会员制连锁超市 S 使用 OTS 来存储消费记录表, 里面包含了客户的 vip 卡号、订单标识, 超市分店标识, 商品标识, 购买时间等, 业务逻辑遵循以下规则: 每个 vip 卡号对应一个客户, 同时一个客户职能拥有一个 vip 卡号; 该连锁超市的不同分店都拥有唯一的分店标识; 订单标识按照时间顺序产生; 为了让数据分布和访问了的分布尽可能均匀, 一下备注主键中哪一组

最合适？ **B**

A 购买时间，vip 卡号

B vip 卡号，购买时间

C vip 卡号

D 订单标识，vip 卡号

备注：https://help.aliyun.com/document_detail/27356.html

主键的第一列是分片键

购买时间，vip 卡号：购买时间，不同时间购买热度不一样

VIP 卡号，购买时间：每天每张卡产生的消费记录数从总体上来讲是均匀的，每一个分片键中的访问压力也应该是均匀的。以 CardID 作为表的分片键可以较好地利用预留读写吞吐量资源

VIP 卡号：不能唯一标识

订单标识，VIP 卡号：作为表的分片键不是一个好的选择。因为订单标识是顺序增长的，因此在同一段时间内产生的消费订单的订单标识的值会集中在一个较小的范围内，这些消费订单记录会集中写入到个别的分片，预留读写吞吐量没有得到高效的利用。

44、对于 ADS 的数据一致性问题，那个说法是错误的？ **B**

A 当 updateType=realtime 时，对于不同主键数据的多次变更，ADS 不保证先执行的变更会比后执行的变更更优先的查询到

B 当 updateType=realtime 时，不支持事务，并且仅遵循会话一致性的设计，所以 ADS 并不能作为 OLTP 系统使用 解释：最终一致性

C 当 updateType=realtime 时，当业务端暂停数据写入的若干时间后，ADS 会保证数据的一致性

D 当 updateType=realtime 时，对于同一主键数据的多次变更，ADS 会遵循 ADS 返回语句执行成功的顺序进行

45、ADS 的高度智能优化策略是指一下哪一个？ **B**

A PBO

B CBO

C RBO

D HBO

解释：分析型数据库拥有高度智能的 CBO(Cost-Based Optimization)优化策略。

!!! 46、关于 ADS 的权限模型，下列哪些描述是正确的？ **A**

A 每个权限级别能聚合其下面级别的所有权限

B 导出数据时需要 DUMP DATA、DESCRIBE 和 SELECT 权限，同时需要数据导出目的地的数据写入相关权限

C 查询表数据需要 SELECT 权限，最小级别是表

D ADS 的授权模型和 MySQL 非常相似，比如 ADS 支持针对用户在 host 上授权

解释：

任何分析型数据库支持的账号类型均可视为一个用户。和 MySQL 略有不同的是，分析型数据库目前不支持针对用户在 host 上授权。

查询数据的权限

查询表数据需要 SELECT 权限，最小级别是列

并非所有查询都需要该权限，例如 SELECT now()

导出数据的权限

导出数据同时需要 DUMP DATA 和 SELECT 权限

同时需要数据导出目的地的数据写入相关权限

聚合： Database -> Table[Group] -> Column，即每个权限级别能聚合其下面级别的所有权限。

47、OTS API 中单行操作不包括以下哪个 Action？ C

A、UpdateRow

B、Putrow

C、PostRow

D、GetRow

备注：Table Store 单行操作单行写入操作 Table Store 的单行写操作有三种:PutRow, UpdateRow 和 DeleteRow

48、ADS 中创建事实表（普通表）时会有一些限制，下列哪个说法是错误的？ D

A、表的名字不能超过 32 个字符

B、一个事实表最多不能超过 1024 列

C、至少有一级 Hash 分区

D、二级分区只能是 Hash 或 List

备注：

分析型数据库支持最多两级分区，并且一级分区仅支持 HASH 分区，二级分区仅支持 LIST 分区；事实表的创建上，有如下限制：（1）一张事实表至少有一级 Hash 分区并且分区数不能小于 8 个；（2）一个事实表组最多可以创建 256 个事实表；（3）一个事实表最多不能超过 1024 个列。

!!!49、关于云盾态势感知的优势，描述错误的是？ A

A、IT 资产划分，软件版本管理 基于威胁情报的安全事件管理对入侵行为造成的损失进行评估

B、建设安全体系 从黑客角度监控云上业务安全 对异常行为进行实时告警 让安全“可见、可控、可管”

C、免安装，免维护 无需安装任何软件，部署任何设备 无需复杂配置和更新，省心省力 安全管控从未如此简单

D、是一个针对阿里云服务器的管家服务

备注：大数据实时分析，还原安全事件全过程，对攻击者进行取证和追溯

50、某公司有一个专门的小团队负责线上系统的安全，一年以来使用云盾的 DDoS 高防 IP 体验很好，成功抵御了 5 次超过 100G 的攻击，但是费用较高，经过技术团队研究后发现公司的平台并不需要 100% 的在线率，做到 99% 即可满足业务要求。CTO 决定让小王寻找一个更合适的替代方案，如果用云盾来满足这个需求，哪项服务是最佳的？ B

A、基础 DDoS 防护

B、安全网络

C、网络安全专家服务

D、服务器安全托管

51、关于 ODPS 中的运算符的说法，正确的是？ B

A、String 类型可以和其他任意类型进行隐式转换：Boolean 不行

- B、除 **boolean** 外，其他类型不允许参与逻辑运算，也不允许其他类型的隐式类型转换
- C、只有 **bigint**、**double** 才能参与算数运算：只有 **string**、**bigint**、**double** 和 **Decimal** 才能参与算术运算
- D、特殊的关系运算符 **rlike** 表示右侧开始匹配

52、哪个字符集是目前 RDS for MySQL 不支持的 B

- A、gbk
- B、big5
- C、latin1
- D、utf8

53、下面列出的参数中哪一个不属于 RDS SQL SERVER 中用的大字段？ C

- A、image
- B、varchar
- C、int
- D、text

54、以下 ODPS SQL 哪个效率最高？ A（见 Join 语句中 where 条件的位置）

- A、`select * from (select * from A where dt=20140301)A left outer join B on B.id=A.id`
- B、没有区别
- C、`select * from A left outer join B on B.id=A.id and A.dt=20140301`
- D、`select * from A left outer join B on B.id=A.id where A.dt=20140301`

疑问：A 与 D 什么区别

当两个表进行 join 操作的时候，主表的 Where 限制可以写在最后，但从表分区限制条件不要写在 Where 条件里，建议写在 ON 条件或者子查询。主表的分区限制条件可以写在 WHERE 条件里(最好先用子查询过滤)。

55、以下哪个是 ODPS 内网的 end point？ B

- A、<http://service.odps.aliyun.com/api>
- B、<http://odps-ext.aliyun-inc.com/api>
- C、<http://service.odps.hangzhou.aliyun.com/api>
- D、<http://odps-int.aliyun-inc.com/api>

备注： 阿里云内网地址：ODPS地址：<http://odps-ext.aliyun-inc.com/api>

公网地址：ODPS地址：<http://service.odps.aliyun.com/api>

56、对于 ADS 的主键，哪个说法是正确的？ D

- A、对于 `updateType=realtime` 的表来讲，主键是必须的，且主键中不能包含作为一级 hash 分区的分区间
- B、对于 ADS 的普通表，无论 `updateType` 如何取值，主键是必要的
- C、对于 `updateType=batch` 的表来讲，主键是必须的，主要是约束导入数据时的重复记录
- D、有主键的表的性能和用法上和没有主键的表之间没有任何区别

备注：对于批量更新表，分析型数据库中主键的概念是弱化的，分析型数据库不要求一个表有主键，有主键的表的性能和用法上和没有主键的表之间没有任何区别

57、某信贷公司推出一款线上产品，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建模并产生预测数据，用户线上申请贷款时，能迅速根据用户模型预测结果进行资质评估、授信等。请根据你对阿里云产品的理解，下述方案中成本最低、性能最好的一个是？ **A**

- A、使用 ODPS 对数据进行建模，然后使用该模型对所有复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 OTS 中，以供实现快速查询
- B、使用 ADS 对数据进行建模，然后使用该模型对所有复合贷款条件的会员进行资质评估和授信评估，将处理结果保存到 ADS 中
- C、使用 ODPS 对数据进行建模，然后使用该模型对所有复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 ODPS 中
- D、使用 ADS 对数据进行建模，然后使用该模型对所有复合贷款条件的会员进行资质评估和授信评估，将处理结果保存到 OTS 中

58、ODPS 项目 prj1 中存在表 t，同时使用下述 SQL 创建视图 v:create view v as select * from t;prj1 的 owner 把视图 v 的读权限赋给了用户 Alice，以下哪个说法是正确的？ **C**

- A、因为没有表 t 的访问权限，所以 Alice 无法访问视图 v
- B、由于权限的依赖关系，Alice 既可以访问视图 v 又可以表 t
- C、Alice 可以通过访问视图 v 来查询表 t 的数据
- D、Alice 可以访问表 t，因为她具有了视图 v 的访问权限，自然获得了表 t 的访问权限

注解：

创建视图时，必须对视图所引用表的读权限。

视图只能包含一个有效的 select 语句。

视图可以引用其它视图，但不能引用自己，也不能循环引用。

不可以向视图写入数据，例如使用 insert into 或者 insert overwrite 操作视图。

■ 当视图建好以后，如果视图的引用表发生了变更，有可能导致视图无法访问，

59、在 project 中已经开启了 label security(列级别授权)的前提下，用户申请了一张表 tbl 的读权限，后来管理员将该表中的字段 x 的安全等级设置为了 4，且用户的安全级别为 1，用户能否读该表中的数据？ **D**

- A、不可以，因为用户的级别较低
- B、可以，因为之前已经有读权限了
- C、提供信息不足以判断
- D、可以访问除了字段 x 之外的数据

60、OTS 的一些表中存在冷数据和热数据的区分，所谓冷数据是指根据时效性判断被查询的可能性不大的数据，反之为热数据。下面关于冷热数据的说法，错误的是？ **D**

- A、可以用不同的表来区分冷热数据，并设置不同的预留读写吞吐量
- B、表中存在大量冷数据会导致数据访问压力不均匀，从而导致表上配置的预留读写吞吐量无法被充分利用
- C、按照冷热数据拆分成不同的表后，可对冷数据对应的表设置较小的预留读写吞吐量
- D、在同一张表可以对不同的分片设置不同的预留读写吞吐量，对于冷数据所在的片，设置较低的预留读写吞吐量

解释：为了解决这种问题，可以用不同的表来区分冷热数据，并设置不同的预留读写吞吐量。例如，将消费记录按月份分表，每一个新的自然月就换一张新的表。当月的消费记录表需要不停写入新的消费记录，同时有查询操作。当月的消费记录表可以设置一个较大的预留读写吞吐量配置来满足访问需求。前几个月的表由于不再写入新数据或者写入的新数据量较少，查询的请求较多，因此前几个月的消费记录表可以设置较小的预留写吞吐量，较大的预留读吞吐量。而历史超过一年的消费记录表，由于再被使用的可能性不大，可以设置较小的预留读写吞吐量配置。已经超出维护年限的消费记录表可以将数据导出，存入 OSS(Open Storage Service)归档，或直接删除。

61、某些在线广告 APP 采用了阿里云的技术，通过对网上搜集的海量数据的分析，给企业主提供精准的广告投放人群。该 APP 的产品经理介绍该系统能实现以下计数指标：

- 1- 能存储海量数据，单标大小超过 100TB
- 2- 支持多张上亿的表进行 Join
- 3- 支持多种筛选条件，包括连续范围筛选、取值筛选、and/or/not 条件组合等
- 4- 支持快速（3 秒以内）对十个维度分别进行技术透视分析

根据以上信息，请判断该 APP 使用的是阿里云的哪款产品？ C

- A、DRDS
- B、OTS
- C、ADS
- D、ODPS

62、关于 ADS 的数据类型，说法不正确的是哪一项？ C

- A、布尔类型 boolean 取值只能为 0 或 1
- B、ADS 所有的数据类型都不支持 unsigned
- C、数值类型 float 可以指定精度，例如：float(5,3),但是 double 类型不支持指定精度
- D、双精度浮点数 double 类型的存储字节数为 4，大整数类型 bigint 存储字节数为 8

63、RDS 和自建数据库相比，以下哪个是错误的？ B

- A、RDS 服务可用性高达 99.95%，但是自建数据库需自行保障，自行搭建主从复制，自建 RAID 等
- B、自建数据库能方便地支持快速部署、弹性扩容
- C、RDS 可以自行备份，自建数据库需要寻找备份存放空间以及定期验证备份是否可恢复
- D、RDS 无需运维，自建数据库需招聘专职 DBA 来维护，花费大量人力成本

64、OTS python CLI 工具需要哪个版本的 python 支持？

- A、2.3 版
- B、2.7 版
- C、2.6 版
- D、3.0 版

65、OTS 的单表可以存储规模较大的数据。目前单表存储的限制为？ C

- A、10T
- B、100T
- C、没有上限

D、1T

66、对于 OTS 表的主键描述正确的是哪一个？ C

- A、由四个属性组成
- B、建表的时候可以暂时不指定，之后用 `alter table` 修改
- C、组成主键的属性只能是 `string` 或者 `integer`
- D、主键本身就是表的分片键

注解：

创建 **Table Store** 表时必须指定表的主键。主键包含 1~4 个主键列。每一个主键列都有名字和类型。**Table Store** 对主键列的名字和类型都有限制，详细信息可以参考 **Table Store** 数据模型的主键一节。**Table Store** 根据表的主键索引数据，表中的行按照它们的主键进行升序排序。属性列的数据类型只能是 `String` 和 `Integer`。如果为 `String` 类型，长度不超过 1KB。

Table Store 会自动把表分成不同的数据分片，以达到对其存储数据的负载均衡。数据分片的划分粒度为主键的第一列，我们将该列称为数据分片键。即，拥有相同数据分片键的行必然在同一个数据分片中。**Table Store** 能够保证对具有同一数据分片键的数据进行更改操作的一致性。

67、在利用 RDS 管理控制台将数据迁移至 RDS-MySQL 中的时候，使用的方法是 MySQL 在线迁移，可以不停服务就能够完成数据库的迁移工作，以下哪个 MySQL 版本不支持在线迁移？ C

- A、MySQL 5.6
- B、MySQL 5.1
- C、MySQL 5.0
- D、MySQL 5.5

!!! 68、ODPS 中一个表可以上限最多拥有多少个分区？ A

- A、60000 个
- B、100000 个
- C、不限制
- D、10000 个

备注：目前 ODPS 支持的分区数量上限为 6 万。

!!! 69、以下哪些应用场景适合 RDS 只读实例来完成？ B

- A、扩大 RDS 的存储能力
- B、分担 RDS 主实例的写压力
- C、缓解 RDS 主实例的写压力
- D、实现两天前的数据回溯

阿里云 RDS 只读实例 分担数据库读写压力

70、ODPS 中负责对等待提交的 `task` 进行排序的模块是？ A

- A、`Scheduler`
- B、`Worker`
- C、`Executor`
- D、`Controller`

备注: scheduler 负责 instance 调度, 包括将 instance 分解为 task、对等待提交的 task 进行排序, 以及向计算集群的伏羲 Master 询问资源占用情况来进行流控;

!!! 71、云盾反欺诈服务是阿里巴巴的什么能力开发的? C

- A、BGP 网络
- B、DDos 攻防经验
- C、大数据风控服务
- D、弹性计算

备注:风险识别服务

通过用户行为信息、软硬件环境信息、设备指纹等综合判定用户请求的风险程度:可信、可疑、风险,业务根据风险结果进行相应处理:直接进入业务流程,图形/短信/邮箱验证,直接拦截,离线分析等。

72、表 t 中有 1000 万行数据,想要随意抽取 10 行数据,在 odpscmd 中执行消耗时间最短的方式是哪种? D

- A、select * from t where rownum<=10
- B、get t 10
- C、select * from t limit 10
- D、read t 10

73、使用 ODPS 动态分区时,SELECT 的分区列放在列表的什么位置? D

- A、中间
- B、开始
- C、任意位置
- D、结尾

1、WEB 应用采用何种方式可以接入云盾反欺诈服务? C

- A: HTMLS
- B: SDK
- C: JavaScript
- D: JavaScript 和 SDK

客户平台集成风险拦截服务方式:前端引入组件 JS 获得滑动通过签名串,服务端调用风险拦截 API 验证签名串,获得签名串验证结果

74、关于 RDS 的续费方式,正确的是: B

- A: 无需续费
- B: 只能手动续费
- C: 只能自动续费
- D: 可以手动续费,也可以设置自动续费

75、下列关于 RDS 可用区的说法正确的是: D

- A: RDS 单可用区轻松实现了同城容灾
- B: RDS 不支持多可用区
- C: RDS 多可用区实现了 ECS 和 RDS 间的网络延迟更小
- D: RDS 支持跨可用区迁移

目前多可用区 RDS 不额外收取任何费用,在已开通多可用区地域的用户可以直接购买多可用区 RDS 实例,也可以通过跨可用区迁移将单可用区 RDS 实例转化成多可用区 RDS 实例。注

意：因为多可用区之间存在一定的网络延迟，因此多可用区 RDS 实例在采用半同步数据复制方案的时候，对于单个更新的响应时间会比单可用区实例长。这种情况最好通过提高并发量的方式来实现整体吞吐量的提高。

76、ADS 的普通表的一级分区数量最多有多少个？ C

A：无限制 B：128 个 C：256 个 D：1000 个

一般来讲，每个分区的数据不超过 800 万条为宜，当然也不绝对，分区数不能超过 256 个

77、OTS 的单个表可以包括多少个属性列？ C

A：0-128 个 B：0-1024 个 C：无限制 D：0-256 个

!!!78、关于 ADS 的 ECU，说法正确的是： A

A：可以通过 DDL 语句 `alter database set ecu_count=N` 来修改 ECU 的个数

B：可以动态的扩容或者缩容，两者都是瞬间的同步操作

C：可通过查询表 `Information_schema.resource_request` 获取目前的 ECU 的状态

D：通过参数设置，可以使同一个 ADS 的库同时使用两种类型的 ECU

解释：缩容和扩容都不是瞬时的同步操作，可以使用元数据查询状态：

```
select * from information_schema.resource_request;
```

79、ODPS 中提供了别名命令 `ALIAS<alias>=<real>`，可以为哪些对象设置别名？ A

A：资源 B：列 C：行 D：表

用途：为资源创建别名。

80、DRDS 和普通的关系型数据库（如 MySQL）相比，SQL 优化要特别考虑的是？ D

A：磁盘 IO 开销 B：内存开销 C：并发度 D：网络 IO 开销

DRDS 是一个高效稳定的分布式关系数据库系统。但是，由于其处理的是分布式的关系查询，因而它对于 SQL 的查询优化与传统的单一数据库（如 mysql，oracle）的查询优化有所不同。后者在查询优化时，主要考虑的就是磁盘 IO 的开销，但前者在优化时还需要考虑另外一个更为重要的 IO 开销—网络。

81、关于 RDS 的配置变更，描述错误的是 A

A：升级过程中服务不中断（备注：闪断 30S） B：MySQL5.1 可以直接升级到 MySQL5.6

C：数据库版本无法降级 D：用户可以将实例根据自身需要升级到相应的规格

82、ADS 的表进行实时插入、更新时，遵循哪种一致性设计？ B

A：强一致性 B：最终一致性 C：会话一致性 D：因果一致性

82:下列关于 RDS 性能的指标说法错误的是

A：IOPS 表示每秒磁盘读写（IO）次数，读写操作包括：数据从磁盘读到内存，或从内存写入磁

盘 B

B: 不同 RDS 实例之间可以资源共享

C: QPS 表示每秒 SQL 语句执行次数

D: 后台管理员通过 top 或 ps 命令可以查看实例的 CPU 利用率

83、假设 A 公司要在两周后进行重量级新产品发布，需要在公司的官网上进行直播，公司领导担心发布过程中会出现大流量攻击，但是公司没有专业的团队做支持，请问这种情况下云盾的哪项服务最适合满足该用户的需求？ D

A: 渗透测试服务 B: 网络安全专家服务 C: 服务器安全托管 D: DDoS 高防 IP

!!! 84、对于 ADS，哪个说法不正确？ A

A: ADS 表存在主键，即可根据主键使用 delete 删除纪录

B: 普通表分区列如果是字符串型，则不支持普通表之间的 join

C: 自查询之间不支持 join

D: ADS 通过 UNION (ALL) 各个分区内的查询结果实现自查询

85、以下关于 ODPS 授权的说法正确的是 B

A: 当一个用户被项目空间 owner 从项目中移除时，所在的授权信息都将被清除，该用户无法再访问当前项目空间中的任何对象

B: 当项目空间的 Owner 决定对另一个用户授权时，需要先将该用户添加到自己的项目空间中来，只有添加到项目空间中的用户才能被授权

C: ODPS 中的角色是一组访问权限的组合。项目创建成功后，自动生成两个角色 owner 和 admin，可以通过命令 list roles 看到

D: 当项目空间中的 owner 授权给一个用户时，ODPS 会先检验该用户是否实际存在，如果不存在，则无法授权

备注：

- 当一个用户被移除后，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。
每一个项目空间在创建时，会自动创建一个 admin 的角色

当项目空间的 Owner Alice 决定对另一个用户授权时，Alice 需要先将该用户添加到自己的项目空间中来。只有添加到项目空间中的用户才能够被授权。

添加用户的命令如下：

```
add user <username> --在项目空间中添加用户
```

86、ODPS 对 SQL DML 语句进行解析后，得到执行计划。构成执行计划的是由多个有相互依赖关系的_____组成？ C

A: 任务 (Task) B: 作业 (Job) C: 执行阶段 (Stage) D: 实例 (Instance)

ODPS 会对其进行解析，得出任务的执行计划。 执行计划是由具有依赖关系的多个执行阶段 (Stage)构成的

!!! 87、下列对于 ADS 中表和表组的描述正确的是 C

A: 维度表在创建时不需要指定表组，但是需要配置分区信息

B: 维度表可以和除维度表组外的任意表组中的表关联

C: 批量插入的表如果只有一级分区，每次导入数据是会覆盖掉已有数据

D: 维度表组有且只有一个，用户可以修改表组属性，但是不能删除

维度表可以和任意表组的任意表进行关联，并且创建时不需要配置分区信息，但是对单表数据量大小有所限制，并且需要消耗更多的存储资源

维度表组有且仅有一个，并且在分析型数据库数据库建立时会自动创建，用户不可修改和删除

88、除了正常数据外，以下哪个数据会占用所购买的 RDS 实例空间？ D

A: 数据库回滚日志 B: 索引 C: 重做日志 D: 以上都是

除了您正常的数据外，还有数据库实例正常运行所需要的空间，比如系统数据库、数据库回滚日志、重做日志、索引等。同时 RDS for MySQL 数据库实例产生的一天内的 Binlog 日志也会占用这个空间

89、对于 ADS 的连接方式，以下说法正确的是： A

A: 以上都是 B: 安装了 php-mysql-5.1.x 模块的 php 环境

C: 使用版本为 5.4 系列的 mysql-jdbc 驱动 D: mysql5.6 的客户端

90、ODPS tunnel 上传或者下载数据时，参数 rd 的默认取值是： B

A: \t B: \r\n C: , D: "

91、云盾网络安全专家服务特别适合以下哪种场景？ B

A: 购买了 100 台 ECS，但没有专业的管理员

B: 每月一次的产品促销活动，担心网络攻击

C: 频繁出现 SQL 注入警告

D: 系统每次重大升级后不清楚有没有漏洞

安全网络是通过 ip 节点组成的，具备流量转发，流量调度能力的一个网络。他需要通过配置和

接入才能使您的应用具备攻击防护的能力。一个 ip 节点就相当于您服务器的一个替身，通过 n

多个替身节点来抵御攻击。黑客无法透过安全节点攻击到您的真实服务器。

92、以下哪个不属于 RODPS 提供的功能 D

A: 将某些 R 的模型直接转换为 SQL 发布

B: 在 R 中存取 ODPS 的数据

C: 在 R 中直接使用分布式的算法包处理 ODPS 中的大数据

D: 在 ODPS 上支持 R 的任务

!!! 93、下面哪一项不是安骑士包含的功能

A: 高危漏洞修复 B: 防 WEB 应用系统密码破解

C: 异地登录报警 D: 木马文件检查

B

安骑士是云盾安全防护体系中的主机安全防护模块，基于云端联动防御，可以为云服务器提供防黑客入侵的服务，包括木马查杀、防密码暴力破解、异地登录提醒、高危漏洞检测修复、体检加固等安全防护功能

94、作为技术人员，小王被 ODPS 的大数据处理能力吸引，想体验一下，他快速注册了阿里云账号，进入官网后试图创建一个名字为 test_project 的 ODSP project，结果没有成功，试分析，小王出错的最有可能的原因是：

- A: 项目名字和现有的项目名字重复 B: 账号未开通 ODPS 的使用权限
C: 账号未充值，余额不足导致出错 D: 未创建合适的 accessID 和 accessKey

95、关于 ODPS 分区的说法，正确的是

A: ODPS 将分区列的每个值作为一个分区

B: 用户最多能指定 2 级分区

C: 分区列只支持 string 和 bigint 两种，且同一个表如果有多个分区列，这些分区列要么全部是 string，要么全部是 bigint，不能混用

D: 对于分区表，在使用数据时必须指定分区列，否则会出错

96、关于 RDS 数据迁移，错误的是

A: ECS 上自建的 MySQL 可以导入至 RDS

B: 用户可以通过 mysqldump 进行分批导入

C: 用在其他 IDC 的 MySQL 数据库也可以导入至 RDS

D: 保存在 OSS 里的图片文件也可以直接导入 RDS

97、云盾加密服务的密钥由谁负责管理？D

A: 阿里云系统管理员 B: 第三方合作伙伴

C: 客户和阿里云系统管理员 D: 用户

借助加密服务，用户能够对密钥进行安全可靠的管理

!!!98、一个 MySQL 数据库实例最多可创建多少个数据库？C

A: 400 B: 300 C: 500 D: 200

同一实例下的数据库共享该实例下的所有资源，其中：

- MySQL 版实例最多可以创建 500 个数据库
- SQL Server 2008 R2 版实例最多可以创建 50 个数据库
- PostgreSQL 版实例和 PPAS 版实例没有数据库个数限制

!!99:ODPS 执行某个 SQL 时返回信息中包括 J3_1_2_Stg1”，以下说法正确的是 B

A: 当前 Task（ID 为 3）依赖于 ID 为 1 的任务，同时触发 ID 为 2 的任务

B: 当前 Task（ID 为 3）依赖 ID 为 1 和 ID 为 2 的两个任务

C: 当前 Task（ID 为 3）依赖于 ID 为 1 的任务，该 Task 的 Worker count 为 2

D: 当前 Task（ID 为 3）有 2 个 Reducer，1 个 mapper

大数据计算服务 ODPS > 工具 > JOB 运行信息查看工具 > logview

在每个 Task 中，可以看到 Task 的名字，对于 M1_stg1，表示这是一个 Map task，stage 为 1。

R2_1_stg1 中间的 1 表示它依赖 M1 执行结束才能开始执行。同理，J5_2_4_Stg2 表示 Join5 这个阶段要依赖 R2 和 M4 两个 task 完全成才能启动运行

100:在 RDS 连接闪断发生的场景中，有可能是由于实例本身故障引起的是哪个？ C

A: 用户升级 RDS 实例配置 B: 用户升级数据库版本

C: 主库出现故障 D: 执行数据回滚操作

备注：RDS 连接闪断发生的场景有：

1. 主库出现故障（操作不当或实例本身故障）
2. 用户触发升级 RDS 实例配置
3. 用户触发升级数据库版本
4. 用户触发执行数据回滚操作

!!!!!! 101、关于 OTS 中分片的说法正确的是： A

A: 一个分片可能包含多个分片键 B: 同一个分片上的纪录拥有相同的分片键

C: 单个分片数据量最大不能超过 1T D: 相同分片键的纪录不一定在同一分片上

A

组成主键的第一个主键列又称为分片键。Table Store 会根据表中每一行分片键的值所属的范围自动将这一行数据分配到对应的分片和机器上，以达到负载均衡的目的。具有相同分片键的行属于同一个分片，一个分片可能包含多个分片键。Table Store 服务会根据特定的规则对分片进行分裂和合并，以达到更好的负载均衡，这个过程是自动的，应用无需关心。单个分片键下所有行的大小总和不能超过 1GB。

102、某公司属于传统行业，一年前刚刚发展互联网线上业务，用了 5 台 ECS，没有专门的系统管理员，最近半年上线系统总是出现各种安全问题，几次高危漏洞被黑客利用，差点将核心数据泄露。为此公司的 CEO 王总非常着急，决定向阿里云求助，请你为王总推荐一款合适的安全产品和服务，快速解决服务器的运维之痛。答案：A

A、服务器安全托管

B、DDos 高防 IP

C、阿里绿网

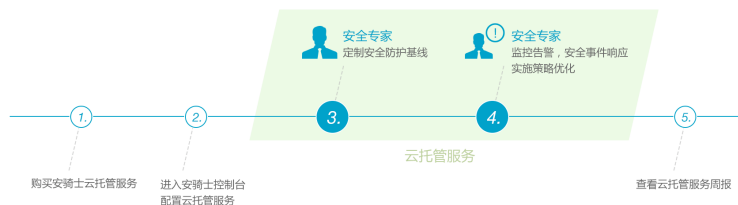
D、安全网络

注解：

一、安骑士云托管服务简介

安骑士云托管服务是一种基于云盾安骑士的安全托管服务，为云服务器提供定制化的安全防护策略、木马文件检测和高危漏洞检测与修复工作。当发生安全事件时，阿里云安全团队提供安全事件分析、响应，并进行系统防护策略的优化。

二、服务购买和使用流程



三、云服务器安全体检加固

云服务器开通安骑士云托管服务后，我们会在 3 个工作日内对完成一次完全体检和系统加固(包括木马病毒检测、高危漏洞检测)，使你的云服务器处于一个相对安全的运行状态，体检服务过程和结果以电话或者邮件告知。

四、报告安全事件

开通安骑士云托管服务的云服务器若发生安全事件，在云托管服务器控制台内，点击"安全事件告知"，将会有专门的安全工程师进行协助。

五、安全事件报告

若服务器发生安全事件，我们将协助你一起分析，最终将为您提供一份高质量的安全事件分析报告

报告内容大纲包括：

- (1) 安全事件描述
- (2) 安全事件溯源
- (3) 漏洞信息、黑客信息
- (4) 漏洞修复加固建议

六、服务报告查看及咨询

- (1) 使用您购买该服务的阿里云帐号，登录云盾控制台，可直接查看或下载服务报告
- (2) 咨询服务报告内容的问题，直接通过工单和服务的安全工程师进行沟通

103、DRDS 实例名称不能以什么开头？ 答案：B

- A、小写字母
- B、中文

1、通过阿里云控制台创建 DRDS 实例，事实上是不需要输入实例名称的，系统自动生成。根据以往经验，实例名称不能中文，实例描述可以是中文。

2、DRDS 的分布式数据库的名称（小写字母、数字、下划线组

- C、大写字母
- D、下划线

注解：

成，字母开头，不超过 24 字符)

104、查询数据需要 select 权限。目前 ADS 中对 select 权限的管理的最小级别是？ 答案：C

- A、表组
- B、表
- C、列
- D、数据库

注解：

查询数据的权限：

查询表数据需要 SELECT 权限，最小级别是列

并非所有查询都需要该权限，例如 SELECT now()

导出数据的权限：

导出数据同时需要 DUMP DATA 和 SELECT 权限

同时需要数据导出目的地的数据写入相关权限

105、下列 ODPS 的那类函数可以用在查询的 where 子句中？ 答案：D

- A、user defined table function

- B、user defined aggregation function
- C、user defined scalar function
- D、user defined function

!!!!!! 106、ODPS 的使用限制包括哪一项。答案：B

- A、单个任务中自定义的 counter 不能超过 128 个
- B、单个任务引用的资源总计字节数不能超过 64M
- C、字符串格式内容不能超过 128 个
- D、单个任务引用的资源个数不能超过 128 个

注解：

ODPS 表 string 列内容长度不允许超过 2MB。

单个任务引用的资源数量不超过 512 个，分区表按照一个单位计算。

单个任务引用的资源总计字节数大小不超过 64MB。

单个任务的输入路数不能超过 128，单个任务的输出路数不能超过 128 路。

单个任务中自定义 Counter 的数量不能超过 64。

单个 Map 或 Reduce Worker 占用 memory 默认为 2048MB，范围[256MB, 12GB]。

单个 Map 或 Reduce Worker 重复读一个资源次数限制 <=64 次。

本地运行模式下，Map Worker 个数不能超过 100；Reduce Worker 个数不能超过 100；默认一路输入下载记录数 100。

107、关于 DDos 高防 IP 的功能，描述错误的是？答案：C

- A、提供实时精准的流量报表及攻击详情，让你及时准确获得当前服务详情
- B、防护多种 DDOS 类型攻击，包括但不限于以下攻击类型 ICMP Flood、UDP Flood、TCP Flood、SYN Flood、ACK Flood 等
- C、DDOS 防护阈值弹性调整，你可以随时升级更高级的防护，调整过程服务中断时间小于等于 30 秒
- D、可随时更换防护 IP

108、由于 RDS 只读实例的数据均由源数据库实例同步而来，以下那些功能只读实例可以使用？

答案：A

- A、独立的白名单配置
- B、备份设置以及临时备份功能
- C、创建、删除数据库
- D、账号授权以及修改账号密码

109、ADS 可以使用 explain 命令查看执行计划。下面说法中正确的是？答案：D

- A、当用户发起一个 explain 查询到 ADS 系统后，ADS 会根据所有活跃节点上的数据来分析执行计划
- B、当用户通过查询的方式，想要获取文本格式的 explain 语句后，将会得到一个 XML 格式的串
- C、逻辑计划汇总，TableExecutor 节点表示参与计算后的信息，JoinExecutor 表示 join 的节点信息
- D、explain 正确执行后返回信息中，第一行为逻辑计划，第二行为物理计划

注解：

当用户通过查询的方式，想要获取文本格式的 explain 语句后，将会得到如下的 json 串。

当用户发起一个 explain 查询到分析型数据库系统后，分析型数据库会抽样一个数据分区来

分析执行计划，并以图形方式展现给用户。

XPLAIN 的 **ResultSet** 记录。其中第一行为逻辑计划，第二行为物理计划。

物理计划各个节点说明：**JoinExecutor** 表示 Join 的节点; **TableExecutor** 表示参与计算的表信息

110、在实际应用中，那种性能指标表示 RDS 实例磁盘空间的使用量？ 答案：A

- A、磁盘空间
- B、CPU 利用率
- C、IOPS
- D、连接数

111、用户 A 是公司的系统管理员，经常去上海出差，每当他在上海远程登录服务器时系统就会发出报警信息，提示“有人异地登录，请注意服务器安全”，有什么方法可以快速、自动解决这个问题？ 答案：A

- A、登录阿里云管理控制台，在安骑士配置项里添加“常用登录地点”
- B、向我的朋友电话求助，他是业界有名的黑客高手
- C、向公司领导请求帮助
- D、立即提交工单，咨询阿里云工程师

!! 112、DRDS 不具备以下哪个特性 答案：D

- A、小表广播
- B、分库分表
- C、平滑扩容
- D、智能索引

113、在 ODPS 表 T 中添加一个新的列 col，类型为 bigint，最简单易行的办法是？ 答案：B

- A、将表 T 删掉重建
- B、alter table t add columns (c bigint)
- C、新建一张表，将原有数据插入
- D、alter table add c bigint

注解：

ALTER TABLE table_name ADD COLUMNS (col_name1 type1, col_name2 type2...)

说明：列的数据类型只能是：bigint，double，boolean，datetime 及 string 类型。

114、以下那个是 RDS 性能上的优势 答案 B

- A、花费人力成本高
- B、无需运维
- C、资源利用率低
- D、需要自建 RAID

115、下面那个特性属于多可用区 RDS 的 答案：A

- A、轻松实现同城容灾
- B、有效控制云产品间的网络延迟
- C、将 RDS 的存储空间扩大 2 倍

D、是为了防止 DDOS 攻击

116、关于 ODPS 用户和权限，说法正确的是？ 答案：A

- A、ODPS 目前为提供在项目空间中彻底移除一个用户及所有权限数据的命令
- B、每个项目空间在创建后，需要手工创建一个 admin 的角色，并且为该角色授予了确定的权限
- C、admin 角色可以将 admin 权限指派给用户
- D、admin 可以设定项目空间的安全配置，但是不能够改项目空间的鉴权模型

注解：

ODPS 目前不支持在项目空间中彻底移除一个用户及其所有权限数据。

每一个项目空间在创建时，会自动创建一个 admin 的角色，并且为该角色授予了确定的权限：能访问项目空间内的所有对象，能进行用户与角色管理，能对用户或角色进行授权。与项目空间 Owner 相比，admin 角色不能将 admin 权限指派给用户，不能设定项目空间的安全配置，不能修改项目空间的鉴权模型。Admin 角色所对应的权限不能被修改。

117、RDS for Mysql 监听的端口是多少？ 答案：A

- A、3306 端口
- B、8080 端口
- C、3600 端口
- D、3060 端口

117、ADS 创建普通表时，关于 updateType 属性那种说法是错误的？ 答案：B

- A、合法的取值为 realtime 或者 batch，该值不显式指定时，默认取 batch
- B、该值为 realtime 时，对应的表必须指定合法的主键并且必须指定二级分区
- C、通过该属性指定数据的更新方式
- D、该值为 batch 时，如果要支持增量导入，必须指定二级分区

注解：

根据表的数据更新方式不同，分析型数据库的表根据 updateType 分为批量更新表（仅能够离线批量更新数据）和实时更新表（能够通过 insert/delete 实时更新数据），用 updateType 以区分，如果 updateType 选项不填则默认为批量更新表。需要注意的是，updateType=realtime 暨为实时更新表时，必须指定合法的主键并且不能有二级分区。

!! 118、关于 ODPS policy 授权，说法不正确的是。答案：C

- A、支持根据 IP 授权
- B、支持授权的时间段
- C、支持根据数据敏感级授权
- D、支持表名通配符授权

https://help.aliyun.com/document_detail/28663.html?spm=5176.doc28664.6.137.IYVXWh

119、云盾 DDOS 基础防护功能如何开启 答案：A

- A、自动开通
- B、需要从管理控制台申请开通
- C、提交工单，24 小时内开通
- D、联系客户经理线下开通

!!!120、从 ODPS 导入到 ADS 中时，可以从哪个表中查询导入状态？答案：B

- A、master.job_status
- B、information_schema.job_instances
- C、user.job_instances
- D、performance_schema.job_status

注解：

数据导入命令发送后，数据并不会立刻导入到分析型数据库中，而是会在后台进行数据的导入工作。用户可以通过使用 SQL 命令或在 iDB Cloud 中进行查询数据导入状态。高级用户也可以直接在 information_schema 中查询全部数据导入的信息（具体见附录）。导入方法有三种：1、通过 SQL 语句查询；2、通过 iDB Cloud 查询数据导入状态；3、通过 Meta DB 进行查询。

121、以下哪种不是加载数据到 ODPS 的合理方法。答案：C

- A、dship
- B、tunnel
- C、DT task
- D、mapreduce

122、Tunnel 服务的公网 end point 是。答案：D

- A、http://service.odps.aliyun.com/tunnel
- B、http://tunnel.odps.aliyun.com
- C、http://tunnel.odps.aliyun.com/api
- D、<http://dt.odps.aliyun.com>

注解：

阿里云内网地址：

ODPS 地址：<http://odps-ext.aliyun-inc.com/api>

Tunnel 地址：<http://dt-ext.odps.aliyun-inc.com>

公网地址：

ODPS 地址：<http://service.odps.aliyun.com/api>

Tunnel 地址：<http://dt.odps.aliyun.com>

!!123、ODPS 中负责对等待提交的 task 进行排序的模块是？答案：A

- A、scheduler
- B、controller
- C、executor
- D、worker

备注：来自 ODPS 权威指南

scheduler 会把该 Instance 分解成多个任务,生成任务流 DAG 图,把可运行的 Task 放到其 taskpool 中。Taskpool 是个优先级队列

!!!124、预付费模式 RDS 实例过期后，系统会自动进行什么操作。答案：A

- A、实例将会被锁定，无法进行写操作
- B、过期当天删除实例
- C、实例将会被锁定，无法进行读写
- D、自动发邮件通知续费，实例可以正常使用

125、关于 RDS 数据转移功能，描述最准确的是 答案：D

- A、支持实例间复制数据库
- B、数据备份文件可以从控制台下载
- C、支持云外 mysql 数据库的迁入
- D、以上功能都支持

126、某图书管理系统使用 ODPS 进行离线数据分析，现在他们要将同一类目的书籍的名称拼成一个字符串，即将属于同一分组的字符串合并在一列展示，以下哪个内置函数可以实现这一功能。

答案:D

- A、regexp_instr
- B、concat
- C、split_part
- D、wm_concat

concat是内置函数，是一个function

注解： WM_concat 是聚合函数，结果是"列"

wm_concat 聚合函数，用途：用指定的 separator 做分隔符，链接 str 中的值。

其输入与输出是多对一的关系，即将多条输入记录聚合成一条输出值。可以与 SQL 中的 group by 语句联用。

concat 用途：返回值是将参数中的所有字符串连接在一起的结果。

127、除了正常数据以外，以下哪个数据会占用所购买的 RDS 实例空间 答案：B

- A、数据库回滚日志
- B、以上都是
- C、索引
- D、重做日志

注解：除了您正常的数据外，还有数据库实例正常运行所需要的空间，比如系统数据库、数据库回滚日志、重做日志、索引等。同时 RDS for MySQL 数据库实例产生的一天内的 Binlog 日志也会占用这个空间。这些文件会保证 RDS 实例正常稳定运行。

128、ADS 的普通表最多包含多少个列？ 答案：B

- A、2048 个
- B、1024 个
- C、4096 个
- D、512 个

注解：

（1）一张事实表至少有一级 Hash 分区并且分区数不能小于 8 个；（2）一个事实表组最多可以创建 256 个事实表；（3）一个事实表最多不能超过 1024 个列。

129、ODPS SQL 任务中，发现 join 操作有数据倾斜时，可以考虑用什么方法优化。答案：C

- A、增加 join 的 instance 数量
- B、调整 split size
- C、用 mapjoin hint
- D、调整左右表的顺序

130、OTS 单个分片下所有行的大小总和不能超过多大。答案：1G

- A、无限制
- B、10G
- C、1M
- D、1T

!!!! 131、某企业使用 ADS 做数据分析，其中部分数据来源于 ODPS。技术人员在使用 LOAD DATA 命令从 ODPS 加载数据到 ADS 的过程中，碰到如下错误信息：ERROR 1105(HY000):You are not the owner of the source table,最有可能出错的原因是那种。答案：C

- A、源表不存在
- B、ADS 的用户 garuda_build@aliyun.com 没有对源表的操作权限
- C、ODPS 的用户 garuda_build@aliyun.com 没有对源表的操作权限
- D、源表中没有数据，为空表

注解：

导入一个新的 ODPS 表时，公有云用户需要在 ODPS 中授权该表给分析型数据库的导入用账号

!!! 132、ODPS 中，允许的 SQL 的最大长度为。答案：B

- A、8M
- B、2M
- C、1M
- D、512K2

!!! 133、DRDS 的读策略中不包含以下哪种 答案：D

- A、主库读
- B、均衡
- C、自定义
- D、主库写

注解：主库读、均衡、只读均衡、自定义四种。

!!!! 134、相比传统数据库支持的数据类型，下列哪个数据类型是 ADS 特有的。答案：D

- A、boolean
- B、UDF_SYS_GEO_DISTANCE
- C、timestamp
- D、multivalue

135、ODPS 中，以下哪种授权方式不能对角色授权。答案：A

- A、label
- B、policy
- C、ACL
- D、package

```
odps@hzjtdsj_odps>set label 1 to role hzjtdsj_zt;  
FAILED: mismatched input 'role' expecting 'USER' or 'TABLE'
```

136、查询数据需要 select 权限，目前 ADS 中对 select 权限的管理的最小级别是。答案：A

- A、列

- B、表组
- C、数据库
- D、表

137、ODPS 中，客户端提交任务成功后，可以根据__A_来出巡作业状态

- A、Instance ID
- B、Task id
- C、SQL id
- D、JOB id

138、SQL 注入攻击就是攻击者通过欺骗什么服务器，执行非授权的任意查询过程。答案：B

- A、备份
- B、数据库
- C、Web
- D、文件

注意：SQL 注入攻击就是攻击者通过欺骗数据库服务器执行非授权的任意查询过程。

!!!!!! 140、以下关于 ODPS MR 中的归并操作（Combiner）的说中，正确的是.答案：C

- A、通常下 combiner 的处理逻辑和 map 一致
- B、Combiner 函数将 shuffle 阶段所有 key 值相等的记录进行归并
- C、当 map 输出数据后，框架会在 map 端对相同 key 值的数据进行本地的归并操作
- D、Combiner 是必须的，可以减少传输带宽

注解：

归并(Combiner)函数将 Shuffle 阶段相邻的 Record 进行归并。用户可以根据不同的业务逻辑选择是否使用归并函数。归并函数是 MapReduce 计算框架的一种优化，通常情况下 Combiner 的逻辑与 reduce 相同。当 map 输出数据后，框架会在 map 端对相同 key 值的数据进行本地的归并操作。

141、ADS 创建表组时，可以指定表组的副本数，关于副本数的说法，以下哪项是正确的。答案：

A

- A、表组的副本数越多，可以一定程度的增加 ADS 的最大承受的 QPS
- B、表组的副本数为 1 时，写入数据会最快，相比副本较多的情况，在数据导入时不可用时间会更短
- C、表组的副本数最小值为 2
- D、表组的副本数越少，表组的可用性会越高

注解：

minRedundancy 表示该表组的副本数，默认为 2，可配置为 1、2、4、8。需要注意的是，如果将一个表组配置为 1 副本，那么这个表组中的表在数据导入时会有不可用的时间。而将表组副本数配置为 4 或更高，可以一定程度的增加分析型数据库的最大承受的 QPS，但是数据存储费用也会相应增加。

executeTimeout 表示该表组的全局 Query 超时时间，默认为 30000，单位毫秒。

142、云盾反欺诈服务可以对哪类应用的风险进行防控 答案：B

- A、移动 APP
- B、WEB 应用和移动 APP
- C、WEB 应用
- D、后台数据分析

注解:

1. 反欺诈服务可支持 WEB 和移动平台。

2. 前端接入，不同的平台，方式不一样：

- Web 平台、移动平台 HTML5 应用，前端页面引入 JS 脚本。
- 移动平台 Native 应用，引入 SDK。

143、若 RDS 用户需要立即备份 RDS 实例，可在不更改备份策略的情况下，对 RDS 实例进行临时备份，临时备份每天可触发次数是多少 答案：C

- A、50 次
- B、10 次
- C、无限次
- D、100 次

!!! 144、ADS 可以使用 explain 命令查看执行计划。下面说法正确的是。答案:A

- A、explain 正确执行后返回信息中，第一行为逻辑计划，第二行为物理计划
- B、当用户发起一个 explain 查询到 ADS 系统后，ADS 会根据所有活跃节点上的数据来分析执行计划
- C、当用户通过查询的方式，想要获取文本格式的 explain 语句后，将会得到一个 XML 格式的串
- D、逻辑计划中，TableExecutor 节点表示参与计算的信息，JoinExecutor 表示 join 的节点信息

145、基于 RDS 系统架构以及只读实例的应用场景，需要满足 MySQL 版本最低为多少，否则是不能满足只读实例的应用场景 答案：D

- A、MySQL 5.2
- B、MySQL 5.1
- C、MySQL 5.5
- D、MySQL 5.6

!!! 146、下列关于 RDS 开发使用说法不正确的是 答案：C

- A、可以使用命令行或图形界面进行逻辑备份
- B、用户可以使用 mysqldump 进行数据导入
- C、支持第三方工具进行物理备份
- D、用户不具有数据库 root 权限

147、对于 ADS 的 hash 分区，哪个说法是正确的。答案：D

- A、ADS 表的第一级分区必须为 hash 分区，第二级没有强制要求，目前可以为 list 或者 hash
- B、hash 分区时一种动态分区类型，需要指定具体的分区列，最多为 2 列

C、ADS 表的一级 hash 分区数最大为 1000

D、若一张表为 batch 更新类型，且仅有一级 hash 分区，则每次导入数据时，会对已有数据进行全量覆盖

148、使用 ODPS tunnel 命令上传数据时，下列哪个参数可以设定是否忽略脏数据。答案：A

A、-dbr

B、-fd

C、-dfp

D、-rd

!!!149、ODPS SQL 的执行计划逻辑上可以被看做是一个有向图，其中边的含义是。答案：D

A、实例

B、任务

C、作业之间的依赖关系

D、执行阶段的依赖关系

stage执行阶段

注解：

执行计划是由具有依赖关系的多个执行阶段(Stage)构成的。目前，执行计划逻辑上可以被看做一个有向图，图中的点是执行阶段，各个执行阶段的依赖关系是图的边。ODPS 会依照图(执行计划)中的依赖关系执行各个阶段。在同一个执行阶段内，会有多个进程，也称之为 Worker，共同完成该执行阶段的计算工作。同一个执行阶段的不同 Worker 只是处理的数据不同，执行逻辑完全相同。计算型任务在执行时，会被实例化，用户可以操作这个 [实例(Instance)] 的信息，例如：获取实例状态(Status Instance)，终止实例运行(Kill Instance)等。

150、从 ODPS 导入数据到 ADS 的时候，导入路径的格式为？ B

A、odps://project_name/table_name:partition_spec

B、odps://project_name/table_name/partition_spec

C、odps://project_name/table_name (partition_spec)

D、odps://project_name/table_name partition_spec

151、OTS 是构建在阿里飞天系统之上的产品，它提供的是什么服务？

A、在线高并发事务服务

B、海量数据的非结构化读写服务

C、在线 NoSQL 数据库服务

D、在线关系型数据库服务

152、以下 acl 授权里，哪种是对 ODPS 的表赋予 select 权限？ A

A、grant select on table table_name to user

B、grant desc on project project_name to user

C、grant read on project project_name to user

D、grant read on table table_name to user

153、在 RDS 只读实例变更配置过程中，只读实例在 RDS 管理控制台中可以进行以下哪种任务？

(D)

A、更改配置

- B、内外网切换
- C、参数设置
- D、查看 SQL 日志

154、以下哪个不属于 ODPS 计算层的模块？

- A、Nuwa
- B、OTS
- C、Pangu
- D、Fuxi

!!! 155、如果 RDS 实例被锁定，则用户不能进行的操作有哪些？ B

- A、select
- B、insert
- C、show view
- D、drop table

解释：如果您的数据库实例忽然只能 select 和 drop，不能做 insert、update、delete 操作，则表明您的数据库被锁定了。

!!! 156、通过 DMS 是否可以管理其他阿里云帐号下的 RDS 实例？

- A、只要有数据库帐号即可管理
- B、通过 DMS “实例授权”功能授权后即可管理
- C、完全不可以
- D、通过 RAM 授权后才能管理

157、使用 OTS Java SDK 进行表操作时需要创建，以下哪项不属于构造此对象时必须提供的信息 V

- A、实例名称
- B、AccountID
- C、Endpoint
- D、AccessKeySecret

158、使用 tunnel 命令上传数据时，如果不指定时间日期格式缺省值为？ C

- A、yyy/MM/dd HH:mm:ss
- B、yyy 年 MM 月 dd 日
- C、yyy-MM-dd HH:mm:ss
- D、yyMMddHHmmss

159、某用户担心突发大流量 DDOS 攻击超过“DDOS 基础防护”的上限影响网站访问，可以选择开通哪个云盾服务，用较低的成本扩大 N 倍的防护能力？

- A、安全网络专家服务
- B、安全网络
- C、阿里绿网
- D、DDOS 高防 IP

160、ADS 存储数据的时候采用的模型是哪种？

- A、雪花型模型
- B、关系型模型
- C、星型模型
- D、多维模型

161、ODPS 表 T 中有 1000 万行数据，需要把这些数据全部导出到本地，下面哪一种是可行的方法？ D

- A、用 select 命令，把结果重定向到文件中
- B、管理控制台中选出该表，下载数据
- C、找到该表的路径，然后把该路径下的文件全部下载
- D、使用 tunnel 命令下载

!!!!!! 162、云盾先知计划的目的是帮助企业建立私有的什么中心？

- A、应急响应中心
- B、病毒防护中心
- C、私有数据中心
- D、远程备份中心

!!!! 163、下列关于 ADS 用户的说法，哪些是正确的？ D

- A、对用户的授权支持 ACL 和 policy 两种
- B、在授权给一个用户之前，需要使用 add user on .*将用户添加到 ADS 库 dbname 中来
- C、数据库拥有者可以删除用户，意味着被删除用户无法再使用 ADS 的任何服务
- D、用户是被授权的数据库用户，由数据库拥有者授权时添加，无须开通 ADS 服务

备注：支持 ACL，不支持 policy 授权

进行授权 SQL 语法

```
GRANT privilege_type [(column_list)] [, privilege_type [(column_list)]] ... ON [object_type]
privilege_level TO user [, user] ...
```

不可以删除用户，可以回收用户

!!!!!! 164、关于 ADS 中的 ECU，哪个说法是正确的？ B

- A、接入节点和计算节点可以指定不同类型的 ECU，并且需要在创建 DB 时就确定，DB 一旦创建，ECU 类型不能修改
- B、ECU 的数量，在 DB 创建后可以修改。即使在 DB 的使用过程中，也可以随时调整（扩容或者缩容）
- C、接入节点和计算节点必须指定相同类型的 ECU，并且需要在创建 DB 时就确定，DB 一旦创建，ECU 类型不能修改
- D、创建 DB 时可以指定 ECU 的数量，必须为偶数，最小为 4

解释：分析型数据库对每个用户的每一个 DB 会分配若干个计算节点（COMPUTENODE），以及若干个接入节点（BUFFERNODE），接入节点用于接收用户的应用前端连接等工作，计算节点用于存储用户的数据和进行计算，另外还有若干个用于放置实时化数据写入缓冲的缓冲节点（BUFFERNODE）。目前分析型数据库仅计算节点是用户可按 ECU 模式配置，分析型数据库会自动根据用户的计算节点的量来配置接入节点等其它角色的数量。

偶数个，至少 2 个

165、使用 ODPS tunnel 命令上传 log.txt 文件到表 t_log 中去,t_log 为分区表(p1 string, p2 string)。
下列命令中正确的是? C

- A、tunnel upload log.txt t_log/p1= “b1” /p2= “b2”
- B、tunnel upload log.txt t_log (p1= “b1”, p2= “b2”)
- C、tunnel upload log.txt t_log/p1= “b1”, p2= “b2”
- D、tunnel upload log.txt t_log/ (p1= “b1”, p2= “b2”)

166、ODPS 通过云帐号服务器，对请求中的签名信息进行验证，返回——给 HTTP Server A

- A、AccountID 还包括md5信息，如果是项目配置身份验证，则是accsee key 和accsesId
- B、MDS 码
- C、AccessID
- D、正确标识

解释：来自 ODPS 权威指南

请求发送给 HTTP server，该请求包括用户的 ACCESSID 和 MD5 签名信息，HTTP SERVER 在接收到请求后，会把 ACCESSID 和 MD5 签名发送给云账户服务进行用户认证，认证通过后，云账号服务会返回该用户的唯一 AccountID,在后续执行逻辑中，发送请求都是包括 accountID,而不是 accessID

167、有关 DDOS 攻击的描述，最准确的是? D

- A、黑客自己购买了大量服务器，专门从事破坏活动
- B、主要是攻击目标是数据库
- C、攻击的目的是让服务超负载，然后窃取机密信息
- D、攻击的主要目的是让制定目标无法提供正常服务，甚至从互联网上消失，是目前最强大，最难防御的攻击之一

168、关于 ODPS 类型之间的类型，说法不正确的是?

- A、隐式类型转换是 ODPS 依据上下文使用环境自动进行的类型转换
- B、ODPS 支持的隐式类型转换规则与显式转换相同
- C、不支持的隐式类型转换会导致异常
- D、部分类型之间不可以通过显式的类型转换，但可以通过 SQL 的隐式转换完成 是sql内建函数！！！！！！

备注：部分类型之间不可以通过显式的类型转换，但可以通过 SQL 内建函数进行转换

隐式类型转换是指在运行时，由 MaxCompute 依据上下文使用环境及类型转换规则自动进行的类型转换。MaxCompute 支持的隐式类型转换规则与显式转换相同：

!!!169、关于 OTS 中数据查询的说法正确的是? D

- A、查询时仅需要提供分片键即可
- B、可以创建索引，然后基于索引列进行查询
- C、查询时可以仅提供主键的一部分，但是和组成主键的列的顺序有关
- D、单行查询必须提供主键

!!!!!!170、ODPS SQL 目前最多支持多少个并发 insert overwrite/into 操作?

- A、1000 个
- B、128 个
- C、64 个
- D、500 个

171、对于 ODPS，下列描述正确的是？

A、在使用多路插入时，可对同一张表同时执行 insert into 和 insert overwrite

B、使用动态分区插入时，动态分区列必须在 select 列表中

C、ODPS SQL 支持 between 子句

D、对于分区表，可指定某个分区的 lifecycle，其中 n 为正整数，单位为天

lifecycle只能加在标准表上

172、ODPS 的云帐号认证使用什么机制？

A、智能卡

B、LDAP

C、数字证书

D、消息签名

173、ADS 一个普通表组最多可以包含多少个表？ A

A、256 个

B、1000 个

C、无限制

D、128 个

备注：事实表的创建上，默认有如下限制：(1) 一张事实表至少有一级 Hash 分区并且分区数不能小于 8 个 (2) 一个事实表组最多可以创建 256 个事实表 (3) 一个事实表最多不能超过 1024 个列。

174、通过备份文件及日志文件，RDS 可以将实例数据恢复至多少天内的任意时刻？

A、5 天

B、7 天

C、6 天

D、3 天

多选题

1、下列哪些 SQL 可以清空表 T？ AB

A、insert overwrite table T select * from T where 1=2;

B、truncate table T;

C、delete from table T;

D、delete table T;

2、对于 OTS 表的预留读写吞吐量，以下各项中描述正确的是？ C

A、1 单位读能力表示该表每秒可以读 1KB 数据

B、操作数据大小不足 1KB 的部分向上取整

C、当表上的操作过于频繁导致预留读/写吞吐量不足以进行更多操作时，OTS 会返回 OTSQuotaExhausted 错误给应用程序

D、1 单位写能力表示每秒可以写 1KB 数据

3、ODPS 的某个带有 Join 的 SQL 运行效率很低，可以采取以下哪些措施来检验或者改善？AC

- A、检查连接键是否存在数据倾斜
- B、检查是否有多对多连接
- C、如小表关联大表，可以考虑使用 mapjoin 优化
- D、检查是否有笛卡尔积

4、开发 ODPS 的 UDF，可以使用哪些语言？B

- A、C++
- B、Java
- C、Python
- D、Ruby

5、以下场景中，哪些适合使用 ODPS？ABC

- A、PB 级离线日志分析
- B、基于海量数据的用户特征和兴趣挖掘
- C、大型互联网企业的数据仓库和 BI 分析
- D、电子商务网站的在线交易

6、关于 ADS 多值列的说法正确的包括哪些？ACD

- A、可以存入 string 类型的多个值，以逗号分隔
- B、可以直接在 select 中使用该列
- C、可以使用 in, contains 条件对该列的单个值进行查询
- D、不能在 group by 中直接使用该列

(多值列数据存入分析型数据库后，可使用 in, contains 条件对该列的单个值进行查询，枚举查询后该列的每个值可像一个普通列一样进行各类操作。但是不允许在没有进行枚举查询时对该列直接 select 或在 group by 中使用该列。)

7、以下 ODPS SQL 中语法错误的是？AB

- A、select col2 from tbl group by substr(col2,2);
- B、select col2 as c from tbl order by col2 limit 100
- C、select substr (col2,2) from tbl group by substr(col2,2);
- D、select col2 as c from tbl order by c limit 100

8、ODPS 的控制层由哪几部分组成？BCD

- A、Master
- B、Worker
- C、Executor
- D、Scheduler

9、RDS 关系型数据库服务支持如下哪些数据库。答案：AC

- A、SQLServer
- B、oracle
- C、MySQL
- D、access

RDS 支持 MySQL、SQL Server、PostgreSQL 和 PPAS (Postgre Plus Advanced Server, 一种高度兼容 Oracle 的数据库) 引擎

10、如何知道一个 ODPS SQL 在运行时占用了多少 instance? **CD**

- A、根据表的大小, 自己计算
- B、使用 `ps -ef` 查看相关进程数
- C、用 `odpscmd` 提交时, 从返回的状态中查看
- D、打开 `logview` 查看

!!!! 11、对于 ADS 的索引说法正确的是哪一项。答案: **ABC**

- A、一个列可以设置列属性 `disableIndex=true`, 用于屏蔽 ADS 的默认列索引
- B、ADS 弱化了索引的概念, 用户无需亲自为自己的数据表配置索引
- C、ADS 拥有高度智能的自动化索引机制, 会搜集相关信息, 去自动创建索引
- D、用户需要对某列进行 `hash Join` 时, 无论是事实表之间的 `join` 还是事实表和维度表的 `join`, 都需要考虑手工处理索引问题。

注解:

列属性上, 一个列可以设置列属性 `disableIndex = true`, 用于屏蔽分析型数据库的默认列索引, 不过需要注意的是, 要如此做, 则这个列应该不在实际查询中所筛选和计算的。

分析型数据库中主键的概念是弱化的, 分析型数据库不要求一个表有主键, 有主键的表的性能和用法上和没有主键的表之间没有任何区别。

分析型数据库拥有高度智能的自动化索引机制, 所以通常用户无需亲自为自己的数据表配置索引。

一种情况例外: 暨用户需要对某列进行 `Hash Join` 时, 无论是事实表之间的 `Join` 还是事实表和维度表的 `Join`, 都需要为事实表的该列建立索引。

12、以下哪些是 ODPS SQL 的优点。答案: **ACD**

- A、相对于使用 API 或者 SDK 编程, ODPS SQL 的学习成本较低
- B、可将其他数据库中的 SQL 语句迁移到 ODPS 上来, 无需修改或者重新开发
- C、提供了丰富的内置函数
- D、用户不需要了解分布式概念

注解:

ODPS 的 SQL 语法与 Oracle, MySQL 有一定差别, 用户无法将其他数据库中的 SQL 语句无缝迁移到 ODPS 上来。

14、在使用 ODPS tunnel 命令 `upload` 数据的时候, 关于一些常见参数, 说法正确的包括以下哪些项。答案: **ACD**

- A、`-c` 本地数据文件编码, 缺省为 `utf8`
- B、`-dfp DateTime` 类型数据格式, 缺省为 `yyyyMMddHHmmss`
- C、`-cp` 指定是否在本地压缩后再上传, 减少网络流量, 缺省为 `true`
- D、`-fd` 本地数据文件的列分隔符, 缺省为逗号

注解:

-bs,--block-size:每次上传至 Tunnel 的数据块大小，默认值：100MiB (MiB=1024*1024B)。

-c,--charset: 指定本地数据文件编码，默认为' UTF-8'；不设定，默认下载源数据。

-cp,--compress: 指定是否在本地压缩后再上传，减少网络流量，默认开启。

-dft: DateTime 类型数据格式，默认为' yyyy-MM-dd HH:mm:ss'；

-dbr: 是否忽略脏数据(多列，少列，列数据类型不匹配等情况)。值为' true' 时，将全部不符合表定义的数据忽略。 值为' false'，若遇到脏数据，则给出错误提示信息，目标表内的原始数据不会被污染。

-e: 指定 odps 的 endpoint.

-fd: 本地数据文件的列分割符，默认为逗号','；

-h: 数据文件是否包括表头，如果为 true，则 dship 会跳过表头从第二行开始上传数据。

-mbr,--max-bad-records: 默认情况下，当上传的脏数据超过 1000 条时，上传动作终止。通过此参数，可以调整可容忍的脏数据量；

-ni: NULL 数据标志符，默认为"(空字符串)；

-rd: 本地数据文件的行分割符，默认为' \r\n'；

-s: 是否扫描本地数据文件，默认值为' false'。值为' true' 时，先扫描数据，若数据格式正确，再导入数据。值为' false'， 不扫描数据，直接进行数据导入。值为' only' 时，仅进行扫描本地数据，扫描结束后不继续导入数据。

-te:指定 tunnel 的 Endpoint.

-tz:指定时区。默认为本地时区：Asia/Shanghai。

15、OTS 支持多种数据类型，包括 答案： ABD

- A、binary
- B、double
- C、varchar
- D、integer

注解：支持多种数据类型(Integer、Boolean、Double、String、Binary)

!!! 16、使用 ODPS 的客户端工具 odpscmd 时，如何进入一个已有的项目.答案： CD

- A、使用 odpscmd 登录时，通过-p 参数，指定登录使用的项目空间
- B、如果账户下就一个 project，登录后就会默认进入该项目
- C、配置文件中指定 project_name，登录后即默认进入项目
- D、进入 odpscmd 后，执行 use

解释：参数：--project=<prj_name> use project

如果账户下就一个 project，登录后就不会默认进入该项目,已动手测试，不可以回默认进入

17、对于 ODPS 的多路输出（multi insert），哪些项描述是不正确的。答案： BD

- A、对于源表为分区表的，相同的一个分区不能出现多次
- B、目标表为分区表的，对于不同的分区，可以混合使用 insert into 和 insert overwrite，对于相同的分区则不可混用
- C、一般情况下，单个 sql 中最多可以写 128 路输出
- D、对于源表为未分区表的，该表可以出现多次

注解：

一般情况下，单个 SQL 里最多可以写 128 路输出，超过 128 路报语法错误。

在一个 multi insert 中，对于分区表，同一个目标分区不可以出现多次；对于未分区表，该表

需要认证的！，不能直接就登录了

不能出现多次。

对于同一张分区表的不同分区，不能同时有 `insert overwrite` 和 `insert into` 操作，否则报错返回。

18、ADS 使用 DUMP DATA 导出数据时，以限制导出行数为 1000 为例（LIMIT 1000），哪些说法是正确的。答案：ABD

- A、实际数据行数可能小于 1000
- B、实际数据行数可能等于 1000
- C、实际数据行数严格等于 1000
- D、实际数据行数可能稍大于 1000

备注：导出方式对海量数据的计算输出具有良好的性能（百万行数据导出在数百毫秒数据级），但是，对于数据精确度有一定牺牲，即实际返回的数据行数，可能是不完全精确。以限制导出行数为 1000 为例（LIMIT 1000）：

实际数据行数**可能稍大于 1000**，例如此时有 120 个数据分片，则等同于每个分区明确指定 "LIMIT 9"，最多肯能返回 1080

实际数据行数**可能稍小于 1000**，如果符合条件的行数的总数小于 1000

实际数据行数**可能稍小于 1000**，如果数据分片很均匀，例如此时有 120 个数据分片，如果某些分片返回数据行小于 9 的话，则等同于每个分区明确指定 "LIMIT 9"

19、ADS 中，同一个表组的表具有以下哪些相同的属性.答案：BC

- A、记录条数
- B、一级 hash 分区数
- C、副本数
- D、二级 List 分区数

注解：

ADS 要求一个表组下所有表的一级分区数目一致。

一个表组所有表副本数一致。

!!!20、使用 ADS 可以很好的支持即席查询，对一些复杂的查询，也能在很短的时间内返回处理结果。以下关于 ADS 中查询涉及的说法正确的有。答案：AD

- A、如果用户的查询的条件会指定聚集列的内容或者范围，那么这样的查询性能便会有较大的提升
- B、**处于关联**方便的考虑，普通表可以放在维度表组上，这样该表和其他任意表关联都不在受表组的限制
- C、ADS 中 join 会优先采用 hash join，如果不满足 hash join 的条件，优化器才会选取其他的关联方法
- D、需要关联查询的普通表，需要放在同一个表组上

!!!21、在 RDS 只读实例变更配置过程中，RDS 主实例无法进行哪些任务 BCD

- A、查看 SQL 审计日志
- B、内外网切换
- C、备份实例
- D、创建实例

!!!22、云盾反欺诈功能可以防范哪些安全威胁 答案：ABC

- A、营销作弊
- B、垃圾注册
- C、盗卡支付
- D、CC 攻击

注解：垃圾注册、暴力破解、撞库、营销作弊、盗卡支付、垃圾广告或内容、恶意登入。

!!! 23、ODPS 通过 eclipse plugin 提供了 MR 的编程框架，开发时通常需要添加以下哪几个类.答案：ACD

- A、reducer
- B、mapreduce driver
- C、combiner
- D、mapper

```
job.setMapperClass(TokenizerMapper.class);
    job.setCombinerClass(SumCombiner.class);
    job.setReducerClass(SumReducer.class);
```

!!! 24、DMS（原 Idb Cloud）是一款访问管理云端数据的 WEB 服务，覆盖 RDS，ADS 等阿里云环境，DMS 具有以下哪些功能 答案：ACD

- A、对象管理
- B、数据库创建、
- C、实例管理
- D、数据管理

解释：数据管理（Data Management）支持 MySQL、SQL Server、PostgreSQL、MongoDB、Redis 等关系型数据库和 NoSQL 的数据库管理，同时还支持 Linux 服务器管理。它是一种集数据管理、结构管理、实例管理、诊断优化、文件管理、命令终端、实时监控和系统管理于一体的数据管理服务。用户使用数据管理服务实现服务器和数据库统一管理入口的同时，还可以获得安全、易用、专业的数据管理体验。

???? 28、关于 ODPS 中小文件的说法，正确的是 AB

- A、在 reduce 计算过程或者实时 tunnel 数据采集过程，会产生大量小文件
- B、小文件越多，会造成浪费资源，影响整体的执行性能，并且影响磁盘空间的利用率
- C、目前提供两种可供选择的小文件合并的方法：ALTER 合并模式和 SQL 合并模式
- D、大小小于盘古文件系统中的块 BLOCK 大小的文件，称之为小文件

!! 29、关于 ODPS 中的 admin 角色，哪些说法是正确的。答案：CD

- A、包含的权限只有 owner 可以调整
- B、可以修改项目空间的鉴权模型
- C、不能将 admin 权限指派给用户
- D、不能设定项目空间的安全配置

注解：

每一个项目空间在创建时，会自动创建一个 admin 的角色，并且为该角色授予了确定的权限：能访问项目空间内的所有对象， 能进行用户与角色管理，能对用户或角色进行授权。与项目空间

Owner 相比，admin 角色不能将 admin 权限指派给用户，不能设定项目空间的安全配置，不能修改项目空间的鉴权模型。Admin 角色所对应的权限不能被修改。

30、odps 授权操作一般涉及到以下几个要素。答案：ABD

- A、操作（action）
- B、客体（object）
- C、选项（option）
- D、主体（subject）

备注：授权操作一般涉及到三个要素：主体(Subject，可以是用户也可以是角色)，客体(Object)和操作(Action)

31、下面那些是 RDS 提供的系统性能监控项目 答案：ABCD

- A、IOPS
- B、网络流量
- C、TPS
- D、CPU 利用率

!!! 32、关于 ODPS 跨项目资源共享的说法正确的是。答案：ABD

- A、package 是一种跨项目空间共享数据及资源的机制，主要用于解决跨项目空间的用户授权问题
- B、如果某个 package 还有人在使用，则该 package 的 owner 无法删除此 package
- C、添加到 package 中的不仅仅是对象本身，还包括相应的操作权限，如不显式指定权限，则默认为读写权限
- D、添加资源时支持的对象类型不包括 project 类型，即不能将 project 添加到 package 里

注解：

- 目前支持的对象类型不包括 Project 类型,也就是不允许通过 Package 在其他 Project 中创对象。
- 添加到 Package 中的不仅仅是对象本身,还包括相应的操作权限。当没有通过[with privileges privileges]来指定操作权限时,默认为只读权限,即 Read/Describe/Select。”对象及其权限”被看作一个整体,添加后不可被更新。若有需要,只能删除和重新添加。

!!! 33、ADS 中不同型号的 ECU 标识着不同的配置。答案：ABD

- A、内存大小
- B、磁盘空间
- C、带宽
- D、Vcpu 核数

注解：

分析型数据库以 ECU（弹性计算单元）作为资源计量的最小单位。ECU（弹性计算单元）拥有多种型号，每种型号的 ECU，标识着不同的 vCPU 核数、内存大小、磁盘空间大小。

34、关于 ADS 的 ECU 说法哪些是正确的。答案：ABC

- A、ECU 的型号在 DB 创建后不可更改
- B、资源计量的最小单位
- C、ECU 数量可以在使用中扩容或者缩容
- D、ECU 数量变化是瞬时的同步操作。

注解:

ECU (弹性计算单元) 作为资源计量的最小单位。ECU 型号 DB 创建后不可修改, ECU 数量可以在使用中随时调整 (扩容/缩容)。缩容和扩容都不是瞬时的同步操作, 可以使用元数据查询状态。

35、DMS (原 Idb Cloud) 是一款访问管理云端数据的 web 服务, 覆盖 RDS、ADS 等阿里云环境, DMS 支持以下哪些数据源。答案: ABCD

- A、SQL Sdserver
- B、MySQL
- C、PostgreSQL
- D、ADS

36、对于数据库类型为 MySQL 的 RDS 实例, 有以下哪些种连接方法 答案: ABCD

- A、Navicat for MySQL
- B、使用 mysql 命令登入
- C、Idb CLOUD (控制台登陆)
- D、mysql-front

37、ODPS 中项目普通用户想要使用 `select * from` 访问一张非本人创建的本项目的表 A, 至少需要以下哪些权限。答案: BD

- A、表 A 的 describe 权限
- B、表 A 的 select 权限
- C、项目 P 的 AccessAllow 权限
- D、项目 P 的 createinstance 权限

注解:

权限描述中 Project 类型对象的 CreateTable 操作, Table 类型的 Select、Alter、Update、Drop 操作需要与 Project 对象的 CreateInstance 操作权限配合使用。单独使用上述几种权限而没有指派 CreateInstance 权限是无法完成对应操作的。这与 ODPS 的内部实现相关。类似的, Table 的 Select 权限也要与 CreateInstance 权限配合使用。使用时请注意。

!!! 38、ADS 中不支持以下哪些表连接方式。答案: ACD

- A、right join
- B、笛卡尔积
- C、semi join
- D、full join

注解:

不支持或语义错误:

SELECT ... FROM A FULL JOIN B: 不支持全连接;

SELECT ... FROM A RIGHT JOIN B: 不支持右连接, 需要转换为左连接;

SELECT ... FROM A SEMI JOIN B: 不支持半连接;

SELECT ... FROM A, B: 单表或多表连接, 但没有 ON 条件

SELECT ... FROM A, B WHERE A.a = B.b: 单表或多表连接在 WHERE 子句中有隐含 ON 条件, 但是没有 on 子句的, 暂不支持

39、某企业用户目前的在线订单系统运行在 mysql 版的 RDS 上, 因为业务量激增, 公司决定采用

DRDS，即将现有系统从 RDS 迁移到 DRDS 上去。从项目实施部署的角度来讲，以下哪些说法是正确的 答案：ABCD

- A、需要评估存储容量以及表的大小，以合理规划初始化 DRDS 的实例和库的个数
- B、需要评估哪些应用不支持，需要做应用改造
- C、需要评估当前应用，以识别哪些应用需要改造
- D、需要收集当前表的大小和具体如何使用等，用来判断哪些表需要拆分，哪些不需要

40、以下哪些场景可以使用 ODPS MR 实现。答案：ABCD

- A、Web 访问日志分析：分析用户访问行为，个性化推荐等
- B、搜索，比如 pagerank、网页爬取等
- C、机器学习：监督学习、无监督学习、分类算法等
- D、广告推荐：用户点击购买行为预测

注解：

搜索：网页爬取、倒排索引、PageRank。

Web 访问日志分析：分析和挖掘用户在 web 上的访问、购物行为特征，实现个性化推荐；分析用户访问行为。

文本统计分析：比如莫言小说的 WordCount、词频 TFIDF 分析；学术论文、专利文献的引用分析和统计；维基百科数据分析等。

海量数据挖掘：非结构化数据、时空数据、图像数据的挖掘。

机器学习：监督学习、无监督学习、分类算法如决策树、SVM 等。

自然语言处理：基于大数据的训练和预测；基于语料库构建单词同现矩阵，频繁项集数据挖掘、重复文档检测等。

广告推荐：用户点击（CTR）和购买行为（CVR）预测。

41、ODPS MR 中哪些操作会因为 Java Sandbox 而报错“access denied”？ 答案：ABCD

- A、起子进程执行 linux 命令
- B、直接访问本地文件
- C、java 反射限制
- D、获取本地 IP

注解：

不允许直接访问本地文件，只能通过 ODPS MapReduce/Graph 提供的接口间接访问： 读取 --resources 选项指定的资源，包括文件、Jar 包和资源表等； 通过 System.out 和 System.err 输出日志信息，可以通过 ODPS 客户端的 Log 命令查看日志信息；

不允许直接访问分布式文件系统，只能透过 ODPS MapReduce/Graph 访问到表的记录：

不允许 JNI 调用限制；

不允许创建 Java 线程，不允许起子进程执行 Linux 命令；

不允许访问网络，包括获取本地 IP 地址等等都会被禁止；

Java 反射限制，“suppressAccessChecks”权限被禁止，无法 setAccessible 某个 private 的属性或方法，以达到读取 private 属性或调用 private 方法的目的。

!! 42、以下哪几项云盾服务可以进行 DDOS 攻击防护 答案：ABD

- A、DDOS 基础防护
- B、DDOS 高防 IP
- C、态势感知

D、安全网络

!!!43、当 ODPS 中某个字段取值为 double，以下手法正确的是。答案：AC

- A、所有数据类型都可以由空值
- B、如果该表为分区表，则该字段可以作为分区列
- C、double 类型中使用 inf 表示无穷大
- D、double 类型中使用 NaN 标识无穷小

注解：

- 1.有效位数 16 位
- 2.上传支持科学计数法表示
- 3.下载仅使用数字表示
- 4.最大值：1.7976931348623157E308
- 5.最小值：4.9E-324
- 6.无穷大：Infinity
- 7.无穷小：-Infinity

44、OTS 以什么样的形式组织数据？ CD

- A: schema
- B: 表组
- C: 表
- D: 实例

Table Store 以实例和表的形式组织数据，通过数据分片和负载均衡技术，达到规模的无缝扩展

!!!45、ADS 提供哪些批量导入数据的方式？ CD

- A: 通过 batchinsert 命令
- B: 通过 rest-like api 方式
- C: 通过 insert
- D: data pipeline 系列命令

46、ODPS SQL 支持的 Join 操作类型包括哪些？ ABCD

- A: INNER JOIN
- B: LEFT OUTER JOIN
- C: FULL OUTER JOIN
- D: RIGHT OUTER JOIN

!!!47、关于 ODPS 中 tunnel upload 的分隔符，说法正确的是 BC

- A: 列分隔符不能包含行分隔符
- B: 支持多个字符的行列分隔符
- C: 转义字符分隔符，在命令行方式下在只支持\r、\n 和\t
- D: 缺省为空格

!!!48、ADS 的数据导入 job 提交后，可以通过以下哪些方法查询当前的导入状态？ BC

- A: 使用 log 命令：log
- B: 使用 SQL: select state from information_schema.current_job where job_ =
- C: 使用 DMS，通过菜单项查看：导入导出=>导入状态
- D: 使用命令：show states for

49、以下哪些方法可以支持 ODPS 中受保护项目空间的数据流出？ BD

- A: 使用 owner 通过 policy 授权
- B: 设置例外（Exception Policy）
- C: 使用 owner 通过 ACL 授权
- D: 设置项目互信（TrustedProject）

!!!!50、ADS 创建普通表的时候有那些限制？ ABC

- A: 一张普通表至少有一级 Hash 分区并且分区数不能小于 8 个
- B: 一个普通表组最多可以创建 256 个普通表
- C: 一个普通表最多不能超过 1024 列

D: 一张普通表至少有一级 Hash 分区并且分区数不能多于 1000 个

51、支持 OLTP 的数据库系统包括哪些特点: A B C D

- A: 数据原子性 B: 擅长事务处理
C: 数据一致性 D: 很好的支持频繁的数据插入和修改

!!!!52、RDS MySQL5.5 新版本特性有哪些? A B C

- A: 插入缓冲的改进提升数据库写性能 B: 支持 InnoDB 数据页压缩
C: 支持多个回滚段增加并发处理能力 D: 分区功能的增强, 但是不可以 drop 分区表
1、提高性能和可扩展性: 支持多个回滚段增加并发处理能力, 插入缓冲的改进提升数据库写性能。
2、提升易管理性和效率: 支持 InnoDB 数据页压缩, 数据页的压缩使数据文件体积变小, 减少磁盘 I/O, 提高吞吐量。
3、提高可用性: 分区功能的增强, 可 drop 分区表。
4、改善检测和诊断: 引入 performancn_shema,用于监控 mysql 监控服务器运行时的性能。

!!!53、ODPS MR 可以通过控制 Reducer 的个数来进行调优, 以下哪些方法可以帮用户调节 Reducer 的个数? ABC

- A: 可以通过调整 Mapper 的个数间接的影响 Reducer 的个数 (Reducer 的个数默认为 Mapper 个数的 1 / 4)
B: 在编程的时候(以 Java 为例), 初始化 Reducer 的方法里, 可手工指定 Reducer 个数
C: 可以通过 setNumReducerTasks 来设置
D: 可以通过 addReducer 随意增加 Reducer 的个数

???54、使用用户管理控制台来创建 ADS 时, 需要提供 ECU 的哪些信息? ABCD

- A: vCPU B: 存储 C: 内存 D: 数量
ECU (弹性计算单元) 拥有多种型号, 每种型号的 ECU, 标识着不同的 vCPU 核数、内存大小、磁盘空间大小
用户在创建数据库时需要根据自己的需求选择这个数据库的 ECU 型号, 以及初始的 ECU 数量 (必须是偶数个, 至少两个), ECU 型号 DB 创建后不可修改, ECU 数量可以在使用中随时调整 (扩容/缩容)

55、RDS 实例空间主要包括哪些文件? A B C D

- A: 日志文件 B: 数据文件 C: 系统文件 D: 临时文件

56、ODPS 的自定义函数包括哪些类? A B C

- A: 自定义标量函数 B: 自定义聚合函数 C: 自定义表值函数 D: 自定义扩展函数

57、ODPS 中, 当一个大表和多个小表做连接时, 可以考虑使用 mapjoin, 以下关于 mapjoin 的限制, 说法正确的是? A B

- A: left outer join 的左表必须是大表, inner join 左表或者右表都可以为大表
B: full outer join 不能使用 mapjoin
C: 最多支持 8 张小表
D: 所有小表占用内存综合不能超过 256M

- ODPS SQL 支持的 Join 操作类型包括：{LEFT OUTER|RIGHT OUTER|FULL OUTER|INNER} JOIN；
- ODPS SQL 目前最多支持 16 个并发 Join 操作；
- 在 mapjoin 中，最多支持 6 张小表的 mapjoin;

!!! 58、云盾 DDoS 防护功能可以防护哪些类型的攻击 A B C D

A: ACK Flood B: SYN Flood C: UDP Flood D: ICMP Flood

阿里云云盾可以防护 SYN Flood，UDP Flood，ACK Flood，ICMP Flood，DNS Flood，CC 攻击等 4 到 7 层 DDoS 的攻击

59、云数据库（RDS）具有以下哪些优势？ ABCD

- A、高可用性，采用主从热备的架构
- B、易用性，控制台自助迁移和管理
- C、高性能、底层使用高端服务器
- D、高安全性、具有网络防护能力和 SQL 审计

60、为了保证应用获得可预期的性能，需要设置 OTS 表的： CD

- A、预留存储空间
- B、预留处理内存
- C、预留写吞吐量
- D、预留读吞吐量

备注：Table Store 根据表的预留读/写吞吐量进行资源的调度和预留，从而保证应用获得可预期的性能。

61、RDS 扩展的功能有哪些？ ABC

- A、备份恢复
- B、安全设置
- C、监控查询
- D、负载均衡

?? 62、某企业因为业务发展带来了数据膨胀，每日产生 4T 数据，现在他们考虑要把现有的分析系统迁移到云上。通过系统分析，针对目前系统的状况去做云产品的选型。他们梳理了一下目前支持的应用类型，大体上有以下几种：提供给一线业务人员前一日以及最近三个月的经营状况的报表分析，同时要向客户提供 30 天内业务数据的即席查询。

请从查询效率和成本两个维度考虑，你认为需要哪些产品组成合适的解决方案？ BC

- A、DRDS
- B、RDS
- C、ODPS
- D、ADS

!!!63、ODPS 的资源类型包括以下哪些类？ ABCD

- A、Archive 类型
- B、Table 类型
- C、File 类型

D、Jar 类型

注释：file/archive/table/jar 表明资源类型，资源类型的介绍请参考[资源\(Resource\)](#)；

64、以下哪些属于 ADS 的特色功能？ ABD

- A、聚集列
- B、多值列
- C、空间列
- D、智能自动索引

65、关于 ADS 和 DMS 功能，说法正确的有？ ABCD

- A、可以进行用户和权限管理
- B、可以导入数据
- C、可以进行 DB 容量管理
- D、可以查看性能报告

66、OTS 写操作具有以下特性： AB

- A、原子性
- B、强一致性
- C、最终一致性
- D、事务一致性

67、RDS FOR MySQL 的高可用架构包括哪些功能？ ABD

- A、同城容灾
- B、异地容灾
- C、防 DDos 攻击
- D、主备架构

68、ODPS SQL 不支持的包括哪些？ ABCD

- A、delete
- B、update
- C、索引
- D、事务

备注：ODPS SQL 不支持事务、索引及 Update/Delete 等操作，同时 ODPS 的 SQL 语法与 Oracle、MySQL 有一定差别，用户无法将其他数据库中的 SQL 语句无缝迁移到 ODPS 上来。

69、OTS API 中提供了哪些 Action 可以进行多行数据操作？ ABC

- A、GetRange
- B、BatchWriteRow
- C、BatchGetRow
- D、GetMultRow

单行数据操作： GetRow / PutRow / UpdateRow / DeleteRow

多行数据操作： GetRange / BatchGetRow / BatchWriteRow

70、ADS 的普通表分区，支持哪些分区方式？ AC

- A、List

1.1 开通分析型数据库...

1.2 创建数据库

1.3 使用DMS创建和管...

1.4 导入数据

1.5 在应用中连接和使...

1.6 新增用户和授权

1.7 删除表或数据库

- B、Interval
- C、Hash
- D、Range

!!!71、云盾先知计划包括以下哪些功能？ ABCD

- A、漏洞收集
- B、审核漏洞
- C、自主设定奖励计划
- D、协助漏洞修复

!!72、以下哪些功能是 RDS 服务天生就具备的？ BCD

- A、自动备份，支持 15 天内任意时刻数据恢复：7 天
- B、数据库主辅备份
- C、数据库安全漏洞实时修复
- D、后台有专业团队提供运维服务

!!73、下列 OLAP 中，无需预先建模的是哪一项？ C

- A、HOLAP
- B、MOLAP
- C、ADS
- D、ROLAP

备注：我们通常将 OLAP 系统分为 MOLAP、ROLAP 和 HOLAP 三种。多维 OLAP（Multi-Dimensional OLAP，简称 MOLAP），是预先根据数据需要分析的维度进行建模，在数据的物理存储层面以“立方体”（Cube）的结构进行存储，具有查询速度快等的优点，但是数据必须预先建模，无法依据使用者的意愿进行即时的修改。而关系型 OLAP（RelationalOLAP，简称 ROLAP），则使用类似关系数据库的模型进行数据存储，通过类似 SQL 等语言进行查询和计算，优点是数据查询计算自由，可以灵活的根据使用者的要求进行分析，但是缺点是在海量数据的情况下分析计算缓慢。至于 HOLAP，则是 MOLAP 和 ROLAP 的混合模式。而阿里云分析型数据库（原名：分析数据库服务 ADS），则是一套 RT-OLAP（Realtime OLAP，实时 OLAP）系统。在数据存储模型上，采用自由灵活的关系模型存储，可以使用 SQL 进行自由灵活的计算分析，无需预先建模，而利用云计算技术，分析型数据库可以在处理百亿条甚至更多量级的数据上达到甚至超越 MOLAP 类系统的处理性能，真正实现百亿数据毫秒级计算。

!!74、ODPS tunnel 命令目前支持的操作包括哪些？ BC

- A、merge
- B、purge
- C、upload
- D、list

备注：支持 upload、download、resume、show、purge、help

75、当 RDS 实例所占用的磁盘容量超过了所购买的最大磁盘容量时，RDS 实例会被系统锁定。实例锁定后，用户只有以下哪些权限？ ABC

- A、select
- B、drop table

C、show view

D、insert

备注：您的实例被锁定后，您只有 select、show view 和 drop table 的权限

!!! 77、RDS 对 SQL server 有哪些限制？ ABCD

A 只能在用户数据库内建立对象，不可再 master、msdb、tempdb、model 等系统数据库中建立用户对象

B 不可更改或销毁 SQL serverMirror 镜像

C 用户不可再 SQL server 中指定备份计划，可在 RDS 管理后台进行相应操作

D 暂时不支持分布式事务（DTC）

RDS for SQL Server 有哪些限制

1、不允许用户在数据库直接修改用户密码或者创建数据库和用户。此功能在数据库中已被屏蔽，需在 RDS 管理页面中操作。

2、不可更改或销毁 SQL Server Mirror 镜像

3、只能在用户数据库内建立对象，不可在 master、msdb、tempdb、model 等系统数据库中建立用户对象

4、不提供 SSIS 服务

5、不提供 SSAS 服务

6、不提供 SSRS 服务

7、用户不可在 SQL Server 中指定备份计划，可在 RDS 管理后台进行相应操作

8、暂不支持分布式事务（DTC）

9、不允许使用 RestoreDB 重置数据库，该操作将导致 Mirror 损坏

10、不提供 broker 服务

!! 78、常见的 OLAP 分为哪三种？ ABD

A ROLAP

B HOLAP

C QOLAP

D MOLAP

备注：我们通常将 OLAP 系统分为 MOLAP、ROLAP 和 HOLAP 三种

79、某餐饮 O2O 公司，立足于分析签约餐厅的用户用餐数据，提供用户菜品推荐，以下说法正确的有？ AB

A 可以利用 ODPS 的海量数据处理能力，对签约餐厅的用户用餐数据进行离线分析

B 可以使用 ADS 的多值列，可以在一条用餐记录中包含多个菜品，减少 join 的次数，提供处理效率

C 可以使用 ADS 的实时插入特性的表，可以支持一些 OLTP 类的场景

D 可以利用 OTS 的高并发低延时的特性，支持一些在线分析，即将产生的业务数据存入 OTS 中，进行一些简单的 join 和计算

80、以下哪些项包含在 ODPS 项目空间的鉴权模型中？ BCD

A LabelSecurity

表级别

B CheckPermissionUsingACL

C SecurityConfiguration

D ProjectProtection

!!! 81、对于 ODPS 的 mapjoin 说法正确的是? BCD

A 如果使用 mapjoin, 则所有小表占用的内存总和不得超过 1G 512M

B inner join 的左表或右表均可以作为大表

C left outer join 的左表必须是大表

D 使用 mapjoin 时需要引用小表或者子查询时, 需要引用别名

- ◆ left outer join 的左表必须是大表;
- ◆ right outer join 的右表必须是大表;
- ◆ inner join 左表或右表均可以作为大表;
- ◆ full outer join 不能使用 mapjoin;
- ◆ mapjoin 支持小表为子查询;
- ◆ 使用 mapjoin 时需要引用小表或是子查询时, 需要引用别名;
- ◆ 在 mapjoin 中, 可以使用不等值连接或者使用 or 连接多个条件;
- ◆ 目前 ODPS 在 mapjoin 中最多支持指定 6 张小表, 否则报语法错误;
- ◆ 如果使用 mapjoin, 则所有小表占用的内存总和不得超过 512MB。请注意由于 ODPS 是压缩存储, 因此小表在被加载到内存后, 数据大小会急剧膨胀。此处的 512MB 限制是加载到内存后的空间大小;
- ◆ 多个表 join 时, 最左边的两个表不能同时是 mapjoin 的表。

82、OTS 的特点非常适合用来存储元数据, 和使用 RDS 作为元数据管理工具相比, OTS 具有以下哪些优点? ABCD

A 支持更大的单表规模, 可以超过百 TB

B 更易扩展, 理论上无限容量

C 更高并发量, 十万级别 QPS

D 更高的可用性, 可达 99.9%

RDS (99.95%)

83、在 RDS 实例下创建数据库的方法有哪些? AC

A 直接调用 RDS 的 openAPI

B 通过 MySQL 客户端工具, 如: MySQL-Front

C 通过管理控制台完成

D 阿里云命令行工具 CLI

84、云监控室阿里巴巴集团多年来服务器监控技术研究积累的成果, 结合阿里云云计算平台强大的数据分析能力为中小网站提供各种应用和服务***** ABC

A 旺旺消息

B 电子邮件

C 手机短信

D 钉钉消息

85、ODPS 的控制是由那几部分组成? ABC

A worker

- B Scheduler
- C executor
- D master

86、ADS 的维度表组中存放的表的特点包括以下哪些项？ AD

- A 数据量较小
- B 为星型或者雪花型
- C 记录条数据不超过 100 万或者不大于 1G
- D 需要和任何表关联

备注：其中维度表组用于存放维度表（特征上是一种数据量较小但是需要和任何表进行关联的表），目前有且仅有一个，并且在分析型数据库数据库建立时会自动创建，用户不可修改和删除

87、以下哪些情况不适合使用 ODPS mapjoin BCD

- A or 连接关联条件
- B 关联表中有临时表
- C 两个表无条件关联
- D 两个大表关联

!!! 88、安骑士的主要功能包括哪几项？ ABD

- A 异地登录提醒
- B 木马文件的查杀
- C 密码破解的发现与处理
- D 高危漏洞检查与修复

备注：包括木马查杀、防密码暴力破解、异地登录提醒、高危漏洞检测修复、体检加固等安全防护功能

89、DMS（原 iDB Cloud）是一款访问管理云端数据的 WEB 服务，覆盖 RDS、ADS 等阿里云环境，DMS 支持一下那些数据源？ ABCD

- A MySQL
- B PostgreSQL
- C ADS
- D SQL server

备注：DMS 支持 MySQL、SQL Server、PostgreSQL 和 PPAS 等数据源。

90、关于 ODPS Policy 授权的特点，说法正确的是？ AD

- A 授权对象支持以通配符 “*”
- B 删除一个对象时，与该对象相关的 policy 授权会被删除
- C policy 支持通过 xml 文件的方式授权
- D 当 Allow 和 deny 同时存在时，则以 deny 优先原则

policy和对象 的主体和客体均无关

??? 91、ODPS 中可以通过配额管理一个项目中可以使用的资源的上限，包括以下哪几种 资源？ ACD

- A 存储
- B 带宽
- C 内存 ??

D CPU

92、RDS 的基本功能有哪些？ ABD

- A 实例管理
- B 账号管理
- C 安全组管理
- D 数据库管理

93、对 OTS 的单行操作，描述正确的有？ ABD

- A getrow: 读取单行数据
- B deleterow: 删除一行
- C updaterow: 更新一行，如果该行不存在，则新增一行
- D putrow: 新插入一行数据，如果数据存在，则删除旧行，在写入新行

解释: putrow 插入数据到指定的行，如果该行不存在，则新增一行；若该行存在，则覆盖原有行。

updaterow: 更新指定行的数据，如果该行不存在，则新增一行；若该行存在，则根据请求的内容在这一行中新增、修改或者删除指定列的值。

94、OPDS 中用户认证(Authentication)的主要功能是检查请求 (Request)发送者的真实身份，通过包括以下哪几项验证？ ABC

- A、正确验证用户状态，包括是否欠费等
- B、正确验证接收到的信息在途中是否被篡改
- C、正确验证消息发送方的真实身份
- D、用户需要提取用户个性化信息时进行用户个性化认证

95、在 ADS 中，两个事实表进行 join 的充要条件包括以下哪些项？ ABCD

- A 两张表的 hash 分区数必须一致
- B 这两个表在一个表组
- C 两张表的 join key 至少有一列建立了 hash map 索引
- D 这两个表的 join key 是 hash 分区列

96、HODPS task 类型包括以下哪些？ ABD

- A join task
- B map task
- C localwork
- D reducetask

97、由于 RDS 只读实例的数据均有源数据库同步而来，以下哪些功能只读实例将无法使用？

- ABC
- A 账号授权及修改账号密码
- B 备份设置及临时备份功能
- C 创建/删除数据库
- D 独立的白名单配置

判断题

1、ODPS 的 LabelSecurity 对敏感数据的粒度可以支持列级别 **对**

2、LoadBalancerID 是识别用户 SLB 实例的唯一标识 **对**

3、RDS for MySQL 的所有备份类型均为全量备份 **否**

4、可以使用一下 ODPS SQL 把两个数据集的结果合并在一起: SELECT * FROM sale_detail WHERE region = 'hangzhou' UNION ALL SELECT * FROM 'shanghai', 此 SQL 可以直接作为子查询应用在其他 SQL 中 **否**

返回的是多列，可能还有重复的数据，所以不能用于子查询集

5、项目空间是 ODPS 的基本组织单元，它类似于传统数据库的 Database 或 Scheme 的概念，是进行多用户隔离和访问控制的主要边界 **对**

!!! 6、当一个用户被从某个 ODPS 项目空间移除后，policy 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。 **错** **ACL 授权仍然会被保留**

6、RDS 只读实例不能自动释放 **对**

7、ODPS 适用于离线数据的处理、分析或挖掘，它同时提供存储和计算两种能力，支持 SQL 和变成 (Map/reduce) 等多种使用方式 **对**

8、在 ADS 中一个表必须从属于一个表组 **对**

9、云盾安骑士可以保护阿里云以外的服务器 **对**

注解：**在非阿里云的服务器上，安骑士提供了木马文件查杀、高危漏洞检测与修复、入侵告警通**

知等功能。

10、ODPS SQL 不支持正则表达式 **错**

11、当前的 DRDS 无法支撑现有业务时，我们可以通过用户管理控制台进行 DRDS 的平滑扩容，以满足业务需求，平滑扩容的流程依次分为配置、迁移、清理、切换四个步骤 **错** **设置-迁移-切换和清理四个部分**

12、非阿里云主机也可以使用云盾 DDoS 高防 IP 服务 **对**

备注：**DDoS 高防 IP 是针对互联网服务器（包括非阿里云主机）**

!!! 13、ADS 的普通表，通常把需进行 join 的列作为一级 hash 分区，把日期作为二级 range 分区列，这样可以进行大表 join，又可以进行每天增量数据导入 **错**

备注：**没有 range 分区**

14、ODPS SQL 不支持在普通 join 的 on 条件中使用不等值表达式、or 逻辑等复杂的 join 条件，但是在 mapjoin 中可以进行如上操作 对

15、用户开启 SLB 健康检查功能后，当后端某个 ECS 健康检查出现问题时会将请求转发到其他健康检查正常的 ECS 上，而当该 ECS 恢复正常运行时，SLB 会将其自动恢复到对外或 XXXXXXX 对用户开启健康检查功能后，当后端某台 ECS 健康检查出现异常时，负载均衡会自动将新的请求分发到其它健康检查正常的 ECS 上；而当该 ECS 恢复正常运行时，负载均衡会将其自动恢复到对外或对内的服务中 对

16、关系型 OLAP 是预先根据数据需要分析的维度进行建模，在数据的物理存储层面以“立方体”(Cube)的结构进行存储，具有查询速度较快等的优点：错 多维 OLAP

17、ODPS 中可以使用 Tunnel 模块上传视图数据：错 目前 Tunnel 仅支持表(不包括视图 View)数据的上传下载

18、ODPS 中只有 string, bigint, datetime 和 double 才能参与算术运算：错，只有 string, bigint, double 才能参与算术运算

!!! 19、在开发 ODPS MR 或者 UDF 的过程中，在用到文本文件、ODPS 表等的时候，只能把它们作为主方法的输入参数，由主方法读取后识别处理 对

20、ADS 的维度表可以和任意表组的任意表进行关联，并且创建时不需要配置分区信息 对
备注：维度表可以和任意表组的任意表进行关联，并且创建时不需要配置分区信息，但是对单表数据量大小有所限制，并且需要消耗更多的存储资源。

21、在创建 ADS 数据库前，需要从业务、数据量、计算复杂度等角度进行仔细分析，然后根据分析结果去合理的设置 ECU 的型号、台数以及接入节点的配置等 否

23、一旦 ODPS 项目空间保护打开 (set security.ProjectProtection=true)，无法将项目空间内的数据转移到项目空间之外，所有的数据都只能在项目空间内流动。 对

24、ODPS SQL 的 where 子句不支持 between 条件查询 对

25、加入云盾先知计划的企业可以自主设定奖励计划，奖励额度没有要求 对
备注：企业可以自由设定高、中、低危漏洞的奖励金额和奖励范围

!! 26、SLB 不允许跨 Region 容灾 对

!!! 27、ODPS 的客户端工具 odpscmd 中可以使用 show functions 查看当前 ODPS 项目空间中所有的自定义函数 错 list functions;

28、RDS 只读实例不能自动释放 对 如果要释放该只读实例，同样在【基本信息】中点击【释放】按钮，输入验证码后即可

29、ODPS Policy 授权文件的格式是 Json 对

30、用户开启 SLB 会话保持功能后，SLB 会把来自同一客户端的访问请求分发到不同的后端 ECS 上进行处理。 错

31、ODPS 向某个存在的表分区插入数据时，分区列不能出现在 select 列表中。 否

32、某大型系统采用了多个云服务器，并将所有服务器的系统日志存储到 OTS 中，考虑到当天日志的访问量比较大，之前的历史日志访问量很少，则可以将日志分成两个，一个用来保存当天日志，设置较高的预留读写吞吐量，另一个表保存之前的历史日志，预留读写吞吐量设置比较低的值，并定时将当前的日志表中的谁归档到历史表中，这样可以在不影响应用的情况下，节省开支。 是

33、云监控除了可以为 aliyun.com 云服务器客户提供安全有效的监控服务外，还能够为其他客户提供功能强大的监控服务，只是配置比较复杂，必须自己开发脚本进行数据收集。 否
备注：云监控除了可以为云服务器 ECS 提供安全有效的监控服务外，还能够为其他自有服务器提供功能强大、简单易用的监控服务。用户不需要自行搭建或者开发监控系统，只需要通过简单的配置即可实现各种监控需求。

34、ADS 中，当 updateType=realtime 时，支持 insert、delete、update 操作。 否
分析型数据库的表根据 updateType 分为批量更新表（仅能够离线批量更新数据）和实时更新表（能够通过 insert/delete 实时更新数据），用 updateType 以区分，如果 updateType 选项不填则默认为批量更新表。需要注意的是，updateType=realtime 暨为实时更新表时，必须指定合法的主键并且不能有二级分区。

35、ADS 的普通表的一级分区只支持两种方式，Hash 和 List，目前不支持 Range 分区。 否
关于分区方面，目前分析型数据库支持最多两级分区，并且一级分区仅支持 HASH 分区，二级分区仅支持 LIST 分区

36、ODPS 的 MapReduce 可以支持在一个 Map 后连续多个 reduce。 是
基于此，ODPS 提供了扩展的 MapReduce 模型，即可以支持 Map 后连接任意多个 Reduce 操作

37、ODPS SQL 支持主键，但是不支持索引和约束。 否
ODPS SQL 采用的是类似于 SQL 的语法，可以看作是标准 SQL 的子集，但不能因此简单的把 ODPS SQL 等价成一个数据库，它在很多方面并不具备数据库的特征，如事务、主键约束、索引等。目前在 ODPS 中允许的最大 SQL 长度是 2MB。

38、如果要查看只读实例的数据延迟，用户只能通过 RDS 管理控制台进入只读实例进行查看。 否

39、用户可以为创建的 SLB 实例指定一个易于识别的名称，如果用户不指定，那么系统将以该 SLB 实例的 LoadBalancerID 作为名称，LoadBalancerID 是识别用户 SLB 实例的唯一标识，无法修改。 是

40、在 ODPS 中，所有的任务都会被实例化（Instance），以 ODPS 实例的形式存在。 否

计算型任务在执行时，会被实例化， 用户可以操作这个 [实例(Instance)] 的信息、

另一方面，部分 ODPS 任务并不是计算型的任务

41、云盾先知计划共享了阿里巴巴应急中心安全专家的资源 and 能力，提供服务时会签署平台用户保密协议。 是

42、ODPS 的 JOIN 支持多路间接，但不支持笛卡尔积，即无 on 条件的链接。答案：是

43、ODPS 的计费是以项目为单位的，对存储、计算和数据下载三个方面分别计费。答案：是

!!! 44、预付费的 RDS 实例可以在不需要的时候主动释放，以节约成本。答案：否

45、odps 分区表可以指定多个分区条件，分区条件顺序无关。答案：否

46、ODPS 中一个用户最多可以创建 10 个项目空间 答案：是



47、用户可以指定 SLB 后端服务器池内各 ECS 的转发权重，权重值越高的 ECS 将被分配到更多的访问请求 答案：否

48、在 ADS 中，数据根据分区列进行分布式的存储和计算。答案：是

?? 49、ODPS 给予标签的安全 (labelsecurity) 是项目空间级别的一种自主访问控制机制 (DAC)。 答案：否

50、云盾先知计划的技术起源于淘宝的运营经验，仅适合对电商类 WEB 应用做风险识别。 答案：否

51、ADS 中，只有相同表组的表才能进行 join 关联（维度表除外）。答案：是

52、在 ODPS 中，所有的请求都会被转化为任务(Task)。 F

学习笔记

容易忘的点

安骑士
态势感知
绿网

讲师试卷 1

一、单选题（共 50 题，每题 1 分，共 50 分）

1. 某电商使用 ADS 进行数据分析，其中商品信息表记录了商品的型号、价格、厂家等信息，从数据分析角度来讲，它是一张维表，该表包含了超过 2000 万条的纪录，大小在 5G 左右，在 ADS 中该如何处理该表？ C

- A、建表时指定为 ADS 的维表，将分区数设置为 1
- B、建表时指定为 ADS 的维表，即不进行拆分，可以和任意表关联
- C、必须建成 ADS 分区表，将会有一些限制，比如不能和不同普通表组上的表进行关联等
- D、可将该表直接打开到和该表关联的事实表中，做成一张宽表，减少 join，提升性能

维度表

维度表与小表

大部分情况下，维度表是小表（<1000w 且 <1G），Garuda 针对维度表的支持：

- 标识为维度表，且分区数设置为 1，即不进行任何切分
- 维度表单表的查询能力（如子查询、Group-By、Order-By 等）不受限制
- 多个维度表之间可以按任意字段 JOIN（但不支持 string 类型列做为 JOIN 列）
- 维度表与事实表（分区表）也可以按任意字段 JOIN

2、DRDS 的读策略中不包含以下哪种？ B

- A、主库读
- B、主库写
- C、自定义
- D、均测

3、在 RDS 只读实例变更配置过程中，只读实例在 RDS 管理控制台中可以进行以下哪种任务？ C

- A、内外网切换
- B、更改配置
- C、查看 SQL 日志
- D、参数设置

4、Tunnel 服务的公网 end point 是？ D

- A、http://tunnel.odps.aliyun.com/api
- B、http://service.odps.aliyun.com/tunnel
- C、http://tunnel.odps.aliyun.com
- D、http://dt.odps.aliyun.com

5、从 ODPS 导入数据到 ADS 的时候，导入路径格式为？ D

- A、odps://project_name/table_name partiton_spec
- B、odps://project_name/table_name:partiton_spec

- C、odps://project_name/table_name(partiton_spec)
- D、odps://project_name/table_name/partiton_spec

6、ODPS 中，以下哪种授权方式不能对角色授权？ D

- A、ACL
- B、Policy
- C、package
- D、label

```
odps@ hzjtdsj_odps>set label 4 to role hzjtdsj_zt;  
FAILED: mismatched input 'role' expecting 'USER' or 'TABLE'
```

7、以下哪种对 RDS 的描述是正确的？ A

- A、一个实例可以创建多个数据库
- B、在实例内数据库命名可重复
- C、数据库不会共享该实例下的资源
- D、以上都不对

8、ORDS 实例名称不能以什么开头？ D

- A、大写字母
- B、下划线
- C、小写字母
- D、中文

9、小王正在使用 ODPS 进行自然语言分析，他从门户网站上搜集了 1000 篇新闻报道，然后每篇存成一条纪录，然后基于这些纪录去做分词，即将每条纪录中的内容拆成很多独立的词语，保存到另一张表中，每个词语一条记录。小王对自定义函数比较熟悉，他想通过编写自定义函数实现分词功能。请帮小王做一下选择，从功能角度来看，哪种自定义函数更适合处理这种场景？ B

- A、User Defined Scalar Function
- B、User Defined Table Valued Function
- C、User Defined Splitted Function
- D、User Defined Aggregation Function

10、在 ADS 中，为了更高效的进行表关联，两个事实表进行 Join 的必要条件不包括以下哪些项？

- A
- A、这两个表的 join key 至少有一列建立了主键
- B、这两个表在一个表组
- C、这两个表的 join key 是 hash 分区列
- D、两个表的 Hash 分区数必须一致

11、关于 DDoS 高防 IP 的功能，描述错误的是？ D

- A、可随时更换防护 IP
- B、提供实时精准的流量报表及攻击详情，让您及时准确获得当前服务详情
- C、防护多种 DDoS 类型攻击，包括但不限于以下攻击类型 ICMP Flood、UDP Flood、TCP Flood、SYN Flood、ACK flood 等
- D、DDoS 防护阈值弹性调整，您可以随时升级更高级别的防护，调整过程服务中断时间小于等于

30 秒

!!!12、以下哪项不是 ODPS MR 的必须组成阶段？ D

- A、Map
- B、Shuffle
- C、Reduce
- D、Combiner

13、关于 RDS For MySQL 的单表尺寸限制，描述错误的是？ B

- A、单一实例下所有表的空间加起来不能超过 2TB
- B、如果实例下有 2 张以上的表，每张表的空间都可以至 2TB
- C、单表的有效最大表尺寸通常受限于操作系统的文件尺寸限制
- D、单表尺寸最大为 2TB

14、哪些监控信息不能从 OTS 的用户管理控制台直接获得？ D

- A、读写吞吐量
- B、平均返回延迟值
- C、QPS
- D、TPS

!!!15、负责轮询 Instance 执行状态的 ODPS 模块是？ B

- A、controller
- B、Worker
- C、Executor
- D、Scheduler

备注：来自 odps 权威指南

Scheduler:负责更新 OTS 中的 instance 状况，包括 running 和 terminated

execute 会查询 task 的执行状态，当 task 执行完成时，它会更新 OTS 中的 task 信息，并把 task 信息汇报给 scheduler。

work:判断该请求需要需要启动 fuxi job,则生成一个作业实例(instance), 发送给 scheduler; scheduler 会把该信息注册到元数据库 OTS 中,把状态设置为 running,worker 把 InstanceId 返回给客户端。

客户端接收到返回的 Instanceid 后，就可以通过该 InstanceId 查询作业的执行状态。

客户端会发送另外一个 rest 请求，查询状态。http server 在账号验证通过后，把请求发送给 worker,work 则根据 InstanceId 查询 OTS 中该作业的执行状态，返回结果。

16、ADS 的普通表最多包含多少个列？ D

- A、512
- B、2048
- C、4096
- D、1024

注解：

(1) 一张事实表至少有一级 Hash 分区并且分区数不能小于 8 个； (2) 一个事实表组最多可以创建 256 个事实表； (3) 一个事实表最多不能超过 1024 个列。



!!!17、关于 RDS 只读实例的限制，说法正确的是？ B

- A、只读实例支持通过备份文件或任意时间点创建临时实例
- B、绑定有只读实例的主实例不支持通过备份集直接覆盖实例
- C、支持创建\删除数据库
- D、支持通过备份集覆盖实例

备注：只读实例有以下功能限制：

- 1 个主实例最多可以创建 5 个只读实例
- 备份设置：不支持备份设置以及临时备份
- 数据迁移：不支持数据迁移至只读实例
- 数据库管理：不支持创建和删除数据库
- 账号管理：不支持创建和删除账号，不支持为账号授权以及修改账号密码功能
- 实例恢复：不支持通过备份文件或任意时间点创建临时实例，不支持通过备份集覆盖实例
- 创建只读实例后，主实例将不支持通过备份集直接覆盖实例来恢复数据

18、渗透测试是通过模拟什么的攻击方法，来评估计算机网络系统安全的一种评估方法？ C

- A、高并发客户
- B、网络专家
- C、恶意黑客
- D、正常客户

解释：渗透测试是通过模拟恶意黑客的攻击方法，来评估计算机网络系统安全的一种评估方法。

19、DRDS 中的表 T 的分库分表键分别为 col1 和 col2，则下列语句中哪个不会进行全表扫描？ D

- A、select* from T where col1=1 or col2=2;
- B、select* from T where col2=2;
- C、select* from T where col1=1;
- D、select* from T where col1=1 and col2=2;

20、关于云盾态势感知的优势，描述错误的是？ B

- A、免安装，免维护 无需安装任何软件，部署任何设备 无需复杂配置和耿鑫，省心省力 安全管控从未如此简单
- B、IT 资产划分，软件版本管理 基于威胁情报的安全事件管理对入侵行为造成的损失进行评估
- C、建设安全体系 从黑客角度监控云上业务安全 对异常行为进行实时告警 让安全“可见、可控、可管”
- D、是一个针对阿里云服务器的管家式服务

备注：大数据实时分析，还原安全事件全过程，对攻击者进行取证和追溯

21、在服务器上手动安装云盾客户端安骑士时需要什么权限？ A

- A、系统管理员权限
- B、FTP 服务权限
- C、数据库权限
- D、系统普通用户权限

登录阿里云帐号，进入服务器安全(安骑士)控制台-安装 Agent 页面，根据页面提示获取最新版本下载地址，以管理员权限在服务器上运行并安装，[点击进入安装页面](#)。

22、用户在 RDS API 调用 ImportDataForSQLServer 进行外部数据迁移操作，RDS 下达任务后，实例状态由“使用中”变为什么状态，表示 API 调用任务执行结束？ B

- A、重启中
- B、恢复中
- C、删除中
- D、创建中

解释：调用本接口，进行外部数据迁移操作，RDS 下达任务，实例状态由“使用中”变为“恢复中”，任务执行结束，实例正常运行

23、某地区级网络提供商使用 ODPS 存储、分析宽带用户的流量信息，用户每登陆一次产生一条纪录，日均记录数在 2000 万条左右，目前最常见为基于区县（该市共有 12 个区县）、日期（天）的各种查询，其中部分查询只带有区县信息。架构师在设计用户登陆信息表时，决定建成分区表，对于分区列的选取，以下性能最好的是哪一项？ B

- A、以字符串型日期信息（天）为分区别
- B、以字段组合（字符串型区县信息，字符串型日期信息（天））为分区列
- C、以字符串型区县信息为分区列
- D、以字符串型日期信息（天）为第一级分区列，以字符串型区县为第二级分区列

24、下列不属于 ODPS 项目空间的对象类型为？ A

- A、Jar 包
- B、资源
- C、表
- D、实例

25、创建 DRDS 库的时候，每个挂载的 RDS 实例默认创建的库的个数为几个？ B

- A、16
- B、8
- C、4
- D、2

26、下面哪一项不是安骑士包含的功能？ D

- A、高危漏洞修复
- B、木马文件检查
- C、异地登陆报警
- D、防 WEB 应用系统密码破解

27、ODPS 执行某个 SQL 时返回信息中包括“J3_1_2_Stg1”，以下说法中正确的是？ D

- A、当前 Task（ID 为 3）依赖于 ID 为 1 的任务，该 Task 的 Worker count 为 2
- B、当前 Task（ID 为 3）依赖于 ID 为 1 的任务，同时触发 ID 为 2 的任务
- C、当前 Task（ID 为 3）有 2 个 Reducer，1 个 mapper
- D、当前 Task（ID 为 3）依赖于 ID 为 1 和 ID 为 2 的两个 Task

28、tbl_p 为一个 ODPS 的分区表，包含一个 “p1=’1” 的分区，现在想把该分区上的数据删除，达到 select* from tbl where p1=’1’查询时，返回结果为空的目的。以下哪个 SQL 可以实现这个功能？ A

- A、alter table tbl drop partition (p1=’1’);
- B、delete from table tbl;
- C、insert into table tbl select * from tbl where 1=2;
- D、truncate table tbl;

29、跨站脚本攻击（Cross-site scripting，通常简称为 XSS）可被用于进行窃取隐私、钓鱼欺骗、偷取密码、传播恶意代码等攻击行为，请问 XSS 攻击主要发生在什么位置？ A

- A、WEB 客户端
- B、WEB 服务器端
- C、数据库端
- D、缓存服务器端

30、ADS 可以使用 explain 命令查看执行计划。下面说法中正确的是？ A

- A、explain 正确执行后返回信息中，第一行为逻辑计划，第二行为物理计划
- B、当用户发起一个 explain 查询到 ADS 系统后，ADS 会根据所有活动节点上的数据来分析执行计划
- C、当用户通过查询的方式，想要获取文本格式的 explain 语句后，将会得到一个 XML 格式的串
- D、逻辑计划中，TableExecutor 节点表示参与计算的信息，JoinExecutor 表示 join 的节点信息

31、ODPS SQL 的执行计划逻辑上可以被看做是一个有向图，其中边的含义是？ A

- A、执行阶段的依赖关系
- B、任务
- C、作业之间的依赖关系
- D、实例

32、一个 ODPS SQL 中最多使用多少个窗口函数？ C

- A、不限制
- B、1
- C、5
- D、10

目前在一个 MaxCompute SQL 语句中，可以使用至多 5 个窗口函数。

33、管理云盾加密服务的密钥时，必须通过哪种方式进行认证？ B

- A、阿里云官网帐号的 AK
- B、身份证（USB Key）
- C、用户名和密码
- D、用户自己上传的 SSL 证书

解释：所有的加密服务实例都需要使用用户手里的身份识别卡（USB KEY）进行管理。此卡仅客户持有，请妥善保管。

34、以下哪种不是加载数据到 ODPS 的合理方法？ C

- A、tunnel
- B、mapreduce
- C、DT task
- D、dship

35、ODPS 表 t1、t2 都是分区表，分区键是 pt，对于 SQL 语句 select t1.id from t1 left outer join t2 on t1.id=t2.id where t1.pt='20140101' and t2.pt='20140101'，哪一个说法是正确的？ A

- A、t1 表会进行分区剪裁
- B、t2 表会进行分区剪裁
- C、两张表都不会进行分区剪裁
- D、两张表都会进行分区剪裁

36、某借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，每天处理的数据在 20P 左右，基于这些数据对用户信用进行建模并产生预测数据，用户线上申请贷款时，能迅速根据用户模型预测结果进行资质评估、授信等。请根据你对阿里云产品的理解，下述方案中成本最低、性能最好的一个是？ B

- A、使用 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 ODPS 中
- B、使用 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 OTS 中，以供实现快速查询
- C、使用 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产评估和授信评估，将处理结果保存在 ADS 中
- D、使用 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产评估和授信评估，将处理结果保存在 OTS 中

37、某图书管理系统使用 ODPS 进行离线数据分析，现在他们需要将同一类目的书籍的名称拼成一个字符串，即将属于同一分组的字符串合并在一列展示，以下哪个内置函数可实现这一功能？ A

- A、wm_concat
- B、regexp_instr
- C、split_part
- D、concat

38、跨站脚本攻击（Cross-site scripting，通常简称为 XSS）可被用于进行窃取隐私、钓鱼欺骗、偷取密码、传播恶意代码等攻击行为，请问 XSS 攻击主要发生在什么位置？ A

- A、WEB 客户端
- B、缓存服务器端
- C、数据库端
- D、WEB 服务器端

解释：跨站脚本攻击（Cross-site scripting，通常简称为 XSS）发生在客户端，可被用于进行窃取隐私、钓鱼欺骗、偷取密码、传播恶意代码等攻击行为

39、OTS 中以什么方式来存储数据？ B

- A、二维表

- B、键值对
- C、对象
- D、聚簇列

40、ODPS 拓展 MR 的主要特点是？ B

- A、支持计算在内存中进行，中间结果不落地
- B、支持 Map 后连续任意多个 Reduce 操作
- C、可以在上一次的 reduce 结束之后，直接增加一次 map
- D、从 Map 直接到 Reduce，省去 Shuffle 阶段

41、ADS 中，对批量导入的表，哪种分区模式是可取的？ D

- A、list+list
- B、list+hash
- C、hash+hash
- D、hash+list

42、使用 OTS Java SDK 进行表操作时需要创建 OTSClient 对象，以下哪项不属于构造此对象时必须提供的信息？ C

- A、AccessKeySecret
- B、Endpoint
- C、AccountID
- D、实例名称

43、网络安全专家服务是云盾什么服务的基础上，推出的安全代维托管服务？ B

- A、基础 Ddos 防护
- B、安骑士
- C、DDoS 高防 IP
- D、阿里绿网

44、一个 SQL Server 数据库实例最多可以创建多少个数据库？ D

- A、400
- B、200
- C、60
- D、50

解释：同一实例下的数据库共享该实例下的所有资源，其中：

- MySQL 版实例最多可以创建 500 个数据库
- SQL Server 2008 R2 版实例最多可以创建 50 个数据库
- PostgreSQL 版实例和 PPAS 版实例没有数据库个数限制

45、ODPS SQL 提供了 CASE...WHEN...，可以进行条件判断。表 tbl 中包含了字段（id bigint，age bigint），目前有三条纪录，分别为（1，20），（2，30），（3，60），执行以下 SQL：

```
select id, case when age>60 then 'A' when age>40 then 'B' when age>20 then 'C' else 'D' end from tbl where id=3;
```

哪个是正确的执行结果？ A

- A、B
- B、C
- C、D
- D、A

46、表 A 为 ADS 中的普通表，列 a 为其中一个普通列，类型为 int，以下 SQL 哪些是正确的？ B

- A、SELECT a + COUNT(*) FROM A
- B、SELECT UDF_SYS_COUNT_COLUMN(a) FROM A
- C、SELECT SUM(COUNT(*)) FROM A
- D、以上都正确

47、ADS 中创建事实表（普通表）时会有一些限制，下列哪个说法是错误的？ A

- A、二级分区只能是 Hash 或者 List
- B、一个事实表最多不能超过 1024 列
- C、表的名字不能超过 32 个字符
- D、至少有一级 Hash 分区

48、某企业使用 ADS 分析数据，其中企业商品库存表（stock_detail）为一张超过 20 亿条纪录的事实表，描述了企业目前为止的库存情况，事实列仅包含库存数一列。商品信息表（time_detail）描述了商品的具体信息，包括商品信息标识、商品基本信息标识、商品通用信息属性等等 20 个属性信息，纪录条数在 10 万条左右，本表中的商品基本信息标识引用另一张存在多个层级维度的维表商品基本信息表（basic_detail，包含 30 个属性）。以下设计方案中，综合成本和性能两个角度来看，哪项是不合理的？ B

- A、企业商品库存表设计为分区事实表，分区数选择 100 个
- B、将商品信息表打平到企业商品库存表，形成一张宽表，使用时可以减少 join
- C、商品信息表适合做成维度表，数据小，不需要建分区表
- D、将商品信息表和商品基本信息表关联取得相关信息后，打平为单张宽维度表解释：
https://help.aliyun.com/document_detail/26425.html

49、ODPS 中哪种操作会使授权实效？ B

- A、上述授权都会失效
- B、对 acl 授权过的表重建
- C、对 policy 授权过的表重建
- D、对 package 授权过的表重建

50、有关 DDoS 攻击的描述，最准确的是？ D

- A、主要是攻击目标是数据库
- B、黑客自己够买了大量服务器，专门从事破坏活动
- C、攻击的目的是让服务超负载，然后窃取机密信息
- D、攻击的主要目的是让指定目标无法提供正常服务，甚至从互联网上消失，是目前最强大、最难防御的攻击之一

二、多选题（共 30 题，每题 1 分，共 30 分）

1、关于 ODPS 跨项目资源共享的说法正确的是？ ABC

- A、Packet 是一种跨项目空间共享数据及资源的机制，主要用于解决跨项目空间的用户授权问题
- B、添加资源时支持的对象类型不包括 Project 类型，即不能将 project 添加到 packet 里
- C、如果某个 package 还有人在使用，则该 package 的 owner 无法删除此 package
- D、添加到 Package 中的不仅仅是对象本身，还包括相应的操作权限，如不显示指定权限，则默认为读写权限

2、ADS 中可以使用 explain 查看执行计划，以下节点哪些属于逻辑计划？ ABC

- A、selectNode
- B、orderByNode
- C、JoinNode
- D、ExecutorNode

逻辑计划各个节点说明：

- SelectNode 表示这 select 中最终输出表达式的相关信息，例如 select 要输出的表达式集合
- GroupNode 表示 GroupBy 语句的相关信息，例如 groupby 的列，having 的表达式等
- OrderByNode 表示 OrderBy 的列信息，例如列名，顺序等。
- JoinNode 表示逻辑 Join 树的信息，例如 join 的 on 条件
- TableNode 表示分区表的信息，例如参与计算的列，表名等。

3、ODPS 由以下哪几部分组成？ ABCD

- A、逻辑层
- B、客户端
- C、接入层
- D、计算层

4、ODPS 中项目普通用户想要使用 select* from 访问一张非本人创建的本项目的表 A，至少需要以下哪些权限？ CD

- A、表 A 的 describe 权限
- B、项目 P 的 AccessAllow 权限
- C、项目 P 的 createinstance 权限
- D、表 A 的 select 权限

注解：

权限描述中 Project 类型对象的 CreateTable 操作，Table 类型的 Select、Alter、Update、Drop 操作需要与 Project 对象的 CreateInstance 操作权限配合使用。单独使用上述几种权限而没有指派 CreateInstance 权限是无法完成对应操作的。这与 ODPS 的内部实现相关。类似的，Table 的 Select 权限也要与 CreateInstance 权限配合使用。使用时请注意。

5、OST API 中提供了哪些 Action 可以进行多行数据操作？ ABD

- A、BatchWriteRow
- B、BatchGetRow
- C、GetMultiRow
- D、GetRange

6、DRDS 的全局唯一 ID 有哪些主要特点？ BCD

- A、全局有序

- B、没有单点性能瓶颈
- C、全局唯一
- D、大并发获取

备注：全局唯一数字序列服务

在分布式环境下，原有的 mysql sequence 生成机制无法高效的生成全局唯一的 sequence。我们借鉴了 oracle 的 sequence 生成样例，实现了一个在 mysql 基础上的高效 sequence 生成器，能做到没有单点性能瓶颈，具备大并发获取和低延迟特性。

DRDS sequence 功能的目标只是为了保证数据的全局唯一，虽然基本上是按时间序列获取的，但并不全局有序。

!!!7、使用 ADS 可以很好的支持即席查询，对一些复杂的查询，也能在很短的时间内返回处理结果。以下关于 ADS 中查询设计的说法正确的有？ AB

- A、需要关联查询的普通表，需要放在同一个表组上
- B、如果用户的查询的条件中会指定聚集列的内容或范围，那么这样的查询性能便会有较大的提升
- C、ADS 中 join 会优先采用 hash join，如果不满足 hash join 的条件，优化器才能选取其他的关联方法
- D、出于关联方便的考虑，普通表可以放在维度表组上，这样该表和其他任意表关联都不再受表组的限制

聚集列

在创建表时，用户可以指定一列或者若干列作为聚集列。在物理上，一个分区内聚集列内容相同的数据会尽可能的分布在同样的区块内存放，因此如果用户的查询 Query 的条件中会指定聚集列的内容或范围，那么这样的查询性能便会有较大的提升。需要注意的是如果用户指定多列为聚集列，那么指定的聚集列的顺序便是比较数据是否相同的顺序。

聚集列可以在建表后进行修改，但是目前修改后下次装载数据完毕后生效。

8、ODPS 中可以通过配额管理一个项目空间中可以使用的资源的上限，包括以下哪几种资源？

BCD

- A、带宽
- B、CPU
- C、存储
- D、内存

!!9、下面的描述中哪些属于 BGP 机房的优点？ ABC

- A、使用 BGP 协议还可以使网络具有很强的扩展性可以将 IDC 网络与其他运营商互联，轻松实现单 IP 多线路，做到所有互联运营商的用户访问都很快。这个是双 IP 双线无法比拟的
- B、由于 BGP 协议本身具有冗余备份、消除环路的特点，所以当 IDC 服务商有多条 BGP 互联线路时可以实现路由的相互备份，在一条线路出现故障时路由会自动切换到其它线路
- C、服务器只需要设置一个 IP 地址，服务器的上行路由与下行路由都能选择最优的路径，所以能真正实现高速的单 IP 高速访问

D、BGP 机房的服务器都具备双机或多机冗余功能，一台服务器宕机后不影响整体的对外服务，具备更高的可用性

备注：BGP 机房的优点：

（1）服务器只需要设置一个 IP 地址，最佳访问路由是由网络上的骨干路由器根据路由跳数与其它技术指标来确定的，不会占用服务器的任何系统资源。服务器的上行路由与下行路由都能选择最优的路径，所以能真正实现高速的单 IP 高速访问。

（2）由于 BGP 协议本身具有冗余备份、消除环路的特点，所以当 IDC 服务商有多条 BGP 互联线路时可以实现路由的相互备份，在一条线路出现故障时路由会自动切换到其它线路。

（3）使用 BGP 协议还可以使网络具有很强的扩展性可以将 IDC 网络与其他运营商互联，轻松实现单 IP 多线路，做到所有互联运营商的用户访问都很快。这个是双 IP 双线无法比拟的。

10、对于 ADS 的索引说法正确的是哪一项？BCD

A、用户需要对某列进行 Hash Join 时，无论是事实表之间的 Join 还是事实表和维度表的 Join，都需要考虑手工处理索引问题

B、ADS 弱化了索引的概念，用户无需亲自为自己的数据配置索引

C、ADS 拥有高度智能的自动化索引机制，会搜集相关信息，去自动创建索引

D、一个列可以设置列属性 `disableIndex=true`，用于屏蔽 ADS 的默认列索引

11、OTS 费用从哪几个维度计量？ACD

A、外网下行流量

B、数据访问次数（万次）

C、数据存储

D、预留读写吞吐量

备注：表格存储从四个维度计量应用程序使用的资源并收取相应的费用：数据存储、预留读写吞吐量、按量读写吞吐量及外网下行流量。

备注：数据存储、行数据大小计算、预留读/写吞吐量、按量读/写吞吐量、外网下行流量

12、如何知道一个 ODPS SQL 在运行时占用了多少 instance？BD

A、根据表的大小，自己计算

B、用 `odpscmd` 提交时，从返回的状态中查看

C、使用 `ps-ef` 查看相关进程数

D、打开 `logview` 查看

13、出于效率和成本的考虑，在使用 ODPS MR 的时候，输入表最好应该只包含需要处理的字段，其余多余字段应该忽略。输入表 `wc_in` 包含 10 个字段，MR 只对其中三个字段（`c1,c2,c3`）进行了处理，以下说法正确的是？CD

A、使用包 `com.aliyun.odps.data` 中的 `TableInfo` 类中的 `setCols()` 方法指定处理的列 `c1, c2, c3` from `wc_in`；同时调用 MR 的时候输入新表 `wc_in_new`

B、MR 的输入为一个 SQL 查询：`select c1,c2,c3 from wc_in`；

C、重建一张表 `wc_in_new`，只包含三个字段（`c1,c2,c3`），加载数据：`insert into table wc_in_new select c1,c2,c3 from wc_in`；同时调用 MR 的时候

D、建一个视图 `v_wc_in`，视图仅包括相应的三个字段：`create view v_wc_in select c1,c2,c3 from wc_in`；然后把视图 `v_wc_in` 作为 MR 的输入

应该选AC吧，MR不支持视图作为输入

!! 14、云监控里的站点监控可以对目标站点服务的可用性以及响应时间进行监控。系统已经默认预置了 8 种监控类型，包括 http 监控、ping 监控、tcp 监控、udp 监控、DNS 监控、pop 监控、smtp 监控、ftp 监控，其中每种监控类型里面包含了两个监控项，分别是_____和_____

CD

- A、响应内容长度：length
- B、请求时间：requesttime
- C、状态码：status
- D、响应时间：responsetime

解释：https://help.aliyun.com/document_detail/28619.html

15、在使用 ODPS tunnel 命令 upload 数据的时候，关于一些常见参数，说法正确的包括以下哪些项？ ACD

- A、-c 本地数据文件编码，缺省为 utf8
- B、-dfp DateTime 类型数据格式，缺省为 yyyyMMddHHmmss 默认是：yyyy-mm-dd hh:mm:ss
- C、-cp 指定是否在本地压缩后再上传，减少网络流量，缺省为 True
- D、-fd 本地数据文件的列分隔符，缺省为逗号

16、关于 ADS 多值列的说法正确的包括哪些？ AD

- A、可以使用 in，contains 条件对该列的单个值进行查询
- B、可以直接在 select 中使用该列
- C、不能在 group by 中使用该列
- D、可以存入 string 类型的多个值，以逗号分隔

17、为了保证应用获得可预期的性能，需要设置 OTS 表的：BD

- A、预留处理内存
- B、预留读吞吐量
- C、预留存储空间
- D、预留写吞吐量

18、以下哪些功能是 RDS 服务天生就具备的？ BCD

- A、自动备份，支持 15 天内任意时刻数据恢复
- B、后台有专业团队提供运维服务
- C、数据库安全漏洞实时修复
- D、数据库主备备份

备注：7 天

!!! 19、ADS 的 DMS 图形界面提供了哪些性能相关的信息？ ABCD

- A、QPS
- B、访问量
- C、磁盘使用率
- D、平均响应时间

20、小王开发了一个 ODPS UDF，打成 jar 包之后，作为资源上传到项目空间里，创建了对应的函

数，但是在执行的时候发现该 jar 包依赖的另一个 jar 包不存在，他应该怎么办才能正常使用？

CD

A、在使用 jar 命令调用的时候，在 classpath 参数中增加依赖 jar 包的路径即可

B、将依赖的 jar 包作为资源上传到项目空间，在 UDF 的初始化函数 setup 中，使用 readJarResource 读取 jar 包资源

C、将依赖的 jar 包作为资源上传到项目空间，同时在创建函数的时候（CREATE FUNCTION），using 子句中增加依赖的包，如果多个包的话，用逗号分隔

D、在对 UDF 打 jar 包的时候，把依赖的 jar 包包含进去，统一打成一个包即可

21、云数据库（RDS）具有以下哪些优势？ ABCD

A、高可用性，采用主从热备的架构

B、易用性，控制台自助迁移和管理

C、高安全性，具有网路防护能力和 SQL 审计

D、高性能，底层使用高端服务器

22、ADS 的数据导入 job 提交后，可以通过以下哪些方法查询当前的导入状态？ CD

A、使用 log 命令：log

B、使用命令：show status for

C、使用 SQL：select state from information_schema.current_job where job_id=

D、使用 DMS，通过菜单项查看：导入导出=>导入状态

23、ODPS Task 类型包括以下哪些？ ABC

A、JoinTask

B、MapTask

C、ReduceTask

D、LocalWork

!!24、哪些对象的操作不需要通过 ODPS 的任务（Task）来完成？ ABCD

A、project

B、UDF

C、instance

D、resource

- 在 MaxCompute 中，并不是所有的请求都会被转化为任务(Task)，例如：[项目空间\(Project\)](#)、[资源\(Resource\)](#)、[自定义函数\(UDF\)](#)及[实例\(Instance\)](#)的操作就不需要通过 MaxCompute 的任务来完成。

!!!25、当 RDS 实例所占用的磁容量超过了所购买的最大磁盘容量时，RDS 实例会被系统锁定。实例被锁定后，用户只有以下哪些权限？ BCD

A、insert

B、drop table

C、select

D、show view

26、ADS 中，同一个表组的表具有以下哪些相同的属性？ BC

- A、二级 List 分区数
- B、一级 hash 分区数
- C、副本数
- D、纪录条数

!!27、当用户的数据库 RDS 实例只能实现下面哪种参数的操作的时候，表明数据库被锁定了？ AB

- A、select
- B、drop
- C、insert
- D、update

注释：如果您的数据库实例忽然只能 select 和 drop，不能做 insert、update、delete 操作，则表明您的数据库被锁定了。

如果您的数据库大小超过这种实例规格的最大允许存储空间的话，则您的数据库实例就会锁住。您的实例被锁定后，您只有 select、show view 和 drop table 的权限。出现这种情况后，建议您尽快升级实例规格或者单独购买空间，当然您也可以删除一些无用的表格以释放空间，只要空间足够后系统会自动解锁。

???28、DRDS 做平滑扩容的时候，现有的数据库中的表需要满足以下哪些要求？ AC

- A、拆分表需要有主键
- B、拆分表需要删除主键
- C、单表需要做小表广播
- D、单表需有主键

29、在 RDS 只读实例变更配置过程中，该只读实例的实例状态为“升降级中”时。只读实例在控制台无法进行以下哪些任务？ ABD

- A、内外网切换
- B、参数设置
- C、查看 SQL 慢日志
- D、重启实例

!!!!30、OPDS 通过 eclipse plugin 提供了 MR 的编程框架，开发时通常需要添加以下哪几个类？

ABD

- A、Reducer
- B、Combiner
- C、MapReduce Driver
- D、Mapper

```
job.setMapperClass(TokenizerMapper.class);
    job.setCombinerClass(SumCombiner.class);
    job.setReducerClass(SumReducer.class);
```

三、判断题（共 20 题，每题 1 分，共 20 分）

1、将数据定时增量加载到 ADS 表中，如果目标表的属性是批量更新，则该表需要保证有二级分区且第二级分区必须是 list 分区，才能保证新增数据不会覆盖原有历史数据 是

2、在 ADS 中，能够实时插入的表一定要定义主键或组合主键，在进行数据插入时，若发现同主键的数据时，默认执行覆盖行为 是

解释：在分析型数据库中，能够实时插入的表一定要定义主键或组合主键，和 MySQL 有一个非常大的不同是，分析型数据库在进行数据插入时，若发现同主键的数据时，默认执行覆盖行为。若使用 insert ignore 语法，则在发现有同主键的数据时，丢弃新插入的数据，保留原有数据。但是无论如何，在主键冲突时分析型数据库无法直接返回错误给 insert 的执行方。需要注意的是，updateType=realtime 暨为实时更新表时，必须指定合法的主键并且不能有二级分区。

3、在 ODPS 中，所有的请求都会被转化为任务（Task） 否

4、预付费的 RDS 实例可以在不需要的时候主动释放，以节约成本 否

5、用户所购买的 RDS 实例不会受其他用户实例的影响。 是

6、RDS 的实例是具备自动主备的，不需要用户额外购买或者配置即可自动完成。是

7、ADS 的维度表可以和任意表组的任意表进行关联，并且创建时不需要配置分区信息。是

8、OTS 保证数据写入强一致，并能保证事务的最终一致性。 否

解释：原子性、强一致

9、当添加的资源为分区表时，ODPS 仅将整张表作为资源，不能将分区视为资源。 否

10、加入云盾先知计划的企业可以自主设定奖励计划，奖励额度没有要求。 是

11、OTS 的 putrow 操作会新插入一行。如果该行内容已经存在，则会报错返回。 否

- 解释：condition 在数据写入前是否进行行存在性检查，可以取下面三个值：
 - IGNORE 表示不做行存在性检查。
 - EXPECT_EXIST 表示期望行存在。
 - EXPECT_NOT_EXIST 表示期望行不存在。
- 若期待该行不存在但该行已存在，则会插入失败，返回错误；反之亦然。

12、DRDS 目前支持强一致分布式事务 否

解释：最终一致性

13、ODPS 基于标签的安全（LabelSecurity）是项目空间级别的一种自主访问控制机制（DAC） 否

14、ADS 的普通表，通常把需进行 Join 的列作为一级 Hash 分区列，把日期作为二级 Range 分区列，这样可以进行大表 Join，又可以每天进行增量数据导入。 否

15、云盾安全网络如果使用“域名接入”的方法，需要 DNS 服务器支持智能解析。 否

16、对于 ADS 的多值列，不允许在不经由 WHERE 子句中用 IN/CONTAINS 进行枚举筛选的情况信下直接在 SELECT/GROUP BY 中使用。 是

17、ODPS 可以使用 Alter 命令修改分区列的列名，修改后对分区没有影响。 否

18、ODPS 的客户端工具 odpscmd 中可以使用 show functions 查看当前 ODPS 项目空间中所有的自定义函数。 否

19、某大型系统使用了多个云服务器，并将所有服务器的系统日志存储到 OTS 中。考虑到当天日志的访问量比较大，之前的历史日志访问量很少，则可以将日志表分成两个，一个用来保存当日日志，设置较高的预留读写存储量，另一个表保存之前的历史日志，预留读写吞吐量设置比较低的值，并定时将当前日志表中的数据归档到历史表中，这样可以在不影响应用的情况下，节省开支。 是

20、当一个用户被从某个 ODPS 项目空间移除后，Policy 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 授权访问限制将被重新激活。 否

讲师试卷 2

一、单选题（共 50 题，每题 1 分，共 50 分）

1、ODPS SQL 目前最多支持多少个并发 insert overwrite / into 操作？ D

- A、1000
- B、500
- C、64
- D、128

2、OTS 的一些表中存在冷数据和热数据的区分，所谓冷数据是指根据时效性判断被查询的可能性不大的数据，反之为热数据。下面关于冷热数据的说法，错误的是？ A

- A、在同一张表可以对不同的分片设置不同的预留读写吞吐量，对于冷数据所在的片，设置较低的预留读写吞吐量
- B、可以用不同的表来区分冷热数据，并设置不同的预留读写吞吐量
- C、表中存在大量冷数据会导致数据访问压力不均匀，从而导致表上配置的预留读写吞吐量无法被充分利用
- D、对冷数据对应的表，可以设置较小的预留读写吞吐量

3、如果在非阿里云的服务器安装了安骑士，如何在管理控制台查看报告？ A

- A、在安装安骑士的过程中需要输入“安装验证 key”，通过此 key 关联用户的管理控制台
- B、通过阿里云官网帐号关联
- C、报告无法体验在阿里云的管理控制台
- D、非阿里云服务器不支持安骑士

备注：对于非阿里云服务器，在安装过程中会提示输入验证 Key，这个验证 Key 用于关联阿里

云账号，通过阿里云账号在安骑士控制台使用相关功能，验证 key 会显示在[安装页面中](#)。

4、ADS 可以使用 explain 命令查看执行计划。下面说法中正确的是？ D

- A、当用户发起一个 explain 查询到 ADS 系统后，ADS 会根据所有活跃节点上的数据来分析执行计划
- B、当用户通过查询的方式，想要获取文本格式的 explain 语句后，将会得到一个 XML 格式的串
- C、逻辑计划中，TableExecutor 节点表示参与计算的信息，JoinExecutor 表示 join 的节点信息
- D、explain 正确执行后返回信息中，第一行为逻辑计划，第二行为物理计划

5、如果我的 ECS 到期释放了，服务器安全托管服务还未到期，正确的处理方法是？ D

- A、既然 ECS 已经到期了，服务器安全托管服务自然也就过期作废
- B、提交工单投诉，把已释放的 ECS 恢复回来
- C、只能申请退款，无别选择
- D、若 ECS 到期释放了，可以通过工单联系我们，重新授权一台新购 ECS，新购 ECS 可以继续使用安全托管服务，服务截止日期不变

6、ADS 的数据库名为 movie_db，该数据库管理员的阿里云帐号为 alice@aliyun.com，密码为 Alice123，accessID 为 aaaa，AccessKey 为 bbbbbb，连接的服务器地址为 movie_db.cn-hangzhou-ads.aliyuncs.com，端口号为 3456，可以使用 mysql 客户端命令连接。下列哪项是正确的？ A

- A、mysql-h movie_db.cn-hangzhou-ads.aliyuncs.com-p3456-uaaaa-pbbbbbb-D movie_db
- B、mysql-h movie_db.cn-hangzhou-ads.aliyuncs.com-p3456-ualice-PAlice-D movie_db
- C、mysql-h movie_db.cn-hangzhou-ads.aliyuncs.com-p3456-uaaaa-Pbbbbbb-D movie_db
- D、mysql-h movie_db.cn-hangzhou-ads.aliyuncs.com-p3456-uALIYUN\$alice@aliyun.com-pAlice123-D movie_db

7、云盾 DDoS 基础防护功能如何开启？ B

- A、提交工单，24 小时之内开通
- B、自动开通
- C、联系客户经理线下开通
- D、需要从管理控制台申请开通

8、DRDS 默认每个实例创建几个数据库？ A

- A、8
- B、16
- C、64
- D、1

9、DRDS 的产品定位不包括以下哪项？ B

- A、基于 RDS 的分布式数据存储和检索产品
- B、海量数据的多维分析
- C、降低用户使用分布式数据库的难度
- D、解决用户单 RDS 无法支撑业务的困难

10、ODPS 中，当一个用户被移除后，与该用户有关的哪些授权会保留？ A

- A、ACL 授权
- B、policy 授权
- C、项目空间授权
- D、package 授权

11、渗透测试是通过模拟什么的攻击方法，来评估计算机网络系统安全的一种评估方法？ C

- A、正常客户
- B、高并发客户
- C、恶意黑客
- D、网络专家

12、如果用户执通过 RDS API 的 CreateDBInstance 接口创建实例，则创建的 RDS 实例类型是什么？

- D
- A、以上都可以实例
- B、只读实例
- C、临时实例
- D、主实例

只读实例是：创建 RDS 只读实例

13、ODPS 的命令行工具 odpscmd 可以使用以下哪个参数调用某个文件中的命令集？ C

- A、k
- B、e
- C、f
- D、c

解释：odpscmd -f xxx.sql

14、DRDS 的服务可用性不低于百分之多少？

- A、100
- B、100
- C、100
- D、100

服务可用性：不低于 99.95%

15、通过 DMS 管理后台或者数据库客户端，连接 RDS 数据库时，提示错误信息“max_user_connections”，代表什么含义？ D

- A、IOPS 超出极限
- B、RDS 空间满了
- C、网络中断
- D、RDS 数据库的连接数满了

16、关于 ODPS 中的任务（Instance）、作业（Job）、作业实例（Instance）以及工作流（Workflow）描述错误的是？ C

- A、Job 是个静态的概念，被提交到系统开始执行时，该 Job 就拥有一个 Instance
- B、工作流描述了 Job 中各个 Instance 之间的依赖关系和运行约束，本质是个 DAG（有向无环图）
- C、单个 SQL query、命令或者 MapReduce 程序统称为一个 Task

D、Instance 是个动态概念，每个 Instance 只能运行一次

17、OTS 的列对长度有限制。对于 string 类型，如果出现在主键里，长度限制为多少？ B

- A、64k
- B、1k
- C、256k
- D、128k

解释：主键是表中每一行的唯一标识。主键由 1 到 4 个主键列组成。应用在创建表的时候，必须明确指定主键的组成，每一个主键列的名字和数据类型以及它们的顺序。属性列的数据类型只能是 String、Integer 和 Binary。如果为 String 或者 Binary 类型，长度不超过 1KB。

18、关于 ODPS 类型之间的类型转化，说法不正确的是？ B

- A、隐式类型转换是 ODPS 依据上下文使用环境自动进行的类型转换
- B、部分类型之间不可以通过显式的类型转换，但可以通过 SQL 的隐式转换完成 显示和yinshi是一致的
- C、不支持的隐式类型转换会导致异常
- D、ODPS 支持的隐式类型转换规则与显式转换相同

19、某图书管理系统使用 ODPS 进行离线数据分析，现在他们要将同一类目的书籍的名称拼成一个字符串，即将属于同一分组的字符串合并在一列展示，以下哪个内置函数可以实现这一功能？

- A
- A、wm_concat
- B、split_part
- C、concat
- D、regexp_instr

20、用户通过 RDS API 能查看可以使用的物理位置，在调用创建的 RDS 实例接口之前，需要调用“查询可用的物理位置”接口进行查询，以下哪个字段信息代表机房的物理区域？ A

- A、RegionID
- B、ZoneID
- C、InstanceID
- D、AccessKeyID

21、ODPS 对 SQL DML 语句进行解析后，得到执行计划。构成执行计划的是由多个有相互依赖关系的__A__组成？

- A、执行阶段（Stage）
- B、作业（Job）
- C、实例（Instance）
- D、任务（Task）

22、在 ADS 中，什么是用户所关心的最大单元，也是用户和 ADS 系统管理员的管理职权的分界点？ C

- A、表组
- B、表

- C、数据库
- D、列

23、某地市级网络提供商使用 ODPS 存储、分析宽带用户的流量信息，用户每登陆一次产生一条纪录，日均记录数在 2000 万条左右，目前最常见需求基于区县（该市共有 12 个区县）、日期（天）各种查询，其中部分查询只带有区县信息。架构师在设计用户登陆信息表时，决定建成分区表，对于分区列的选取，以下性能最好的是哪一项？D

- A、以字符串型日期信息（天）为第一级分区列，以字符串型区县为第二级分区列
- B、以字符串型区县信息为分区列
- C、以字符串型日期信息（天）为分区列
- D、以字段组合（字符串型区县信息，字符串型日期信息（天）为分区列）

24、某电商使用 ADS 进行数据分析，其中商品信息表纪录了商品的型号、价格、厂家等等信息，从数据分析角度来讲，它是一张维表。该表包含了超过 2000 万条的纪录，大小在 5G 左右，在 ADS 中如何处理该表？A

- A、必须建成 ADS 分区表，将会有一些限制，比如不能和不同普通表组上的表进行关联等
- B、建表时指定为 ADS 的维表，将分区数设置为 1
- C、建表时指定为 ADS 的维表，即不进行拆分，可以和任意表关联
- D、可将该表直接打平到和该表关联的事实表中，做成一张宽表，减少 join，提升性能

25、ODPS 表 t1、t2 都是分区表，分区键是 pt，对于 SQL 语句 select t1.id from t1 left outer join t2 on t1.id=t2.id where t1.pt='20140101' and t2.pt='20140101'，哪一个说法是正确的？C

- A、t2 表会进行分区剪裁
- B、两张表都会进行分区剪裁
- C、t1 表会进行分区剪裁
- D、两张表都不会进行分区剪裁

26、关于 ADS 的聚集列，下列说法不正确的是哪一项？B

- A、如果用户的查询 Query 的条件中会指定聚集列的内容或范围，那么这样的查询性能便会有较大的提升
- B、聚集列可以在建表后进行修改，修改后 DB 会在后台自动刷新数据
- C、一个分区内聚集列内容相同的数据会尽可能的分布在同样的区块内存
- D、创建表时可以指定一列或者若干列作为聚集列

备注：聚集列

在创建表时，用户可以指定一列或者若干列作为聚集列。在物理上，一个分区内聚集列内容相同的数据会尽可能的分布在同样的区块内存放，因此如果用户的查询 Query 的条件中会指定聚集列的内容或范围，那么这样的查询性能便会有较大的提升。需要注意的是如果用户指定多列为聚集列，那么指定的聚集列的顺序便是比较数据是否相同的顺序。

聚集列可以在建表后进行修改，但是目前修改后下次装载数据完毕后生效。

27、某企业使用 ADS 做数据分析，其中部分数据来源于 ODPS。技术人员在使用 LOAD DATA 命令从 ODPS 加载数据到 ADS 的过程中，碰到如下错误信息：

ERROR 1105 (HY000): You are not the owner of the source table

最有可能的出错原因是哪种？ A

- A、ODPS 的用户 garuda_build@aliyun.com 没有对源表的操作权限
- B、ADS 的用户 garuda_build@aliyun.com 没有对源表的操作权限
- C、源表不存在
- D、源表中没有数据，为空表

28、云盾网络安全专家服务特别适合以下哪种场景？ C

- A、系统每次重大升级后不清楚有没有漏洞
- B、够买了 100 台 ECS，但没有专业的管理同
- C、每月一次的产品促销活动，担心网络攻击
- D、频繁出现 SQL 注入警告

29、基于 RDS 系统架构以及只读实例的应用场景，需要满足 MySQL 版本最低为多少，否则是不能满足只读实例的应用场景？ A

- A、MySQL 5.6
- B、MySQL 5.5
- C、MySQL 5.1
- D、MySQL 5.2

30、ODPS 中一个表可以上限最多拥有多少个分区？ C

- A、10000
- B、不限制
- C、60000
- D、100000

31、作为技术人员，小王被 ODPS 强大的数据处理能力吸引，想去体验一下。他快速注册了一个阿里云帐号，进入官网后试图创建一个名字为 test_project 的 ODPS project，结果没有成功。分析，小王出错的最有可能的原因是哪一个？ D

- A、帐号未充值，余额不足导致出错
- B、帐号未开通 ODPS 的使用权限
- C、未创建合适的 accessID 和 accessKey
- D、项目名字和现有的其他帐户的名字重复

32、表 t 中有 1000 万行数据，想要随意抽取 10 行数据，在 odpscmd 中执行消耗时间最短的方式是哪种？ C

- A、select * from t where rownum<=10
- B、select * from t limit 10
- C、read t 10
- D、get t 10

33、关于 ADS 的 ECU，说法正确的是？ D

- A、可以动态的扩容或者缩容，两者都是瞬间的同步操作
- B、通过参数设置，可以使同一个 ADS 的库同时使用两种类型的 ECU
- C、可以通过查询表 `information_schema.resource_request` 获得目前的 ECU 的状态
- D、可以通过 DDL 语句 `alter database set ecu_count=N` 来修改 ECU 的个数

!!34、ADS 中两个普通表关联，对于 join key，哪个说法是不正确的？ D

- A、如果两张表的 join key 都没有 hashmap 索引，则会报错，修正方法为：至少一个 join key 上手工添加 hashmap 索引，且修改过的表需要重新加载数据，重新进行关联
- B、两张表的 Hash 分区数必须一致，且分区列一致
- C、两张表的 Join Key 至少有一列建立了 HashMap 索引
- D、如果选择只在一张表上建 HashMap 索引的话，推荐建立在数据量较大表的一侧

备注：为了更高效的进行表关联，分析型数据库对表关联操作在分析型数据库中，两个事实表进行 Join 的充要条件是：(1) 这两个表在一个表组；(2) 这两个表的 Join Key 是 Hash 分区列；(3) 两张表的 Hash 分区数必须一致，否则 Join 结果不准确；(4) 两张表的 Join Key 至少有一列建立了 HashMap 索引，推荐建立在数据量较小的一侧。

35、跨站脚本攻击（Cross-site scripting，通常简称为 XSS）可被用于进行窃取隐私、钓鱼欺骗、偷取密码、传播恶意代码等攻击行为，请问 XSS 攻击主要发生在什么位置？ C

- A、WEB 服务器端
- B、数据库端
- C、WEB 客户端
- D、缓存服务器端

36、在 project 中已经开启了 label security（列级别授权）的前提下，用户申请了一张表 tbl 的读权限，后来管理员将该表中的字段 x 的安全等级设置为了 4，且用户的安全级别为 1，用户能否读该表中的数据？ C

- A、提供信息不足以判断
- B、可以，因为之前已经有读权限了
- C、可以访问除了字段 x 之外的数据
- D、不可以，因为用户的级别较低

37、tbl_p 为一个 ODPS 的分区表，包含一个 `p1='1'` 的分区，现在想把该分区上的数据删除，达到使用 `select * from tbl where p1='1'` 查询时，返回结果为空的目的。以下哪个 SQL 可以实现该功能？ B

- A、`truncate table tbl;`
- B、`alter table tbl drop partition (p1='1');`
- C、`insert into table tbl select * from tbl where 1=2;`
- D、`delete from table tbl;`

38、A 为 ADS 中的普通表，a 为 A 的分区列，类型为 bigint，b 为 A 中的普通列，类型为 bigint，请判断以下相关 SQL 中哪句语法是错误的？ B

- A、`SELECT a FROM A ORDER BY a LIMIT 100`
- B、`SELECT a+b, count(a) FROM A group by a+b ORDER BY a`

- C、SELECT a,b FROM A ORDER BY a+b LIMIT 100
- D、SELECT a, count(a) FROM A group by a ORDER BY a

0.8 版本中不支持但是 0.9 版本的 Full MPP Mode 支持（详见 4.4 节）：

- `SELECT a + b, COUNT(1) AS x FROM A GROUP BY a + b ORDER BY x` : 0.8 版本下按照

非分区列表达式 (`a+b`) 分组聚合再取 TOP-N 是非精确的（详见 4.3）

39、以下哪个不是由 ODPS 接入层提供的服务？ A

- A、用户空间管理操作
- B、HTTP 服务
- C、Load Balance
- D、用户认证

40、在 RDS 只读实例变更配置过程中，只读实例在 RDS 管理控制台中可以进行以下哪种任务？

- A
- A、查看 SQL 日志
- B、内外网切换
- C、更改配置
- D、参数设置

41、ODPS SQL 提供了 CASE...WHEN...，可以进行条件判断。表 tbl 中包含了字段 (id bigint, age bigint)，目前有三条纪录，分别为 (1, 20)，(2, 30)，(3, 60)，执行以下 SQL：

```
select id, case when age>60 then 'A' when age>40 then 'B' when age>20 then 'C' else 'D' end from tbl
where id=3;
```

哪个是正确的执行结果？ B

- A、A
- B、B
- C、C
- D、D

42、关于阿里云云监控的功能，描述错误的是？ C

- A、自定义监控，对用户开放自定义监控的服务，允许用户自定义个性化监控需求
- B、支持站点监控，包括对 http、ping、dns、tcp、udp、smtp、pop、ftp 等服务的可用性和响应时间的统计、监控、报警服务
- C、是安装在云主机 ECS 里面的一个客户端，可自动监测各项云服务是否有异常，还可以实时报警
- D、支持云服务监控，包括对 ecs、rds、slb、cdn、ocs、oss 等云服务的监控报警服务

43、OTS 种以什么方式来存储数据？ C

- A、对象
- B、二维表

- C、键值对
- D、聚簇列

44、ODPS 项目空间 prj1 中存在表 t1，owner 打开 LabelSecurity 设置后，执行如下命令：

```
SET LABEL 1 TO TABLE t1;  
SET LABEL 2 TO TABLE t1(id);  
SET LABEL 3 TO TABLE t1;
```

此时，t1 表中的列 id 的敏感等级标签是什么？ C

- A、0
- B、1
- C、2
- D、3

45、RDS 采用的主备 M-M 的高可用架构，其主备之间的数据同步依靠日志的方式实现，MySQL 数据库和 SQL Server 分别使用的日志是什么？ C

- A、均为事务日志
- B、MySQL 数据库为事务日志，SQL Server 为 BINLOG 日志
- C、MySQL 数据库为 BINLOG 日志,SQL Server 为事务日志
- D、均为 BINLOG 日志

46、关于 ODPS policy 授权，说法不正确的是？ D

- A、支持根据 ip 授权
- B、支持表名通配符授权
- C、支持授权的时间段
- D、支持根据数据敏感级授权

47、ADS 中创建事实表（普通表）时会有一些限制，下列哪个说法是错误的？ D

- A、表的名字不能超过 32 个字符
- B、至少有一级 Hash 分区
- C、一个事实表最多不能超过 1024 列
- D、二级分区只能是 Hash 或者 List

48、管理云盾加密服务的密钥时，必须通过哪种方式进行认证？ A

- A、身份证（USB Key）
- B、用户名和密码
- C、用户自己上传的 SSL 证书
- D、阿里云官网帐号的 AK

49、SQL 注入攻击就是攻击者通过欺骗什么服务器，执行非授权的任意查询过程？B

- A、文件
- B、数据库
- C、备份
- D、WEB

50、下面哪一项不是安骑士包含的功能？ C

- A、高危漏洞修复
- B、异地登陆报警
- C、防 WEB 应用系统密码破解
- D、木马文件检查

二、多选题（共 30 题，每题 1 分，共 30 分）

1、某在线系统使用的是 DRDS，其中订单表保存了用户订单信息，目前订单表的分库键为订单标识。事实上基于该订单表的常见查询中经常用到的查询条件除了订单标识外，还有用户标识，为了减少 库的扫描，保证这两类常用查询的性能，可以采取以下哪些方案？ CD

- A、如果同一个查询中只出现一个字段，则可以使用异构索引表
- B、如果同一个查询中同时出现这两个字段，则创建基于这两个字段的索引
- C、如果同一个查询中同时出现这两个字段，可调整表结构，其中一个字段作为分库键，另一个字段作为分表键
- D、如果同一个查询中只出现一个字段，则新建一张结构相同但是分库键为用户标识的表，将数据从源表加载进来，同时维护两张表的数据

2、关于 ODPS 跨项目资源共享的说法正确的是？ ACD

- A、添加资源时支持的对象类型不包括 Project 类型，即不能将 project 添加到 packet 里
- B、添加到 Package 中的不仅仅是对象本身，还包括相应的操作权限，如不显示指定权限，则默认为读写权限
- C、Packet 是一种跨项目空间共享数据及资源的机制，主要用于解决跨项目空间的用户授权问题
- D、如果某个 package 还有人在使用，则该 package 的 owner 无法删除此 package

3、以下场景中，哪些适合使用 ODPS？ BCD

- A、电子商务网站的在线交易
- B、大型互联网企业的数据仓库和 BI 分析
- C、PB 级离线日志分析
- D、基于海量数据的用户特征和兴趣挖掘

4、下面哪些是 RDS 提供的系统性能监控项目？ ABCD

- A、IOPS
- B、CPU 利用率
- C、TPS
- D、网络流量

5、OTS API 中提供了哪些 Action 可以进行多行数据操作？ ABD

- A、BatchGetRow
- B、GetRange
- C、GetMultiRow
- D、BatchWriteRow

6、ADS 提供哪些批量导入数据的方式？ AC

- A、data pipeline 系列命令
- B、通过 rest-like api 方式
- C、通过 insert
- D、通过 batchinsert 命令

7、OTS 费用从哪几个维度计量？ABC

- A、数据存储
- B、预留读写吞吐量
- C、外网下行流量
- D、数据访问次数（万次）

8、对于 OTS 表的预留读写吞吐量，以下各项中描述正确的是？A

- A、当表上的操作过于频繁导致预留读 / 写吞吐量不足以进行更多操作时，OTS 会返回 OTSQuotaExhausted 错误给应用程序
- B、1 单位读能力表示该表每秒可以读 1KB 数据
- C、1 单位写能力表示每秒可以写 1KB 数据
- D、操作数据大小不足 1KB 的部分向上取整

9、OPDS 中用户认证(Authentication)的主要功能是检查请求 (Request)发送者的真实身份，通过包括以下哪几项验证？ABC

- A、正确验证用户状态，包括是否欠费等
- B、正确验证接收到的信息在途中是否被篡改
- C、正确验证消息发送方的真实身份
- D、用户需要提取用户个性化信息时进行用户个性化认证

10、RDS 中慢 SQL 包括哪些操作？ACD

- A、全表扫描
- B、内存排序
- C、扫描数据量过大
- D、磁盘排序

11、云盾安全网络除了 DDoS 防护功能外，还具备哪些安全功能？ABD

- A、WAF 应用防火墙
- B、CC 防护
- C、网站内容检查
- D、隐藏源站

解释：一、相同点

- 1、具备防 DDoS、CC 和 WAF 功能。
- 2、支持阿里云 ECS、SLB 和非阿里云 IDC 机房服务器。

二、不同点，主要体现在 DDoS 防御方案及效果上

- 1、高防 IP 是海量带宽硬扛，支持 300G 按天付费，保证在清洗过程中对用户业务体验无损失。支持 HTTPS 业务的 WAF 防御。
- 2、安全网络是多节点切换防御，每个节点的防御带宽小，防御过程中对用户业务有一定影响，

可通过集成 SDK 把业务影响时间控制在 1 秒内。暂时不支持 HTTPS 业务的 WAF 防御。

12、在 RDS 只读实例变更配置过程中，RDS 主实例无法进行哪些任务 ACD

- A、备份实例
- B、查看 SQL 审计日志
- C、创建实例
- D、内外网切换

13、对于 ODPS 项目空间的描述，正确的是？ ACD

- A、一个用户可以拥有多个项目空间的权限
- B、所有的对象都会属于某个项目空间
- C、通过安全授权，可以在一个项目空间中访问另一个项目空间
- D、一个帐号可以创建多个项目空间，最多不能超过 10 个

!!! 14、ODPS 的控制层主要功能包括？ ABCD

- A、用户空间和对象的管理
- B、命令的解析
- C、执行逻辑
- D、数据对象的访问控制与授权

备注：来自 ODPS 权威指南

(1)用户控件(project)管理

(2)对象管理，包括对表、资源、作业的管理

(3)命令的解析和执行

(4)元数据的管理

执行逻辑：计算层

!!!15、用 Java 写 ODPS 的自定义函数时，以下哪几种类型是可以用在接口中的类型？ AC

- A、long
- B、int
- C、string
- D、bigint

对应ODPS的bigint类型

UDF 支持 ODPS SQL 的数据类型有：bigint, string, double, boolean 以及 datetime 类型。

ODPS 数据类型与 Java 类型的对应关系如下：

ODPS SQL Type	Bigint	String	Double	Boolean	Decimal
Java Type	Long	String	Double	Boolean	Decimal

16、ODPS SQL 支持的 Join 操作类型包括哪些？ ABCD

- A、LEFT OUTER JOIN
- B、RIGHT OUTER JOIN
- C、INNER JOIN

D、FULL OUTER JOIN

17、ADS 支持一下哪些连接方式？ ABCD

- A、安装了 php-mysql-5.1.x 模块的 php 环境
- B、使用版本为 5.4 系列的 mysql-jdbc 驱动
- C、用户管理控制台中的 DMS
- D、mysql5.6 的客户端

18、DRDS 全表扫描目前支持哪些聚合函数？ ABCD

- A、SUM
- B、MIN
- C、COUNT
- D、MAX

19、某猎头公司启动了一个人才管理系统，从网上搜集人才简历，保存到自己的 OTS 系统的表 t_resume 中，部分纪录包含了一些大小超过 64G 的信息，比如扫描的证书照片、个人照片等信息。作为架构师，请为这些大字段选择合适的处理方式 CD

- A、将超过 64KB 的单个值拆分成多个列存储在 OTS 中
- B、根据大字段内容生成 MD5，将 MD5 码替代原有内容保存到字段中
- C、使用 OSS 存储大字段，并将对应的 url 存入到 OTS 的表 t_resume 中
- D、将属性列压缩之后再以 Binary 类型存储到 OTS 的表 t_resume 中

20、在管理控制台的云盾配置页里可以设置哪些“报警方式”？通过这些方式，用户可以在出现风险的第一时刻收到事件通知。BCD

- A、钉钉消息
- B、实时电话
- C、手机短信
- D、电子邮箱

21、DRDS 则将整个扩容的过程分为多个阶段，包括以下哪些步骤？ AC

- A、切换数据库
- B、增量同步
- C、全量迁移
- D、重启实例

22、某企业用户目前的在线订单系统运行在 MySQL 版的 RDS 上，因为业务量激增，公司决定采用 DRDS，即将现有系统从 RDS 迁移到 DRDS 上去。从项目实施部署的角度来看，以下哪些说法是正确的？ ABCD

- A、需要搜集当前表的大小和具体如何使用等，用来判断哪些表需要拆分，哪些不需要
- B、需要评估当前应用，以识别哪些应用需要改造
- C、需要评估哪些应用不能支持，需要做应用改造
- D、需要评估存储容量以及表的大小，以合理规划初始化 DRDS 的实例和库的个数

23、连接 RDS 的 MySQL 数据库的方法有哪些？ ABC

- A、客户端：Mysql-Front
- B、Mysql 命令登陆
- C、Navicat_mysql
- D、SQL Server Management Studio

24、使用用户管理控制台创建 ADS 时，需要提供 ECU 的哪些信息？BC

- A、存储
- B、数量
- C、类型
- D、vCPU

25、对于数据库类型为 MySQL 的 RDS 实例，有以下哪些种连接方法？ABCD

- A、iDB Cloud（控制台登陆）
- B、MySQL-Front
- C、Navicat for MySQL
- D、使用 mysql 命令登陆

26、ODPS 中项目普通用户想要使用 `select * from` 访问一张非本人创建的本项目的表 A，至少需要以下哪些权限？BC

- A、项目 P 的 `select` 权限
- B、表 A 的 `select` 权限
- C、项目 P 的 `createinstance` 权限
- D、表 A 的 `decribe` 权限

27、使用 MySQL 客户端连接 ADS 时，以下哪些信息是必须的？ABC

- A、端口
- B、服务器地址
- C、数据库名称
- D、AccountID

28、对于 ADS 的特色功能，以下说法正确的是？ABC

- A、智能自动索引会自动为导入数据的每一列创建符合该列情况的索引类型，无需用户显式指定创建索引或索引类型
- B、聚集列能智能提升查询性能
- C、多值列支持 `in`、`contains` 等查询
- D、ADS 采用高职能的基于规则的优化策略

29、支持 OLTP 的数据库系统包括一下哪些特点？ABCD

- A、擅长事务处理
- B、数据原子性
- C、很好的支持频繁的数据插入和修改
- D、数据一致性

30、ODPS MR 中哪些操作会因为 Java Sandbox 而报错 “access denied”？ABCD

- A、获取本地 IP
- B、起子进程执行 Linux 命令
- C、直接访问本地文件
- D、Java 反射限制

三、判断题（共 20 题，每题 1 分，共 20 分）

- 1、ODPS 中可以使用 Tunnel 模块上传视图数据。 否 **支持分区和表**
- 2、ODPS Policy 授权文件的格式是 Json。 是
- 3、将数据定时增量加载到 ADS 表中，如果目标表的属性是批量更新，则该表需要保证有两级分区且第二级分区必须是 list 分区，才能保证新增数据不会覆盖原有历史数据。 是
- 4、如果要查看只读实例的数据延迟，用户只能通过 RDS 管理控制台进入 RDS 只读实例进行查看。 否
- 5、OTS 管理便捷，应用程序无需关心软硬件升级，配置更新，集群扩容等繁琐运维任务。 是
- 6、OTS 的 BatchWriteRow 用于插入、修改、删除一个表或者多个表中的多行纪录。 是
- 备注：批量插入，修改或删除一个或多个表中的若干行数据
- 7、ADS 中，当 updateType=realtime 时，支持 insert、delete、update 操作。 否
- 8、ADS 建表时如果选定实时写入，则可以使用 insert 插入数据，可以作为 OLTP 系统使用。 否
- 9、关系型数据库服务 RDS(Relational Database Service)是阿里云提供了一种稳定可靠、可弹性伸缩的在线数据库服务。 是
- 10、在 ODPS 中，所有的请求都会被转化为任务（Task）。 否
- 11、ODP 的 LabeSecurity 对敏感数据的粒度可以支持列级别。 是
- 12、ADS 的两张维度关联的时候，两个 join key 可以均不创建 hashmap 的索引。 否
- 13、在开发 ODPS MR 或者 UDP 的过程中，在用到文本文件、ODPS 表等的时候，只能把它们作为主方法的输入参数，由主方法读取后识别处理。 是
- 14、云监视除了可以为 aliyun.com 云服务器客户提供安全有效的监控服务外，还能够为其他自由服务器客户提供功能强大的监控服务，只是配置比较复杂，必须自己开发脚本进行数据收集。 否
- 15、目前 DRDS 可以基于 MySQL 以及 SQLServer 版的 RDS，但是不支持 PostGreSQL 版的 RDS 否
- 16、任务（task）是 ODPS 的基本计算单元。SQL 以及 MapReduce 等都是通过任务完成的。 是

17、ODPS 的分区只支持两种方式，对于数值分区键支持 mod，对于字符串类型的分区键则使用 hash 值作为分区的依据。 否

18、未经客户允许云盾先知计划不会对外公开任何漏洞标题及细节。 是

19、在 ADS 中，数据根据分区列进行分布式的存储和计算。 是

20、RDS 只读实例数据库版本必须大于等于 RDS 主实例版本，并且 RDS 主实例版本最低是 MySQL 5.6。 是

3、Data IDE 中，周期性调度任务的调度规则是：若任务有上游任务则上游任务必须执行成功，且当前时间点已经等于或超过任务的定时时间，两个条件都满足后当前任务才可以运行；若该任务没有上游任务，则定时时间满足即可运行。Task2 是按天调度的周期任务，配置定时时间 12:00，依赖属性的上游任务配置为任务 task1，task1 是按天调度的周期任务，定时时间配置为 0:00，关于 task2 每天调度情况的说法中正确的是_____。

- a. task1 执行成功后，如果时间没到 12 点，那么等到 12 点 task2 可以运行
- b. task1 执行成功后，task2 会立即运行
- c. task1 执行成功后，如果时间超过 12 点，那么 task2 任务已经无法运行
- d. task1 任务执行成功时，时间必须刚好是 12 点 task2 才能运行

4、在大数据计算服务（MaxCompute，原 ODPS）中，使用 MapReduce 开发一个分词程序，即输入为保存在表中的多篇文章（每篇文章一行纪录），参考一个字典文件（该文件会经常更新）进行分词。简而言之，在执行该程序的时候，既需要输入待分词的文章（已经保存在表中了），又需要输入字典（为经常变动的本地文件），以下最合理的解决方案为_____。

- a. 将这个字典与源代码一起进行打包，直接执行即可
- b. 每次都把字典文件作为资源上传到大数据计算服务上，然后 MR 中通过资源的方式来访问
- c. 将字典打包 jar 包，使用 jar 命令执行 MR 程序时，通过 resources 参数指定该 jar 包
- d. 将字典文件保存在本地，在 MR 中通过访问本地最新的字典文件实现

5、大数据计算服务（MaxCompute，原 ODPS）项目 prj1 中，用户 owner 创建表 T1，通过 ACL 方式将读权限给用户 U1（此前用户 U1 无项目空间 prj1 的任何权限），现在使用 ACL 授权 U1 可以通过 select* from prj1.T1 来访问 T1 中的数据。接下来 owner 将表 T1 删除，重新创建一张新表，名字仍然叫 T1，以下说法中正确的是_____。

- a. 如果新建后的 T1 和原来的表 T1 表结构一致，则 U1 仍可以访问 T1
- b. 用户 U1 仍然可以访问表 T1
- c. 用户 U1 无法访问当前表 T1
- d. 如果新建后的 T1 和原来的表 T1 表结构、数据内容均一致，则 U1 仍可以访问 T1

##在添加用户或创建角色之后，需要对用户或角色进行授权。MaxCompute 授权是一种基

于对象的授权。通过授权的权限数据(即访问控制列表, Access Control List)被看做是该对象的一种子资源。只有当对象已经存在时,才能进行授权操作;当对象被删除时,通过授权的权限数据会被自动删除。

##https://help.aliyun.com/document_detail/27935.html?spm=5176.doc27932.6.67

9.gPnGt8

6、在 Data IDE 中配置数据同步任务时,将云数据库 RDS (MySQL) 中的数据同步至大数据计算服务 (MaxCompute, 原 ODPS) 分区表对应的日期分区中,采用数据同步任务周期性的调度向不同分区中写入数据,把 yyyymmdd 的数据写入对应的 yyyy-mm-dd 分区中,如 20160308 的数据需要写入 MaxCompute 表的'2016-03-08'分区中,MaxCompute 的分区可以通过 () 实现。

- a. 在 MaxCompute 侧中使用系统内置参数\$bizdate
- b. 在 MaxCompute 侧中使用变量\${datetime},并在节点参数出赋值 datetime=\${YYYY/MM/DD}
- c. 在 MaxCompute 侧的分区中写入常量值 2016-03-18 即可
- d. 在 MaxCompute 侧中使用变量\${datetime},并在节点参数出赋值 datetime=\${YYYY-MM-DD}

7、在 Data IDE 中,某数据同步任务执行成功,但是发现目标表数据为空。以下说法正确的又 ()。(正确答案的数量: 2 个)

- a. 查看具体运行日志,看日志里具体源表信息,造成目标表为空的原因是源表不存在
- b. 查看具体运行日志,看日志里具体源表信息,经检查满足抽取条件的数据记录数为 0
- c. 运行任务帐号只有目标表的读权限却没有写权限
- d. 当配置了最大容错条数,即“当出错的纪录数超过”的值,如果源表中满足抽取条件的纪录数小于当前配置的最大容错条数的值,并且所有的纪录全部同步出错,则整个同步任务的状态为执行成功,目标表数据为空

8、某在线广告 APP 采用了阿里云的技术,通过对网上搜索的海量数据的分析,给企业主提供精准的广告投放人群。该 APP 的产品经理介绍该系统能实现以下技术指标: 1-能存储海量数据,单表大小超过 10TB 2-支持多张上亿的表进行 Join 3-支持多种筛选条件,包括连续范围筛选、取值筛选、and/or/not 条件组合等 4-支持快速 (3 秒以内) 对数十个维度分别进行透视分析。根据以上信息,请判断该 APP 最有可能使用的是阿里云的哪一款产品?

- a. 分布式关系型数据库 (DRDS)
- b. 表格存储 (Table Store, 原 OTS)
- c. 分析型数据库 (AnalyticDB, 原 ADS)
- d. 大数据计算服务 (MaxCompute, 原 ODPS)

9、某企业因为业务发展带来了数据膨胀,每日产生 4T 数据,现在他们考虑要把现有的分析系统迁移到云上。通过系统分析,针对目前系统的状况去做云产品的选型。他们梳理了以下

目前支持的应用类型，大体上有以下几种：提供给一线业务人员前一日以及最近三个月的经营状况的报表分析，同时要向客户提供 30 天内业务数据的即席查询。请从查询效率和成本两个维度考虑，需要哪两款产品组成合适的解决方案？（正确答案的数量：2 个）

- a. ADS
- b. DRDS
- c. RDS
- d. MaxCompute

10、大数据计算服务（MaxCompute，原 ODPS）提供了完整的安全策略和授权策略，来保护我们的数据，在想下载数据到本地的时候，除了要有相应的权限之外，还要选择合适的下载方式。现在大数据计算服务的某表中有 1000 万行数据，需要把这些数据全部导出到本地。下面哪一种是可行的方法？

- a. 在客户端工具 odpscmd 中用 select 命令，把结果重定向到文件中
- b. 管理控制台中读出该表，下载数据
- c. 在命令行工具 odpscmd 中使用 tunnel 命令下载
- d. 找到该表在大数据计算服务上的保存路径，然后把该路径下的文件全部下载

11、大数据计算服务（MaxCompute，原 ODPS）中关于 tunnel 命令行工具的说法，正确的是？（正确答案的数量：2 个）

- a. tunnel 上传到分区表时，会把数据直接上传到指定的分区中去，不保证业务逻辑
- b. tunnel 命令上传文件时，不能进行压缩
- c. tunnel 上传的数据是先写到临时目录，最后确定成功后才写到结果目录
- d. tunnel 上传过程中不加密，数据以明文方式在网络中传输

##https://help.aliyun.com/document_detail/27833.html

- ##-cp, -compress：指定是否在本地上压缩后再上传，减少网络流量，默认开启。

##自动机密传输，无法控制

12、分布式文件系统可以有效的解决大数据存储和管理的难题，阿里云飞天平台中哪个系统提供了这样的功能？

- a. 伏羲
- b. 盘古
- c. 女娲
- d. 大禹

13、项目空间（project）是大数据计算服务（MaxCompute，原 ODPS）的基本组织单元，它类似于传统数据库的 Database。以下关于项目空间的说法中正确的有_____。（正确答案的数量：3 个）

- a. 所有的对象都会属于某个项目空间
- b. 一个帐号可以拥有多个项目空间的权限
- c. 一个帐号可以创建多个项目空间，最多不能超过 10 个
- d. 通过安全授权，可以在一个项目空间中访问另一个项目空间中的数据

这个是错的

这个是对的

14、Data IDE 中不同角色拥有不同的权限，可以通过角色授权来控制成员的操作权限。角色

主要有组织管理员、项目管理员、开发者、访客等，其中开发角色不具备（）权限。

- a. 新建任务
- b. 添加项目成员
- c. 上传资源
- d. 新建函数

#https://help.aliyun.com/document_detail/30277.html?spm=5176.doc30285.6.557.ISCVUi

17、组织是 Data IDE 的特有概念，组织的基本对象为项目空间，以下说法正确的有（）。（正确答案的数量：3 个）

- a. 一个帐号可以加入同一个组织的多个项目空间
- b. 一个组织可以创建多个项目空间
- c. 一个帐号只能加入一个组织
- d. 一个项目空间可以属于多个组织

#组织可以理解为公司

18、Data IDE 中，项目空间 prj1 只能配置一个成员为项目管理员

对
错

#项目管理员：默认为当前组织管理员，后续到项目管理中可以添加其他成员为管理员。

#https://help.aliyun.com/document_detail/30263.html?spm=5176.doc30277.6.544.FYzMbi

FYzMbi

19、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数据同步。DataX 任务配置文件使用（）文件格式。

- a. java
- b. bson
- c. xml
- d. json

#https://help.aliyun.com/knowledge_detail/42846.html

24、在 Data IDE 中，当从云数据库 RDS(MySQL)抽取数据时，可配置切分键进行数据分片，提高数据同步效率，以下关于切分键的说法中正确的有（）。（正确答案的数量：2 个）

- a. 可以选择任意字段作为切分键
- b. 从云数据库 RDS 导出数据到 MaxCompute 配置切分键可以实现并发，提高数据同步速度
- c. 推荐使用表的主键作为切分键
- d. 配置切分键容易出现数据热点

25、Data IDE 中的数据同步任务只支持全量数据同步。

对
错

#https://help.aliyun.com/document_detail/30269.html

26、在 Data IDE 中使用数据同步任务时，可以将某个自定义变量的值写入目标表中，自定义变量采用“变量名=参数”的格式进行赋值。

对
错

#自定义变量和常量的写入方法：如果需要把常量或者变量导入 ODPS 中表的某个字段，只需要点击插入按钮，然后输入常量或者变量的值，并且用英文单引号包起来即可，如变量 `'${yesterday}'`，然后再参数配置组件配置给变量赋值如 `yesterday=${yyyymmdd}`

27、在 Data IDE 中配置数据同步任务将云数据库 RDS(MySQL)同步至大数据计算服务（MaxCompute，原 ODPS）中，需要确认在云数据库 RDS 实例的 IP 白名单中是否包含了 Data IDE 的 IP 段，如未包含，则需要添加。

对
错

30、Data IDE 中，事实表 A 需要按天增量更新（配置成任务 task1），事实表 B 需要按天增量更新（配置成任务 task2），每天事实表 A 中的部分数据信息来自事实表 B 前一天的数据。task1 和 task2 需要如何配置才能实现上述场景？

- a. task1 的依赖属性中的上游任务配置为 task2
- b. 不需要配置依赖关系，在 task1 中直接使用表 B 生成表 A 即可
- c. task1 的跨周期依赖属性设置为“等待自定义任务的上一周期结束”，自定义任务选择 task2
- d. task2 的依赖属性中的上游任务配置为 task1

31、Data IDE 中，周期性调度任务是指按配置的时间周期进行自动调度的任务，以下场景中，（）不适合配置成周期性调度任务。

- a. MaxCompute 中基于商品库存表每隔一小时生成一张库存盘点的汇总表
- b. 每隔 5 分钟对 MaxCompute 中的访问日志表进行一次数据清洗
- c. 今天市场部经理临时要求分析师基于 MaxCompute 中的用户交易表生成一份昨天有成交纪录的品牌列表
- d. 把云数据库 RDS 中的用户登录记录表每天 0 点 30 分同步到 MaxCompute 表中

32、Data IDE 中，调度系统提供了一些常用的时间参数，周期性调度任务使用这些参数后，在调度运行时参数会自动赋值为具体的时间，以下哪些参数是调度系统内置时间参数？（正确答案的数量：2 个）

- a. \${yyyymmdd}
- b. \${bdp.system.bizdate}
- c. \${yyyy-mm-dd}
- d. \${bdp.system.cyctime}

33、Data IDE 中，调度系统可以根据配置好的任务属性进行自动调度。属性中包含的生效日期是指任务在有效期内可以进行调度运行，任务 task1 的生效日期为 2016-01-01 至 2016-06-30，以下关于任务 task1 的说法中正确的有（ ）。（正确答案的数量：2 个）

a. task1 在 2016 年 7 月 1 日当天在调度上无法操作测试

b.task1 在 2016 年 4 月 1 日那天操作补数据时无法补业务日期为 2016 年 1 月 1 日的数据

c. task1 在 2016 年 6 月 30 日当天就不会再自动调度运行

d. task1 如果在 2016 年 1 月 1 日前添加，则 2016 年 1 月 1 日当天才会自动调度运行

34、Data IDE 中，周期性调度任务在配置依赖属性时，上游任务只能是同个项目空间的任務。

对

错

35、Data IDE 中，task1 和 task2 都是按天调度的周期性调度任务，task1 设置定时时间为 02:00，task2 设置定时时间为 00:00，task2 依赖属性的上游任务配置为 task1，每天 task2 何时可以运行？

a. 只要 task1 执行成功，task2 就马上运行

b. task1 在 2:00 之后才执行成功，则 task2 不再运行

c. task1 执行成功时，时间若已到 2:00，则 task2 可以运行

d. task1 执行成功时，时间没到 2:00，则 task2 不再运行

36、Data IDE 中，调度任务每次运行前都先将任务实例化，即生成一个实例，调度运行时实际上在执行相应的实例。任务配置不同的调度周期，自动调度生成的实例个数也就不一样，以下说法正确的是？（正确答案的数量：3 个）

a. 调度周期配置成按天调度的任务，每天生成一个自动调度实例

b. 调度周期配置成在 00:00 至 23:59 之间按分钟调度，每隔 30 分钟执行一次的任务，每天生成 48 个自动调度实例

c. 调度周期配置为按周调度，每周一调度一次的任务，只有周一才会生成一个自动调度实例，周二到周日都不会生成实例

d. 调度周期配置成在 00:00 至 23:59 之间按小时调度，每天生成 24 个自行调度实例

37、Data IDE 中，虚拟节点类型是空跑节点，即调度时直接返回成功状态不会真正执行。两个周期调度任务 task1、task2 之间没有依赖关系，业务上也没有共同上游任务，但是希望可以有一个共同的父任务进行统筹管理，那么可以创建一个虚拟节点作为 task1 和 task2 的上游任务。

对

错

38、Data IDE 中，调度任务每次运行前都先将任务实例化，即生成一个实例，调度运行时实际上在执行相应的实例。可以通过运维中心的任务运维管理暂停任务自动调度的实例。如果按天调度的周期性任务 task1 的 2016 年 4 月 9 日自动调度的实例被操作暂停，以下关于 task1 的说法中正确的是_____。

a. 仅 2016 年 4 月 9 日的自动调度实例被暂停

b. 2016 年 4 月 8 日的自动调度的实例也会被暂停

c. 所有在 2016 年 4 月 9 日之后生成的自动调度的实例都会被暂停

d. 2016 年 4 月 10 日自动调度的实例也会被暂停

（跨周期依赖

配置节点/ workflow 任务的跨周期依赖，如，天调度任务今天需要执行的数据依赖本任务昨天执行的数据，那么可以配置依赖任务昨天的周期，如此，昨天的实例必须先执行成功，今天的实例才可以调度起来，这种依赖主要是体现在任务调度实例的依赖。

- 不依赖上一调度周期。所有任务默认选择该选项，即不依赖任何任务的上周期实例。
- 自依赖，等待上一调度周期结束，才能继续运行。使用场景：任务 A 当前周期数据来源依赖与任务 A 上周期执行的结果；或者小时/分钟调度任务 A 不允许实例并行。
- 等待下游任务的上一周期结束，才能继续运行。依赖第一层子任务的上周期。这种使用场景不多，选择此项，后续该任务一旦被其他任务直接依赖则实例都依赖所有第一层子任务的上周期实例。
- 等待自定义任务的上一周期结束，才能继续运行。常见场景：天任务 A 依赖一个数据是天任务 B 昨天产出的。

【说明】

1) 依赖属性配置的调度依赖是同周期依赖和跨周期依赖不冲突。任务 A 可以配置依赖性依赖任务 B 也可以配置跨周期依赖依赖 B，如此任务 A 即依赖任务 B 本周期也依赖任务 B 上周期。

2) 若任务 A 是小时/分钟调度，任务 B 为天调度，任务 B 配置依赖任务 A 的上周期，那么当天 B 任务的实例会依赖 A 任务昨天所有实例。

3) 若任务 A 和 B 都是小时/任务调度，调度周期一样，任务 B 配置依赖任务 A 的上周期，则任务 B 每个实例都将依赖 A 昨天所有实例和 A 与 B 同周期的前一个周期实例。

4)如天任务 A 配置跨周期依赖 B 的上周期,那么对 A 任务进行补数据的时候,补数据实例会去依赖 B 任务自动调度上周期实例,如果自动调度的上周期实例不存在则不依赖)

39、Data IDE 中,运维中心的“任务管理视图”以 DAG 图(有向无环图)的方式显示调度任务之间的依赖关系和运行约束。通过任务管理视图可以查看任务的定义,但不能查看任务的运行状态,以下关于任务管理视图说法错误的是_____。

- a. 可以查看任务配置的调度周期粒度
- b. 可以查看任务 owner
- c. 可以查看任务影响的记录条数
- d. 可以查看任务描述

40、Data IDE 中,调度任务每次运行前都先将任务实例化,即生成一个实例,调度运行时实际上在执行相应的实例。在调度的不同阶段,实例会处于不同的状态,包括未运行、运行中、运行成功、运行失败、等待时间、等待资源等。当实例处于_____状态时不能手工停止。

- a. 未运行
- b. 运行中
- c. 等待时间
- d. 等待资源

#https://help.aliyun.com/document_detail/35161.html?spm=5176.doc30279.6.569.7Gb6vd

41、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。在调度的不同阶段,实例会处于不同的状态,包括_____。(正确答案的数量: 3 个)

- a. 未完成
- b. 运行中
- c. 未运行
- d. 失败

#https://help.aliyun.com/document_detail/35161.html?spm=5176.doc30279.6.569.7Gb6vd

42、Data IDE 中,调度任务每次运行前都先将任务实例化,即生产实例,调度运行时实际上在执行相应的实例。运维中心的任务运维视图显示的实例是任务自动调度的实例,对这些实例可以进行_____。(正确答案的数量: 2 个)

- a. 补数据
- b. 重跑并恢复调度
- c. 查看节点运行日志
- d. 测试

43、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。

对
错

44、在大数据计算服务（MaxCompute，原 ODPS）的表中存有一批网站及其相互之间链接关系，开发人员希望通过 pagerank 算法来分析这批网站的排名，即把网站看做节点，把网站之间的连接关系看做权重，从而实现对网站的评分。_____计算框架更适合实现这个场景。

- a. SQL
- b. MapReduce
- c. Graph
- d. Tunnel

45、大数据计算服务（MaxCompute，原 ODPS）的图计算引擎是分布式的，在加载数据时，系统默认如何将数据分配到不同的 Worker 中？

- a. 根据边的权值进行 hash 并对 Worker 数取模
- b. 完全随即分配
- c. 根据顶点 ID 的 hash 值对 Worker 数取模
- d. 按照点 ID 值排序，再平均分配给 Worker

46、大数据计算服务（MaxCompute，原 ODPS）图计算框架再计算过程中通过对图进行编辑、演化最终求解出结果，在计算过程中用户可以对图进行哪些编辑操作？（正确答案的数量：4）

- a. 读取本地数据修改点的权值
- b. 增加边
- c. 删除点
- d. 修改边的权值
- e. 增加点

47、MaxCompute Graph 是 MaxCompute 提供的编程框架之一，主要是用来进行图像处理。

对
错

48、大数据计算服务（MaxCompute，原 ODPS）的 MapReduce 提供了分布式的编程框架，以下哪类大数据计算服务可以作为 MR 的输入和输出？

- a. 表
- b. 视图
- c. 资源
- d. 项目空间

49、大数据计算服务（MaxCompute，原 ODPS）提供了 MapReduce 编程接口，用户可以使用 MapReduce 提供的 SDK 编写程序处理大数据计算服务的中的数据。目前提供了_____语言的 SDK。

- a. Perl

- b. C++
- c. Python
- d. Java

50、使用大数据计算服务（MaxCompute，原 ODPS）的 odpscmd 客户端中的 jar 命令调用 MapReduce 时，需要指定两个参数：classpath 和 resources。比如以下命令执行了一个 MapReduce 程序，输入为表 t_in，输出为表 t_out：jar-resources test.jar-classpath test.jar test t_in t_out；其中，test.jar 出现了两次，请判断以下说法正确的是_____。

- a. 两次出现的意义相同
- b. 两个引用的 test.jar 都是保存在大数据计算服务上的资源
- c. 两个引用的 test.jar 都是保存在本地

d. classpath 引用的是 main 函数，运行于本地，指定的 jar 包路径，即本地文件路径

#https://help.aliyun.com/document_detail/27878.html?spm=5176.doc44626.6.624.uP5Ph7

51、大数据计算服务（MaxCompute，原 ODPS）提出了安全沙箱的概念，为了保证整个计算环境的安全，某些特殊操作会被禁止。以下哪些操作会受到沙箱的影响？（正确答案的数量：2 个）

- a. 使用内置函数
- b. 使用 MR
- c. 使用 DDL 创建表
- d. 使用 UDF

（安全限制：ODPS MapReduce 及 UDF 程序在分布式环境中运行时受到 Java 沙箱的限制。但在本地运行时，则没有此限制；）

https://help.aliyun.com/document_detail/27882.html

52、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 reduce 的输入数据可能来自于多个 map 的输出
- b. 一个 MR 处理可以不包括任何 map
- c. 一个 MR 处理可以不包括任何 reduce
- d. 一个 map 的输出结果可能会被分配到多个 reduce 上去

53、大数据计算服务（MaxCompute，原 ODPS）执行 MapReduce 时，每个 Map Worker 在读入各自的数据后，需要为每一条数据指定一个 Key，这个 Key 的值决定了记录会被发送到那一个 Reduce Worker。单个 Reduce Worker 有可能接收到多个 Key 值的记录，一个 Key 值对应的记录也有可能发送到多个 Reduce Work。

对
错

（具有相同 Key 的数据会被发送给同一个 Reduce Worker，单个 Reduce Worker 有可能会

接收到多个 Key 值的数据)

54、大数据计算服务 (MaxCompute, 原 ODPS) 的 MapReduce 由多个步骤组成, 以下哪个步骤是必须的?

- a. combine
- b. shuffle
- c. map
- d. reduce

55、在开发大数据计算服务 (MaxCompute, 原 ODPS) 的 MapReduce 时, 可以使用 JobClient 类中的接口提交 MapReduce 作业。目前提交方式支持_____。(正确答案的数量: 2 个)

- a. 非阻塞方式
- b. 远程提交
- c. 阻塞方式
- d. 本地提交

56、大数据计算服务 (MaxCompute, 原 ODPS) 的 MapReduce 的输入可以为空, 也可以为视图。

对

错

#输入可以为空, 不支持视图(View)作为输入;

#https://help.aliyun.com/document_detail/27880.html?spm=5176.doc27875.6.626.i7KKbo

57、某销售企业基于自己顾客的若干个指标 (价值、成本、使用的产品等) 对已有的用户群进行划分: 特征相似的用户做为一类, 特征不同的用户分属于不同的分类。然后会针对不同类型的客户进行不同的营销策略。这个案例属于大数据分析的哪种类型的问题。

- a. 分类
- b. 关联分析
- c. 聚类
- d. 时序分析

58、某超市通过分析顾客的购买记录后发现, 购买纸尿裤的人由较大的概率也会购买啤酒, 这个案例数据大数据分析的哪种类型的问题?

- a. 聚类
- b. 时序分析
- c. 分类
- d. 关联分析

59、数据仓库是一个面向主题的 (Subject Oriented)、集成的 (Integrate)、相对稳定的 (Non-Volatile)、反映历史变化 (Time Variant) 的数据集合, 主要用于支持决策、面向分析型数据处理。大数据计算服务 (MaxCompute, 原 ODPS) 出了能很好的处理大数据语境下的分析功能外, 也是传统数据仓库上云的解决方案之一。下列哪些传统数据仓库的功能可以通过大数据计算服务实现? (正确答案的数量: 3 个)

- a. 定时的 ETL 任务 (数据抽取、转换和加载)

- b. 固定报表分析
- c. 数据挖掘
- d. 即席查询

60、通过阿里云的数据集成服务（Data Integration，原云道 CDP），可以将云数据库 RDS、分布式关系型数据库 DRDS、分析型数据库（AnalyticDB，原 ADS）中存储的结构化数据同步到大数据计算服务（MaxCompute，原 ODPS）中去。

对
错

61、大数据计算服务（MaxCompute，原 ODPS）中的日志表 log 是一张分区表，分区表是 dt，每天产生一个分区用于存储当天新增的数据，数据是通过 tunnel 接口上传，运维人员需要在每天上传数据之前生成好当天的分区，假设当前日期是 20160301，语句_____可以创建合适的分区。

- a. add partition log (dt='20160301')
- b. insert overwrite table log partition(dt='20160301')
- c. insert into table log partition(dt='20160301')
- d. alter table log add partition(dt='20160301')

#https://help.aliyun.com/document_detail/27862.html?spm=5176.doc27800.6.606.JtUt0M

62、大数据计算服务（MaxCompute，原 ODPS）中的订单表 fact_order 是一张分区表，分区键是 dt，每天新增的订单存储在当天的分区中，如果要更新 dt='20160310'这个分区中的数据，对应的源表是 ods_order，则正确的 SQL 语句结构是_____。

- a. update fact_order select_from ods_order where dt='20160310'
- b. insert overwrite table fact_order ds='20160310' select_from ods_order
- c. insert overwrite table fact_order partition(dt='20160310') select_from ods_order
- d. insert overwrite table fact_order select_from ods_order where dt='20160310'

63、小王正在使用大数据计算服务（MaxCompute，原 ODPS）进行自然语言分析，他从门户网站上搜集了 1000 篇新闻报道，然后每隔存成一条记录，然后基于这些记录去做分词，即将每条记录中的内容拆成很多独立的词语，保存到另一张表中，每个词语一条记录。小王对自定义函数比较熟悉，他想通过编写自定义函数实现分词功能。请帮小王做一下选择，从功能角度来看，哪种自定义函数更适合处理这种场景？

- a. User Defined Aggregation Function（用户自定义聚组函数）
- b. User Defined Scalar Function（用户自定义标量函数）
- c. User Defined Splitted Function（用户自定义分组函数）
- d. User Defined Table Valued Function（用户自定义表值函数）

65、大数据计算服务（MaxCompute，原 ODPS）的运维人员想要知道一张日志表 log 已经占用了多少存储空间，可以在命令行工具 odpscmd 中使用_____命令查看。

- a. ls log;
- b. desc log;
- c. select* from log;

d. size log;

66、运营人员为了搞促销，需要找出目标客户名单，这批客户的信息存储在大数据计算服务（MaxCompute，原 ODPS）的 user 表中，并且用户名字段 username 中包含“vip”字样，通过执行 SQL 语句_____可以找出这批客户。

- a. select* from user where username contains('vip')
- b. select* from user where username='vip'
- c. select* from user where username like '*vip*'
- d. select* from user where username like '%vip%'

67、大数据计算服务（MaxCompute，原 ODPS）中的日志表 log 是一张分区表，分区键是 dt，每天产生一个分区用于存储当天新增的数据，现在共有 dt='20160101'至 dt='20160131'共 31 个分区的数据，为了删除 20160101 当天新增的数据，可以通过_____方式实现。

- a. delete from log where dt='20160101'
- b. truncate table where dt='20160101'
- c. drop partition log (dt='20160101')
- d. alter table log drop partition (dt='20160101')

#将指定的非分区表中的数据清空，该命令不支持分区表。对于分区表，可以用 ALTER

TABLE table_name DROP PARTITION 的方式将分区里的数据清除。

68、大数据计算服务（MaxCompute，原 ODPS）中的用户表 dim_user 是一张非分区表，另外有一部分新增用户的数据存在 user_delta 表中，两张表的结构相同，为了在 dim_user 表中得到全量的用户数据，可以采用_____方式。

- a. select* from user_delta into dim_user;
- b. insert into dim_user select* from user_delta union all select*from dim_user
- c. insert overwrite table dim_user select*from user_delta
- d. insert into table dim_user select* from user_delta

69、大数据计算服务（MaxCompute，原 ODPS）中的用户基本信息表 dim_user 和扩展信息表 dim_user_ext 都有一个整型字段 user_id(业务上该字段的每个取值都唯一标识一条用户记录)。表 dim_user 中共有 10000 条记录，其中 user_id 从 2000 到 11999，dim_user_ext 表中共有 5000 条记录，其中 user_id 从 0 到 4999。开发人员在建立用户模型时需要综合用到两张表中所有用户的数据，如果有扩展信息则加上，如果没有则只用基本信息，_____可以实现该需求。

- a. 使用全关联：select t1.*,t2.* from dim_user t1 full outer join dim_user_ext t2 on t1.user_id=t2.user_id;
- b. 使用右关联：select t1.*,t2.* from dim_user t1 right join dim_user_ext t2 on t1.user_id=t2.user_id;
- c. 使用左关联：select t1.*,t2.* from dim_user t1 left join dim_user_ext t2 on t1.user_id=t2.user_id;
- d. 使用内关联：select t1.*,t2.* from dim_user t1 join dim_user_ext t2 on t1.user_id=t2.user_id;

70、对于大数据计算服务（MaxCompute，原 ODPS ）内置绝对值函数 abs 描述正确的有_____。（正确答案的数量：2 个）

- a. 当输入参数是 bigint 时，返回值是 bigint 类型
- b. 当输入参数是 string 类型时，一定会导致异常
- c. 输入类型是 boolean 则返回值是 True
- d. 当输入参数是 double 时，返回值是 double 类型

71、大数据计算服务（MaxCompute，原 ODPS ）中，表 ods_sale_detail 存放的是销售明细记录，建表语句如下：create table ods_sale_detail(sale_date datetime, region string, amt decimal); 其中 sale_date 字段表示销售的日期, amt 表示销售全额, 想要找出所有汇总销售额大于 10000 的日期以及当日的销售额，以下哪几个 SQL 可以达到这个目的？（正确答案的数量：2 个）

- a. select sale_date, sum(amt) from ods_sale_detail group by sale_date where amt>10000;
- b. select* from(select sale_date,sum(amt) s_amt from sale_date group by sale_date)sub where s_amt>10000;
- c. select sale_date, sum(amt) from ods_sale_detail group by sale_date having sum(amt)>10000;
- d. select sale_date, sum(amt) from ods_sale_detail group by sale_date where sum(amt)>10000;

72、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，可以使用_____从日志表 log 中找出所有符合这种特征的访问。（正确答案的数量：3 个）

- a. select* from log where tolower(log_content) like '%select%';
- b. select* from log where instr(tolower(log_content),'select')>0;
- c. select* from log where 'select' in tolower(log_content);
- d. select* from log where regexp_instr(tolower(log_content),'select',1)>0;

73、大数据计算服务（MaxCompute，原 ODPS ）中，使用 CREATE TABLE LIKE 语句创建新表时，原表的哪几种属性会被复制到新表上？（正确答案的数量：3 个）

- a. 列注释
- b. 分区
- c. 表的生命周期属性
- d. 表注释

74、在一个大数据计算服务（MaxCompute，原 ODPS ）的 SQL 语句中，可以读多张表中的数据，但是只能更新一张目标表中的数据。

对

错

75、大数据计算服务（MaxCompute，原 ODPS ）的内置函数 substr 只能操作 string 类型的输入值，当输入的值是其他任何类型时，会导致异常。

对

错

76、大数据计算服务（MaxCompute，原 ODPS）的 insert overwrite 操作中，select 中的表达式类型与对应的目标中的数据类型必须一致，否则会导致异常。

对

错

77、大数据计算服务（MaxCompute，原 ODPS）中的客户表 customer 是一张非分区表，现在想要清空表的内容然后重新上传数据，可以通过_____将数据清空。

- a. delete from customer;
- b. insert overwrite table customer select* from customer;
- c. truncate table customer;
- d. drop table customer;

78、大数据计算服务（MaxCompute，原 ODPS）中客户信息表的建表语句如下：create table customer(cust_id bigint, province string, addr string, full_addr string); 其中 full_addr 字段需要根据源表中的 province 和 addr 拼成一个完整的地址，源表中的一条记录的 province 值为空（NULL），addr 为‘文一西路 969 号’，则 concat（province，addr）的值为_____。

- a. 运行出错
- b. “<NULL>文一西路 969 号”
- c. ‘文一西路 969 号’
- d. 空值（NULL）

79、大数据计算服务（MaxCompute，原 ODPS）中的 log 表保存了日志内容，都被存到了 log_content 字段中，以访问者的 IP 地址开头，示例格式如下：127.0.0.1—[22/Mar/2016:03:47:43-0700]“GET/favicon.ico HTTP/1.1” 404 209
"Mozilla/5.0(X11;Linux x86_64) AppleWebKit/537.36(KHTML,like Gecko)Chrome/49.0.2623.87 Safari/537.36"为了提取出访问者的 IP 地址，可以用哪些内置的函数？（正确答案的数量：3 个）

- a. regexp_substr
- b. regexp_extract
- c. substr 和 instr
- d. repeat

80、在大数据计算服务（MaxCompute，原 ODPS）的客户端工具 odpscmd 中开发 SQL 时，用户对 project1 和 project2 都有足够的操作权限，当前所在的项目空间是 project1，想要查看 project2 中有哪些表，可以通过以下哪几种操作？（正确答案的数量：2 个）

- a. show tables in project2;
- b. show project2.tables;
- c. use project2;show tables;
- d. show tables project2;

81、以下哪几个大数据计算服务（MaxCompute，原 ODPS）的 SQL 语句可以用用户表 user 中找出用户名 username 中包含‘hu’的记录？（正确答案的数量：2 个）

- a. select* from user where username rlike ‘.*hu.*’

b. select* from user where username like '%hu%'

c. select* from user where username contains('hu')

d. select* from user where username like '*hu*'

82、大数据计算服务（MaxCompute，原 ODPS ）中，在命令行工具 odpscmd 中使用 select 语句查看表中的记录时，可以结合 limit 子句来限制输出结果的行数。

对

错

83、在使用 java 创建大数据计算服务（MaxCompute，原 ODPS ）的 UDF 时，可以在函数定义中引用资源，并且在函数中读取这些资源。

对

错

84、在大数据计算服务（MaxCompute，原 ODPS ）的数据仓库中的订单表 fact_order，建表语句如下：create table fact_order(order_id string, order_amt double, order_dt string)partitioned by (dt string); 此表中的数据是从 ods_order 加工而来，ods_order 建表语句如下：create table ods_order(order_id string, order_amt bigint, order_dt string);ods_order 中有一条记录数据值是 order_id order_amt order_dt 0001 100 20160301 运行 SQL 语句将数据从 ods_order 加载到 fact_order 中：insert overwrite table fact_order partition(dt='20160301') select* from ods_order; 对此语句的执行结果描述正确的是_____。

a. 语句可以执行，但是这条数据会被当作脏数据丢弃

b. 目标表与源表中的数据类型不一致，执行出错

c. 提交 SQL 时会报语法错误

d. 语句可以执行，order_amt 的值会被自动的转为 double 类型

#https://help.aliyun.com/document_detail/27860.html?spm=5176.doc27867.6.604.nyZFfK

85、大数据计算服务（MaxCompute，原 ODPS ）的开发人员需要创建一张用户消费习惯表，其中一个字段想要命名为 like，则以下说法正确的是_____。

a. 可以将字段命名为 like，无需特殊处理

b. 可以将字段命名为 like，在创建表及引用该字段时，需要用括起来，即表示为'like'

c. 在创建表时无需特殊处理，但是在引用该字段时需要表示为'like'

d. 无法将字段命名为 like，因为这是关键字

MaxCompute 将 SQL 语句的关键字作为保留字。在对表、列或是分区命名时请不要使

用，否则会报错。保留字不区分大小写。下面只给出常用的保留字列表，完整的保留字列表

请参阅 [MaxCompute SQL 保留字](#)。

86、大数据计算服务（MaxCompute，原 ODPS ）SQL 中支持 join、left join、right join、full join 等，对于连接条件描述正确的有_____。（正确答案的数量：3 个）

a. 只能有一个连接条件

- b. 没有 mapjoin hint 时，必须是等值连接
- c. 没有 mapjoin hint 时，多个连接条件之间必须是 AND 关系
- d. 使用 mapjoin hint 时，语法上可以允许没有连接条件

87、在大数据计算服务（MaxCompute，原 ODPS）SQL 中可以用逻辑运算连接多个条件，以下逻辑运算结果正确的有_____。（正确答案的数量：4 个）

- a. True or NULL = True
- b. False OR False = False
- c. True AND False = False
- d. True AND True = True
- e. NULL and False = NULL

#https://help.aliyun.com/document_detail/27861.html?spm=5176.doc27863.6.605.B29b8J

88、大数据计算服务（MaxCompute，原 ODPS）的开发人员原来已经获得了视图 V 的读权限，但是这个视图被管理员使用 create or replace view 语句修改了视图的定义，则这位开发人员必须重新申请权限，否则无法读取这个视图。

对

错

#如果没有指定 if not exists，在视图已经存在时用 create view 会导致异常。这种情况可以用 create or replace view 来重建视图，重建后视图本身的权限保持不变。

89、大数据计算服务（MaxCompute，原 ODPS）tunnel 命令提供了丰富的参数，可以通过参数设置，对上传或者下载操作进行控制，现在有一个大文件要上传至 odps 某表中，想先检查一下该文件内容是否和表的定义匹配，并不实际上传，以免文件中有错误记录会导致上传出错。可以使用哪个参数实现这个功能？

- a. 错误记录丢弃：-dbr
- b. 字段分隔符：-fd
- c. 扫描选项：-s
- d. 空字符串显示：-ni

90、大数据计算服务（MaxCompute，原 ODPS）的 tunnel 命令行工具对上传的数据有格式上的要求，它支持以下哪种格式的数据？

- a. JSON 文件
- b. xml 文件
- c. xls 文件
- d. csv 文件

91、以下哪些场景可以使用大数据计算服务（MaxCompute，原 ODPS）的 Tunnel SDK 方式实现？（正确答案的数量：3）

- a. 定时批量上传数据
- b. 下载某个分区表的一个分区的数据
- c. 一次性上传本地格式化文本文件

d. 随时变化的数据，通过事件（消息）驱动触发

92、使用大数据计算服务（MaxCompute，原 ODPS）的 Tunnel SDK 上传下载数据时，任务会被提交至 Tunnel 服务器（tunnel endpoint），由其进行响应和处理。

对

错

93、大数据计算服务（MaxCompute，原 ODPS）中的角色（Role）是一组访问权限的集合。当需要读一组用户赋予相同的权限时，可以使用角色来授权。基于角色的授权可以大大简化授权流程，降低授权管理成本，当需要对用户授权时，应当优先考虑是否应该使用角色来完成。关于角色的说法中正确的是_____。

a. 可以将 Owner 视为一个角色，赋给其他用户

b. Admin 是项目空间自动创建的角色，可以为该角色添加权限

c. 删除一个角色时，会先检查该角色内是否还存在其他用户。若存在，则删除该角色失败

d. 创建完角色后，需要给角色添加权限，仅支持以 ACL 授权的方式添加

#https://help.aliyun.com/document_detail/27934.html?spm=5176.doc27933.6.678.iOUYx5

94、大数据计算服务（MaxCompute，原 ODPS）中的 ACL 授权时类似 SQL92 定义的 GRANT/REVOKE 语法来实现，更接近于传统关系型数据库的授权模式，Policy 授权主要解决 ACL 授权机制无法解决的一些复杂授权场景。以下场景中，哪些既能使用 ACL 授权又能使用 Policy 授权实现？

a. 一次操作对一组对象进行授权

b. 只允许用户使用 SQL（而不允许其他类型的 Task）来访问某张表

c. 将一个对象的多个权限一次性授予一个用户

d. 带限制条件的授权

95、大数据计算服务（MaxCompute，原 ODPS）中的授权是指把某些对象的某些权限赋给一些指定的用户，必须包括以下哪三个要素？（正确答案的数量：3 个）

a. 操作（Action，比如读、写等）

b. 效果（Effect，比如同意、拒绝等）

c. 主体（Subject，可以是用户、角色等）

d. 客体（Object，可以是表、资源等）

#https://help.aliyun.com/document_detail/27935.html?spm=5176.doc27939.6.679.6oNN28

96、使用大数据计算服务（MaxCompute，原 ODPS）中的 ACL 授权方式进行授权时，可以使用通配符，但是，目前仅支持星号（*）。

对

错

97、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. 大数据的分析建模

b. 海量数据仓库

c. OLTP

OLTP是交易系统类的"运行"事务

d. 网站日志离线分析

98、大数据计算服务（MaxCompute，原 ODPS ）有别于我们常见的关系型数据库，以下关系型数据库中支持的功能（feature），在大数据计算服务中也支持的是哪一项？

a. 索引

b. 事务

c. 视图

d. 主键

99、大数据计算服务（MaxCompute，原 ODPS ）提供的是海量数据的存储和计算能力，和我们熟悉的关系型数据库存在较大的差别。以下说法中正确的是？（正确答案的数量：3 个）

a. 不支持事务

b. 不支持 delete

c. 不支持索引

d. 不支持压缩

100、大数据计算服务（MaxCompute，原 ODPS ）的分区（Partition）是指一张表下，根据分区字段（一个或多个字段的组合）对数据存储进行划分。也就是说，如果表没有分区，数据是直接放在表所在的目录下；而如果表有分区，每个分区对应表下的一个目录，数据是分别存储在不同的分区目录下。

对

错

#分区表指的是在创建表时指定分区空间，即指定表内的某几个字段作为分区列。在大多数情况下，用户可以将分区类比为文件系统下的目录。MaxCompute 将分区列的每个值作为一个分区(目录)。用户可以指定多级分区，即将表的多个字段作为表的分区，分区之间正如多级目录的关系。

#[https://help.aliyun.com/document_detail/27820.html?spm=5176.doc27862.6.561.](https://help.aliyun.com/document_detail/27820.html?spm=5176.doc27862.6.561.PfnS56)

PfnS56

1、试卷

单选题

1、云盾加密服务采用了什么机制来保证密钥的安全？ B

A 防篡改硬件密码机 前期

B 符合国家密码管理局要求的算法

C 符合银监会对密钥强度的要求

D 支持国际通用密码算法 加密算法支持全面支持国产算法以及部分国际通用密码算法

2、RDS 采用的备 M-M 的高可用架构，其备之间的数据同步依靠日志的方式实现，MySQL 数据和 SQL server 分别使用的日志是什么？ D

A 均为实物日志

B MySQL 为事务日志，SQL server 为 Binlog 日志

C 均为 Binlog 日志

D MySQL 为 Binlog 日志，SQL server 为事务日志

3、ODPS 的屏显行数可以通过管理控制台实现，最大的显示条数是多少？ D

A 100 条

B 10000 条

C 1000 条

D 5000 条

注解：

limit number]的 number 是常数，限制输出行数。当使用无 limit 的 select 语句直接从屏幕输出查看结果时，最多只输出 5000 行

4、阿里云云监控产品里的监控数据可以采用哪种方式下载？ B

A 通过 openAPI 环境

B 支持管理控制台和 OpenAPI 两种方式下载

C 只能查看报表，无法下载

D 从管理控制台直接下载

5、在实际应用中，那种性能指标表示 RDS 实例磁盘空间的使用量？ A

A 磁盘空间

B 连接数

C CPU 利用率

D iops

6、ODPS 表 T 中某列 C 的数据类型为 bigint，需要修改为 double，以下哪种方式可以实现？

A

A 将表 T 删掉重建

B ALTER TABLE T COLUMN C RENAME TO C DOUBLE;

C ALTER TABLE T DROP COLUMN C;ALTER TABLE T ADD C DOUBLE;

D ALTER TABLE T CHANGE COLUMNS C BIGINT

7、以下哪个功能，是 ODPS 的 mapreduce 不支持的？ D

A 统计信息

B 任务进度汇报

C 推测执行

D 文件压缩

8、ODPS 的 ACL 授权操作一般涉及三个要素，不包括下列哪个？ C

A 操作

B 客体

C 限制条件

D 主体

9、以下哪个数据库是 RDS 支持的？ A

A MySQL

B db2

C MariaDB

D Sybase

10、下面哪个特性属于多可用区 RDS 的？ A

A 轻松实现同城容灾

B 将 RDS 的存储空间扩大 2 倍

C 有效控制云产品间的网络延迟

D 是为了防止 DDoS 攻击

11、用户在购买 RDS 实例时所选择的 __A__ 决定了该实例的最大连接数

A 内存大小

B CPU 利用率大小

C 网络流量

D 磁盘容量大小

12、ADS 中表组在创建时需要制定哪两个参数？ D

A numhashPartitions & partiton Type

B numhashPartitions & minRedundancy

C numhashPartitions & executeTimeout

D executeTimeout & minRedundancy

13、通过 DMS 管理后台或者数据库客户端，连接 RDS 数据库时，提示错误信息“max_user_connections”，代表什么含义？ D

A IOPS 超出极限

B RDS 空间满了

C 网络中断

D RDS 数据库的连接数满了

14、OTS API 中单行操作不包括以下哪个？ A

A Postrow

B putrow

C updaterow

D getrow

15、ADS 建表可指定的更新方式为 B

A 导入更新和插入更新

B 批量更新和实时更新

C 自动更新和手工更新

D 全量更新和增量更新

备注：分析型数据库的表根据 **updateType** 分为批量更新表（仅能够离线批量更新数据）和实时更新表（能够通过 **insert/delete** 实时更新数据）

16、以下哪个不属于 ODPS 的访问策略的访问控制元素？ A

A 效果

B 操作

C 权限

D 主体

17、RDS for MySQL 不支持以下哪种引擎 B

A 以上都不支持

B memory

C InnoDB

D MyISAM

目前仅支持 **MyISAM**（新创建实例已不支持，只支持部分存量实例）、**InnoDB**、**TokuDB** 三种。

18、开发 ODPS 的用户自定义标量函数，主要是实现其中的哪方法

A evaluate

B main

C iterate

D process

!!!19、ODPS 中，哪个命令可查看角色 **role_test** 的权限 C

A list role role_test

B show grants for role role_test

C desc role role_test

D show acl for role rol_test

20、下列关于 ADS 的说法中，哪一项是正确的？ B

A 表组的属性可以修改，修改后马上生效；

B 对于批量插入的表可以建立二级 list 分区，二级 list 分区是非动态分区，分区值由用户导入时指定

C minredundancy 是表组的一个属性，可以指定不同的值没，比如 1、2、3、4 等，缺省值为 2 ；

D 表组是一个逻辑概念，同一个表组中的表可以选择同时上线，再删除表组时，会把表组中的表一起删除

注解：

minRedundancy 表示该表组的副本数，默认为 2，可配置为 1、2、4、8。

仅允许删除没有任何表的空表组，维度表组不允许删除。

需要注意的是，**minRedundancy** 修改后需要下次装载数据时才会生效

21、ODPS 项目空间 **prj1** 中存在表 **t1**，owner 打开 **LabelSecurity** 设置后，执行如下命令：SET LABEL 1 TO TABLE t1；SET LABEL 2 TO TABLE t1(id)；SET LABEL 3 TO TABLE t1；此时，t1 表

中的列 ID 的敏感等级标签是什么? B

- A、0 级
- B、2 级
- C、1 级
- D、3 级

!! 22、对于 ODPS DDL 的操作，一般都是由那层完成? A

- A worker
- B executor
- C Scheduler
- D controller

!! 23、如果需要从 ODPS 同步一张表到 ADS，则该表的访问权限需要授予 ODPS 中的那个用户?

- A aliyun\$gurada_build@aliyun.com
- B aliyun\$gurada_data@aliyun.com
- C aliyun\$gurada_ads@aliyun.com
- D aliyun\$ads_data@aliyun.com

24、使用 OTS Java SDK 进行表操作时需要创建 OTSClient 对象，以下哪项不属于构造此对象时必须提供的信息 A

- A AccountID
- B 实例名称
- C accessKeySecret
- D endpoint

注解：创建 OTSClient 对象，在构造函数中指定 Endpoint，AccessKeyID，AccessKeySecret，实例名字。

25、下列不属于 ODPS 项目空间的对象类型为? C

- A 表
- B 实例
- C jar 包
- D 资源

解释：以操作该项目空间下的对象，例如：[表\(Table\)](#)，[资源\(Resource\)](#)，[函数\(Function\)](#)，[实例\(Instance\)](#)等，而不需要关心操作对象所在的项目空间

注解：DPS 提供了项目空间、表、资源、函数等多种对象类型。Jar 包属于资源的具体文件。

26、DRDS 默认每个实例创建几个数据库? D

- A 16 个
- B 64 个
- C 1 个
- D 8 个

备注：默认每个实例创建 8 个数据库，也就是说 2 个实例就会有 16 个分库，3 个实例 24 个分库以此类推。

https://help.aliyun.com/document_detail/29664.html

27、关于 ODPS SQL，说法不正确的是？ D

- A 在很多方面并不具备数据库的特征
- B 适用于海量数据，实时性要求不高的场合
- C 每个作业的准备，提交等阶段要花费较长时间
- D 大部分标准 SQL 的功能都支持，但是不支持窗口函数、rownum 等

!!!!!! 28、网络安全专家服务是阿里云云盾什么服务的基础上，推出的安全代为托管服务？ C

- A 阿里绿网
- B DDOS 高防 IP
- C 安骑士
- D 基础 DDOS 防护

29、ODPS 中，当一个用户被移除后，与该用户有关的那些授权会保留？ C

- A package 授权
- B policy 授权
- C ACL 授权
- D 项目空间授权

30、在 ADS 中，什么是用户所关心的最大单元，也是用户和 ADS 系统管理员的管理职权的分界点？ B

- A 列
- B 数据库
- C 表
- D 表组

解释：在分析型数据库中，数据库是用户所关心的最大单元，也是用户和分析型数据库系统管理员的管理职权的分界点。分析型数据库系统管理员最小可管理数据库粒度的参数，而无法未经用户授权来查看和管理数据库内部的结构和信息。

31、对于 ADS 的 list 分区，哪个说法是正确的？ C

- A list 分区可以通过制定最大保留的二级分区数来进行生命周期管理。一旦分区数大于这个值，将会根据导入时间进行排序，将最早导入的分区下线
- B list 分区信息定义时可以指定一个现有数据中的列，但是对数据类型有要求
- C list 分区为非动态分区，即分区值不由数据本身决定，而是每次导入时用户指定的
- D list 分区列的类型目前仅支持数值型和字符串型两种

解释：将经常需要进行 Join 的列（例如买家 ID）作为一级 Hash 分区列，而将日期列作为二级分区列。这样的表既可以进行大表 Join 的加速，又可以每天进行增量数据导入，并且指定保留若干天的数据在线上来进行生命周期管理。

二级 List 分区为非动态分区，暨分区值不是由数据本身决定的，而是由每次导入数据时用户指定的。所以在进行分区信息定义时需要指定一个和现有数据中的列不同的新列名，以及这个列的

类型（目前仅支持 long）。二级分区有一个可选属性，available_partition_num，即为最大保留的二级分区数，当新的数据装载进来后，若线上存在的二级分区数大于这个值，那么会根据二级分区的值进行排序，下线最小的若干分区的数据。实时更新表不支持二级分区。

32、阿里绿网具有什么功能？ B

A 绿网的目标是建立低碳高校的绿色 IDC，帮助用户降低服务器能耗

B 帮助用户检查网站内容的合规性，包括文字内容和图片

C 用来限制访问 ECS 云主机的源 IP，防止黑客窃取数据

D 用来限制服务器对外访问的资源，防止 ECS 云主机中毒

34、下列关于 RDS 的使用，标书正确的是？ D

A 只读实例的配置必须与主实例完全相同

B 允许第三方工具进行物理备份

C MySQL 存储引擎仅支持 InnoDB

D 创建数据库时必须使用 open api 或 RDS 管理控制台进行操作

!!! 33、针对 ODPS SQL 做逻辑分析的时候，会进行一些优化，不包括以下哪种优化？ A

A 扫描方式

B 列映射

C 谓语动词下沉

D 常量表达式

备注：来自 ODPS 权威指南

1、常量表达式的计算

2、列裁剪

3、谓词下推

!!! 34、OTS 中可以使用拼接的分片键，拼接就是把几个属性拼接成一个属性。拼接是为了解决什么问题？ C

A 解决数据生命周期问题

B 解决写压力过于集中的问题

C 解决单个分片键下数据量过大的问题

D 解决 OTS 预留读写量不足的问题

备注：Table Store 建议单个分片下的数据量大小不超过 1GB。如果您的表中单个分片键的所有行的总数据量大小可能超过 1GB，在设计表时可以将原来的多个主键列拼接成分片键。

35、ADS 中，对批量导入的表，那种分区模式是可取的？ D

A hash+hash

B list +list

C list+hash

D hash+list

36、OTS 中某秒内发生了 10 次读操作，其中 5 次每次操作读出当行数据大小为 2.78KB，另外 5 次操作读出当行数据大小为 5.36KB，则共消耗多少*** D

A 11CU

- B 20CU
- C 10CU
- D 15CU

各操作消耗的写服务能力单元的计算规则:

PutRow - 本次消耗的写 CU 为修改的行主键数据大小与属性列数据大小之和除以 4KB 向上取整, 若指定条件检查不为 IGNORE, 还需消耗该行主键数据大小除以 4KB 向上取整的读 CU。如果操作不满足应用指定的行存在性检查条件, 则操作失败并消耗 1 单位写服务能力单元和 1 单位读服务能力单元 **PutRow** 详解

UpdateRow - 本次消耗的写 CU 为修改的行主键数据大小与属性列数据大小之和除以 4KB 向上取整, **UpdateRow** 中包含的需要删除的属性列, 只有其列名计入该属性列数据大小, 若指定条件检查不为 IGNORE, 还需消耗该行主键数据大小除以 4KB 向上取整的读 CU。如果操作不满足应用指定的行存在性检查条件, 则操作失败并消耗 1 单位写服务能力单元和 1 单位读服务能力单元 **UpdateRow** 详解

DeleteRow - 被删除的行主键数据大小除以 4KB 向上取整。若指定条件检查不为 IGNORE, 还需消耗该行主键数据大小除以 4KB 向上取整的读 CU。如果操作不满足应用程序指定的行存在性检查条件, 则操作失败并消耗 1 单位写服务能力单元 **DeleteRow** 详解

37、ODPS 中哪种操作会使授权失效? D

- A 对 package 授权过的表重建
- B 上述授权都会失效
- C 对 policy 授权过的表重建
- D 对 ACL 授权过的表重建

38、使用 ODPS SQL: create table t1 like t2; 建表时, 表 t1 不会具有表 t2 的哪种属性?

- B
- A 分区
- B 生命周期
- C 二级分区
- D 字段的注释

注解:

lifecycle 指明此表的生命周期, **create table like** 语句不会复制源表的生命周期属性。除生命周期属性外, 列名、列注释以及表注释等均相同。

39、在 RDS 只读实例变更配置过程中, 并且只读实例的对应源数据库实例状态为”升降级中“。主实例可以进行下列哪种任务? C

- A 重启实例
- B 内外网切换
- C 系统资源监控
- D 创建实例

40、ODPS 表 t1、t2 都是分区表, 分区键是 pt, 对于 sql 语句 select t1.id from t1 left outer join t2 on t1.id=t2.id where t1.pt='20140101' and t2.pt='20140101', 哪一个说法是正确的。答案: C

- A、两张表都会进行分区裁剪
- B、t2 表会进行分区裁剪

- C、t1 表会进行分区裁剪
- D、两张表都不会进行分区裁剪

备注：当两个表进行 join 操作的时候，主表的 Where 限制可以写在最后，从表要先完成分区筛选后再进行 Join 操作；从表分区限制条件不允许写在 Where 条件里，要求写在 ON 条件或者子查询，主表可以写在 WHERE 条件里(最好先用子查询过滤)。

41、DRDS 事务支持哪种一致性？ A

- A 最终一致性事务
- B 会话一致性事务
- C 强一致性事务
- D 因果一致性事务

42、关于 ODPS MR 哪项说法是正确的？ A

- A mapworker 在输出数据时，需要为每一条输出数据指定一个 key
- B 进入 reduce 前，数据需要进行合并操作，然后按照 key 排序
- C map 和 reduce 前都需要对数据进行分片
- D reducer 的个数和 mapper 的个数一致

43、!!!大型会员制连锁超市 S 使用 OTS 来存储消费记录表，里面包含了客户的 vip 卡号、订单标识，超市分店标识，商品标识，购买时间等，业务逻辑遵循以下规则：每个 vip 卡号对应一个客户，同时一个客户只能拥有一个 vip 卡号；该连锁超市的不同分店都拥有唯一的分店标识；订单标识按照时间顺序产生；为了让数据分布和访问的分布尽可能均匀，一下备注主键中哪一组最合适？ B

- A 购买时间，vip 卡号
- B vip 卡号，购买时间
- C vip 卡号
- D 订单标识，vip 卡号

备注：https://help.aliyun.com/document_detail/27356.html

主键的第一列是分片键

购买时间，vip 卡号：购买时间，不同时间购买热度不一样

VIP 卡号，购买时间：每天每张卡产生的消费记录数从总体上来讲是均匀的，每一个分片键中的访问压力也应该是均匀的。以 CardID 作为表的分片键可以较好地利用预留读写吞吐量资源

VIP 卡号：不能唯一标识

订单标识，VIP 卡号：作为表的分片键不是一个好的选择。因为订单标识是顺序增长的，因此在同一时间段内产生的消费订单的订单标识的值会集中在一个较小的范围内，这些消费订单记录会集中写入到个别的分片，预留读写吞吐量没有得到高效的利用。

44、对于 ADS 的数据一致性问题，那个说法是错误的？ B

A 当 updateType=realtime 时，对于不同主键数据的多次变更，ADS 不保证先执行的变更会比后执行的变更更优秀的查询到

B 当 updateType=realtime 时，不支持事务，并且仅遵循会话一致性的设计，所以 ADS 并不能作

为 OLTP 系统使用 解释：最终一致性

C 当 updateType=realtime 时，当业务端暂停数据写入的若干时间后，ADS 会保证数据的一致性

D 当 updateType=realtime 时，对于同一主键数据的多次变更，ADS 会遵循 ADS 返回语句执行成功的顺序进行

45、ADS 的高度智能优化策略是指一下哪一个？ B

A PBO

B CBO

C RBO

D HBO

解释：分析型数据库拥有高度智能的 CBO(Cost-Based Optimization)优化策略。

!!! 46、关于 ADS 的权限模型，下列哪些描述是正确的？ A

A 每个权限级别能聚合其下面级别的所有权限

B 导出数据时需要 DUMP DATA、DESCRIBE 和 SELECT 权限，同时需要数据导出目的地的数据写入相关权限

C 查询表数据需要 SELECT 权限，最小级别是表

D ADS 的授权模型和 MySQL 非常相似，比如 ADS 支持针对用户在 host 上授权

解释：

任何分析型数据库支持的账号类型均可视为一个用户。和 MySQL 略有不同的是，分析型数据库目前不支持针对用户在 host 上授权。

查询数据的权限

查询表数据需要 SELECT 权限，最小级别是列

并非所有查询都需要该权限，例如 SELECT now()

导出数据的权限

导出数据同时需要 DUMP DATA 和 SELECT 权限

同时需要数据导出目的地的数据写入相关权限

聚合： Database -> Table[Group] -> Column，即每个权限级别能聚合其下面级别的所有权限。

47、OTS API 中单行操作不包括以下哪个 Action？ C

A、UpdateRow

B、Putrow

C、PostRow

D、GetRow

备注：Table Store 单行操作单行写入操作 Table Store 的单行写操作有三种:PutRow，UpdateRow 和 DeleteRow

48、ADS 中创建事实表（普通表）时会有一些限制，下列哪个说法是错误的？ D

E、表的名字不能超过 32 个字符

F、一个事实表最多不能超过 1024 列

G、至少有一级 Hash 分区

H、二级分区只能是 Hash 或 List

备注：

分析型数据库支持最多两级分区，并且一级分区仅支持 HASH 分区，二级分区仅支持 LIST 分区；

事实表的创建上，有如下限制：（1）一张事实表至少有一级 Hash 分区并且分区数不能小于 8 个；（2）一个事实表组最多可以创建 256 个事实表；（3）一个事实表最多不能超过 1024 个列。

49、关于云盾态势感知的优势，描述错误的是？ A

- E、IT 资产划分，软件版本管理 基于威胁情报的安全事件管理对入侵行为造成的损失进行评估
- F、建设安全体系 从黑客角度监控云上业务安全 对异常行为进行实时告警 让安全“可见、可控、可管”
- G、免安装，免维护 无需安装任何软件，部署任何设备 无需复杂配置和更新，省心省力 安全管控从未如此简单
- H、是一个针对阿里云服务器的管家服务

备注：大数据实时分析，还原安全事件全过程，对攻击者进行取证和追溯

50、某公司有一个专门的小团队负责线上系统的安全，一年以来使用云盾的 DDoS 高防 IP 体验很好，成功抵御了 5 次超过 100G 的攻击，但是费用较高，经过技术团队研究后发现公司的平台并不需要 100% 的在线率，做到 99% 即可满足业务要求。CTO 决定让小王寻找一个更合适的替代方案，如果用云盾来满足这个需求，哪项服务是最佳的？ B

- E、基础 DDoS 防护
- F、安全网络
- G、网络安全专家服务
- H、服务器安全托管

51、关于 ODPS 中的运算符的说法，正确的是？ B

- E、String 类型可以和其他任意类型进行隐式转换：Boolean 不行
- F、除 boolean 外，其他类型不允许参与逻辑运算，也不允许其他类型的隐式类型转换
- G、只有 bigint、double 才能参与算数运算：只有 string、bigint、double 和 Decimal 才能参与算术运算
- H、特殊的关系运算符 rlike 表示右侧开始匹配

52、哪个字符集是目前 RDS for MySQL 不支持的？ B

- E、gbk
- F、big5
- G、latin1
- H、utf8

53、下面列出的参数中哪一个不属于 RDS SQL SERVER 中用的大字段？ C

- E、image
- F、varchar
- G、int
- H、text

54、以下 ODPS SQL 哪个效率最高？ A（见 Join 语句中 where 条件的位置）

- E、select * from (select * from A where dt=20140301)A`left outer join B on B.id=A.id
- F、没有区别

G、select * from A left outer join B on B.id=A.id and A.dt=20140301
H、select * from A left outer join B on B.id=A.id where A.dt=20140301

疑问：A 与 D 什么区别

当两个表进行 join 操作的时候，主表的 Where 限制可以写在最后，但从表分区限制条件不要写在 Where 条件里，建议写在 ON 条件或者子查询。主表的分区限制条件可以写在 WHERE 条件里(最好先用子查询过滤)。

55、以下哪个是 ODPS 内网的 end point? B

E、<http://service.odps.aliyun.com/api>

F、<http://odps-ext.aliyun-inc.com/api>

G、<http://service.odps.hangzhou.aliyun.com/api>

H、<http://odps-int.aliyun-inc.com/api>

备注： 阿里云内网地址：ODPS地址：<http://odps-ext.aliyun-inc.com/api>

公网地址: ODPS地址：<http://service.odps.aliyun.com/api>

!!!56、对于 ADS 的主键，哪个说法是正确的? D

E、对于 updateType=realtime 的表来讲，主键是必须的，且主键中不能包含作为一级 hash 分区的分区列

F、对于 ADS 的普通表，无论 updateType 如何取值，主键是必要的

G、对于 updateType=batch 的表来讲，主键是必须的，主要是约束导入数据时的重复记录

H、有主键的表的性能和用法上和没有主键的表之间没有任何区别

备注：对于批量更新表，分析型数据库中主键的概念是弱化的，分析型数据库不要求一个表有主键，有主键的表的性能和用法上和没有主键的表之间没有任何区别

57、某信贷公司推出一款线上产品，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建模并产生预测数据，用户线上申请贷款时，能迅速根据用户模型预测结果进行资质评估、授信等。请根据你对阿里云产品的理解，下述方案中成本最低、性能最好的一个是? A

E、使用 ODPS 对数据进行建模，然后使用该模型对所有复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 OTS 中，以供实现快速查询

F、使用 ADS 对数据进行建模，然后使用该模型对所有复合贷款条件的会员进行资质评估和授信评估，将处理结果保存到 ADS 中

G、使用 ODPS 对数据进行建模，然后使用该模型对所有复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 ODPS 中

H、使用 ADS 对数据进行建模，然后使用该模型对所有复合贷款条件的会员进行资质评估和授信评估，将处理结果保存到 OTS 中

58、ODPS 项目 prj1 中存在表 t，同时使用下述 SQL 创建视图 v:create view v as select * from t;prj1 的 owner 把视图 v 的读权限赋给了用户 Alice，以下哪个说法是正确的? C

E、因为没有表 t 的访问权限，所以 Alice 无法访问视图 v

F、由于权限的依赖关系，Alice 既可以访问视图 v 又可以表 t

G、Alice 可以通过访问视图 v 来查询表 t 的数据

H、Alice 可以访问表 t，因为她具有了视图 v 的访问权限，自然获得了表 t 的访问权限

注解：

创建视图时，必须对视图所引用表的读权限。

视图只能包含一个有效的 select 语句。

视图可以引用其它视图，但不能引用自己，也不能循环引用。

不可以向视图写入数据，例如使用 insert into 或者 insert overwrite 操作视图。

当视图建好以后，如果视图的引用表发生了变更，有可能导致视图无法访问，

59、在 project 中已经开启了 label security(列级别授权)的前提下，用户申请了一张表 tbl 的读权限，后来管理员将该表中的字段 x 的安全等级设置为了 4，且用户的安全级别为 1，用户能否读该表中的数据？ D

E、不可以，因为用户的级别较低

F、可以，因为之前已经有读权限了

G、提供信息不足以判断

H、可以访问除了字段 x 之外的数据

60、OTS 的一些表中存在冷数据和热数据的区分，所谓冷数据是指根据时效性判断被查询的可能性不大的数据，反之为热数据。下面关于冷热数据的说法，错误的是？ D

E、可以用不同的表来区分冷热数据，并设置不同的预留读写吞吐量

F、表中存在大量冷数据会导致数据访问压力不均匀，从而导致表上配置的预留读写吞吐量无法被充分利用

G、按照冷热数据拆分成不同的表后，可对冷数据对应的表设置较小的预留读写吞吐量

H、在同一张表可以对不同的分片设置不同的预留读写吞吐量，对于冷数据所在的片，设置较低的预留读写吞吐量

解释：为了解决这种问题，可以用不同的表来区分冷热数据，并设置不同的预留读写吞吐量。

例如，将消费记录按月份分表，每一个新的自然月就换一张新的表。当月的消费记录表需要不停写入新的消费记录，同时有查询操作。当月的消费记录表可以设置一个较大的预留读写吞吐量配置来满足访问需求。前几个月的表由于不再写入新数据或者写入的新数据量较少，查询的请求较多，因此前几个月的消费记录表可以设置较小的预留写吞吐量，较大的预留读吞吐量。而历史超过一年的消费记录表，由于再被使用的可能性不大，可以设置较小的预留读写吞吐量配置。已经超出维护年限的消费记录表可以将数据导出，存入 OSS(Open Storage Service)归档，或直接删除。

61、某些在线广告 APP 采用了阿里云的技术，通过对网上搜集的海量数据的分析，给企业主提供精准的广告投放人群。该 APP 的产品经理介绍该系统能实现以下计数指标：

5- 能存储海量数据，单标大小超过 100TB

6- 支持多张上亿的表进行 Join

7- 支持多种筛选条件，包括连续范围筛选、取值筛选、and/or/not 条件组合等

8- 支持快速（3 秒以内）对十个维度分别进行技术透视分析

根据以上信息，请判断该 APP 使用的是阿里云的哪款产品？ C

E、DRDS

F、OTS

G、ADS

H、ODPS

62、关于 ADS 的数据类型，说法不正确的是哪一项？ C

E、布尔类型 `boolean` 取值只能为 0 或 1

F、ADS 所有的数据类型都不支持 `unsigned`

G、数值类型 `float` 可以指定精度，例如：`float(5,3)`，但是 `double` 类型不支持指定精度

H、双精度浮点数 `double` 类型的存储字节数为 4，大整数类型 `bigint` 存储字节数为 8

63、RDS 和自建数据库相比，以下哪个是错误的？ B

E、RDS 服务可用性高达 99.95%，但是自建数据库需自行保障，自行搭建主从复制，自建 RAID 等

F、自建数据库能方便地支持快速部署、弹性扩容

G、RDS 可以自行备份，自建数据库需要寻找备份存放空间以及定期验证备份是否可恢复

H、RDS 无需运维，自建数据库需招聘专职 DBA 来维护，花费大量人力成本

64、OTS python CLI 工具需要哪个版本的 python 支持？

A、2.3 版

B、2.7 版

C、2.6 版

D、3.0 版

65、OTS 的单表可以存储规模较大的数据。目前单表存储的限制为？ C

A、10T

B、100T

C、没有上限

D、1T

66、对于 OTS 表的主键描述正确的是哪一个？ C

E、由四个属性组成

F、建表的时候可以暂时不指定，之后用 `alter table` 修改

G、组成主键的属性只能是 `string` 或者 `integer`

H、主键本身就是表的分片键

注解：

创建 `Table Store` 表时必须指定表的主键。主键包含 1~4 个主键列。每一个主键列都有名字和类型。`Table Store` 对主键列的名字和类型都有限制，详细信息可以参考 `Table Store` 数据模型的主键一节。`Table Store` 根据表的主键索引数据，表中的行按照它们的主键进行升序排序。属性列的数据类型只能是 `String` 和 `Integer`。如果为 `String` 类型，长度不超过 1KB。

`Table Store` 会自动把表分成不同的数据分片，以达到对其存储数据的负载均衡。数据分片的划分粒度为主键的第一列，我们将该列称为数据分片键。即，拥有相同数据分片键的行必然在同一个数据分片中。`Table Store` 能够保证对具有同一数据分片键的数据进行更改操作的一致性。

67、在利用 RDS 管理控制台将数据迁移至 RDS-MySQL 中的时候，使用的方法是 MySQL 在线迁移，可以不停服务就能够完成数据库的迁移工作，以下哪个 MySQL 版本不支持在线迁移？ C

E、MySQL 5.6

F、MySQL 5.1

G、MySQL 5.0

H、MySQL 5.5

!!! 68、ODPS 中一个表可以上限最多拥有多少个分区? A

A、60000 个

B、100000 个

C、不限制

D、10000 个

备注：目前 ODPS 支持的分区数量上限为 6 万。

!!! 69、以下哪些应用场景适合 RDS 只读实例来完成? B

E、扩大 RDS 的存储能力

F、分担 RDS 主实例的写压力

G、缓解 RDS 主实例的写压力

H、实现两天前的数据回溯

阿里云 RDS 只读实例 分担数据库读写压力

70、ODPS 中负责对等待提交的 task 进行排序的模块是? A

E、Scheduler

F、Worker

G、Executor

H、Controller

备注：scheduler 负责 instance 调度，包括将 instance 分解为 task、对等待提交的 task 进行排序，以及向计算集群的伏羲 Master 询问资源占用情况进行流控；

!!! 71、云盾反欺诈服务是阿里巴巴的什么能力开发的? C

E、BGP 网络

F、DDos 攻防经验

G、大数据风控服务

H、弹性计算

beizhu:风险识别服务

通过用户行为信息、软硬件环境信息、设备指纹等综合判定用户请求的风险程度：可信、可疑、风险，业务根据风险结果进行相应处理：直接进入业务流程，图形/短信/邮箱验证，直接拦截，离线分析等。

72、表 t 中有 1000 万行数据，想要随意抽取 10 行数据，在 odpscmd 中执行消耗时间最短的方式是哪种? D

E、select * from t where rownum<=10

F、get t 10

G、select * from t limit 10

H、read t 10

73、使用 ODPS 动态分区时，SELECT 的分区列放在列表的什么位置？ **D**
E、中间
F、开始
G、任意位置
H、**结尾**

1、WEB 应用采用何种方式可以接入云盾反欺诈服务？ **C**
A: HTMLS B: SDK C: JavaScript D: JavaScript 和 SDK

客户平台集成风险拦截服务方式：前端引入组件 JS 获得滑动通过签名串，服务端调用风险拦截 API 验证签名串，获得签名串验证结果

74、关于 RDS 的续费方式，正确的是： **B**

A: 无需续费 B: 只能手动续费
C: 只能自动续费 D: 可以手动续费，也可以设置自动续费

75、下列关于 RDS 可用区的说法正确的是： **D**

A: RDS 单可用区轻松实现了同城容灾
B: RDS 不支持多可用区
C: RDS 多可用区实现了 ECS 和 RDS 间的网络延迟更小
D: **RDS 支持跨可用区迁移**

目前多可用区 RDS 不额外收取任何费用，在已开通多可用区地域的用户可以直接购买多可用区 RDS 实例，也可以通过跨可用区迁移将单可用区 RDS 实例转化成多可用区 RDS 实例。注意：因为多可用区之间存在一定的网络延迟，因此多可用区 RDS 实例在采用半同步数据复制方案的时候，对于单个更新的响应时间会比单可用区实例长。这种情况最好通过提高并发量的方式来实现整体吞吐量的提高。

76、ADS 的普通表的一级分区数量最多有多少个？ **C**

A: 无限制 B: 128 个 C: 256 个 D: 1000 个

一般来讲，每个分区的数据不超过 800 万条为宜，当然也不绝对，分区数不能超过 256 个

77、OTS 的单个表可以包括多少个属性列？ **C**

A: 0-128 个 B: 0-1024 个 C: 无限制 D: 0-256 个

!!!78、关于 ADS 的 ECU，说法正确的是： **A**

A: 可以通过 DDL 语句 alter database set ecu_count=N 来修改 ECU 的个数
B: 可以动态的扩容或者缩容，两者都是瞬间的同步操作
C: 可通过查询表 information_schema.resource_request 获取目前的 ECU 的状态
D: 通过参数设置，可以使同一个 ADS 的库同时使用两种类型的 ECU

解释：缩容和扩容都不是瞬时的同步操作，可以使用元数据查询状态：

```
select * from information_schema.resource_request;
```

79、ODPS 中提供了别名命令 `ALIAS<alias>=<real>`，可以为哪些对象设置别名？ A

A: 资源 B: 列 C: 行 D: 表

用途：为资源创建别名。

80、DRDS 和普通的关系型数据库（如 MySQL）相比，SQL 优化要特别考虑的是？ D

A: 磁盘 IO 开销 B: 内存开销 C: 并发度 D: 网络 IO 开销

DRDS 是一个高效稳定的分布式关系数据库系统。但是，由于其处理的是分布式的关系查询，因而它对于 SQL 的查询优化与传统的单一数据库（如 mysql，oracle）的查询优化有所不同。后者在查询优化时，主要考虑的就是磁盘 IO 的开销，但前者在优化时还需要考虑另外一个更为重要的 IO 开销—网络。

81、关于 RDS 的配置变更，描述错误的是 A

A: 升级过程中服务不中断（备注：闪断 30S） B: MySQL5.1 可以直接升级到 MySQL5.6

C: 数据库版本无法降级 D: 用户可以将实例根据自身需要升级到相应的规格

82、ADS 的表进行实时插入、更新时，遵循哪种一致性设计？ B

A: 强一致性 B: 最终一致性 C: 会话一致性 D: 因果一致性

82:下列关于 RDS 性能的指标说法错误的是

A: IOPS 表示每秒磁盘读写（IO）次数，读写操作包括：数据从磁盘读到内存，或从内存写入磁盘 B

B: 不同 RDS 实例之间可以资源共享

C: QPS 表示每秒 SQL 语句执行次数

D: 后台管理员通过 top 或 ps 命令可以查看实例的 CPU 利用率

83、假设 A 公司要在两周后进行重量级新产品发布，需要在公司的官网上进行直播，公司领导担心发布过程中会出现大流量攻击，但是公司没有专业的团队做支持，请问这种情况下云盾的哪项服务最适合满足该用户的需求？ D

A: 渗透测试服务 B: 网络安全专家服务 C: 服务器安全托管 D: DDoS 高防 IP

!!! 84、对于 ADS，哪个说法不正确？ A

A: ADS 表存在主键，即可根据主键使用 delete 删除纪录

B: 普通表分区列如果是字符串型，则不支持普通表之间的 join

C: 自查询之间不支持 join

D: ADS 通过 UNION（ALL）各个分区内的查询结果实现自查询

85、以下关于 ODPS 授权的说法正确的是 B

A: 当一个用户被项目空间 owner 从项目中移除时，所在的授权信息都将被清除，该用户无法再访问当前项目空间中的任何对象

B: 当项目空间的 Owner 决定对另一个用户授权时，需要先将该用户添加到自己的项目空间中来，只有添加到项目空间中的用户才能被授权

C: ODPS 中的角色是一组访问权限的组合。项目创建成功后，自动生成两个角色 owner 和 admin，可以通过命令 list roles 看到

D: 当项目空间中的 owner 授权给一个用户时，ODPS 会先检验该用户是否实际存在，如果不存在，则无法授权

备注:

- 当一个用户被移除后，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。
每一个项目空间在创建时，会自动创建一个 admin 的角色

当项目空间的 Owner Alice 决定对另一个用户授权时，Alice 需要先将该用户添加到自己的项目空间中。只有添加到项目空间中的用户才能够被授权。

添加用户的命令如下：

```
add user <username> --在项目空间中添加用户
```

86、ODPS 对 SQL DML 语句进行解析后，得到执行计划。构成执行计划的是由多个有相互依赖关系的_____组成？ C

A: 任务（Task） B: 作业（Job） C: 执行阶段（Stage） D: 实例（Instance）

ODPS 会对其进行解析，得出任务的执行计划。 执行计划是由具有依赖关系的多个执行阶段 (Stage)构成的

!!! 87、下列对于 ADS 中表和表组的描述正确的是 C

A: 维度表在创建时不需要指定表组，但是需要配置分区信息

B: 维度表可以和除维度表组外的任意表组中的表关联

C: 批量插入的表如果只有一级分区，每次导入数据是会覆盖掉已有数据

D: 维度表组有且只有一个，用户可以修改表组属性，但是不能删除

维度表可以和任意表组的任意表进行关联，并且创建时不需要配置分区信息，但是对单表数据量大小有所限制，并且需要消耗更多的存储资源

维度表组有且仅有一个，并且在分析型数据库数据库建立时会自动创建，用户不可修改和删除

88、除了正常数据外，以下哪个数据会占用所购买的 RDS 实例空间？ D

A: 数据库回滚日志 B: 索引 C: 重做日志 D: 以上都是

除了您正常的数据外，还有数据库实例正常运行所需要的空间，比如系统数据库、数据库回滚日志、重做日志、索引等。同时 RDS for MySQL 数据库实例产生的一天内的 Binlog 日志也会占用这个空间

89、对于 ADS 的连接方式，以下说法正确的是： A

A: 以上都是 B: 安装了 php-mysql-5.1.x 模块的 php 环境

C: 使用版本为 5.4 系列的 mysql-jdbc 驱动 D: mysql5.6 的客户端

90、ODPS tunnel 上传或者下载数据时，参数 rd 的默认取值是： B

A: \t B: \r\n C: , D: ”

91、云盾网络安全专家服务特别适合以下哪种场景？ B

A: 购买了 100 台 ECS，但没有专业的管理员

B: 每月一次的产品促销活动，担心网络攻击

C: 频繁出现 SQL 注入警告

D: 系统每次重大升级后不清楚有没有漏洞

安全网络是通过 ip 节点组成的，具备流量转发，流量调度能力的一个网络。他需要通过配置和

接入才能使您的应用具备攻击防护的能力。一个 ip 节点就相当于您服务器的一个替身，通过 n

多个替身节点来抵御攻击。黑客无法透过安全节点攻击到您的真实服务器。

92、以下哪个不属于 RODPS 提供的功能 D

A: 将某些 R 的模型直接转换为 SQL 发布

B: 在 R 中存取 ODPS 的数据

C: 在 R 中直接使用分布式的算法包处理 ODPS 中的大数据

D: 在 ODPS 上支持 R 的任务

!!!93、下面哪一项不是安骑士包含的功能

A: 高危漏洞修复 B: 防 WEB 应用系统密码破解

C: 异地登录报警 D: 木马文件检查

B

安骑士是云盾安全防护体系中的主机安全防护模块，基于云端联动防御，可以为云服务器提供防黑客入侵的服务，包括木马查杀、防密码暴力破解、异地登录提醒、高危漏洞检测修复、体检加固等安全防护功能

94、作为技术人员，小王被 ODPS 的大数据处理能力吸引，想体验一下，他快速注册了阿里云账号，进入官网后试图创建一个名字为 test_project 的 ODSP project，结果没有成功，试分析，小王出错的最有可能的原因是：

A: 项目名称和现有的项目名称重复 B: 账号未开通 ODPS 的使用权限

C: 账号未充值，余额不足导致出错 D: 未创建合适的 accessID 和 accessKey

A

95、关于 ODPS 分区的说法，正确的是

A: ODPS 将分区列的每个值作为一个分区

B、用户最多能指定 2 级分区

C: 分区列只支持 string 和 bigint 两种，且同一个表如果有多个分区列，这些分区列要么全部是 string，要么全部是 bigint，不能混用

D: 对于分区表，在使用数据时必须指定分区列，否则会出错

96、关于 RDS 数据迁移，错误的是

A: ECS 上自建的 MySQL 可以导入至 RDS

B: 用户可以通过 `mysqldump` 进行分批导入

C: 用在其他 IDC 的 MySQL 数据库也可以导入至 RDS

D: 保存在 OSS 里的图片文件也可以直接导入 RDS

97、云盾加密服务的密钥由谁负责管理? D

A: 阿里云系统管理员 B: 第三方合作伙伴

C: 客户和阿里云系统管理员 D: 用户

借助加密服务, 用户能够对密钥进行安全可靠的管理

!!!98、一个 MySQL 数据库实例最多可创建多少个数据库? C

A: 400 B: 300 C: 500 D: 200

同一实例下的数据库共享该实例下的所有资源, 其中:

- MySQL 版实例最多可以创建 500 个数据库
- SQL Server 2008 R2 版实例最多可以创建 50 个数据库
- PostgreSQL 版实例和 PPAS 版实例没有数据库个数限制

!! 99:ODPS 执行某个 SQL 时返回信息中包括 J3_1_2_Stg1", 以下说法正确的是 B

A: 当前 Task (ID 为 3) 依赖于 ID 为 1 的任务, 同时触发 ID 为 2 的任务

B: 当前 Task (ID 为 3) 依赖 ID 为 1 和 ID 为 2 的两个任务

C: 当前 Task (ID 为 3) 依赖于 ID 为 1 的任务, 该 Task 的 Worker count 为 2

D: 当前 Task (ID 为 3) 有 2 个 Reducer, 1 个 mapper

大数据计算服务 ODPS > 工具 > JOB 运行信息查看工具 > logview

在每个 Task 中, 可以看到 Task 的名字, 对于 M1_stg1, 表示这是一个 Map task, stage 为 1。R2_1_stg1 中间的 1 表示它依赖 M1 执行结束才能开始执行。同理, J5_2_4_Stg2 表示 Join5 这个阶段要依赖 R2 和 M4 两个 task 完全成才能启动运行

100:在 RDS 连接闪断发生的场景中, 有可能是由于实例本身故障引起的是哪个? C

A: 用户升级 RDS 实例配置 B: 用户升级数据库版本

C: 主库出现故障 D: 执行数据回滚操作

备注: RDS 连接闪断发生的场景有:

1. 主库出现故障 (操作不当或实例本身故障)
2. 用户触发升级 RDS 实例配置
3. 用户触发升级数据库版本
4. 用户触发执行数据回滚操作

!!!!!! 101、关于 OTS 中分片的说法正确的是: A

A: 一个分片可能包含多个分片键 B: 同一个分片上的纪录拥有相同的分片键

C: 单个分片数据量最大不能超过 1T D: 相同分片键的纪录不一定在同一分片上

A

组成主键的第一个主键列又称为分片键。Table Store 会根据表中每一行分片键的值所属的范围自动将这一行数据分配到对应的分片和机器上, 以达到负载均衡的目的。具有相同分片键的行属于同一个分片, 一个分片可能包含多个分片键。Table Store 服务会根据特定的规则对分片进

行分裂和合并，以达到更好的负载均衡，这个过程是自动的，应用无需关心。单个分片键下所有行的大小总和不能超过 1GB。

102、某公司属于传统行业，一年前刚刚发展互联网线上业务，用了 5 台 ECS，没有专门的系统管理员，最近半年来上线系统总是出现各种安全问题，几次高危漏洞被黑客利用，差点将核心数据泄露。为此公司的 CEO 王总非常着急，决定向阿里云求助，请你为王总推荐一款合适的安全产品和服务，快速解决服务器的运维之痛。答案：A

A、服务器安全托管

B、DDos 高防 IP

C、阿里绿网

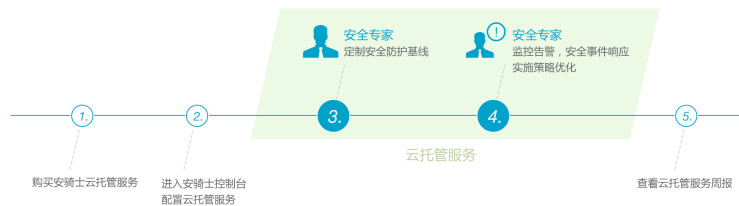
D、安全网络

注解：

一、安骑士云托管服务简介

安骑士云托管服务是一种基于云盾安骑士的安全托管服务，为云服务器提供定制化的安全防护策略、木马文件检测和高危漏洞检测与修复工作。当发生安全事件时，阿里云安全团队提供安全事件分析、响应，并进行系统防护策略的优化。

二、服务购买和使用流程



三、云服务器安全体检加固

云服务器开通安骑士云托管服务后，我们会在 3 个工作日内对完成一次完全体检和系统加固(包括木马病毒检测、高危漏洞检测)，使你的云服务器处于一个相对安全的运行状态，体检服务过程和结果以电话或者邮件告知。

四、报告安全事件

开通安骑士云托管服务的云服务器若发生安全事件，在云托管服务器控制台内，点击“安全事件告知”，将会有专门的安全工程师进行协助。

五、安全事件报告

若服务器发生安全事件，我们将协助你一起分析，最终将为您提供一份高质量的安全事件分析报告

报告内容大纲包括：

- (1) 安全事件描述
- (2) 安全事件溯源
- (3) 漏洞信息、黑客信息
- (4) 漏洞修复加固建议

六、服务报告查看及咨询

- (1) 使用您购买该服务的阿里云帐号，登录云盾控制台，可直接查看或下载服务报告
- (2) 咨询服务报告内容的问题，直接通过工单和服务的安全工程师进行沟通

103、DRDS 实例名称不能以什么开头？答案：B

- A、小写字母
- B、中文
- C、大写字母
- D、下划线

注解：

1、通过阿里云控制台创建 DRDS 实例，事实上是不需要输入实例名称的，系统自动生成。根据以往经验，实例名称不能中文，实例描述可以是中文。

2、DRDS 的分布式数据库的名称（小写字母、数字、下划线组成，字母开头，不超过 24 字符）

104、查询数据需要 select 权限。目前 ADS 中对 select 权限的管理的最小级别是？ 答案： C

- A、表组
- B、表
- C、列
- D、数据库

注解：

查询数据的权限：

查询表数据需要 SELECT 权限，最小级别是列

并非所有查询都需要该权限，例如 SELECT now()

导出数据的权限：

导出数据同时需要 DUMP DATA 和 SELECT 权限

同时需要数据导出目的地的数据写入相关权限

105、下列 ODPS 的那类函数可以用在查询的 where 子句中？ 答案： D

- A、user defined table function
- B、user defined aggregation function
- C、user defined scalar function
- D、user defined function

!!!!!! 106、ODPS 的使用限制包括哪一项。 答案： B

- A、单个任务中自定义的 counter 不能超过 128 个
- B、单个任务引用的资源总计字节数不能超过 64M
- C、字符串格式内容不能超过 128 个
- D、单个任务引用的资源个数不能超过 128 个

注解：

ODPS 表 string 列内容长度不允许超过 2MB。

单个任务引用的资源数量不超过 512 个，分区表按照一个单位计算。

单个任务引用的资源总计字节数大小不超过 64MB。

单个任务的输入路数不能超过 128，单个任务的输出路数不能超过 128 路。

单个任务中自定义 Counter 的数量不能超过 64。

单个 Map 或 Reduce Worker 占用 memory 默认为 2048MB，范围[256MB, 12GB]。

单个 Map 或 Reduce Worker 重复读一个资源次数限制 <=64 次。

本地运行模式下，Map Worker 个数不能超过 100；Reduce Worker 个数不能超过 100；默认一路输入下载记录数 100。

107、关于 DDos 高防 IP 的功能，描述错误的是？ 答案：C

A、提供实时精准的流量报表及攻击详情，让你及时准确获得当前服务详情

B、防护多种 DDOS 类型攻击，包括但不限于以下攻击类型 ICMP Flood、UDP Flood、TCP Flood、SYN Flood、ACK Flood 等

C、DDOS 防护阈值弹性调整，你可以随时升级更高级的防护，调整过程服务中断时间小于等于 30 秒

D、可随时更换防护 IP

108、由于 RDS 只读实例的数据均由源数据库实例同步而来，以下那些功能只读实例可以使用？

答案：A

A、独立的白名单配置

B、备份设置以及临时备份功能

C、创建、删除数据库

D、账号授权以及修改账号密码

109、ADS 可以使用 explain 命令查看执行计划。下面说法中正确的是？ 答案：D

A、当用户发起一个 explain 查询到 ADS 系统后，ADS 会根据所有活跃节点上的数据来分析执行计划

B、当用户通过查询的方式，想要获取文本格式的 explain 语句后，将会得到一个 XML 格式的串

C、逻辑计划汇总，TableExecutor 节点表示参与计算后的信息，JoinExecutor 表示 join 的节点信息

D、explain 正确执行后返回信息中，第一行为逻辑计划，第二行为物理计划

注解：

当用户通过查询的方式，想要获取文本格式的 explain 语句后，将会得到如下的 json 串。

当用户发起一个 explain 查询到分析型数据库系统后，分析型数据库会抽样一个数据分区来分析执行计划，并以图形方式展现给用户。

XPLAIN 的 ResultSet 记录。其中第一行为逻辑计划，第二行为物理计划。

物理计划各个节点说明：JoinExecutor 表示 Join 的节点；TableExecutor 表示参与计算的表信息

110、在实际应用中，那种性能指标表示 RDS 实例磁盘空间的使用量？ 答案：A

A、磁盘空间

B、CPU 利用率

C、IOPS

D、连接数

111、用户 A 是公司的系统管理员，经常去上海出差，每当他在上海远程登录服务器时系统就会发出报警信息，提示“有人异地登录，请注意服务器安全”，有什么方法可以快速、自动解决这个问题？ 答案：A

A、登录阿里云管理控制台，在安骑士配置项里添加“常用登录地点”

B、向我的朋友电话求助，他是业界有名的黑客高手

C、向公司领导请求帮助

D、立即提交工单，咨询阿里云工程师

!! 112、DRDS 不具备以下哪个特性 答案：D

- A、小表广播
- B、分库分表
- C、平滑扩容
- D、智能索引

113、在 ODPS 表 T 中添加一个新的列 col，类型为 bigint，最简单易行的办法是？ 答案：B

- A、将表 T 删掉重建
- B、alter table t add columns (c bigint)
- C、新建一张表，将原有数据插入
- D、alter table add c bigint

注解：

ALTER TABLE table_name ADD COLUMNS (col_name1 type1, col_name2 type2...)

说明：列的数据类型只能是：bigint，double，boolean，datetime 及 string 类型。

114、以下那个是 RDS 性能上的优势 答案 B

- A、花费人力成本高
- B、无需运维
- C、资源利用率低
- D、需要自建 RAID

115、下面那个特性属于多可用区 RDS 的 答案：A

- A、轻松实现同城容灾
- B、有效控制云产品间的网络延迟
- C、将 RDS 的存储空间扩大 2 倍
- D、是为了防止 DDOS 攻击

116、关于 ODPS 用户和权限，说法正确的是？ 答案：A

- A、ODPS 目前为提供在项目空间中彻底移除一个用户及所有权限数据的命令
- B、每个项目空间在创建后，需要手工创建一个 admin 的角色，并且为该角色授予了确定的权限
- C、admin 角色可以将 admin 权限指派给用户
- D、admin 可以设定项目空间的安全配置，但是不能够改项目空间的鉴权模型

注解：

ODPS 目前不支持在项目空间中彻底移除一个用户及其所有权限数据。

每一个项目空间在创建时，会自动创建一个 admin 的角色，并且为该角色授予了确定的权限：能访问项目空间内的所有对象，能进行用户与角色管理，能对用户或角色进行授权。与项目空间 Owner 相比，admin 角色不能将 admin 权限指派给用户，不能设定项目空间的安全配置，不能修改项目空间的鉴权模型。Admin 角色所对应的权限不能被修改。

117、RDS for Mysql 监听的端口是多少？ 答案：A

- A、3306 端口
- B、8080 端口
- C、3600 端口
- D、3060 端口

117、ADS 创建普通表时，关于 updateType 属性那种说法是错误的？ 答案： B

- A、合法的取值为 realtime 或者 batch，该值不显式指定时，默认取 batch
- B、该值为 realtime 时，对应的表必须指定合法的主键并且必须指定二级分区
- C、通过该属性指定数据的更新方式
- D、该值为 batch 时，如果要支持增量导入，必须指定二级分区

注解：

根据表的数据更新方式不同，分析型数据库的表根据 updateType 分为批量更新表（仅能够离线批量更新数据）和实时更新表（能够通过 insert/delete 实时更新数据），用 updateType 以区分，如果 updateType 选项不填则默认为批量更新表。需要注意的是，updateType=realtime 暨为实时更新表时，必须指定合法的主键并且不能有二级分区。

!! 118、关于 ODPS policy 授权，说法不正确的是。答案： C

- A、支持根据 IP 授权
- B、支持授权的时间段
- C、支持根据数据敏感级授权
- D、支持表名通配符授权

https://help.aliyun.com/document_detail/28663.html?spm=5176.doc28664.6.137.IYVXWh

119、云盾 DDOS 基础防护功能如何开启 答案： A

- A、自动开通
- B、需要从管理控制台申请开通
- C、提交工单，24 小时内开通
- D、联系客户经理线下开通

!!!120、从 ODPS 导入到 ADS 中时，可以从哪个表中查询导入状态？ 答案： B

- A、master.job_status
- B、information_schema.job_instances
- C、user.job_instances
- D、performance_schema.job_status

注解：

数据导入命令发送后，数据并不会立刻导入到分析型数据库中，而是会在后台进行数据的导入工作。用户可以通过使用 SQL 命令或在 iDB Cloud 中进行查询数据导入状态。高级用户也可以直接在 information_schema 中查询全部数据导入的信息（具体见附录）。导入方法有三种：1、通过 SQL 语句查询；2、通过 iDB Cloud 查询数据导入状态；3、通过 Meta DB 进行查询。

121、以下哪种不是加载数据到 ODPS 的合理方法。 答案： C

- A、dship
- B、tunnel
- C、DT task
- D、mapreduce

122、Tunnel 服务的公网 end point 是。 答案： D

- A、<http://service.odps.aliyun.com/tunnel>
- B、<http://tunnel.odps.aliyun.com>

C、<http://tunnel.odps.aliyun.com/api>

D、<http://dt.odps.aliyun.com>

注解：

阿里云内网地址：

ODPS 地址：<http://odps-ext.aliyun-inc.com/api>

Tunnel 地址：<http://dt-ext.odps.aliyun-inc.com>

公网地址：

ODPS 地址：<http://service.odps.aliyun.com/api>

Tunnel 地址：<http://dt.odps.aliyun.com>

!!! 123、ODPS 中负责对等待提交的 task 进行排序的模块是？ 答案：A

A、scheduler

B、controller

C、executor

D、worker

备注：来自 ODPS 权威指南

scheduler 会把该 Instance 分解成多个任务,生成任务流 DAG 图,把可运行的 Task 放到其 taskpool 中。Taskpool 是个优先级队列

!!! 124、预付费模式 RDS 实例过期后，系统会自动进行什么操作。 答案：A

A、实例将会被锁定，无法进行写操作

B、过期当天删除实例

C、实例将会被锁定，无法进行读写

D、自动发邮件通知续费，实例可以正常使用

125、关于 RDS 数据转移功能，描述最准确的是 答案：D

A、支持实例间复制数据库

B、数据备份文件可以从控制台下载

C、支持云外 mysql 数据库的迁入

D、以上功能都支持

126、某图书管理系统使用 ODPS 进行离线数据分析，现在他们要将同一类目的书籍的名称拼成一个字符串，即将属于同一分组的字符串合并在一列展示，以下哪个内置函数可以实现这一功能。

答案:D

A、`regexp_instr`

B、`concat`

C、`split_part`

D、`wm_concat`

注解：

`wm_concat` 聚合函数，用途：用指定的 `separator` 做分隔符，链接 `str` 中的值。

其输入与输出是多对一的关系，即将多条输入记录聚合成一条输出值。可以与 SQL 中的 `group by` 语句联用。

`concat` 用途：返回值是将参数中的所有字符串连接在一起的结果。

127、除了正常数据以外，以下哪个数据会占用所购买的 RDS 实例空间 答案：B

- A、数据库回滚日志
- B、以上都是
- C、索引
- D、重做日志

注解：除了您正常的数据外，还有数据库实例正常运行所需要的空间，比如系统数据库、数据库回滚日志、重做日志、索引等。同时 RDS for MySQL 数据库实例产生的一天内的 Binlog 日志也会占用这个空间。这些文件会保证 RDS 实例正常稳定运行。

128、ADS 的普通表最多包含多少个列？ 答案：B

- A、2048 个
- B、1024 个
- C、4096 个
- D、512 个

注解：

(1) 一张事实表至少有一级 Hash 分区并且分区数不能小于 8 个； (2) 一个事实表组最多可以创建 256 个事实表； (3) 一个事实表最多不能超过 1024 个列。

129、ODPS SQL 任务中，发现 join 操作有数据倾斜时，可以考虑用什么方法优化。 答案：C

- A、增加 join 的 instance 数量
- B、调整 split size
- C、用 mapjoin hint
- D、调整左右表的顺序

130、OTS 单个分片下所有行的大小总和不能超过多大。 答案：1G

- A、无限制
- B、10G
- C、1M
- D、1T

!!!! 131、某企业使用 ADS 做数据分析，其中部分数据来源于 ODPS。技术人员在使用 LOAD DATA 命令从 ODPS 加载数据到 ADS 的过程中，碰到如下错误信息：ERROR 1105(HY000):You are not the owner of the source table,最有可能出错的原因是那种。 答案：C

- A、源表不存在
- B、ADS 的用户 garuda_build@aliyun.com 没有对源表的操作权限
- C、ODPS 的用户 garuda_build@aliyun.com 没有对源表的操作权限
- D、源表中没有数据，为空表

注解：

导入一个新的 ODPS 表时，公有云用户需要在 ODPS 中授权该表给分析型数据库的导入用账号

!!! 132、ODPS 中，允许的 SQL 的最大长度为。 答案：B

- A、8M
- B、2M
- C、1M

D、512K

!!!! 133、DRDS 的读策略中不包含以下哪种 答案: D

- A、主库读
- B、均衡
- C、自定义
- D、主库写

注解: 主库读、均衡、只读均衡、自定义四种。

!!!! 134、相比传统数据库支持的数据类型, 下列哪个数据类型是 ADS 特有的。答案: D

- A、boolean
- B、UDF_SYS_GEO_DISTANCE
- C、timestamp
- D、multivalue

135、ODPS 中, 以下哪种授权方式不能对角色授权。答案: A

- A、label
- B、policy
- C、ACL
- D、package

```
odps@hzjtdsj_odps>set label 4 to role hzjtdsj_zt;  
FAILED: mismatched input 'role' expecting 'USER' or 'TABLE'
```

136、查询数据需要 select 权限, 目前 ADS 中对 select 权限的管理的最小级别是。答案: A

- A、列
- B、表组
- C、数据库
- D、表

137、ODPS 中, 客户端提交任务成功后, 可以根据__A__来出巡作业状态

- A、Instance ID
- B、Task id
- C、SQL id
- D、JOB id

138、SQL 注入攻击就是攻击者通过欺骗什么服务器, 执行非授权的任意查询过程。答案: B

- A、备份
- B、数据库
- C、Web
- D、文件

注意: SQL 注入攻击就是攻击者通过欺骗数据库服务器执行非授权的任意查询过程。

!!!!!! 140、以下关于 ODPS MR 中的归并操作 (Combiner) 的说中, 正确的是。答案: C

- A、通常下 combiner 的处理逻辑和 map 一致

- B、Combiner 函数将 shuffle 阶段所有 key 值相等的记录进行归并
- C、当 map 输出数据后，框架会在 map 端对相同 key 值的数据进行本地的归并操作
- D、Combiner 是必须的，可以减少传输带宽

注解：

归并(Combiner)函数将 Shuffle 阶段相邻的 Record 进行归并。用户可以根据不同的业务逻辑选择是否使用归并函数。归并函数是 MapReduce 计算框架的一种优化，通常情况下 Combiner 的逻辑与 reduce 相同。当 map 输出数据后，框架会在 map 端对相同 key 值的数据进行本地的归并操作。

141、ADS 创建表组时，可以指定表组的副本数，关于副本数的说法，以下哪项是正确的。答案：

A

- A、表组的副本数越多，可以一定程度的增加 ADS 的最大承受的 QPS
- B、表组的副本数为 1 时，写入数据会最快，相比副本较多的情况，在数据导入时不可用时间会更短
- C、表组的副本数最小值为 2
- D、表组的副本数越少，表组的可用性会越高

注解：

minRedundancy 表示该表组的副本数，默认为 2，可配置为 1、2、4、8。需要注意的是，如果将一个表组配置为 1 副本，那么这个表组中的表在数据导入时会有不可用的时间。而将表组副本数配置为 4 或更高，可以一定程度的增加分析型数据库的最大承受的 QPS，但是数据存储费用也会相应增加。

executeTimeout 表示该表组的全局 Query 超时时间，默认为 30000，单位毫秒。

142、云盾反欺诈服务可以对哪类应用的风险进行防控 答案：B

- A、移动 APP
- B、WEB 应用和移动 APP
- C、WEB 应用
- D、后台数据分析

注解：

1. 反欺诈服务可支持 WEB 和移动平台。

2. 前端接入，不同的平台，方式不一样：

- Web 平台、移动平台 HTML5 应用，前端页面引入 JS 脚本。
- 移动平台 Native 应用，引入 SDK。

143、若 RDS 用户需要立即备份 RDS 实例，可在不更改备份策略的情况下，对 RDS 实例进行临时备份，临时备份每天可触发次数是多少 答案：C

- A、50 次
- B、10 次
- C、无限次
- D、100 次

!!! 144、ADS 可以使用 explain 命令查看执行计划。下面说法正确的是。答案:A

- A、explain 正确执行后返回信息中，第一行为逻辑计划，第二行为物理计划
- B、当用户发起一个 explain 查询到 ADS 系统后，ADS 会根据所有活跃节点上的数据来分析执行计划
- C、当用户通过查询的方式，想要获取文本格式的 explain 语句后，将会得到一个 XML 格式的串
- D、逻辑计划中，TableExecutor 节点表示参与计算的信息，JoinExecutor 表示 join 的节点信息

145、基于 RDS 系统架构以及只读实例的应用场景，需要满足 MySQL 版本最低为多少，否则是不能满足只读实例的应用场景 答案: D

- A、MySQL 5.2
- B、MySQL 5.1
- C、MySQL 5.5
- D、MySQL 5.6

!!! 146、下列关于 RDS 开发使用说法不正确的是 答案: C

- A、可以使用命令行或图形界面进行逻辑备份
- B、用户可以使用 mysqldump 进行数据导入
- C、支持第三方工具进行物理备份
- D、用户不具有数据库 root 权限

147、对于 ADS 的 hash 分区，哪个说法是正确的。答案: D

- A、ADS 表的第一级分区必须为 hash 分区，第二级没有强制要求，目前可以为 list 或者 hash
- B、hash 分区时一种动态分区类型，需要指定具体的分区列，最多为 2 列
- C、ADS 表的一级 hash 分区数最大为 1000
- D、若一张表为 batch 更新类型，且仅有一级 hash 分区，则每次导入数据时，会对已有数据进行全量覆盖

148、使用 ODPS tunnel 命令上传数据时，下列哪个参数可以设定是否忽略脏数据。答案: A

- A、-dbr
- B、-fd
- C、-dfp
- D、-rd

!!! 149、ODPS SQL 的执行计划逻辑上可以被看做是一个有向图，其中边的含义是。答案: D

- A、实例
- B、任务
- C、作业之间的依赖关系
- D、执行阶段的依赖关系

注解:

执行计划是由具有依赖关系的多个执行阶段(Stage)构成的。目前，执行计划逻辑上可以被看做一个有向图，图中的点是执行阶段，各个执行阶段的依赖关系是图的边。ODPS 会依照图(执行计划)中的依赖关系执行各个阶段。在同一个执行阶段内，会有多个进程，也称之为 Worker，共同完成该执行阶段的计算工作。同一个执行阶段的不同 Worker 只是处理的数据不同，执行逻辑完

全相同。计算型任务在执行时,会被实例化,用户可以操作这个 [实例(instance)] 的信息,例如:获取实例状态(Status Instance),终止实例运行(Kill Instance)等。

150、从 ODPS 导入数据到 ADS 的时候,导入路径的格式为?

- A、odps://project_name/table_name:partition_spec
- B、odps://project_name/table_name/partition_spec
- C、odps://project_name/table_name (partition_spec)
- D、odps://project_name/table_name partition_spec

151、OTS 是构建在阿里飞天系统之上的产品,它提供的是什么服务?

- A、在线高并发事务服务
- B、海量数据的非结构化读写服务
- C、在线 NoSQL 数据库服务
- D、在线关系型数据库服务

152、以下 acl 授权里,哪种是对 ODPS 的表赋予 select 权限? A

- A、grant select on table table_name to user
- B、grant desc on project project_name to user
- C、grant read on project project_name to user
- D、grant read on table table_name to user

153、在 RDS 只读实例变更配置过程中,只读实例在 RDS 管理控制台中可以进行以下哪种任务?
(D)

- A、更改配置
- B、内外网切换
- C、参数设置
- D、查看 SQL 日志

154、以下哪个不属于 ODPS 计算层的模块?

- A、Nuwa
- B、OTS
- C、Pangu
- D、Fuxi

!!! 155、如果 RDS 实例被锁定,则用户不能进行的操作有哪些? B

- A、select
- B、insert
- C、show view
- D、drop table

解释:如果您的数据库实例忽然只能 select 和 drop,不能做 insert、update、delete 操作,则表明您的数据库被锁定了。

!!! 156、通过 DMS 是否可以管理其他阿里云帐号下的 RDS 实例?

- A、只要有数据库帐号即可管理
- B、通过 DMS “实例授权”功能授权后即可管理

- C、完全不可以
- D、通过 RAM 授权后才能管理

157、使用OTS Java SDK 进行表操作时需要创建，以下哪项不属于构造此对象时必须提供的信息

- V
- A、实例名称
- B、AccountID
- C、Endpoint
- D、AccessKeySecret

158、使用 tunnel 命令上传数据时，如果不指定时间日期格式缺省值为？ C

- A、yyy/MM/dd HH:mm:ss
- B、yyy 年 MM 月 dd 日
- C、yyy-MM-dd HH:mm:ss
- D、yyyMMddHHmmss

159、某用户担心突发大流量 DDOS 攻击超过“DDOS 基础防护”的上限影响网站访问，可以选择开通哪个云盾服务，用较低的成本扩大 N 倍的防护能力？

- A、安全网络专家服务
- B、安全网络
- C、阿里绿网
- D、DDOS 高防 IP

160、ADS 存储数据的时候采用的模型是哪种？

- A、雪花型模型
- B、关系型模型
- C、星型模型
- D、多维模型

161、ODPS 表 T 中有 1000 万行数据，需要把这些数据全部导出到本地，下面哪一种是可行的方法？ D

- A、用 select 命令，把结果重定向到文件中
- B、管理控制台中选出该表，下载数据
- C、找到该表的路径，然后把该路径下的文件全部下载
- D、使用 tunnel 命令下载

!!!!!! 162、云盾先知计划的目的是帮助企业建立私有的什么中心？

- A、应急响应中心
- B、病毒防护中心
- C、私有数据中心
- D、远程备份中心

!!! 163、下列关于 ADS 用户的说法，哪些是正确的？ D

- A、对用户的授权支持 ACL 和 policy 两种
- B、在授权给一个用户之前，需要使用 add user on .*将用户添加到 ADS 库 dbname 中来

- C、数据库拥有者可以删除用户，意味着被删除用户无法再使用 ADS 的任何服务
- D、用户是被授权的数据库用户，由数据库拥有者授权时添加，无须开通 ADS 服务
- 备注：支持 ACL，不支持 policy 授权

进行授权 **SQL 语法**

```
GRANT privilege_type [(column_list)] [, privilege_type [(column_list)]] ... ON [object_type]
privilege_level TO user [, user] ...
```

不可以删除用户，可以回收用户

!!!!164、关于 ADS 中的 ECU，哪个说法是正确的？ B

A、接入节点和计算节点可以指定不同类型的 ECU，并且需要在创建 DB 时就确定，DB 一旦创建，ECU 类型不能修改

B、ECU 的数量，在 DB 创建后可以修改。即使在 DB 的使用过程中，也可以随时调整（扩容或者缩容）

C、接入节点和计算节点必须指定相同类型的 ECU，并且需要在创建 DB 时就确定，DB 一旦创建，ECU 类型不能修改

D、创建 DB 时可以指定 ECU 的数量，必须为偶数，最小为 4

解释：分析型数据库对每个用户的每一个 DB 会分配若干个计算节点（COMPUTENODE），以及若干个接入节点（BUFFERNODE），接入节点用于接收用户的应用前端连接等工作，计算节点用于存储用户的数据和进行计算，另外还有若干个用于放置实时化数据写入缓冲的缓冲节点（BUFFERNODE）。目前分析型数据库仅计算节点是用户可按 ECU 模式配置，分析型数据库会自动根据用户的计算节点的量来配置接入节点等其它角色的数量。

偶数个，至少 2 个

165、使用 ODPS tunnel 命令上传 log.txt 文件到表 t_log 中去，t_log 为分区表（p1 string, p2 string）。下列命令中正确的是？ C

A、tunnel upload log.txt t_log/p1= “b1” /p2= “b2”

B、tunnel upload log.txt t_log（p1= “b1”，p2= “b2”）

C、tunnel upload log.txt t_log/p1= “b1”，p2= “b2”

D、tunnel upload log.txt t_log/（p1= “b1”，p2= “b2”）

166、ODPS 通过云帐号服务器，对请求中的签名信息进行验证，返回——给 HTTP Server A

A、AccountID

B、MDS 码

C、AccessID

D、正确标识

解释：来自 ODPS 权威指南

请求发送给 HTTP server，该请求包括用户的 ACCESSID 和 MD5 签名信息，HTTP SERVER 在接收到请求后，会把 ACCESSID 和 MD5 签名发送给云账户服务进行用户认证，认证通过后，云账户服务会返回该用户的唯一 AccountID，在后续执行逻辑中，发送请求都是包括 accountID，而不是 accessID

167、有关 DDOS 攻击的描述，最准确的是？ D

- A、黑客自己购买了大量服务器，专门从事破坏活动
- B、主要是攻击目标是数据库
- C、攻击的目的是让服务超负载，然后窃取机密信息
- D、攻击的主要目的是让制定目标无法提供正常服务，甚至从互联网上消失，是目前最强大，最难防御的攻击之一

168、关于 ODPS 类型之间的类型，说法不正确的是？

- A、隐式类型转换是 ODPS 依据上下文使用环境自动进行的类型转换
- B、ODPS 支持的隐式类型转换规则与显式转换相同
- C、不支持的隐式类型转换会导致异常
- D、部分类型之间不可以通过显式的类型转换，但可以通过 SQL 的隐式转换完成

备注：部分类型之间不可以通过显式的类型转换，但可以通过 SQL 内建函数进行转换

隐式类型转换是指在运行时，由 MaxCompute 依据上下文使用环境及类型转换规则自动进行的类型转换。MaxCompute 支持的隐式类型转换规则与显式转换相同：

!!!169、关于 OTS 中数据查询的说法正确的是？ D

- A、查询时仅提供分片键即可
- B、可以创建索引，然后基于索引列进行查询
- C、查询时可以仅提供主键的一部分，但是和组成主键的列的顺序有关
- D、单行查询必须提供主键

!!!!!!170、ODPS SQL 目前最多支持多少个并发 insert overwrite/into 操作？

- A、1000 个
- B、128 个
- C、64 个
- D、500 个

171、对于 ODPS，下列描述正确的是？

- A、在使用多路插入时，可对同一张表同时执行 insert into 和 insert overwrite
- B、使用动态分区插入时，动态分区列必须在 select 列表中
- C、ODPS SQL 支持 between 子句
- D、对于分区表，可指定某个分区的 lifecycle，其中 n 为正整数，单位为天

172、ODPS 的云帐号认证使用什么机制？

- A、智能卡
- B、LDAP
- C、数字证书
- D、消息签名

!!!!!!173、ADS 一个普通表组最多可以包含多少个表？ A

- A、256 个
- B、1000 个
- C、无限制
- D、128 个

备注：事实表的创建上，默认有如下限制：(1) 一张事实表至少有一级 Hash 分区并且分区数不能小于 8 个；(2) 一个事实表组最多可以创建 256 个事实表；(3) 一个事实表最多不能超过 1024 个列。

174、通过备份文件及日志文件，RDS 可以将实例数据恢复至多少天内的任意时刻？

- A、5 天
- B、7 天
- C、6 天
- D、3 天

多选题

1、下列哪些 SQL 可以清空表 T？ AB

- A、insert overwrite table T select * from T where 1=2;
- B、truncate table T;
- C、delete from table T;
- D、delete table T;

2、对于 OTS 表的预留读写吞吐量，以下各项中描述正确的是？ C

- A、1 单位读能力表示该表每秒可以读 1KB 数据
- B、操作数据大小不足 1KB 的部分向上取整
- C、当表上的操作过于频繁导致预留读/写吞吐量不足以进行更多操作时，OTS 会返回 OTSQuotaExhausted 错误给应用程序
- D、1 单位写能力表示每秒可以写 1KB 数据

3、ODPS 的某个带有 Join 的 SQL 运行效率很低，可以采取以下哪些措施来检验或者改善？ AC

- A、检查连接键是否存在数据倾斜
- B、检查是否有多对多连接
- C、如小表关联大表，可以考虑使用 mapjoin 优化
- D、检查是否有笛卡尔积

4、开发 ODPS 的 UDF，可以使用哪些语言？ B

- A、C++
- B、Java
- C、Python
- D、Ruby

5、以下场景中，哪些适合使用 ODPS？ABC

- A、PB 级离线日志分析
- B、基于海量数据的用户特征和兴趣挖掘
- C、大型互联网企业的数据仓库和 BI 分析
- D、电子商务网站的在线交易

6、关于 ADS 多值列的说法正确的包括哪些？ACD

- A、可以存入 string 类型的多个值，以逗号分隔
- B、可以直接在 select 中使用该列
- C、可以使用 in, contains 条件对该列的单个值进行查询
- D、不能在 group by 中直接使用该列

(多值列数据存入分析型数据库后，可使用 in, contains 条件对该列的单个值进行查询，枚举查询后该列的每个值可像一个普通列一样进行各类操作。但是不允许在没有进行枚举查询时对该列直接 select 或在 group by 中使用该列。)

7、以下 ODPS SQL 中语法错误的是？ AB

- A、select col2 from tbl group by substr(col2,2);
- B、select cols as c from tbl order by col2 limit 100
- C、select substr (col2,2) from tbl group by substr(col2,2);
- D、select col2 as c from tbl order by c limit 100

group By 只能是前后一致

8、ODPS 的控制层由哪几部分组成？ BCD

- A、Master
- B、Worker
- C、Executor
- D、Scheduler

9、RDS 关系型数据库服务支持如下哪些数据库。答案：AC

- A、SQLServer
- B、oracle
- C、MySQL
- D、access

10、如何知道一个 ODPS SQL 在运行时占用了多少 instance？ CD

- A、根据表的大小，自己计算
- B、使用 ps -ef 查看相关进程数
- C、用 odpscmd 提交时，从返回的状态中查看
- D、打开 logview 查看

!!!! 11、对于 ADS 的索引说法正确的是哪一项。答案：ABC

- A、一个列可以设置列属性 disableIndex=true，用于屏蔽 ADS 的默认列索引
- B、ADS 弱化了索引的概念，用户无需亲自为自己的数据表配置索引
- C、ADS 拥有高度智能的自动化索引机制，会搜集相关信息，去自动创建索引
- D、用户需要对某列进行 hash Join 时，无论是事实表之间的 join 还是事实表和维度表的 join，都需要考虑手工处理索引问题。

注解：

列属性上，一个列可以设置列属性 disableIndex = true，用于屏蔽分析型数据库的默认列索引，不过需要注意的是，要如此做，则这个列应该不在实际查询中所筛选和计算的。

分析型数据库中主键的概念是弱化的，分析型数据库不要求一个表有主键，有主键的表的性能和用法上和没有主键的表之间没有任何区别。

分析型数据库拥有高度智能的自动化索引机制，所以通常用户无需亲自为自己的数据表配置索

引。

一种情况例外：暨用户需要对某列进行 Hash Join 时，无论是事实表之间的 Join 还是事实表和维度表的 Join，都需要为事实表的该列建立索引。

12、以下哪些是 ODPS SQL 的优点。答案：ACD

- A、相对于使用 API 或者 SDK 编程，ODPS SQL 的学习成本较低
- B、可将其他数据库中的 SQL 语句迁移到 ODPS 上来，无需修改或者重新开发
- C、提供了丰富的内置函数
- D、用户不需要了解分布式概念

注解：

ODPS 的 SQL 语法与 Oracle，MySQL 有一定差别，用户无法将其他数据库中得 SQL 语句无缝迁移到 ODPS 上来。

14、在使用 ODPS tunnel 命令 upload 数据的时候，关于一些常见参数，说法正确的包括以下哪些项。答案：ACD

- A、-c 本地数据文件编码，缺省为 utf8
- B、-dfp DateTime 类型数据格式，缺省为 yyyyMMddHHmmss
- C、-cp 指定是否在本地压缩后再上传，减少网络流量，缺省为 true
- D、-fd 本地数据文件的列分隔符，缺省为逗号

注解：

- bs,--block-size:每次上传至 Tunnel 的数据块大小，默认值：100MiB (MiB=1024*1024B)。
- c,--charset: 指定本地数据文件编码，默认为 'UTF-8'；不设定，默认下载源数据。
- cp,--compress: 指定是否在本地压缩后再上传，减少网络流量，默认开启。
- dfp: DateTime 类型数据格式，默认为 'yyyy-MM-dd HH:mm:ss'；
- dbr: 是否忽略脏数据(多列，少列，列数据类型不匹配等情况)。值为 'true' 时，将全部不符合表定义的数据忽略。 值为 'false'，若遇到脏数据，则给出错误提示信息，目标表内的原始数据不会被污染。
- e: 指定 odps 的 endpoint.
- fd: 本地数据文件的列分割符，默认为逗号 ','；
- h: 数据文件是否包括表头，如果为 true，则 dship 会跳过表头从第二行开始上传数据。
- mbr,--max-bad-records: 默认情况下，当上传的脏数据超过 1000 条时，上传动作终止。通过此参数，可以调整可容忍的脏数据量；
- ni: NULL 数据标志符，默认为 ""(空字符串)；
- rd: 本地数据文件的行分割符，默认为 '\r\n'；
- s: 是否扫描本地数据文件，默认值为 'false'。值为 'true' 时，先扫描数据，若数据格式正确，再导入数据。值为 'false'， 不扫描数据，直接进行数据导入。值为 'only' 时，仅进行扫描本地数据，扫描结束后不继续导入数据。
- te:指定 tunnel 的 Endpoint.
- tz:指定时区。默认为本地时区：Asia/Shanghai。

15、OTS 支持多种数据类型，包括 答案：ABD

- A、binary
- B、double
- C、varchar

D、integer

注解：支持多种数据类型(Integer、Boolean、Double、String、Binary)

!!! 16、使用 ODPS 的客户端工具 odpscmd 时，如何进入一个已有的项目.答案：CD

A、使用 odpscmd 登录时，通过 -p 参数，指定登录使用的项目空间

B、如果账户下就一个 project，登录后就会默认进入该项目

C、配置文件中指定 project_name，登录后即默认进入项目

D、进入 odpscmd 后，执行 use

解释：参数：--project=<prj_name> use project

如果账户下就一个 project，登录后就不会默认进入该项目,已动手测试，不可以回默认进入

17、对于 ODPS 的多路输出 (multi insert)，哪些项描述是不正确的。答案：BD

A、对于源表为分区表的，相同的一个分区不能出现多次

B、目标表为分区表的，对于不同的分区，可以混合使用 insert into

和 insert overwrite，对于相同的分区则不可混用

C、一般情况下，单个 sql 中最多可以写 128 路输出

D、对于源表为未分区表的，该表可以出现多次

注解：

一般情况下，单个 SQL 里最多可以写 128 路输出，超过 128 路报错语法错误。

在一个 multi insert 中，对于分区表，同一个目标分区不可以出现多次; 对于未分区表，该表不能出现多次。

对于同一张分区表的不同分区，不能同时有 insert overwrite 和 insert into 操作，否则报错返回。

18、ADS 使用 DUMP DATA 导出数据时，以限制导出行数为 1000 为例 (LIMIT 1000)，哪些说法是正确的。答案：ABD

A、实际数据行数可能小于 1000

B、实际数据行数可能等于 1000

C、实际数据行数严格等于 1000

D、实际数据行数可能稍大于 1000

备注：导出方式对海量数据的计算输出具有良好的性能 (百万行数据导出在数百毫秒数据级)，但是，对于数据精确度有一定牺牲，即实际返回的数据行数，可能是不完全精确。以限制导出行数为 1000 为例 (LIMIT 1000)：

实际数据行数可能稍大于 1000，例如此时有 120 个数据分片，则等同于每个分区明确指定 "LIMIT 9"，最多肯能返回 1080

实际数据行数可能稍小于 1000，如果符合条件的行数的总数小于 1000

实际数据行数可能稍小于 1000，如果数据分片很均匀，例如此时有 120 个数据分片，如果某些分片返回数据行小于 9 的话，则等同于每个分区明确指定 "LIMIT 9"

19、ADS 中，同一个表组的表具有以下哪些相同的属性.答案：BC

A、记录条数

B、一级 hash 分区数

C、副本数

D、二级 List 分区数

注解:

ADS 要求一个表组下所有表的一级分区数目一致。

一个表组所有表副本数一致。

!!!20、使用 ADS 可以很好的支持即席查询，对一些复杂的查询，也能在很短的时间内返回处理结果。以下关于 ADS 中查询涉及的说法正确的有。答案：AD

- A、如果用户的查询的条件会指定聚集列的内容或者范围，那么这样的查询性能便会有较大的提升
- B、处于关联方便的考虑，普通表可以放在维度表组上，这样该表和其他任意表关联都不在受表组的限制
- C、ADS 中 join 会优先采用 hash join，如果不满足 hash join 的条件，优化器才会选取其他的关联方法
- D、需要关联查询的普通表，需要放在同一个表组上

!!!21、在 RDS 只读实例变更配置过程中，RDS 主实例无法进行哪些任务 BCD

- A、查看 SQL 审计日志
- B、内外网切换
- C、备份实例
- D、创建实例

!!!22、云盾反欺诈功能可以防范哪些安全威胁 答案：ABC

- A、营销作弊
- B、垃圾注册
- C、盗卡支付
- D、CC 攻击

注解：垃圾注册、暴力破解、撞库、营销作弊、盗卡支付、垃圾广告或内容、恶意登入。

!!! 23、ODPS 通过 eclipse plugin 提供了 MR 的编程框架，开发时通常需要添加以下哪几个类.答案：ACD

- A、reducer
- B、mapreduce driver
- C、combiner
- D、mapper

```
job.setMapperClass(TokenizerMapper.class);
    job.setCombinerClass(SumCombiner.class);
    job.setReducerClass(SumReducer.class);
```

!!! 24、DMS（原 Idb Cloud）是一款访问管理云端数据的 WEB 服务，覆盖 RDS，ADS 等阿里云环境，DMS 具有以下哪些功能 答案：ACD

- A、对象管理
- B、数据库创建
- C、实例管理
- D、数据管理

解释：数据管理（Data Management）支持 MySQL、SQL Server、PostgreSQL、MongoDB、Redis 等关系型数据库和 NoSQL 的数据库管理，同时还支持 Linux 服务器管理。它是一种集数据管理、结构管理、实例管理、诊断优化、文件管理、命令终端、实时监控和系统管理于一体的数据管理服务。用户使用数据管理服务实现服务器和数据库统一管理入口的同时，还可以获得安全、易用、专业的数据管理体验。

28、关于 ODPS 中小文件的说法，正确的是 AB

- A、在 reduce 计算过程或者实时 tunnel 数据采集过程，会产生大量小文件
- B、小文件越多，会造成浪费资源，影响整体的执行性能，并且影响磁盘空间的利用率
- C、目前提供两种可供选择的小文件合并的方法：ALTER 合并模式和 SQL 合并模式
- D、大小小于盘古文件系统中的块 BLOCK 大小的文件，称之为小文件

29、关于 ODPS 中的 admin 角色，哪些说法是正确的。答案：CD

- A、包含的权限只有 owner 可以调整
- B、可以修改项目空间的鉴权模型
- C、不能将 admin 权限指派给用户
- D、不能设定项目空间的安全配置

注解：

每一个项目空间在创建时，会自动创建一个 admin 的角色，并且为该角色授予了确定的权限：能访问项目空间内的所有对象，能进行用户与角色管理，能对用户或角色进行授权。与项目空间 Owner 相比，admin 角色不能将 admin 权限指派给用户，不能设定项目空间的安全配置，不能修改项目空间的鉴权模型。Admin 角色所对应的权限不能被修改。

30、odps 授权操作一般涉及到以下几个要素。答案：ABD

- A、操作（action）
- B、客体（object）
- C、选项（option）
- D、主体（subject）

备注：授权操作一般涉及到三个要素：主体(Subject，可以是用户也可以是角色)，客体(Object)和操作(Action)

31、下面那些是 RDS 提供的系统性能监控项目 答案：ABCD

- A、IOPS
- B、网络流量
- C、TPS
- D、CPU 利用率

32、关于 ODPS 跨项目资源共享的说法正确的是。答案：ABD

- A、package 是一种跨项目空间共享数据及资源的机制，主要用于解决跨项目空间的用户授权问题
- B、如果某个 package 还有人在使用，则该 package 的 owner 无法删除此 package
- C、添加到 package 中的不仅仅是对象本身，还包括相应的操作权限，如不显式指定权限，则默认为读写权限
- D、添加资源时支持的对象类型不包括 project 类型，即不能将 project 添加到 package 里

注解:

- 目前支持的对象类型不包括 Project 类型,也就是不允许通过 Package 在其他 Project 中创对象。
- 添加到 Package 中的不仅仅是对象本身,还包括相应的操作权限。当没有通过[with privileges privileges]来指定操作权限时,默认为只读权限,即 Read/Describe/Select。”对象及其权限 ”被看作一个整体,添加后不可被更新。若有需要,只能删除和重新添加。

!!! 33、ADS 中不同型号的 ECU 标识着不同的配置。答案: ABD

- A、内存大小
- B、磁盘空间
- C、带宽
- D、Vcpu 核数

注解:

分析型数据库以 ECU (弹性计算单元) 作为资源计量的最小单位。ECU (弹性计算单元) 拥有多种型号, 每种型号的 ECU, 标识着不同的 vCPU 核数、内存大小、磁盘空间大小。

34、关于 ADS 的 ECU 说法哪些是正确的。答案: ABC

- A、ECU 的型号在 DB 创建后不可更改
- B、资源计量的最小单位
- C、ECU 数量可以在使用中扩容或者缩容
- D、ECU 数量变化是瞬时的同步操作。

注解:

ECU (弹性计算单元) 作为资源计量的最小单位。ECU 型号 DB 创建后不可修改, ECU 数量可以在使用中随时调整 (扩容/缩容)。缩容和扩容都不是瞬时的同步操作, 可以使用元数据查询状态。

35、DMS (原 Idb Cloud) 是一款访问管理云端数据的 web 服务, 覆盖 RDS、ADS 等阿里云环境, DMS 支持以下哪些数据源。答案: ABCD

- A、SQL Sdserver
- B、MySQL
- C、PostgreSQL
- D、ADS

36、对于数据库类型为 MySQL 的 RDS 实例, 有以下哪些种连接方法 答案: ABCD

- A、Navicat for MySQL
- B、使用 mvsql 命令登入
- C、Idb CLOUD (控制台登陆)
- D、mysql-front

37、ODPS 中项目普通用户想要使用 select * from 访问一张非本人创建的本项目的表 A, 至少需要以下哪些权限。答案:BD

- A、表 A 的 describe 权限
- B、表 A 的 select 权限
- C、项目 P 的 AccessAllow 权限
- D、项目 P 的 createinstance 权限

注解:

权限描述中 **Project** 类型对象的 **CreateTable** 操作, **Table** 类型的 **Select**、**Alter**、**Update**、**Drop** 操作需要与 **Project** 对象的 **CreateInstance** 操作权限配合使用。单独使用上述几种权限而没有指派 **CreateInstance** 权限是无法完成对应操作的。这与 ODPS 的内部实现相关。类似的, **Table** 的 **Select** 权限也要与 **CreateInstance** 权限配合使用。使用时请注意。

!!!38、ADS 中不支持以下哪些表连接方式。答案: ACD

- A、right join
- B、笛卡尔积
- C、semi join
- D、full join

注解:

不支持或语义错误:

SELECT ... FROM A FULL JOIN B: 不支持全连接;

SELECT ... FROM A RIGHT JOIN B: 不支持右连接, 需要转换为左连接;

SELECT ... FROM A SEMI JOIN B: 不支持半连接;

SELECT ... FROM A, B: 单表或多表连接, 但没有 ON 条件

SELECT ... FROM A, B WHERE A.a = B.b: 单表或多表连接在 WHERE 子句中有隐含 ON 条件, 但是没有 on 子句的, 暂不支持

39、某企业用户目前的在线订单系统运行在 mysql 版的 RDS 上, 因为业务量激增, 公司决定采用 DRDS, 即将现有系统从 RDS 迁移到 DRDS 上去。从项目实施部署的角度来讲, 以下哪些说法是正确的 答案: ABCD

- A、需要评估存储容量以及表的大小, 以合理规划初始化 DRDS 的实例和库的个数
- B、需要评估哪些应用不支持, 需要做应用改造
- C、需要评估当前应用, 以识别哪些应用需要改造
- D、需要收集当前表的大小和具体如何使用等, 用来判断哪些表需要拆分, 哪些不需要

40、以下哪些场景可以使用 ODPS MR 实现。答案: ABCD

- A、Web 访问日志分析: 分析用户访问行为, 个性化推荐等
- B、搜索, 比如 pagerank、网页爬取等
- C、机器学习: 监督学习、无监督学习、分类算法等
- D、广告推荐: 用户点击购买行为预测

注解:

搜索: 网页爬取、倒排索引、PageRank。

Web 访问日志分析: 分析和挖掘用户在 web 上的访问、购物行为特征, 实现个性化推荐; 分析用户访问行为。

文本统计分析: 比如莫言小说的 WordCount、词频 TFIDF 分析; 学术论文、专利文献的引用分析和统计; 维基百科数据分析等。

海量数据挖掘: 非结构化数据、时空数据、图像数据的挖掘。

机器学习: 监督学习、无监督学习、分类算法如决策树、SVM 等。

自然语言处理: 基于大数据的训练和预测; 基于语料库构建单词同现矩阵, 频繁项集数据挖掘、重复文档检测等。

广告推荐: 用户点击 (CTR) 和购买行为 (CVR) 预测。

41、ODPS MR 中哪些操作会因为 Java Sandbox 而报错“access denied”? 答案: ABCD

- A、起子进程执行 linux 命令
- B、直接访问本地文件
- C、java 反射限制
- D、获取本地 IP

注解:

不允许直接访问本地文件, 只能通过 ODPS MapReduce/Graph 提供的接口间接访问: 读取--resources 选项指定的资源, 包括文件、Jar 包和资源表等; 通过 System.out 和 System.err 输出日志信息, 可以通过 ODPS 客户端的 Log 命令查看日志信息;

不允许直接访问分布式文件系统, 只能透过 ODPS MapReduce/Graph 访问到表的记录:

不允许 JNI 调用限制;

不允许创建 Java 线程, 不允许起子进程执行 Linux 命令;

不允许访问网络, 包括获取本地 IP 地址等等都会被禁止;

Java 反射限制, “suppressAccessChecks”权限被禁止, 无法 setAccessible 某个 private 的属性或方法, 以达到读取 private 属性或调用 private 方法的目的。

!! 42、以下哪几项云盾服务可以进行 DDOS 攻击防护 答案: ABD

- A、DDOS 基础防护
- B、DDOS 高防 IP
- C、态势感知
- D、安全网络

!!! 43、当 ODPS 中某个字段取值为 double, 以下手法正确的是。答案: AC

- A、所有数据类型都可以由空值
- B、如果该表为分区表, 则该字段可以作为分区列
- C、double 类型中使用 inf 表示无穷大
- D、double 类型中使用 NaN 标识无穷小

注解:

1.有效位数 16 位

2.上传支持科学计数法表示

3.下载仅使用数字表示

4.最大值 : 1.7976931348623157E308

5.最小值 : 4.9E-324

6.无穷大 : Infinity

7.无穷小 : -Infinity

44、OTS 以什么样的形式组织数据? CD

A: schema B: 表组 C: 表 D: 实例

Table Store 以实例和表的形式组织数据, 通过数据分片和负载均衡技术, 达到规模的无缝扩展

!!! 45、ADS 提供哪些批量导入数据的方式? CD

A: 通过 batchinsert 命令 B: 通过 rest-like api 方式
C: 通过 insert D: data pipeline 系列命令

46、ODPS SQL 支持的 Join 操作类型包括哪些？ A B C D

A: INNER JOIN B: LEFT OUTER JOIN C: FULL OUTER JOIN D: RIGHT OUTER JOIN

!!!47、关于 ODPS 中 tunnel upload 的分隔符，说法正确的是 B C

A: 列分隔符不能包含行分隔符

B: 支持多个字符的行列分隔符

C: 转义字符分隔符，在命令行方式下在只支持\r、\n 和\t D: 缺省为空格

!!!48、ADS 的数据导入 job 提交后，可以通过以下哪些方法查询当前的导入状态？ BC

A: 使用 log 命令：log

B: 使用 SQL: select state from information_schema.current_job where job_ =

C: 使用 DMS，通过菜单项查看：导入导出=>导入状态

D: 使用命令：show states for

49、以下哪些方法可以支持 ODPS 中受保护项目空间的数据流出？ BD

A: 使用 owner 通过 policy 授权

B: 设置例外（Exception Policy）

C: 使用 owner 通过 ACL 授权

D: 设置项目互信（TrustedProject）

!!!!50、ADS 创建普通表的时候有那些限制？ ABC

A: 一张普通表至少有一级 Hash 分区并且分区数不能小于 8 个

B: 一个普通表组最多可以创建 256 个普通表

C: 一个普通表最多不能超过 1024 列

D: 一张普通表至少有一级 Hash 分区并且分区数不能多于 1000 个

51、支持 OLTP 的数据库系统包括哪些特点： A B C D

A: 数据原子性

B: 擅长事务处理

C: 数据一致性

D: 很好的支持频繁的数据插入和修改

!!!!52、RDS MySQL5.5 新版本特性有哪些？ A B C

A: 插入缓冲的改进提升数据库写性能 B: 支持 InnoDB 数据页压缩

C: 支持多个回滚段增加并发处理能力 D: 分区功能的增强，但是不可以 drop 分区表

1、提高性能和可扩展性：支持多个回滚段增加并发处理能力，插入缓冲的改进提升数据库写性能。

2、提升易管理性和效率：支持 InnoDB 数据页压缩，数据页的压缩使数据文件体积变小，减少磁盘 I/O，提高吞吐量。

3、提高可用性：分区功能的增强，可 drop 分区表。

4、改善检测和诊断：引入 performancn_shema,用于监控 mysql 监控服务器运行时的性能。

!!!53、ODPS MR 可以通过控制 Reducer 的个数来进行调优，以下哪些方法可以帮用户调节 Reducer 的个数？ ABC

A: 可以通过调整 Mapper 的个数间接的影响 Reducer 的个数（Reducer 的个数默认为 Mapper 个数的 1 / 4）

B: 在编程的时候(以 Java 为例)，初始化 Reducer 的方法里，可手工指定 Reducer 个数

C: 可以通过 setNumReducerTasks 来设置

D: 可以通过 addReducer 随意增加 Reducer 的个数

54、使用用户管理控制台来创建 ADS 时，需要提供 ECU 的哪些信息？ ABCD

A: vCPU B: 存储 C: 内存 D: 数量

ECU（弹性计算单元）拥有多种型号，每种型号的 ECU，标识着不同的 vCPU 核数、内存大小、磁盘空间大小

用户在创建数据库时需要根据自己的需求选择这个数据库的 ECU 型号，以及初始的 ECU 数量（必须是偶数个，至少两个），ECU 型号 DB 创建后不可修改，ECU 数量可以在使用中随时调整（扩容/缩容）

55、RDS 实例空间主要包括哪些文件？ A B C D

A: 日志文件 B: 数据文件 C: 系统文件 D: 临时文件

56、ODPS 的自定义函数包括哪些类？ A B C

A: 自定义标量函数 B: 自定义聚合函数 C: 自定义表值函数 D: 自定义扩展函数

57、ODPS 中，当一个大表和多个小表做连接时，可以考虑使用 mapjoin，以下关于 mapjoin 的限制，说法正确的是？ A B

A: left outer join 的左表必须是大表，inner join 左表或者右表都可以为大表

B: full outer join 不能使用 mapjoin

C: 最多支持 8 张小表

D: 所有小表占用内存综合不能超过 256M

- ODPS SQL 支持的 Join 操作类型包括：{LEFT OUTER|RIGHT OUTER|FULL OUTER|INNER} JOIN；
- ODPS SQL 目前最多支持 16 个并发 Join 操作；
- 在 mapjoin 中，最多支持 6 张小表的 mapjoin;

58、云盾 DDoS 防护功能可以防护哪些类型的攻击 A B C D

A: ACK Flood B: SYN Flood C: UDP Flood D: ICMP Flood

阿里云云盾可以防护 SYN Flood，UDP Flood，ACK Flood，ICMP Flood，DNS Flood，CC 攻击等 4 到 7 层 DDoS 的攻击

59、云数据库（RDS）具有以下哪些优势？ ABCD

A、高可用性，采用主从热备的架构

B、易用性，控制台自助迁移和管理

C、高性能、底层使用高端服务器

D、高安全性、具有网络防护能力和 SQL 审计

60、为了保证应用获得可预期的性能，需要设置 OTS 表的： CD

A、预留存储空间

B、预留处理内存

C、预留写吞吐量

D、预留读吞吐量

备注：Table Store 根据表的预留读/写吞吐量进行资源的调度和预留，从而保证应用获得可预期的性能。

61、RDS 扩展的功能有哪些？ABC

- A、备份恢复
- B、安全设置
- C、监控查询
- D、负载均衡

62、某企业因为业务发展带来了数据膨胀，每日产生 4T 数据，现在他们考虑要把现有的分析系统迁移到云上。通过系统分析，针对目前系统的状况去做云产品的选型。他们梳理了一下目前支持的应用类型，大体上有以下几种：提供给一线业务人员前一日以及最近三个月的经营状况的报表分析，同时要向客户提供 30 天内业务数据的即席查询。

请从查询效率和成本两个维度考虑，你认为需要哪些产品组成合适的解决方案？BC

- E、DRDS
- F、RDS
- G、ODPS
- H、ADS

63、ODPS 的资源类型包括以下哪些类？ABCD

- A、Archive 类型
- B、Table 类型
- C、File 类型
- D、Jar 类型

注释：file/archive/table/jar 表明资源类型，资源类型的介绍请参考[资源\(Resource\)](#)；

64、以下哪些属于 ADS 的特色功能？ABD

- A、聚集列
- B、多值列
- C、空间列
- D、智能自动索引

65、关于 ADS 和 DMS 功能，说法正确的有？ABCD

- A、可以进行用户和权限管理
- B、可以导入数据
- C、可以进行 DB 容量管理
- D、可以查看性能报告

66、OTS 写操作具有以下特性：AB

- A、原子性
- B、强一致性
- C、最终一致性
- D、事务一致性

67、RDS FOR MySQL 的高可用架构包括哪些功能？ABD

- A、同城容灾
- B、异地容灾
- C、防 DDos 攻击

1.1 开通分析型数据库...

1.2 创建数据库

1.3 使用DMS创建和管...

1.4 导入数据

1.5 在应用中连接和使...

1.6 新增用户和授权

1.7 删除表或数据库

D、主备架构

68、ODPS SQL 不支持的包括哪些？ **ABCD**

- A、delete
- B、update
- C、索引
- D、事务

备注：ODPS SQL 不支持事务、索引及 Update/Delete 等操作，同时 ODPS 的 SQL 语法与 Oracle，MySQL 有一定差别，用户无法将其他数据库中的 SQL 语句无缝迁移到 ODPS 上来。

69、OTS API 中提供了哪些 Action 可以进行多行数据操作？ **ABC**

- A、GetRange
- B、BatchWriteRow
- C、BatchGetRow
- D、GetMultiRow

单行数据操作：GetRow / PutRow / UpdateRow / DeleteRow

多行数据操作：GetRange / BatchGetRow / BatchWriteRow

70、ADS 的普通表分区，支持哪些分区方式？ **AC**

- A、List
- B、Interval
- C、Hash
- D、Range

!!!71、云盾先知计划包括以下哪些功能？ **ABCD**

- A、漏洞收集
- B、审核漏洞
- C、自主设定奖励计划
- D、协助漏洞修复

!!72、以下哪些功能是 RDS 服务天生就具备的？ **BCD**

- A、自动备份，支持 15 天内任意时刻数据恢复：7 天
- B、数据库主辅备份
- C、数据库安全漏洞实时修复
- D、后台有专业团队提供运维服务

!!73、下列 OLAP 中，无需预先建模的是哪一项？ **C**

- A、HOLAP
- B、MOLAP
- C、ADS
- D、ROLAP

备注：我们通常将 OLAP 系统分为 MOLAP、ROLAP 和 HOLAP 三种。多维 OLAP (Multi-Dimensional OLAP，简称 MOLAP)，是预先根据数据需要分析的维度进行建模，在数据的物理存储层面以“立方体”(Cube)的结构进行存储，具有查询速度快等的优点，但是数据必须预先建模，无法依据使

用者的意愿进行即时的修改。而关系型 OLAP (RelationalOLAP, 简称 ROLAP), 则使用类似关系数据库的模型进行数据存储, 通过类似 SQL 等语言进行查询和计算, 优点是数据查询计算自由, 可以灵活的根据使用者的要求进行分析, 但是缺点是在海量数据的情况下分析计算缓慢。至于 HOLAP, 则是 MOLAP 和 ROLAP 的混合模式。而阿里云分析型数据库 (原名: 分析数据库服务 ADS), 则是一套 RT-OLAP (Realtime OLAP, 实时 OLAP) 系统。在数据存储模型上, 采用自由灵活的关系模型存储, 可以使用 SQL 进行自由灵活的计算分析, 无需预先建模, 而利用云计算技术, 分析型数据库可以在处理百亿条甚至更多量级的数据上达到甚至超越 MOLAP 类系统的处理性能, 真正实现百亿数据毫秒级计算。

!! 74、ODPS tunnel 命令目前支持的操作包括哪些? BC

- A、merge
- B、purge
- C、upload
- D、list

备注: 支持 upload、download、resume、show、purge、help

75、当 RDS 实例所占用的磁盘容量超过了所购买的最大磁盘容量时, RDS 实例会被系统锁定。实例锁定后, 用户只有以下哪些权限? ABC

- A、select
- B、drop table
- C、show view
- D、insert

备注: 您的实例被锁定后, 您只有 select、show view 和 drop table 的权限

!!! 77、RDS 对 SQL server 有哪些限制? ABCD

- A 只能在用户数据库内建立对象, 不可在 master、msdb、tempdb、model 等系统数据库中建立用户对象
- B 不可更改或销毁 SQL serverMirror 镜像
- C 用户不可在 SQL server 中指定备份计划, 可在 RDS 管理后台进行相应操作
- D 暂时不支持分布式事务 (DTC)

RDS for SQL Server 有哪些限制

- 1、不允许用户在数据库直接修改用户密码或者创建数据库和用户。此功能在数据库中已被屏蔽, 需在 RDS 管理页面中操作。
- 2、不可更改或销毁 SQL Server Mirror 镜像
- 3、只能在用户数据库内建立对象, 不可在 master、msdb、tempdb、model 等系统数据库中建立用户对象
- 4、不提供 SSIS 服务
- 5、不提供 SSAS 服务
- 6、不提供 SSRS 服务
- 7、用户不可在 SQL Server 中指定备份计划, 可在 RDS 管理后台进行相应操作
- 8、暂不支持分布式事务 (DTC)
- 9、不允许使用 RestoreDB 重置数据库, 该操作将导致 Mirror 损坏
- 10、不提供 broker 服务

!! 78、常见的 OLAP 分为哪三种? ABD

- A ROLAP
- B HOLAP
- C QOLAP
- D MOLAP

备注: 我们通常将 OLAP 系统分为 MOLAP、ROLAP 和 HOLAP 三种

79、某餐饮 O2O 公司,立足于分析签约餐厅的用户用餐数据,提供用户菜品推荐,以下说法正确的有? AB

- A 可以利用 ODPS 的海量数据处理能力,对签约餐厅的用户用餐数据进行离线分析
- B 可以使用 ADS 的多值列,可以在一条用餐记录中包含多个菜品,减少 join 的次数,提供处理效率
- C 可以使用 ADS 的实时插入特性的表,可以支持一些 OLTP 类的场景
- D 可以利用 OTS 的高并发低延时的特性,支持一些在线分析,即将产生的业务数据存入 OTS 中,进行一些简单的 join 和计算

80、以下哪些项包含在 ODPS 项目空间的鉴权模型中? BCD

- A LabelSecurity
- B CheckPermissionUsingACL
- C SecurityConfiguration
- D ProjectProtection

!!! 81、对于 ODPS 的 mapjoin 说法正确的是? BCD

A 如果使用 mapjoin,则所有小表占用的内存总和不得超过 1G 512M

- B inner join 的左表或右表均可以作为大表
- C left outer join 的左表必须是大表
- D 使用 mapjoin 时需要引用小表或者子查询时,需要引用别名

- ◆ left outer join 的左表必须是大表;
- ◆ right outer join 的右表必须是大表;
- ◆ inner join 左表或右表均可以作为大表;
- ◆ full outer join 不能使用 mapjoin;
- ◆ mapjoin 支持小表为子查询;
- ◆ 使用 mapjoin 时需要引用小表或是子查询时,需要引用别名;
- ◆ 在 mapjoin 中,可以使用不等值连接或者使用 or 连接多个条件;
- ◆ 目前 ODPS 在 mapjoin 中最多支持指定 6 张小表,否则报语法错误;
- ◆ 如果使用 mapjoin,则所有小表占用的内存总和不得超过 512MB。请注意由于 ODPS 是压缩存储,因此小表在被加载到内存后,数据大小会急剧膨胀。此处的 512MB 限制是加载到内存后的空间大小;
- ◆ 多个表 join 时,最左边的两个表不能同时是 mapjoin 的表。

82、OTS 的特点非常适合用来存储元数据,和使用 RDS 作为元数据管理工具相比,OTS 具有以下哪些优点? ABCD

- A 支持更大的单表规模,可以超过百 TB

B 更易扩展，理论上无限容量
C 更高并发量，十万级别 QPS
D 更高的可用性，可达 99.9%
RDS（99.95%）

83、在 RDS 实例下创建数据库的方法有哪些？ AC
A 直接调用 RDS 的 openAPI
B 通过 MySQL 客户端工具，如：MySQL-Front
C 通过管理控制台完成
D 阿里云命令行工具 CLI

84、云监控室阿里巴巴集团多年来服务器监控技术研究积累的成果，结合阿里云云计算平台强大的数据分析能力为中小网站提供各种应用和服务***** ABC
A 旺旺消息
B 电子邮件
C 手机短信
D 钉钉消息

85、ODPS 的控制是由那几部分组成？ ABC
A worker
B Scheduler
C executor
D master

86、ADS 的维度表组中存放的表的特点包括以下哪些项？ AD
A 数据量较小
B 为星型或者雪花型
C 记录条数据不超过 100 万或者不大于 1G
D 需要和任何表关联
备注：其中维度表组用于存放维度表（特征上是一种数据量较小但是需要和任何表进行关联的表），目前有且仅有一个，并且在分析型数据库数据库建立时会自动创建，用户不可修改和删除

87、以下哪些情况不适合使用 ODPS mapjoin BCD
A or 连接关联条件
B 关联表中有临时表
C 两个表无条件关联
D 两个大表关联

!!! 88、安骑士的主要功能包括哪几项？ ABD
A 异地登录提醒
B 木马文件的查杀
C 密码破解的发现与处理
D 高危漏洞检查与修复
备注：包括木马查杀、防密码暴力破解、异地登录提醒、高危漏洞检测修复、体检加固等安全防

护功能

89、DMS（原 iDB Cloud）是一款访问管理云端数据的 WEB 服务，覆盖 RDS、ADS 等阿里云环境，DMS 支持一下那些数据源？ ABCD

- A MySQL
- B PostGresSQL
- C ADS
- D SQL server

备注：DMS 支持 MySQL、SQL Server、PostgreSQL 和 PPAS 等数据源。

90、关于 ODPS Policy 授权的特点，说法正确的是？ AD

- A 授权对象支持以通配符 “*”
- B 删除一个对象时，与该对象相关的 policy 授权会被删除
- C policy 支持通过 xml 文件的方式授权
- D 当 Allow 和 deny 同时存在时，则以 deny 优先原则

?? ? 91、ODPS 中可以通过配额管理一个项目中可以使用的资源的上限，包括以下哪几种 资源？ ACD

- A 存储
- B 带宽
- C 内存 ??
- D CPU

92、RDS 的基本功能有哪些？ ABD

- A 实例管理
- B 账号管理
- C 安全组管理
- D 数据库管理

93、对 OTS 的单行操作，描述正确的有？ ABD

- A getrow: 读取单行数据
- B deleterow: 删除一行
- C updaterow: 更新一行，如果该行不存在，则新增一行
- D putrow: 新插入一行数据，如果数据存在，则删除旧行，在写入新行

解释：putrow 插入数据到指定的行，如果该行不存在，则新增一行；若该行存在，则覆盖原有行。

updaterow: 更新指定行的数据，如果该行不存在，则新增一行；若该行存在，则根据请求的内容在这一行中新增、修改或者删除指定列的值。

?94、OPDS 中用户认证(Authentication)的主要功能是检查请求 (Request)发送者的真实身份，通过包括以下哪几项验证？ ABC

- A、正确验证用户状态，包括是否欠费等
- B、正确验证接收到的信息在途中是否被篡改
- C、正确验证消息发送方的真实身份
- D、用户需要提取用户个性化信息时进行用户个性化认证

95、在 ADS 中，两个事实表进行 join 的充要条件包括以下哪些项？ ABCD

A 两张表的 hash 分区数必须一致

B 这两个表在一个表组

C 两张表的 join key 至少有一列建立了 hash map 索引

D 这两个表的 join key 是 hash 分区列

96、ODPS task 类型包括以下哪些？ ABD

A join task

B map task

C localwork

D reducetask

97、由于 RDS 只读实例的数据均有源数据库同步而来，以下哪些功能只读实例将无法使用？

ABC

A 账号授权及修改账号密码

B 备份设置及临时备份功能

C 创建/删除数据库

D 独立的白名单配置

判断题

1、ODPS 的 LabelSecurity 对敏感数据的粒度可以支持列级别 对

2、LoadBalancerID 是识别用户 SLB 实例的唯一标识 对

3、RDS for MySQL 的所有备份类型均为全量备份 否

4、可以使用一下 ODPS SQL 把两个数据集的结果合并在一起：SELECT * FROM sale_ditail WHERE region = 'hangzhou' UNION ALL SELECT * FROM 'shanghai'，此 SQL 可以直接作为子查询应用在其他 SQL 中 否

5、项目空间是 ODPS 的基本组织单元，它类似于传统数据库的 Database 或 Scheme 的概念，是进行多用户隔离和访问控制的主要边界 对

114 6、当一个用户被从某个 ODPS 项目空间移除后，policy 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。 错 ACL 授权仍然会被保留

6、RDS 只读实例不能自动释放 对

7、ODPS 适用于离线数据的处理、分析或挖掘，它同时提供存储和计算两种能力，支持 SQL 和变成（Map/reduce）等多种使用方式 对

8、在 ADS 中一个表必须从属于一个表组 对

9、云盾安骑士可以保障阿里云以外的服务器 对

注解：在非阿里云的服务器上，安骑士提供了木马文件查杀、高危漏洞检测与修复、入侵告警通

知等功能。

10、ODPS SQL 不支持正则表达式 错

11、当前的 DRDS 无法支撑现有业务时，我们可以通过用户管理控制台进行 DRDS 的平滑扩容，以满足业务需求，平滑扩容的流程依次分为配置、迁移、清理、切换四个步骤 错 设置-迁移-切换和清理四个部分

12、非阿里云主机也可以使用云盾 DDoS 高防 IP 服务 对

备注：DDoS 高防 IP 是针对互联网服务器（包括非阿里云主机）

!!! 13、ADS 的普通表，通常把需进行 join 的列作为一级 hash 分区，把日期作为二级 range 分区列，这样可以进行大表 join，又可以进行每天增量数据导入 错

备注：没有 range 分区

14、ODPS SQL 不支持在普通 join 的 on 条件中使用不等值表达式、or 逻辑等复杂的 join 条件，但是在 mapjoin 中可以进行如上操作 对

15、用户开启 SLB 健康检查功能后，当后端某个 ECS 健康检查出现问题时会将请求转发到其他健康检查正常的 ECS 上，而当该 ECS 恢复正常运行时，SLB 会将其自动恢复到对外或 XXXXXXX 对用户开启健康检查功能后，当后端某台 ECS 健康检查出现异常时，负载均衡会自动将新的请求分发到其它健康检查正常的 ECS 上；而当该 ECS 恢复正常运行时，负载均衡会将其自动恢复到对外或对内的服务中 对

16、关系型 OLAP 是预先根据数据需要分析的维度进行建模，在数据的物理存储层面以“立方体”(Cube)的结构进行存储，具有查询速度较快等的优点：错 多维 OLAP

17、ODPS 中可以使用 Tunnel 模块上传视图数据：错 目前 Tunnel 仅支持表（不包括视图 View）数据的上传下载

18、ODPS 中只有 string、bigint、datetime 和 double 才能参与算术运算：错，只有 string、bigint、double 才能参与算术运算

!!! 19、在开发 ODPS MR 或者 UDF 的过程中，在用到文本文件、ODPS 表等的时候，只能把它们作为主方法的输入参数，由主方法读取后识别处理 对

20、ADS 的维度表可以和任意表组的任意表进行关联，并且创建时不需要配置分区信息 对
备注：维度表可以和任意表组的任意表进行关联，并且创建时不需要配置分区信息，但是对单表数据量大小有所限制，并且需要消耗更多的存储资源。

21、在创建 ADS 数据库前，需要从业务、数据量、计算复杂度等角度进行仔细分析，然后根据分

析结果去合理的设置 ECU 的型号、台数以及接入节点的配置等 否

23、一旦 ODPS 项目空间保护打开（set security.ProjectProtection=true），无法将项目空间内的数据转移到项目空间之外，所有的数据都只能在项目空间内流动。 对

24、ODPS SQL 的 where 子句不支持 between 条件查询 对

25、加入云盾先知计划的企业可以自主设定奖励计划，奖励额度没有要求 对

备注：企业可以自由设定高、中、低危漏洞的奖励金额和奖励范围

!! 26、SLB 不允许跨 Region 容灾 对

!!! 27、ODPS 的客户端工具 odpscmd 中可以使用 show functions 查看当前 ODPS 项目空间中所有的自定义函数 错 list functions;

28、RDS 只读实例不能自动释放 对 如果要释放该只读实例，同样在【基本信息】中点击【释放】按钮，输入验证码后即可

29、ODPS Policy 授权文件的格式是 Json 对

30、用户开启 SLB 会话保持功能后，SLB 会把来自同一客户端的访问请求分发到不同的后端 ECS 上进行处理 错

31、ODPS 向某个存在的表分区插入数据时，分区列不能出现在 select 列表中。否

32、某大型系统采用了多个云服务器，并将所有服务器的系统日志存储到 OTS 中，考虑到当天日志的访问量比较大，之前的历史日志访问量很少，则可以将日志分成两个，一个用来保存当天日志，设置较高的预留读写吞吐量，另一个表保存之前的历史日志，预留读写吞吐量设置比较低，并定时将当前的日志表中的谁归档到历史表中，这样可以在不影响应用的情况下，节省开支。
是

33、云监控除了可以为 aliyun.com 云服务器客户提供安全有效的监控服务外，还能够为其他客户提供功能强大的监控服务，只是配置比较复杂，必须自己开发脚本进行数据收集。 否
备注：云监控除了可以为云服务器 ECS 提供安全有效的监控服务外，还能够为其他自有服务器提供功能强大、简单易用的监控服务。用户不需要自行搭建或者开发监控系统，只需要通过简单的配置即可实现各种监控需求。

34、ADS 中，当 updateType=realtime 时，支持 insert、delete、update 操作。否
分析型数据库的表根据 updateType 分为批量更新表（仅能够离线批量更新数据）和实时更新表（能够通过 insert/delete 实时更新数据），用 updateType 以区分，如果 updateType 选项不填则默认为批量更新表。需要注意的是，updateType=realtime 暨为实时更新表时，必须指定合法的主键并且不能有二级分区。

35、ADS 的普通表的一级分区只支持两种方式，Hash 和 List，目前不支持 Range 分区。 否

关于分区方面，目前分析型数据库支持最多两级分区，并且一级分区仅支持 HASH 分区，二级分区仅支持 LIST 分区

36、ODPS 的 MapReduce 可以支持在一个 Map 后连续多个 reduce。 是
基于此，ODPS 提供了扩展的 MapReduce 模型，即可以支持 Map 后连接任意多个 Reduce 操作

37、ODPS SQL 支持主键，但是不支持索引和约束。 否
ODPS SQL 采用的是类似于 SQL 的语法，可以看作是标准 SQL 的子集，但不能因此简单的把 ODPS SQL 等价成一个数据库，它在很多方面并不具备数据库的特征，如事务、主键约束、索引等。目前在 ODPS 中允许的最大 SQL 长度是 2MB。

38、如果要查看只读实例的数据延迟，用户只能通过 RDS 管理控制台进入只读实例进行查看。 否

39、用户可以为创建的 SLB 实例指定一个易于识别的名称，如果用户不指定，那么系统将以该 SLB 实例的 LoadBalancerID 作为名称，LoadBalancerID 是识别用户 SLB 实例的唯一标识，无法修改。 是

40、在 ODPS 中，所有的任务都会被实例化（Instance），以 ODPS 实例的形式存在。 否
计算型任务在执行时，会被实例化， 用户可以操作这个 [实例(Instance)] 的信息、

另一方面，部分 ODPS 任务并不是计算型的任务

41、云盾先知计划共享了阿里巴巴应急中心安全专家的资源和能力，提供服务时会签署平台用户保密协议。 是

42、ODPS 的 JOIN 支持多路间接，但不支持笛卡尔积，即无 on 条件的链接。 答案：是

43、ODPS 的计费是以项目为单位的，对存储、计算和数据下载三个方面分别计费。 答案：是

!!! 44、预付费的 RDS 实例可以在不需要的时候主动释放，以节约成本。 答案：否

45、odps 分区表可以指定多个分区条件，分区条件顺序无关。 答案：否

46、ODPS 中一个用户最多可以创建 10 个项目空间 答案：是



47、用户可以指定 SLB 后端服务器池内各 ECS 的转发权重，权重值越高的 ECS 将被分配到更少的访问请求 答案：否

48、在 ADS 中，数据根据分区列进行分布式的存储和计算。 答案：是

?? 49、ODPS 给予标签的安全 (labelsecurity) 是项目空间级别的一种自主访问控制机制 (DAC)。

答案：否

50、云盾先知计划的技术起源于淘宝的运营经验，仅适合对电商类 WEB 应用做风险识别。 答

案：否

51、ADS 中，只有相同表组的表才能进行 join 关联（维度表除外）。答案：是

52、在 ODPS 中，所有的请求都会被转化为任务(Task)。F

学习笔记

容易忘的点

安骑士

态势感知

绿网

labelsecurity

讲师试卷 1

一、单选题（共 50 题，每题 1 分，共 50 分）

1、某电商使用 ADS 进行数据分析，其中商品信息表记录了商品的型号、价格、厂家等信息，从数据分析角度来讲，它是一张维表，该表包含了超过 2000 万条的纪录，大小在 5G 左右，在 ADS 中该如何处理该表？ C

A、建表时指定为 ADS 的维表，将分区数设置为 1

B、建表时指定为 ADS 的维表，即不进行拆分，可以和任意表关联

C、必须建成 ADS 分区表，将会有一些限制，比如不能和不同普通表组上的表进行关联等

D、可将该表直接打开到和该表关联的事实表中，做成一张宽表，减少 join，提升性能

维度表

维度表与小表

大部分情况下，维度表是小表（<1000w 且 <1G），Garuda 针对维度表的支持：

- 标识为维度表，且分区数设置为 1，即不进行任何切分
- 维度表单表的查询能力（如子查询、Group-By、Order-By 等）不受限制
- 多个维度表之间可以按任意字段 JOIN（但不支持 string 类型列做为 JOIN 列）
- 维度表与事实表（分区表）也可以按任意字段 JOIN

2、DRDS 的读策略中不包含以下哪种？ B

A、主库读

- B、主库写
- C、自定义
- D、均测

3、在 RDS 只读实例变更配置过程中，只读实例在 RDS 管理控制台中可以进行以下哪种任务？ C

- A、内外网切换
- B、更改配置
- C、查看 SQL 日志
- D、参数设置

4、Tunnel 服务的公网 end point 是？ D

- A、http://tunnel.odps.aliyun.com/api
- B、http://service.odps.aliyun.com/tunnel
- C、http://tunnel.odps.aliyun.com
- D、http://dt.odps.aliyun.com

5、从 ODPS 导入数据到 ADS 的时候，导入路径格式为？ D

- A、odps://project_name/table_name partiton_spec
- B、odps://project_name/table_name:partiton_spec
- C、odps://project_name/table_name(partiton_spec)
- D、odps://project_name/table_name/partiton_spec

6、ODPS 中，以下哪种授权方式不能对角色授权？ D

- A、ACL
- B、Policy
- C、package
- D、label

```
odps@hzjtdsj_odps>set label to role hzjtdsj_zt;  
FAILED: mismatched input 'role' expecting 'USER' or 'TABLE'
```

7、以下哪种对 RDS 的描述是正确的？ A

- A、一个实例可以创建多个数据库
- B、在实例内数据库命名可重复
- C、数据库不会共享该实例下的资源
- D、以上都不对

8、ORDS 实例名称不能以什么开头？ D

- A、大写字母
- B、下划线
- C、小写字母
- D、中文

9、小王正在使用 ODPS 进行自然语言分析，他从门户网站上搜集了 1000 篇新闻报道，然后每篇存成一条纪录，然后基于这些纪录去做分词，即将每条纪录中的内容拆成很多独立的词语，保存到另一张表中，每个词语一条记录。小王对自定义函数比较熟悉，他想通过编写自定义函数实现

分词功能。请帮小王做一下选择，从功能角度来看，哪种自定义函数更适合处理这种场景？ B

- A、User Defined Scalar Function
- B、User Defined Table Valued Function
- C、User Defined Splitted Function
- D、User Defined Aggregation Function

10、在 ADS 中，为了更高效的进行表关联，两个事实表进行 Join 的必要条件不包括以下哪些项？

- A
- A、这两个表的 join key 至少有一列建立了主键
- B、这两个表在一个表组
- C、这两个表的 join key 是 hash 分区列
- D、两个表的 Hash 分区数必须一致

11、关于 DDoS 高防 IP 的功能，描述错误的是？ D

- A、可随时更换防护 IP
- B、提供实时精准的流量报表及攻击详情，让您及时准确获得当前服务详情
- C、防护多种 DDoS 类型攻击，包括但不限于以下攻击类型 ICMP Flood、UDP Flood、TCP Flood、SYN Flood、ACK flood 等
- D、DDoS 防护阈值弹性调整，您可以随时升级更高级别的防护，调整过程服务中断时间小于等于 30 秒

!!!12、以下哪项不是 ODPS MR 的必须组成阶段？ D

- A、Map
- B、Shuffle
- C、Reduce
- D、Combiner

13、关于 RDS For MySQL 的单表尺寸限制，描述错误的是？ B

- A、单一实例下所有表的空间加起来不能超过 2TB
- B、如果实例下有 2 张以上的表，每张表的空间都可以至 2TB
- C、单表的有效最大表尺寸通常受限于操作系统的文件尺寸限制
- D、单表尺寸最大为 2TB

14、哪些监控信息不能从 OTS 的用户管理控制台直接获得？ D

- A、读写吞吐量
- B、平均返回延迟值
- C、QPS
- D、TPS

!!!15、负责轮询 Instance 执行状态的 ODPS 模块是？ B

- A、controller
- B、Worker
- C、Executor
- D、Scheduler



备注：来自 odps 权威指南

Scheduler:负责更新 OTS 中的 instance 状况，包括 running 和 terminated

execute 会查询 task 的执行状态，当 task 执行完成时，它会更新 OTS 中的 task 信息，并把 task 信息汇报给 scheduler。

work:判断该请求是否需要启动 fuxi job,则生成一个作业实例(instance), 发送给 scheduler; scheduler 会把该信息注册到元数据库 OTS 中,把状态设置为 running,worker 把 InstanceId 返回给客户端。

客户端接收到返回的 InstanceId 后，就可以通过该 InstanceId 查询作业的执行状态。

客户端会发送另外一个 rest 请求，查询状态。http server 在账号验证通过后，把请求发送给 worker,worker 则根据 InstanceId 查询 OTS 中该作业的执行状态，返回结果。

16、ADS 的普通表最多包含多少个列？ D

- A、512
- B、2048
- C、4096
- D、1024

注解：

（1）一张事实表至少有一级 Hash 分区并且分区数不能小于 8 个；（2）一个事实表组最多可以创建 256 个事实表；（3）一个事实表最多不能超过 1024 个列。

!!!17、关于 RDS 只读实例的限制，说法正确的是？ B

- A、只读实例支持通过备份文件或任意时间点创建临时实例
- B、绑定有只读实例的主实例不支持通过备份集直接覆盖实例
- C、支持创建\删除数据库
- D、支持通过备份集覆盖实例

备注：只读实例有以下功能限制：

- 1 个主实例最多可以创建 5 个只读实例
- 备份设置：不支持备份设置以及临时备份
- 数据迁移：不支持数据迁移至只读实例
- 数据库管理：不支持创建和删除数据库
- 账号管理：不支持创建和删除账号，不支持为账号授权以及修改账号密码功能
- 实例恢复：不支持通过备份文件或任意时间点创建临时实例，不支持通过备份集覆盖实例
- 创建只读实例后，主实例将不支持通过备份集直接覆盖实例来恢复数据

18、渗透测试是通过模拟什么的攻击方法，来评估计算机网络系统安全的一种评估方法？ C

- A、高并发客户
- B、网络专家
- C、恶意黑客
- D、正常客户

解释：渗透测试是通过模拟恶意黑客的攻击方法，来评估计算机网络系统安全的一种评估方法。

19、DRDS 中的表 T 的分库分表键分别为 col1 和 col2，则下列语句中哪个不会进行全表扫描？ D

- A、select* from T where col1=1 or col2=2;

- B、select* from T where col2=2;
- C、select* from T where col1=1;
- D、select* from T where col1=1 and col2=2;

20、关于云盾态势感知的优势，描述错误的是？ B

- A、免安装，免维护 无需安装任何软件，部署任何设备 无需复杂配置和耿鑫，省心省力 安全管控从未如此简单
- B、IT 资产划分，软件版本管理 基于威胁情报的安全事件管理对入侵行为造成的损失进行评估
- C、建设安全体系 从黑客角度监控云上业务安全 对异常行为进行实时告警 让安全“可见、可控、可管”
- D、是一个针对阿里云服务器的管家式服务

备注：大数据实时分析，还原安全事件全过程，对攻击者进行取证和追溯

21、在服务器上手动安装云盾客户端安骑士时需要什么权限？ A

- A、系统管理员权限
- B、FTP 服务权限
- C、数据库权限
- D、系统普通用户权限

登录阿里云帐号，进入服务器安全(安骑士)控制台-安装 Agent 页面，根据页面提示获取最新版本下载地址，以管理员权限在服务器上运行并安装，[点击进入安装页面](#)。

22、用户在 RDS API 调用 ImportDataForSQLServer 进行外部数据迁移操作，RDS 下达任务后，实例状态由“使用中”变为什么状态，表示 API 调用任务执行结束？ B

- A、重启中
- B、恢复中
- C、删除中
- D、创建中

解释：调用本接口，进行外部数据迁移操作，RDS 下达任务，实例状态由“使用中”变为“恢复中”，任务执行结束，实例正常运行

23、某地区级网络提供商使用 ODPS 存储、分析宽带用户的流量信息，用户每登陆一次产生一条纪录，日均记录数在 2000 万条左右，目前最常见为基于区县（该市共有 12 个区县）、日期（天）的各种查询，其中部分查询只带有区县信息。架构师在设计用户登陆信息表时，决定建成分区表，对于分区列的选取，以下性能最好的是哪一项？ B

- A、以字符串型日期信息（天）为分区别
- B、以字段组合（字符串型区县信息，字符串型日期信息（天））为分区列
- C、以字符串型区县信息为分区列
- D、以字符串型日期信息（天）为第一级分区列，以字符串型区县为第二级分区列

24、下列不属于 ODPS 项目空间的对象类型为？ A

- A、Jar 包
- B、资源
- C、表

D、实例

25、创建 DRDS 库的时候，每个挂载的 RDS 实例默认创建的库的个数为几个？ B

- A、16
- B、8
- C、4
- D、2

26、下面哪一项不是安骑士包含的功能？ D

- A、高危漏洞修复
- B、木马文件检查
- C、异地登陆报警
- D、防 WEB 应用系统密码破解

27、ODPS 执行某个 SQL 时返回信息中包括 “J3_1_2_Stg1”，以下说法中正确的是？ D

- A、当前 Task（ID 为 3）依赖于 ID 为 1 的任务，该 Task 的 Worker count 为 2
- B、当前 Task（ID 为 3）依赖于 ID 为 1 的任务，同时触发 ID 为 2 的任务
- C、当前 Task（ID 为 3）有 2 个 Reducer，1 个 mapper
- D、当前 Task（ID 为 3）依赖于 ID 为 1 和 ID 为 2 的两个 Task

28、tbl_p 为一个 ODPS 的分区表，包含一个 “p1=’1’” 的分区，现在想把该分区上的数据删除，达到 select* from tbl where p1=’1’查询时，返回结果为空的目的。以下哪个 SQL 可以实现这个功能？ A

- A、alter table tbl drop partition (p1=’1’);
- B、delete from table tbl;
- C、insert into table tbl select * from tbl where 1=2;
- D、truncate table tbl;

29、跨站脚本攻击（Cross-site scripting，通常简称为 XSS）可被用于进行窃取隐私、钓鱼欺骗、偷取密码、传播恶意代码等攻击行为，请问 XSS 攻击主要发生在什么位置？ A

- A、WEB 客户端
- B、WEB 服务器端
- C、数据库端
- D、缓存服务器端

30、ADS 可以使用 explain 命令查看执行计划。下面说法中正确的是？ A

- A、explain 正确执行后返回信息中，第一行为逻辑计划，第二行为物理计划
- B、当用户发起一个 explain 查询到 ADS 系统后，ADS 会根据所有活动节点上的数据来分析执行计划
- C、当用户通过查询的方式，想要获取文本格式的 explain 语句后，将会得到一个 XML 格式的串
- D、逻辑计划中，TableExecutor 节点表示参与计算的信息，JoinExecutor 表示 join 的节点信息

31、ODPS SQL 的执行计划逻辑上可以被看做是一个有向图，其中边的含义是？ A

- A、执行阶段的依赖关系

- B、任务
- C、作业之间的依赖关系
- D、实例

32、一个 ODPS SQL 中最多使用多少个窗口函数？ C

- A、不限制
- B、1
- C、5
- D、10

目前在一个 MaxCompute SQL 语句中，可以使用至多 5 个窗口函数。

33、管理云盾加密服务的密钥时，必须通过哪种方式进行认证？ B

- A、阿里云官网帐号的 AK
- B、身份证（USB Key）
- C、用户名和密码
- D、用户自己上传的 SSL 证书

解释：所有的加密服务实例都需要使用用户手里的身份识别卡（USB KEY）进行管理。此卡仅客户持有，请妥善保管。

34、以下哪种不是加载数据到 ODPS 的合理方法？ C

- A、tunnel
- B、mapreduce
- C、DT task
- D、dship

35、ODPS 表 t1、t2 都是分区表，分区键是 pt，对于 SQL 语句 `select t1.id from t1 left outer join t2 on t1.id=t2.id where t1.pt='20140101' and t2.pt='20140101'`，哪一个说法是正确的？ A

- A、t1 表会进行分区剪裁
- B、t2 表会进行分区剪裁
- C、两张表都不会进行分区剪裁
- D、两张表都会进行分区剪裁

36、某借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，每天处理的数据在 20P 左右，基于这些数据对用户信用进行建模并产生预测数据，用户线上申请贷款时，能迅速根据用户模型预测结果进行资质评估、授信等。请根据你对阿里云产品的理解，下述方案中成本最低、性能最好的一个是？ B

- A、使用 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 ODPS 中
- B、使用 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 OTS 中，以供实现快速查询
- C、使用 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产评估和授信评估，将处理结果保存在 ADS 中
- D、使用 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产评估和授信评估，将处理结果保存在 OTS 中

37、某图书管理系统使用 ODPS 进行离线数据分析，现在他们需要将同一类目的书籍的名称拼成一个字符串，即将属于同一分组的字符串合并在一列展示，以下哪个内置函数可实现这一功能？ A

- A、wm_concat
- B、regexp_instr
- C、split_part
- D、concat

38、跨站脚本攻击（Cross-site scripting，通常简称为 XSS）可被用于进行窃取隐私、钓鱼欺骗、偷取密码、传播恶意代码等攻击行为，请问 XSS 攻击主要发生在什么位置？ A

- A、WEB 客户端
- B、缓存服务器端
- C、数据库端
- D、WEB 服务器端

解释：跨站脚本攻击（Cross-site scripting，通常简称为 XSS）发生在客户端，可被用于进行窃取隐私、钓鱼欺骗、偷取密码、传播恶意代码等攻击行为

39、OTS 中以什么方式来存储数据？ B

- A、二维表
- B、键值对
- C、对象
- D、聚簇列

40、ODPS 拓展 MR 的主要特点是？ B

- A、支持计算在内存中进行，中间结果不落地
- B、支持 Map 后连续任意多个 Reduce 操作
- C、可以在上一次的 reduce 结束之后，直接增加一次 map
- D、从 Map 直接到 Reduce，省去 Shuffle 阶段

41、ADS 中，对批量导入的表，哪种分区模式是可取的？ D

- A、list+list
- B、list+hash
- C、hash+hash
- D、hash+list

42、使用 OTS Java SDK 进行表操作时需要创建 OTSClient 对象，以下哪项不属于构造此对象时必须提供的信息？ C

- A、AccessKeySecret
- B、Endpoint
- C、AccountID
- D、实例名称

43、网络安全专家服务是云盾什么服务的基础上，推出的安全代维托管服务？ B

- A、基础 Ddos 防护

- B、安骑士
- C、DDoS 高防 IP
- D、阿里绿网

44、一个 SQL Server 数据库实例最多可以创建多少个数据库？ D

- A、400
- B、200
- C、60
- D、50

解释：同一实例下的数据库共享该实例下的所有资源，其中：

- MySQL 版实例最多可以创建 500 个数据库
- SQL Server 2008 R2 版实例最多可以创建 50 个数据库
- PostgreSQL 版实例和 PPAS 版实例没有数据库个数限制

45、ODPS SQL 提供了 CASE...WHEN...，可以进行条件判断。表 tbl 中包含了字段（id bigint，age bigint），目前有三条纪录，分别为（1，20），（2，30），（3，60），执行以下 SQL：

```
select id, case when age>60 then 'A' when age>40 then 'B' when age>20 then 'C' else 'D' end from tbl
where id=3;
```

哪个是正确的执行结果？ A

- A、B
- B、C
- C、D
- D、A

46、表 A 为 ADS 中的普通表，列 a 为其中一个普通列，类型为 int，以下 SQL 哪些是正确的？ B

- A、SELECT a + COUNT(*) FROM A
- B、SELECT UDF_SYS_COUNT_COLUMN(a) FROM A
- C、SELECT SUM(COUNT(*)) FROM A
- D、以上都正确

47、ADS 中创建事实表（普通表）时会有一些限制，下列哪个说法是错误的？ A

- A、二级分区只能是 Hash 或者 List
- B、一个事实表最多不能超过 1024 列
- C、表的名字不能超过 32 个字符
- D、至少有一级 Hash 分区

48、某企业使用 ADS 分析数据，其中企业商品库存表（stock_detail）为一张超过 20 亿条纪录的事实表，描述了企业目前为止的库存情况，事实列仅包含库存数一列。商品信息表（time_detail）描述了商品的具体信息，包括商品信息标识、商品基本信息标识、商品通用信息属性等等 20 个属性信息，纪录条数在 10 万条左右，本表中的商品基本信息标识引用另一张存在多个层级维度的维表商品基本信息表（basic_detail，包含 30 个属性）。以下设计方案中，综合成本和性能两个角度来看，哪项是不合理的？ B

- A、企业商品库存表设计为分区事实表，分区数选择 100 个
- B、将商品信息表打平到企业商品库存表，形成一张宽表，使用时可以减少 join

C、商品信息表适合做成维度表，数据小，不需要建分区表
D、将商品信息表和商品基本信息表关联取得相关信息后，打平为单张宽维度表解释：
https://help.aliyun.com/document_detail/26425.html

49、ODPS 中哪种操作会使授权实效？ B

- A、上述授权都会失效
- B、对 acl 授权过的表重建
- C、对 policy 授权过的表重建
- D、对 package 授权过的表重建

50、有关 DDoS 攻击的描述，最准确的是？ D

- A、主要是攻击目标是数据库
- B、黑客自己够买了大量服务器，专门从事破坏活动
- C、攻击的目的是让服务超负载，然后窃取机密信息
- D、攻击的主要目的是让指定目标无法提供正常服务，甚至从互联网上消失，是目前最强大、最难防御的攻击之一

二、多选题（共 30 题，每题 1 分，共 30 分）

1、关于 ODPS 跨项目资源共享的说法正确的是？ ABC

- A、Packet 是一种跨项目空间共享数据及资源的机制，主要用于解决跨项目空间的用户授权问题
- B、添加资源时支持的对象类型不包括 Project 类型，即不能将 project 添加到 packet 里
- C、如果某个 package 还有人在使用，则该 package 的 owner 无法删除此 package
- D、添加到 Package 中的不仅仅是对象本身，还包括相应的操作权限，如不显示指定权限，则默认为读写权限

2、ADS 中可以使用 explain 查看执行计划，以下节点哪些属于逻辑计划？ ABC

- A、selectNode
- B、orderByNode
- C、JoinNode
- D、ExecutorNode

逻辑计划各个节点说明：

- SelectNode 表示这 select 中最终输出表达式的相关信息，例如 select 要输出的表达式集合
- GroupNode 表示 GroupBy 语句的相关信息，例如 groupby 的列，having 的表达式等
- OrderByNode 表示 OrderBy 的列信息，例如列名，顺序等。
- JoinNode 表示逻辑 Join 树的信息，例如 join 的 on 条件
- TableNode 表示分区表的信息，例如参与计算的列，表名等。

3、ODPS 由以下哪几部分组成？ ABCD

- A、逻辑层
- B、客户端
- C、接入层
- D、计算层

4、ODPS 中项目普通用户想要使用 `select* from` 访问一张非本人创建的本项目的表 A，至少需要以下哪些权限？ CD

- A、表 A 的 describe 权限
- B、项目 P 的 AccessAllow 权限
- C、项目 P 的 createinstance 权限
- D、表 A 的 select 权限

注解：

权限描述中 Project 类型对象的 CreateTable 操作，Table 类型的 Select、Alter、Update、Drop 操作需要与 Project 对象的 CreateInstance 操作权限配合使用。单独使用上述几种权限而没有指派 CreateInstance 权限是无法完成对应操作的。这与 ODPS 的内部实现相关。类似的，Table 的 Select 权限也要与 CreateInstance 权限配合使用。使用时请注意。

5、OST API 中提供了哪些 Action 可以进行多行数据操作？ ABD

- A、BatchWriteRow
- B、BatchGetRow
- C、GetMultiRow
- D、GetRange

6、DRDS 的全局唯一 ID 有哪些主要特点？ BCD

- A、全局有序
- B、没有单点性能瓶颈
- C、全局唯一
- D、大并发获取

备注：全局唯一数字序列服务

在分布式环境下，原有的 mysql sequence 生成机制无法高效的生成全局唯一的 sequence。我们借鉴了 oracle 的 sequence 生成样例，实现了一个在 mysql 基础上的高效 sequence 生成器，能做到没有单点性能瓶颈，具备大并发获取和低延迟特性。

DRDS sequence 功能的目标只是为了保证数据的全局唯一，虽然基本上是按时间序列获取的，但并不全局有序。

!!!!7、使用 ADS 可以很好的支持即席查询，对一些复杂的查询，也能在很短的时间内返回处理结果。以下关于 ADS 中查询设计的说法正确的有？ AB

- A、需要关联查询的普通表，需要放在同一个表组上
- B、如果用户的查询的条件中会指定聚集列的内容或范围，那么这样的查询性能便会有较大的提升
- C、ADS 中 join 会优先采用 hash join，如果不满足 hash join 的条件，优化器才能选取其他的关联方法
- D、出于关联方便的考虑，普通表可以放在维度表组上，这样该表和其他任意表关联都不再受表组的限制

聚集列

在创建表时，用户可以指定一列或者若干列作为聚集列。在物理上，一个分区内聚集列内容相同的数据会尽可能的分布在同样的区块内存放，因此如果用户的查询 Query 的条件中会指定聚集列的内容或范围，那么这样的查询性能便会有较大的提升。需要注意的是如果用户指定多列为聚集列，那么指定的聚集列的顺序便是比较数据是否相同的顺序。
聚集列可以在建表后进行修改，但是目前修改后下次装载数据完毕后生效。

8、ODPS 中可以通过配额管理一个项目空间中可以使用的资源的上限，包括以下哪几种资源？
BCD

- A、带宽
- B、CPU
- C、存储
- D、内存

!!9、下面的描述中哪些属于 BGP 机房的优点？ABC

- A、使用 BGP 协议还可以使网络具有很强的扩展性可以将 IDC 网络与其他运营商互联，轻松实现单 IP 多线路，做到所有互联运营商的用户访问都很快。这个是双 IP 双线无法比拟的
- B、由于 BGP 协议本身具有冗余备份、消除环路的特点，所以当 IDC 服务商有多条 BGP 互联线路时可以实现路由的相互备份，在一条线路出现故障时路由会自动切换到其它线路
- C、服务器只需要设置一个 IP 地址，服务器的上行路由与下行路由都能选择最优的路径，所以能真正实现高速的单 IP 高速访问
- D、BGP 机房的服务器都具备双机或多机冗错功能，一台服务器宕机后不影响整体的对外服务，具备更高的可用性

备注：BGP 机房的优点：

（1）服务器只需要设置一个 IP 地址，最佳访问路由是由网络上的骨干路由器根据路由跳数与其它技术指标来确定的，不会占用服务器的任何系统资源。服务器的上行路由与下行路由都能选择最优的路径，所以能真正实现高速的单 IP 高速访问。

（2）由于 BGP 协议本身具有冗余备份、消除环路的特点，所以当 IDC 服务商有多条 BGP 互联线路时可以实现路由的相互备份，在一条线路出现故障时路由会自动切换到其它线路。

（3）使用 BGP 协议还可以使网络具有很强的扩展性可以将 IDC 网络与其他运营商互联，轻松实现单 IP 多线路，做到所有互联运营商的用户访问都很快。这个是双 IP 双线无法比拟的。

10、对于 ADS 的索引说法正确的是哪一项？BCD

- A、用户需要对某列进行 Hash Join 时，无论是事实表之间的 Join 还是事实表和维度表的 Join，都需要考虑手工处理索引问题
- B、ADS 弱化了索引的概念，用户无需亲自为自己的数据配置索引
- C、ADS 拥有高度智能的自动化索引机制，会搜集相关信息，去自动创建索引
- D、一个列可以设置列属性 disableIndex=true，用于屏蔽 ADS 的默认列索引

11、OTS 费用从哪几个维度计量？ACD

- A、外网下行流量
- B、数据访问次数（万次）
- C、数据存储
- D、预留读写吞吐量

备注：表格存储从四个维度计量应用程序使用的资源并收取相应的费用：数据存储、预留读写吞吐量、按量读写吞吐量及外网下行流量。

备注：数据存储、行数据大小计算、预留读/写吞吐量、按量读/写吞吐量、外网下行流量

12、如何知道一个 ODPS SQL 在运行时占用了多少 instance？ BD

- A、根据表的大小，自己计算
- B、用 `odpscmd` 提交时，从返回的状态中查看
- C、使用 `ps -ef` 查看相关进程数
- D、打开 `logview` 查看

13、出于效率和成本的考虑，在使用 ODPS MR 的时候，输入表最好应该只包含需要处理的字段，其余多余字段应该忽略。输入表 `wc_in` 包含 10 个字段，MR 只对其中三个字段（`c1,c2,c3`）进行了处理，以下说法正确的是？ CD

- A、使用包 `com.aliyun.odps.data` 中的 `TableInfo` 类中的 `setCols()` 方法指定处理的列 `c1, c2, c3` from `wc_in`；同时调用 MR 的时候输入新表 `wc_in_new`
- B、MR 的输入为一个 SQL 查询：`select c1,c2,c3 from wc_in;`
- C、重建一张表 `wc_in_new`，只包含三个字段（`c1,c2,c3`），加载数据：`insert into table wc_in_new select c1,c2,c3 from wc_in;`同时调用 MR 的时候
- D、建一个视图 `v_wc_in`，视图仅包括相应的三个字段：`create view v_wc_in select c1,c2,c3 from wc_in;`然后把视图 `v_wc_in` 作为 MR 的输入

!! 14、云监控里的站点监控可以对目标站点服务的可用性以及响应时间进行监控。系统已经默认预置了 8 种监控类型，包括 `http` 监控、`ping` 监控、`tcp` 监控、`udp` 监控、`DNS` 监控、`pop` 监控、`smtp` 监控、`ftp` 监控，其中每种监控类型里面包含了两个监控项，分别是_____和_____ CD

- A、响应内容长度：`length`
- B、请求时间：`requesttime`
- C、状态码：`status`
- D、响应时间：`responsetime`

解释：https://help.aliyun.com/document_detail/28619.html

15、在使用 ODPS tunnel 命令 `upload` 数据的时候，关于一些常见参数，说法正确的包括以下哪些项？ ACD

- A、`-c` 本地数据文件编码，缺省为 `utf8`
- B、`-dfp` `DateTime` 类型数据格式，缺省为 `yyyyMMddHHmmss`
- C、`-cp` 指定是否在本地压缩后再上传，减少网络流量，缺省为 `True`
- D、`-fd` 本地数据文件的列分隔符，缺省为逗号

16、关于 ADS 多值列的说法正确的包括哪些？ AD

- A、可以使用 `in`，`contains` 条件对该列的单个值进行查询
- B、可以直接在 `select` 种使用该列
- C、不能在 `group by` 种直接使用该列
- D、可以存入 `string` 类型的多个值，以逗号分隔

17、为了保证应用获得可预期的性能，需要设置 OTS 表的：BD

- A、预留处理内存
- B、预留读吞吐量
- C、预留存储空间
- D、预留写吞吐量

18、以下哪些功能是 RDS 服务天生就具备的？BCD

- A、自动备份，支持 15 天内任意时刻数据恢复
- B、后台有专业团队提供运维服务
- C、数据库安全漏洞实时修复
- D、数据库主辅备份

备注：7 天

!!! 19、ADS 的 DMS 图形界面提供了哪些性能相关的信息？ABCD

- A、QPS
- B、访问量
- C、磁盘使用率
- D、平均响应时间

20、小王开发了一个 ODPS UDF，打成 jar 包之后，作为资源上传到项目空间里，创建了对应的函数，但是在执行的时候发现该 jar 包依赖的另一个 jar 包不存在，他应该怎么办才能正常使用？

CD

- A、在使用 jar 命令调用的时候，在 classpath 参数中增加依赖 jar 包的路径即可
- B、将依赖的 jar 包作为资源上传到项目空间，在 UDF 的初始化函数 setup 中，使用 readJarResource 读取 jar 包资源
- C、将依赖的 jar 包作为资源上传到项目空间，同时在创建函数的时候（CREATE FUNCTION），using 子句中增加依赖的包，如果多个包的话，用逗号分隔
- D、在对 UDF 打 jar 包的时候，把依赖的 jar 包包含进去，统一打成一个包即可

21、云数据库（RDS）具有以下哪些优势？ABCD

- A、高可用性，采用主从热备的架构
- B、易用性，控制台自助迁移和管理
- C、高安全性，具有网路防护能力和 SQL 审计
- D、高性能，底层使用高端服务器

22、ADS 的数据导入 job 提交后，可以通过以下哪些方法查询当前的导入状态？CD

- A、使用 log 命令：log
- B、使用命令：show status for
- C、使用 SQL：select state from information_schema.current_job where job_id=
- D、使用 DMS，通过菜单项查看：导入导出=>导入状态

23、ODPS Task 类型包括以下哪些？ABC

- A、JoinTask
- B、MapTask

- C、ReduceTask
- D、LocalWork

!!24、哪些对象的操作不需要通过 ODPS 的任务（Task）来完成？ ABCD

- A、project
- B、UDF
- C、instance
- D、resource

- 在 MaxCompute 中，并不是所有的请求都会被转化为任务(Task)，例如：[项目空间\(Project\)](#)、[资源\(Resource\)](#)、[自定义函数\(UDF\)](#)及[实例\(Instance\)](#)的操作就不需要通过 MaxCompute 的任务来完成。

!!!25、当 RDS 实例所占用的磁容量超过了所购买的最大磁盘容量时，RDS 实例会被系统锁定。实例被锁定后，用户只有以下哪些权限？ BCD

- A、insert
- B、drop table
- C、select
- D、show view

26、ADS 中，同一个表组的表具有以下哪些相同的属性？ BC

- A、二级 List 分区数
- B、一级 hash 分区数
- C、副本数
- D、纪录条数

!!27、当用户的数据库 RDS 实例只能实现下面哪种参数的操作的时候，表明数据库被锁定了？ AB

- A、select
- B、drop
- C、insert
- D、update

注释：如果您的数据库实例忽然只能 select 和 drop，不能做 insert、update、delete 操作，则表明您的数据库被锁定了。

如果您的数据库大小超过这种实例规格的最大允许存储空间的话，则您的数据库实例就会锁住。您的实例被锁定后，您只有 select、show view 和 drop table 的权限。出现这种情况后，建议您尽快升级实例规格或者单独购买空间，当然您也可以删除一些无用的表格以释放空间，只要空间足够后系统会自动解锁。

???28、DRDS 做平滑扩容的时候，现有的数据库中的表需要满足以下哪些要求？ AC

- A、拆分表需要有主键
- B、拆分表需要删除主键
- C、单表需要做小表广播
- D、单表需有主键

29、在 RDS 只读实例变更配置过程中，该只读实例的实例状态为“升降级中”时。只读实例在控制台无法进行以下哪些任务？ ABD

- A、内外网切换
- B、参数设置
- C、查看 SQL 慢日志
- D、重启实例

!!!!30、OPDS 通过 eclipse plugin 提供了 MR 的编程框架，开发时通常需要添加以下哪几个类？

ABD

- A、Reducer
- B、Combiner
- C、MapReduce Driver
- D、Mapper

```
job.setMapperClass(TokenizerMapper.class);
    job.setCombinerClass(SumCombiner.class);
    job.setReducerClass(SumReducer.class);
```

三、判断题（共 20 题，每题 1 分，共 20 分）

1、将数据定时增量加载到 ADS 表中，如果目标表的属性是批量更新，则该表需要保证有两级分区且第二级分区必须是 list 分区，才能保证新增数据不会覆盖原有历史数据 是

2、在 ADS 中，能够实时插入的表一定要定义主键或组合主键，在进行数据插入时，若发现同主键的数据时，默认执行覆盖行为 是

解释：在分析型数据库中，能够实时插入的表一定要定义主键或组合主键，和 MySQL 有一个非常大的不同是，分析型数据库在进行数据插入时，若发现同主键的数据时，默认执行覆盖行为。若使用 insert ignore 语法，则在发现有同主键的数据时，丢弃新插入的数据，保留原有数据。但是无论如何，在主键冲突时分析型数据库无法直接返回错误给 insert 的执行方。需要注意的是，updateType=realtime 暨为实时更新表时，必须指定合法的主键并且不能有二级分区。

3、在 ODPS 中，所有的请求都会被转化为任务（Task） 否

4、预付费的 RDS 实例可以在不需要的时候主动释放，以节约成本 否

5、用户所购买的 RDS 实例不会受其他用户实例的影响。 是

6、RDS 的实例是具备自动主备的，不需要用户额外购买或者配置即可自动完成。是

7、ADS 的维度表可以和任意表组的任意表进行关联，并且创建时不需要配置分区信息。是

8、OTS 保证数据写入强一致，并能保证事务的最终一致性。 否

解释：原子性、强一致

9、当添加的资源为分区表时，ODPS 仅将整张表作为资源，不能将分区视为资源。 否

10、加入云盾先知计划的企业可以自主设定奖励计划，奖励额度没有要求。 是

11、OTS 的 putrow 操作会新插入一行。如果该行内容已经存在，则会报错返回。 否

- 解释：condition 在数据写入前是否进行行存在性检查，可以取下面三个值：
 - IGNORE 表示不做行存在性检查。
 - EXPECT_EXIST 表示期望行存在。
 - EXPECT_NOT_EXIST 表示期望行不存在。
- 若期待该行不存在但该行已存在，则会插入失败，返回错误；反之亦然。

12、DRDS 目前支持强一致分布式事务 否

解释：最终一致性

13、ODPS 基于标签的安全（LabelSecurity）是项目空间级别的一种自主访问控制机制（DAC） 否

14、ADS 的普通表，通常把需进行 Join 的列作为一级 Hash 分区列，把日期作为二级 Range 分区列，这样可以进行大表 Join，又可以每天进行增量数据导入。 否

15、云盾安全网络如果使用“域名接入”的方法，需要 DNS 服务器支持智能解析。 否

16、对于 ADS 的多值列，不允许在不经由 WHERE 子句中用 IN/CONTAINS 进行枚举筛选的情况信下直接在 SELECT/GROUP BY 中使用。 是

17、ODPS 可以使用 Alter 命令修改分区列的列名，修改后对分区没有影响。 否

18、ODPS 的客户端工具 odpscmd 中可以使用 show functions 查看当前 ODPS 项目空间中所有的自定义函数。 否

19、某大型系统使用了多个云服务器，并将所有服务器的系统日志存储到 OTS 中。考虑到当天日志的访问量比较大，之前的历史日志访问量很少，则可以将日志表分成两个，一个用来保存当日日志，设置较高的预留读写存储量，另一个表保存之前的历史日志，预留读写吞吐量设置比较低的值，并定时将当前日志表中的数据归档到历史表中，这样可以在不影响应用的情况下，节省开支。 是

20、当一个用户被从某个 ODPS 项目空间移除后，Policy 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 授权访问限制将被重新激活。 否

讲师试卷 2

一、单选题（共 50 题，每题 1 分，共 50 分）

1、ODPS SQL 目前最多支持多少个并发 insert overwrite / into 操作？ D

A、1000

- B、500
- C、64
- D、128

2、OTS 的一些表中存在冷数据和热数据的区分，所谓冷数据是指根据时效性判断被查询的可能性不大的数据，反之为热数据。下面关于冷热数据的说法，错误的是？ A

- A、在同一张表可以对不同的分片设置不同的预留读写吞吐量，对于冷数据所在的片，设置较低的预留读写吞吐量
- B、可以用不同的表来区分冷热数据，并设置不同的预留读写吞吐量
- C、表中存在大量冷数据会导致数据访问压力不均匀，从而导致表上配置的预留读写吞吐量无法被充分利用
- D、对冷数据对应的表，可以设置较小的预留读写吞吐量

3、如果在非阿里云的服务器安装了安骑士，如何在管理控制台查看报告？ A

- A、在安装安骑士的过程中需要输入“安装验证 key”，通过此 key 关联用户的管理控制台
- B、通过阿里云官网帐号关联
- C、报告无法体验在阿里云的管理控制台
- D、非阿里云服务器不支持安骑士

备注：对于非阿里云服务器，在安装过程中会提示输入验证 Key，这个验证 Key 用于关联阿里

云帐号，通过阿里云帐号在安骑士控制台使用相关功能，验证 key 会显示在[安装页面中](#)。

4、ADS 可以使用 explain 命令查看执行计划。下面说法中正确的是？ D

- A、当用户发起一个 explain 查询到 ADS 系统后，ADS 会根据所有活跃节点上的数据来分析执行计划
- B、当用户通过查询的方式，想要获取文本格式的 explain 语句后，将会得到一个 XML 格式的串
- C、逻辑计划中，TableExecutor 节点表示参与计算的信息，JoinExecutor 表示 join 的节点信息
- D、explain 正确执行后返回信息中，第一行为逻辑计划，第二行为物理计划

5、如果我的 ECS 到期释放了，服务器安全托管服务还未到期，正确的处理方法是？ D

- A、既然 ECS 已经到期了，服务器安全托管服务自然也就过期作废
- B、提交工单投诉，把已释放的 ECS 恢复回来
- C、只能申请退款，无别选择
- D、若 ECS 到期释放了，可以通过工单联系我们，重新授权一台新购 ECS，新购 ECS 可以继续使用安全托管服务，服务截止日期不变

6、ADS 的数据库名为 movie_db，该数据库管理员的阿里云帐号为 alice@aliyun.com，密码为 Alice123，accessID 为 aaaa，AccessKey 为 bbbbbb，连接的服务器地址为 movie_db.cn-hangzhou-ads.aliyuncs.com，端口号为 3456，可以使用 mysql 客户端命令连接。下列哪项是正确的？ A

- A、mysql-h movie_db.cn-hangzhou-ads.aliyuncs.com-p3456-uaaaa-pbbbbbb-D movie_db
- B、mysql-h movie_db.cn-hangzhou-ads.aliyuncs.com-p3456-ualice-PAlice-D movie_db
- C、mysql-h movie_db.cn-hangzhou-ads.aliyuncs.com-p3456-uaaaa-Pbbbbbb-D movie_db
- D、mysql-h movie_db.cn-hangzhou-ads.aliyuncs.com-p3456-uALIYUN\$alice@aliyun.com-pAlice123-D movie_db

7、云盾 DDoS 基础防护功能如何开启？ B

- A、提交工单，24 小时之内开通
- B、自动开通
- C、联系客户经理线下开通
- D、需要从管理控制台申请开通

8、DRDS 默认每个实例创建几个数据库？ A

- A、8
- B、16
- C、64
- D、1

9、DRDS 的产品定位不包括以下哪项？ B

- A、基于 RDS 的分布式数据存储和检索产品
- B、海量数据的多维分析
- C、降低用户使用分布式数据库的难度
- D、解决用户单 RDS 无法支撑业务的困难

10、ODPS 中，当一个用户被移除后，与该用户有关的哪些授权会保留？ A

- A、ACL 授权
- B、policy 授权
- C、项目空间授权
- D、package 授权

11、渗透测试是通过模拟什么的攻击方法，来评估计算机网络系统安全的一种评估方法？ C

- A、正常客户
- B、高并发客户
- C、恶意黑客
- D、网络专家

12、如果用户执通过 RDS API 的 CreateDBInstance 接口创建实例，则创建的 RDS 实例类型是什么？

- D
- A、以上都可以实例
- B、只读实例
- C、临时实例
- D、主实例

只读实例是：创建 RDS 只读实例

13、ODPS 的命令行工具 odpscmd 可以使用以下哪个参数调用某个文件中的命令集？ C

- A、k
- B、e
- C、f
- D、c

解释: `odpscmd -f xxx.sql`

14、DRDS 的服务可用性不低于百分之多少？

- A、100
- B、100
- C、100
- D、100

服务可用性：不低于 99.95%

15、通过 DMS 管理后台或者数据库客户端，连接 RDS 数据库时，提示错误信息“max_user_connections”，代表什么含义？D

- A、IOPS 超出极限
- B、RDS 空间满了
- C、网络中断
- D、RDS 数据库的连接数满了

16、关于 ODPS 中的任务（Instance）、作业（Job）、作业实例（Instance）以及工作流（Workflow）描述错误的是？C

- A、Job 是个静态的概念，被提交到系统开始执行时，该 Job 就拥有一个 Instance
- B、工作流描述了 Job 中各个 Instance 之间的依赖关系和运行约束，本质是个 DAG（有向无环图）
- C、单个 SQL query、命令或者 MapReduce 程序统称为一个 Task
- D、Instance 是个动态概念，每个 Instance 只能运行一次

17、OTS 的列对长度有限制。对于 string 类型，如果出现在主键里，长度限制为多少？B

- A、64k
- B、1k
- C、256k
- D、128k

解释：主键是表中每一行的唯一标识。主键由 1 到 4 个主键列组成。应用在创建表的时候，必须明确指定主键的组成，每一个主键列的名字和数据类型以及它们的顺序。属性列的数据类型只能是 String、Integer 和 Binary。如果为 String 或者 Binary 类型，长度不超过 1KB。

18、关于 ODPS 类型之间的类型转化，说法不正确的是？B

- A、隐式类型转换是 ODPS 依据上下文使用环境自动进行的类型转换
- B、部分类型之间不可以通过显式的类型转换，但可以通过 SQL 的隐式转换完成
- C、不支持的隐式类型转换会导致异常
- D、ODPS 支持的隐式类型转换规则与显式转换相同

19、某图书管理系统使用 ODPS 进行离线数据分析，现在他们要将同一类目的书籍的名称拼成一个字符串，即将属于同一分组的字符串合并在一列展示，以下哪个内置函数可以实现这一功能？

- A
- A、wm_concat
- B、split_part
- C、concat

D、regexp_instr

20、用户通过 RDS API 能查看可以使用的物理位置，在调用创建的 RDS 实例接口之前，需要调用“查询可用的物理位置”接口进行查询，以下哪个字段信息代表机房的物理区域？ A

- A、RegionID
- B、ZoneID
- C、InstanceID
- D、AccessKeyID

21、ODPS 对 SQL DML 语句进行解析后，得到执行计划。构成执行计划的是由多个有相互依赖关系的__A__组成？

- A、执行阶段（Stage）
- B、作业（Job）
- C、实例（Instance）
- D、任务（Task）

22、在 ADS 中，什么是用户所关心的最大单元，也是用户和 ADS 系统管理员的管理职权的分界点？ C

- A、表组
- B、表
- C、数据库
- D、列

23、某地市级网络提供商使用 ODPS 存储、分析宽带用户的流量信息，用户每登陆一次产生一条纪录，日均记录数在 2000 万条左右，目前最常见需求基于区县（该市共有 12 个区县）、日期（天）各种查询，其中部分查询只带有区县信息。架构师在设计用户登陆信息表时，决定建成分区表，对于分区列的选取，以下性能最好的是哪一项？ D

- A、以字符串型日期信息（天）为第一级分区列，以字符串型区县为第二级分区列
- B、以字符串型区县信息为分区列
- C、以字符串型日期信息（天）为分区列
- D、以字段组合（字符串型区县信息，字符串型日期信息（天）为分区列）

24、某电商使用 ADS 进行数据分析，其中商品信息表纪录了商品的型号、价格、厂家等等信息，从数据分析角度来讲，它是一张维表。该表包含了超过 2000 万条的纪录，大小在 5G 左右，在 ADS 中如何处理该表？ A

- A、必须建成 ADS 分区表，将会有一些限制，比如不能和不同普通表组上的表进行关联等
- B、建表时指定为 ADS 的维表，将分区数设置为 1
- C、建表时指定为 ADS 的维表，即不进行拆分，可以和任意表关联
- D、可将该表直接打平到和该表关联的事实表中，做成一张宽表，减少 join，提升性能

25、ODPS 表 t1、t2 都是分区表，分区键是 pt，对于 SQL 语句 select t1.id from t1 left outer join t2 on t1.id=t2.id where t1.pt='20140101' and t2.pt='20140101'，哪一个说法是正确的？ C

- A、t2 表会进行分区剪裁
- B、两张表都会进行分区剪裁

- C、t1 表会进行分区剪裁
- D、两张表都不会进行分区剪裁

26、关于 ADS 的聚集列，下列说法不正确的是哪一项？ B

- A、如果用户的查询 Query 的条件中会指定聚集列的内容或范围，那么这样的查询性能便会有较大的提升
- B、聚集列可以在建表后进行修改，修改后 DB 会在后台自动刷新数据
- C、一个分区内聚集列内容相同的数据会尽可能的分布在同样的区块内存
- D、创建表时可以指定一列或者若干列作为聚集列

备注：聚集列

在创建表时，用户可以指定一列或者若干列作为聚集列。在物理上，一个分区内聚集列内容相同的数据会尽可能的分布在同样的区块内存放，因此如果用户的查询 Query 的条件中会指定聚集列的内容或范围，那么这样的查询性能便会有较大的提升。需要注意的是如果用户指定多列为聚集列，那么指定的聚集列的顺序便是比较数据是否相同的顺序。

聚集列可以在建表后进行修改，但是目前修改后下次装载数据完毕后生效。

27、某企业使用 ADS 做数据分析，其中部分数据来源于 ODPS。技术人员在使用 LOAD DATA 命令从 ODPS 加载数据到 ADS 的过程中，碰到如下错误信息：

ERROR 1105 (HY000): You are not the owner of the source table

最有可能的出错原因是哪种？ A

- A、ODPS 的用户 garuda_build@aliyun.com 没有对源表的操作权限
- B、ADS 的用户 garuda_build@aliyun.com 没有对源表的操作权限
- C、源表不存在
- D、源表中没有数据，为空表

28、云盾网络安全专家服务特别适合以下哪种场景？ C

- A、系统每次重大升级后不清楚有没有漏洞
- B、够买了 100 台 ECS，但没有专业的管理同
- C、每月一次的产品促销活动，担心网络攻击
- D、频繁出现 SQL 注入警告

29、基于 RDS 系统架构以及只读实例的应用场景，需要满足 MySQL 版本最低为多少，否则是不能满足只读实例的应用场景？ A

- A、MySQL 5.6
- B、MySQL 5.5
- C、MySQL 5.1
- D、MySQL 5.2

30、ODPS 中一个表可以上限最多拥有多少个分区？ C

- A、10000

- B、不限制
- C、60000
- D、100000

31、作为技术人员，小王被 ODPS 强大的数据处理能力吸引，想去体验一下。他快速注册了一个阿里云帐号，进入官网后试图创建一个名字为 test_project 的 ODPS project，结果没有成功。分析，小王出错的最有可能的原因是哪一个？ D

- A、帐号未充值，余额不足导致出错
- B、帐号未开通 ODPS 的使用权限
- C、未创建合适的 accessID 和 accessKey
- D、项目名字和现有的其他帐户的名字重复

32、表 t 中有 1000 万行数据，想要随意抽取 10 行数据，在 odpscmd 中执行消耗时间最短的方式是哪种？ C

- A、select * from t where rownum<=10
- B、select * from t limit 10
- C、read t 10
- D、get t 10

33、关于 ADS 的 ECU，说法正确的是？ D

- A、可以动态的扩容或者缩容，两者都是瞬间的同步操作
- B、通过参数设置，可以使同一个 ADS 的库同时使用两种类型的 ECU
- C、可以通过查询表 information_schema.resource_request 获得目前的 ECU 的状态
- D、可以通过 DDL 语句 alter database set ecu_count=N 来修改 ECU 的个数

!!34、ADS 中两个普通表关联，对于 join key，哪个说法是不正确的？ D

- A、如果两张表的 join key 都没有 hashmap 索引，则会报错，修正方法为：至少一个 join key 上手工添加 hashmap 索引，且修改过的表需要重新加载数据，重新进行关联
- B、两张表的 Hash 分区数必须一致，且分区列一致
- C、两张表的 Join Key 至少有一列建立了 HashMap 索引
- D、如果选择只在一张表上建 HashMap 索引的话，推荐建立在数据量较大表的一侧

备注：为了更高效的进行表关联，分析型数据库对表关联操作在分析型数据库中，两个事实表进行 Join 的充要条件是：(1) 这两个表在一个表组；(2) 这两个表的 Join Key 是 Hash 分区列；(3) 两张表的 Hash 分区数必须一致，否则 Join 结果不准确；(4) 两张表的 Join Key 至少有一列建立了 HashMap 索引，推荐建立在数据量较小的一侧。

35、跨站脚本攻击（Cross-site scripting，通常简称为 XSS）可被用于进行窃取隐私、钓鱼欺骗、偷取密码、传播恶意代码等攻击行为，请问 XSS 攻击主要发生在什么位置？ C

- A、WEB 服务器端
- B、数据库端
- C、WEB 客户端
- D、缓存服务器端

36、在 project 中已经开启了 label security（列级别授权）的前提下，用户申请了一张表 tbl 的读权限，后来管理员将该表中的字段 x 的安全等级设置为了 4，且用户的安全级别为 1，用户能否读该表中的数据？ C

- A、提供信息不足以判断
- B、可以，因为之前已经有读权限了
- C、可以访问除了字段 x 之外的数据
- D、不可以，因为用户的级别较低

37、tbl_p 为一个 ODPS 的分区表，包含一个 p1='1' 的分区，现在想把该分区上的数据删除，达到使用 select * from tbl where pi='1' 查询时，返回结果为空的目的。以下哪个 SQL 可以实现该功能？ B

- A、truncate table tbl;
- B、alter table tbl drop partition (p1='1');
- C、insert into table tbl select * from tbl where 1=2;
- D、delete from table tbl;

38、A 为 ADS 中的普通表，a 为 A 的分区列，类型为 bigint，b 为 A 中的普通列，类型为 bigint，请判断以下相关 SQL 中哪句语法是错误的？ B

- A、SELECT a FROM A ORDER BY a LIMIT 100
- B、SELECT a+b, count(a) FROM A group by a+b ORDER BY a
- C、SELECT a,b FROM A ORDER BY a+b LIMIT 100
- D、SELECT a, count(a) FROM A group by a ORDER BY a

0.8 版本中不支持但是 0.9 版本的 Full MPP Mode 支持（详见 4.4 节）：

- `SELECT a + b, COUNT(1) AS x FROM A GROUP BY a + b ORDER BY x`：0.8 版本下按照

非分区列表表达式（`a+b`）分组聚合再取 TOP-N 是非精确的（详见 4.3）

39、以下哪个不是由 ODPS 接入层提供的服务？ A

- A、用户空间管理操作
- B、HTTP 服务
- C、Load Balance
- D、用户认证

40、在 RDS 只读实例变更配置过程中，只读实例在 RDS 管理控制台中可以进行以下哪种任务？

- A
- A、查看 SQL 日志
- B、内外网切换
- C、更改配置
- D、参数设置

41、ODPS SQL 提供了 CASE...WHEN...，可以进行条件判断。表 tbl 中包含了字段（id bigint，age

bigint)，目前有三条纪录，分别为 (1, 20)，(2, 30)，(3, 60)，执行以下 SQL：

```
select id, case when age>60 then 'A' when age>40 then 'B' when age>20 then 'C' else 'D' end from tbl
where id=3;
```

哪个是正确的执行结果？ B

- A、A
- B、B
- C、C
- D、D

42、关于阿里云云监控的功能，描述错误的是？ C

- A、自定义监控，对用户开放自定义监控的服务，允许用户自定义个性化监控需求
- B、支持站点监控，包括对 http、ping、dns、tcp、udp、smtp、pop、ftp 等服务的可用性和响应时间的统计、监控、报警服务
- C、是安装在云主机 ECS 里面的一个客户端，可自动监测各项云服务是否有异常，还可以实时报警
- D、支持云服务监控，包括对 ecs、rds、slb、cdn、ocs、oss 等云服务的监控报警服务

43、OTS 种以什么方式来存储数据？ C

- A、对象
- B、二维表
- C、键值对
- D、聚簇列

44、ODPS 项目空间 prj1 中存在表 t1，owner 打开 LabelSecurity 设置后，执行如下命令：

```
SET LABEL 1 TO TABLE t1;
SET LABEL 2 TO TABLE t1(id);
SET LABEL 3 TO TABLE t1;
```

此时，t1 表中的列 id 的敏感等级标签是什么？ C

- A、0
- B、1
- C、2
- D、3

45、RDS 采用的主备 M-M 的高可用架构，其主备之间的数据同步依靠日志的方式实现，MySQL 数据库和 SQL Server 分别使用的日志是什么？ C

- A、均为事务日志
- B、MySQL 数据库为事务日志，SQL Server 为 BINLOG 日志
- C、MySQL 数据库为 BINLOG 日志,SQL Server 为事务日志
- D、均为 BINLOG 日志

46、关于 ODPS policy 授权，说法不正确的是？ D

- A、支持根据 ip 授权
- B、支持表名通配符授权
- C、支持授权的时间段
- D、支持根据用户授权

D、支持根据数据敏感级授权

47、ADS 中创建事实表（普通表）时会有一些限制，下列哪个说法是错误的？D

- A、表的名字不能超过 32 个字符
- B、至少有一级 Hash 分区
- C、一个事实表最多不能超过 1024 列
- D、二级分区只能是 Hash 或者 List

48、管理云盾加密服务的密钥时，必须通过哪种方式进行认证？A

- A、身份证（USB Key）
- B、用户名和密码
- C、用户自己上传的 SSL 证书
- D、阿里云官网帐号的 AK

49、SQL 注入攻击就是攻击者通过欺骗什么服务器，执行非授权的任意查询过程？B

- A、文件
- B、数据库
- C、备份
- D、WEB

50、下面哪一项不是安骑士包含的功能？C

- A、高危漏洞修复
- B、异地登陆报警
- C、防 WEB 应用系统密码破解
- D、木马文件检查

二、多选题（共 30 题，每题 1 分，共 30 分）

1、某在线系统使用的是 DRDS，其中订单表保存了用户订单信息，目前订单表的分库键为订单标识。事实上基于该订单表的常见查询中经常用到的查询条件除了订单标识外，还有用户标识，为了减少 库的扫描，保证这两类常用查询的性能，可以采取以下哪些方案？CD

- A、如果同一个查询中只出现一个字段，则可以使用异构索引表
- B、如果同一个查询中同时出现这两个字段，则创建基于这两个字段的索引
- C、如果同一个查询中同时出现这两个字段，可调整表结构，其中一个字段作为分库键，另一个字段作为分表键
- D、如果同一个查询中只出现一个字段，则新建一张结构相同但是分库键为用户标识的表，将数据从源表加载进来，同时维护两张表的数据

2、关于 ODPS 跨项目资源共享的说法正确的是？ACD

- A、添加资源时支持的对象类型不包括 Project 类型，即不能将 project 添加到 packet 里
- B、添加到 Package 中的不仅仅是对象本身，还包括相应的操作权限，如不显示指定权限，则默认为读写权限
- C、Packet 是一种跨项目空间共享数据及资源的机制，主要用于解决跨项目空间的用户授权问题
- D、如果某个 package 还有人在使用，则该 package 的 owner 无法删除此 package

3、以下场景中，哪些适合使用 ODPS？ BCD

- A、电子商务网站的在线交易
- B、大型互联网企业的数据仓库和 BI 分析
- C、PB 级离线日志分析
- D、基于海量数据的用户特征和兴趣挖掘

4、下面哪些是 RDS 提供的系统性能监控项目？ ABCD

- A、IOPS
- B、CPU 利用率
- C、TPS
- D、网络流量

5、OTS API 中提供了哪些 Action 可以进行多行数据操作？ ABD

- A、BatchGetRow
- B、GetRange
- C、GetMultiRow
- D、BatchWriteRow

6、ADS 提供哪些批量导入数据的方式？ AC

- A、data pipeline 系列命令
- B、通过 rest-like api 方式
- C、通过 insert
- D、通过 batchinsert 命令

7、OTS 费用从哪几个维度计量？ ABC

- A、数据存储
- B、预留读写吞吐量
- C、外网下行流量
- D、数据访问次数（万次）

8、对于 OTS 表的预留读写吞吐量，以下各项中描述正确的是？ A

- A、当表上的操作过于频繁导致预留读 / 写吞吐量不足以进行更多操作时，OTS 会返回 OTSQuotaExhausted 错误给应用程序
- B、1 单位读能力表示该表每秒可以读 1KB 数据
- C、1 单位写能力表示每秒可以写 1KB 数据
- D、操作数据大小不足 1KB 的部分向上取整

9、OPDS 中用户认证(Authentication)的主要功能是检查请求 (Request)发送者的真实身份，通过包括以下哪几项验证？ ABC

- A、正确验证用户状态，包括是否欠费等
- B、正确验证接收到的信息在途中是否被篡改
- C、正确验证消息发送方的真实身份
- D、用户需要提取用户个性化信息时进行用户个性化认证

10、RDS 中慢 SQL 包括哪些操作？ ACD

- A、全表扫描
- B、内存排序
- C、扫描数据量过大
- D、磁盘排序

11、云盾安全网络除了 DDoS 防护功能外，还具备哪些安全功能？ ABD

- A、WAF 应用防火墙
- B、CC 防护
- C、网站内容检查
- D、隐藏源站

解释：一、相同点

- 1、具备防 DDoS、CC 和 WAF 功能。
- 2、支持阿里云 ECS、SLB 和非阿里云 IDC 机房服务器。

二、不同点，主要体现在 DDoS 防御方案及效果上

- 1、高防 IP 是海量带宽硬扛，支持 300G 按天付费，保证在清洗过程中对用户业务体验无损失。支持 HTTPS 业务的 WAF 防御。
- 2、安全网络是多节点切换防御，每个节点的防御带宽小，防御过程中对用户业务有一定影响，可通过集成 SDK 把业务影响时间控制在 1 秒内。暂时不支持 HTTPS 业务的 WAF 防御。

12、在 RDS 只读实例变更配置过程中，RDS 主实例无法进行哪些任务 ACD

- A、备份实例
- B、查看 SQL 审计日志
- C、创建实例
- D、内外网切换

13、对于 ODPS 项目空间的描述，正确的是？ ACD

- A、一个用户可以拥有多个项目空间的权限
- B、所有的对象都会属于某个项目空间
- C、通过安全授权，可以在一个项目空间中访问另一个项目空间
- D、一个帐号可以创建多个项目空间，最多不能超过 10 个

14、ODPS 的控制层主要功能包括？ ABCD

- A、用户空间和对象的管理
- B、命令的解析
- C、执行逻辑
- D、数据对象的访问控制与授权

备注：来自 ODPS 权威指南

- (1)用户控件(project)管理
- (2)对象管理，包括对表、资源、作业的管理
- (3)命令的解析和执行
- (4)元数据的管理

执行逻辑：计算层

!!!15、用 Java 写 ODPS 的自定义函数时，以下哪几种类型是可以用在接口中的类型？ AC

- A、long
- B、int
- C、string
- D、bigint

UDF 支持 ODPS SQL 的数据类型有：bigint, string, double, boolean 以及 datetime 类型。

ODPS 数据类型与 Java 类型的对应关系如下：

ODPS SQL Type	Bigint	String	Double	Boolean	Decimal
Java Type	Long	String	Double	Boolean	Decimal

16、ODPS SQL 支持的 Join 操作类型包括哪些？ ABCD

- A、LEFT OUTER JOIN
- B、RIGHT OUTER JOIN
- C、INNER JOIN
- D、FULL OUTER JOIN

17、ADS 支持一下哪些连接方式？ ABCD

- A、安装了 php-mysql-5.1.x 模块的 php 环境
- B、使用版本为 5.4 系列的 mysql-jdbc 驱动
- C、用户管理控制台中的 DMS
- D、mysql5.6 的客户端

18、DRDS 全表扫描目前支持哪些聚合函数？ ABCD

- A、SUM
- B、MIN
- C、COUNT
- D、MAX

19、某猎头公司启动了一个人才管理系统，从网上搜集人才简历，保存到自己的 OTS 系统的表 t_resume 中，部分纪录包含了一些大小超过 64G 的信息，比如扫描的证书照片、个人照片等信息。作为架构师，请为这些大字段选择合适的处理方式 CD

- A、将超过 64KB 的单个值拆分成多个列存储在 OTS 中
- B、根据大字段内容生成 MD5，将 MD5 码替代原有内容保存到字段中
- C、使用 OSS 存储大字段，并将对应的 url 存入到 OTS 的表 t_resume 中
- D、将属性列压缩之后再以 Binary 类型存储到 OTS 的表 t_resume 中

20、在管理控制台的云盾配置页里可以设置哪些“报警方式”？通过这些方式，用户可以在出现风险的第一时刻收到事件通知。BCD

- A、钉钉消息
- B、实时电话
- C、手机短信
- D、电子邮箱

21、DRDS 则将整个扩容的过程分为多个阶段，包括以下哪些步骤？ AC

- A、切换数据库
- B、增量同步
- C、全量迁移
- D、重启实例

22、某企业用户目前的在线订单系统运行在 MySQL 版的 RDS 上，因为业务量激增，公司决定采用 DRDS，即将现有系统从 RDS 迁移到 DRDS 上去。从项目实施部署的角度来看，以下哪些说法是正确的？ ABCD

- A、需要搜集当前表的大小和具体如何使用等，用来判断哪些表需要拆分，哪些不需要
- B、需要评估当前应用，以识别哪些应用需要改造
- C、需要评估哪些应用不能支持，需要做应用改造
- D、需要评估存储容量以及表的大小，以合理规划初始化 DRDS 的实例和库的个数

23、连接 RDS 的 MySQL 数据库的方法有哪些？ ABC

- A、客户端：Mysql-Front
- B、Mysql 命令登陆
- C、Navicat_mysql
- D、SQL Server Management Studio

24、使用用户管理控制台创建 ADS 时，需要提供 ECU 的哪些信息？ BC

- A、存储
- B、数量
- C、类型
- D、vCPU

25、对于数据库类型为 MySQL 的 RDS 实例，有以下哪些种连接方法？ ABCD

- A、iDB Cloud（控制台登陆）
- B、MySQL-Front
- C、Navicat for MySQL
- D、使用 mysql 命令登陆

26、ODPS 中项目普通用户想要使用 `select * from` 访问一张非本人创建的本项目的表 A，至少需要以下哪些权限？ BC

- A、项目 P 的 select 权限
- B、表 A 的 select 权限
- C、项目 P 的 createinstance 权限
- D、表 A 的 describe 权限

27、使用 MySQL 客户端连接 ADS 时，以下哪些信息是必须的？ABC

- A、端口
- B、服务器地址
- C、数据库名称
- D、AccountID

28、对于 ADS 的特色功能，以下说法正确的是？ABC

- A、智能自动索引会自动为导入数据的每一列创建符合该列情况的索引类型，无需用户显式指定创建索引或索引类型
- B、聚集列能智能提升查询性能
- C、多值列支持 in、contains 等查询
- D、ADS 采用高职能的基于规则的优化策略

29、支持 OLTP 的数据库系统包括一下哪些特点？ABCD

- A、擅长事务处理
- B、数据原子性
- C、很好的支持频繁的数据插入和修改
- D、数据一致性

30、ODPS MR 中哪些操作会因为 Java Sandbox 而报错“access denied”？ABCD

- A、获取本地 IP
- B、起子进程执行 Linux 命令
- C、直接访问本地文件
- D、Java 反射限制

三、判断题（共 20 题，每题 1 分，共 20 分）

1、ODPS 中可以使用 Tunnel 模块上传视图数据。 否

2、ODPS Policy 授权文件的格式是 Json。 是

3、将数据定时增量加载到 ADS 表中，如果目标表的属性是批量更新，则该表需要保证有两级分区且第二级分区必须是 list 分区，才能保证新增数据不会覆盖原有历史数据。 是

4、如果要查看只读实例的数据延迟，用户只能通过 RDS 管理控制台进入 RDS 只读实例进行查看。
否

5、OTS 管理便捷，应用程序无需关心软硬件升级，配置更新，集群扩容等繁琐运维任务。 是

6、OTS 的 BatchWriteRow 用于插入、修改、删除一个表或者多个表中的多行纪录。 是

备注：批量插入，修改或删除一个或多个表中的若干行数据

7、ADS 中，当 updateType=realtime 时，支持 insert、delete、update 操作。 否

- 8、ADS 建表时如果选定实时写入，则可以使用 insert 插入数据，可以作为 OLTP 系统使用。 否
- 9、关系型数据库服务 RDS(Relational Database Service)是阿里云提供的一种稳定可靠、可弹性伸缩的在线数据库服务。 是
- 10、在 ODPS 中，所有的请求都会被转化为任务（Task）。 否
- 11、ODP 的 LabeSecurity 对敏感数据的粒度可以支持列级别。 是
- 12、ADS 的两张维度关联的时候，两个 join key 可以均不创建 hashmap 的索引。 否
- 13、在开发 ODPS MR 或者 UDP 的过程中，在用到文本文件、ODPS 表等的时候，只能把它们作为主方法的输入参数，由主方法读取后识别处理。 是
- 14、云监视除了可以为 aliyun.com 云服务器客户提供安全有效的监控服务外，还能够为其他自由服务器客户提供功能强大的监控服务，只是配置比较复杂，必须自己开发脚本进行数据收集。 否
- 15、目前 DRDS 可以基于 MySQL 以及 SQLServer 版的 RDS，但是不支持 PostGreSQL 版的 RDS 否
- 16、任务（task）是 ODPS 的基本计算单元。SQL 以及 MapReduce 等都是通过任务完成的。 是
- 17、ODPS 的分区只支持两种方式，对于数值分区键支持 mod，对于字符串类型的分区键则使用 hash 值作为分区的依据。 否
- 18、未经客户允许云盾先知计划不会对外公开任何漏洞标题及细节。 是
- 19、在 ADS 中，数据根据分区列进行分布式的存储和计算。 是
- 20、RDS 只读实例数据库版本必须大于等于 RDS 主实例版本，并且 RDS 主实例版本最低是 MySQL 5.6。 是

大数据

单选：

- 1、Data IDE 的数据管理模块主要是对项目空间中的表的元数据、权限申请等进行管理操作，以下说法错误的是（D）
- A. 可以通过数据管理模块查看项目空间中的表的生命周期
 - B. 可以通过数据管理模块查看项目空间中的表的结构
 - C. 可以通过数据管理模块申请项目空间中的表的查询权限
 - D. 可以通过数据管理模块查看项目空间中的表的数据
- 2、在数据仓库中，随着表数量越来越多，对于表的管理会越来越复杂，希望可以按一定的

规则划分类目，将不同的表归到不同类目中去，方便管理，即在业务层面上定义类目，对表进行类目划分，以下关于表类说法正确的是（D）

- A. Data IDE 的组织管理模块实现了类目管理
- B. 可以通过 Data IDE 的项目管理模块实现
- C. 大数据计算服务（MaxCompute，原 ODPS）中提供了类目管理功能
- D. Data IDE 的数据管理模块实现了类目管理

3、出于业务需要，要把云数据库 RDS 中的某张表的数据同步到大数据计算服务（MaxCompute，原 ODPS）的某张表中，希望通过 Data IDE 的同步任务来实现，以下关于实现流程的说法正确的是（C）

- A. 拥有项目管理角色的成员先配置 RDS 数据源，然后拥有运维角色的成员创建 MaxCompute 表，再创建、配置数据同步任务
- B. 拥有开发角色的成员先配置 RDS 数据源，然后创建 MaxCompute 表，再创建、配置数据同步任务
- C. 拥有项目管理员角色的成员先配置 RDS 数据源，然后拥有开发角色的成员创建 MaxCompute 表，再创建、配置数据同步任务
- D. 拥有开发角色的成员先配置 RDS 数据源，然后拥有运维角色的成员创建 MaxCompute 表，再创建、配置数据同步任务

4、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数据同步，DataX 任务配置文件使用（D）

- A. xml
- B. java
- C. bson
- D. json

!!!5、在 Data IDE 中使用数据同步任务将云数据库 RDS（MySQL）的用户日志表同步至大数据计算服务（MaxCompute，原 ODPS）中，其中 RDS 表中的用户访问时间字段为 visit-time（datetime 类型），格式为 YYYY-MM-DD HH:MM:SS，需要将该数据转化并存储在 MaxCompute 对应 string 类型列中，存储的日期格式为 YYYYMMDD，以下做法正确的是（B）

- A. 在 RDS Mysql 数据同步至 MaxCompute 中，RDS Mysql 侧不支持使用函数
- B. 在数据同步字段配置界面中增加字段 'DATE-FORMAT(visit_tim, '%Y%m%d')' 并与 MaxCompute 时间列对应
- C. 在数据同步字段配置界面中增加字段 'DATE-FORMAT(visit_tim, '%Y-%m-%d')' 并与 MaxCompute 时间列对应
- D. 在数据同步字段配置界面中增加字段 DATE-FORMAT(visit_tim, '%Y%m%d') 并与 MaxCompute 时间列对应

6、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数据同步。使用 DataX 将存储在对象存储 OSS 中的数据同步至阿里云端大数据计算服务（MaxCompute，原 ODPS）时，以（B）为粒度进行切分实现并发处理

- A. 没有任何切分
- B. OSS 的 object
- C. OSS 的 endpoint

D. OSS 的 Bucket

解释: <https://github.com/alibaba/DataX/blob/master/ossreader/doc/ossreader.md>

7、Data IDE 的数据同步任务 Task1，将数据从大数据计算服务（MaxCompute，原 ODPS）同步到云数据库 RDS（MySQL）中，源数据表中存在数据质量问题，加载的时候会有部分出错记录，Task1 中已经配置了容错记录条数，为了保证正确记录能够保留，在数据加载与控制中，还需要对 Task1 配置（B）

- A. 数据过滤条件
- B. 视为脏数据，保留原有数据（insert into）
- C. 作业速率上限
- D. 源表切分主键

!!!!8、Data IDE 上某个数据同步任务执行失败，运行日志中提示：“Code:[OdpsReader-33],Description:[分区配置错误.]-分区配置错误，根据您所配置的分区没有匹配到源头表中的分区.源头表所有分区是:[ds='2010328' ds='20160328'.done' ds='20160329' ds='20160330' ds='20160330.done'], 您配置的分区是:[ds=\${daytime}], 请您根据实际情况作出修改。”（B）会造成类似错误。

- A. 改任务在运行的时候，分区 ds=\${daytime}被删除了
- B. 配置任务的源表时，制定分区值时使用了自定义变量\${daytime}，但在数据同步节点的参数配置中未给变量 daytime 赋值

C. 该任务配置的源表分区还没有生成，该任务就已经开始运行了。可能的原因是该任务没有配置合适的依赖关系

- D. 配置任务的源表时，指定分区值时应该写成：ds='20160328' ds='20160328.done' ds='20160329' ds='20160329.done' ds='20160330' ds='20160330.done'

9、Data IDE 中，事实表 A 需要按天增量更新（配置成任务 task1），事实表 B 需要按天增量更新（配置成任务 task2），每天事实表 A 中的部分数据信息来自事实表 B 前一天的数据。Task1 和 task2 需要如何配置才能实现上述场景。（D）

- E. task1 的依赖属性中的上游任务配置为 task2
- F. 不需要配置以来关系，在 task1 直接使用表 B 生成表 A 即可
- G. task2 的依赖属性中的上游任务配置为 task1
- H. task1 的跨周期依赖属性设置为“等待自定义任务的上一周期结束”，自定义任务选择 task2

10、Data IDE 中，通过数据开发面板的脚本开发，可以创建 SQL 脚本文件并直接编辑、执行。

关于脚本文件，以下说法正确的是（B）

- E. 脚本文件可以自动调度
- F. 脚本文件不能发布到目标项目空间
- G. 脚本文件可以设置依赖属性
- H. 脚本文件不能保存

11、Data IDE 中，项目空间 prj1 的 SQL 任务 task1 中，部分数据来源于项目空间 prj2 的任务 task2 生成的表 table2，task1 和 task2 都是按天调度的周期性调度任务，task1 每次执行要使用 table2 同周期生成的数据，为了保证 task1 能正常执行，即每次运行时所需的 table2 数据

已经生成，以下 task1 和 task2 配置说法正确的是？（B）

- E. task1 代码中表 tabke2 的写法为“table2”，task1 依赖属性的上游任务配置为 prj2 中的 task2
- F. task1 代码中表 tabke2 的写法为“prj2.table2”，task1 依赖属性的上游任务配置为 prj2 中的 task2
- G. task1 代码中表 tabke2 的写法为“prj2.table2”，但是 task1 和 task2 不在个项目无法配置成依赖关系
- H. task1 代码中表 tabke2 的写法为“table2”，但是 task1 和 task2 不在个项目无法配置成依赖关系

12、Data IDE 中，task1 和 task2 都是按天调度的周期性调度任务，task1 生成分区表 table1，task2 生成分区表 table2，每天 task2 取 table1 最新分区的数据进行统计，而 task1 有部分数据来源是 table2 前一天分区的数据。Task1 和 task2 的依赖关系需要如何配置？（C）

- E. 只需要 task1 依赖属性的上游任务配置为 task2
- F. task1 和 task2 都不需要配置依赖属性
- G. task1 跨周期依赖选择“等待自定义任务的上一周期结束，才能继续运行”，自定义任务填写 task2，task2 依赖属性的上游任务配置为 task1
- H. 只需要 task2 依赖属性的上游任务配置为 task1

13、Data IDE 中表 A 的数据是通过按天调度的数据同步任务产生的，表 A 中的分区字段 dt 为 string 类型，填充的是数据产生的日期（格式为“yyyy-mm-dd”），即任务自动调度日期的前一天，比如 2016 年 4 月 9 日运行调度任务的时候，dt 的值为“2016-04-08”。在配置周期性同步任务的时候，A 表的 dt 字段需要（ ）才使得任务每次调度运行时自动替换日期。（C）

- E. 直接使用系统时间变量 \${bdp.system.cyctime}
- F. 引入自定义变量 var，令 var=\${yyyy-mm-dd-1}，使用 var 对 dt 进行赋值
- G. 直接使用系统时间变量 \${bdp.system.bizdate}
- H. 引入自定义变量 var，令 var=\${yyyy-mm-dd}，使用 var 对 dt 进行赋值

14、Data IDE 中，SQL 任务 task1 是按小时调度的周期性调度任务，任务配置为每 1 小时执行一次，结果会插入到表 table 中去，table1 为一个有两个分区键（ds 和 hh）的分区表，其中 ds 取前一个小时的年月日（格式为 yyyymmdd），hh 取前一小时的小时（格式为 hh24），要使用调度系统提供的时间参数定义分区，使得 task1 每次调度运行中自动替换时间值，task1 应该如何配置？（B）

- E. task1 代码里 table1 的分区 ds=\${daytime}，hh=\${hh}，任务参数配置：daytime=\${yyyymmdd},hh=\${hh24-1/24}
- F. task1 代码里 table1 的分区 ds=\${bdp.system.bizdate},hh=\${hh}，任务参数配置：hh=\${hh24-1/24}
- G. task1 代码里 table1 的分区 ds=\${bdp.system.bizdate},hh=\${hh}，任务参数配置：hh=\${hh24}
- H. task1 代码里 table1 的分区 ds=\${daytime}，hh=\${hh}，任务参数配置：daytime=\${yyyymmdd-1/24},hh=\${hh24-1/24}

!!!! 15、Data IDE 中，通过运维中心的任务运维视图可以对调度任务操作重跑并恢复调度，即重跑该任务同时把下游未运行的任务恢复调度，任务 task3 依赖任务 task2，任务 task2 依赖任务 task1，对 task2 操作重跑并恢复调度，以下哪个场景会使 task2 和 task3 执行？（A）

- E. task1 为成功状态, task2 为失败状态, task3 为未运行状态
- F. task1、task2、task3 都是未运行状态
- G. task1、task2、task3 都是运行成功状态
- H. task1 为失败状态, task2、task3 都是未运行状态

16、Data IDE 中, 调度任务每次运行前都先将任务实例化, 即生成一个实例, 调度运行时实际上在执行相应的实例。运维中心中的任务运维视图有运维、测试、补数据 3 种类型实例, 运维实例是任务日常自动调度生成的实例, 测试实例和补数据实例是通过任务管理视图对任务进行测试和补数据生成的实例。以下哪种类型任务不会生成自动调度实例? (C)

- E. 按天调度的周期调度 workflow 任务
- F. 按月调度的周期调度 workflow 任务
- G. 一次性 workflow 任务
- H. 按周调度的周期调度 workflow 任务

17、Data IDE 中, 调度任务每次运行前都先将任务实例化, 即生成一个实例, 调度运行时实际上在执行相应的实例。通过运维中心的任务运维视图可以查看任务自动调度的实例。任务 task1 是按天调度的周期性任务, 2016 年 4 月 9 日 11 点 30 分成功提交到调度系统。2016 年 4 月 10 日到任务运维视图查看 task1 的实例时可以查到____C__。

- E. task1 业务日期为 2016-04-08 的实例
- F. task1 业务日期为 2016-04-10 的实例
- G. task1 业务日期为 2016-04-09 的实例
- H. 无相关实例

18、Data IDE 中, 项目空间 prj1 随着调度任务量的不断增多, 经常出现任务调度时长时间处于等待资源状态。为了对自己负责的周期性任务的调度时间进行规划调整, prj1 的开放角色成员 A 想统计 prj1 每天执行时间最长且任务负责人为自己的 10 个任务, 通过__B 可以快速查到。原答案 A, 修改后为 B

- E. 运维中心的任务管理列表
- F. 运维中心的任务运维列表
- G. 数据开发面板中的节点属性
- H. 项目管理中的血缘分布

19、大数据计算服务的图计算引擎是分布式的, 在加载数据时, 系统默认如何将数据分配到不同的 Worker 中? (D)

- A. 完全随机分配
- B. 按照点 ID 值排序, 再平均分配给 Worker
- C. 根据边的权值进行 hash 并对 Worker 数取模
- D. 根据顶点 ID 的 hash 值对 Worker 数取值

20、开发人员在使用大数据计算服务的图计算时, 主要的输入数据在表中, 另外还有一些非结构化的配置信息存在一个本地文件中, 则可以采用哪种方式使得这些配置信息在图计算过程中可用? (A)

- E. 将这个配置信息做为 resource 上传到大数据计算服务中, 在图计算过程中可以读取 resource 中的数据

- F. 输入输出只能是 Table，必须要把这部分数据放到表中
- G. 在 group 程序中直接读取本地的文件信息
- H. 输入只能是一张表，因此无法满足这个场景

21、传统的 MapReduce 模型要求每一轮 MapReduce 操作之后，数据必须落地到分布式文件系统上。而一般的 MapReduce 应用通常由多个 MapReduce 作业组成，每个作业结束之后需要写入磁盘，接下去的 Map 任务很多情况下只是读一遍数据，为后续的 Shuffle 阶段做准备，这样其实造成了冗余的 IO 操作。为了解决这一问题，提供更优的性能，大数据计算服务提供了扩展的 MapReduce 模型，该模型区别于普通 MapReduce 模型的主要特点是？（A）

- F. 支持 Map 后连接任意多个 Reduce 操作，如 Map-Reduce-Reduce
- G. 支持 Map 后不连接 Reduce，而是连接另一个 map，如 Map-Map-Reduce
- H. 支持 Chain Mapper / Reducer，即支持 Map-Reduce-Map-Reduce
- I. 支持没有 Map，直接进入 Reduce

!!!! 22、大数据计算服务的 MapReduce 中，归并函数是 MapReduce 计算框架的一种优化，即框架会在 map 端对相同 key 值的数据进行本地的合并操作。关于 Combiner（归并），以下说法正确的是？（A）

- D. Combiner 的合理使用，可以减少传输带宽，提升整体性能
- E. Combiner 是必须的
- F. Combiner 的合理使用，可以减少传输带宽，提升整体性能
- G. Combiner 需要和 partitioner 配合使用

23、从优化角度来看，大数据计算服务的 MapReduce 对于哪个节点处理哪个分片这个问题，采用了___A___策略。

- F. 本地性原理，即数据在哪台机器上，就在哪台机器上启动 map
- G. 能者多劳原理，将数据移动到计算能力最强的节点上，启动 map
- H. 基于历史处理信息的优化原理，即将数据移到从历史执行信息看处理效果最好的机器上，启动 map
- I. 负载均衡原理，即将数据放在当前负载最小的节点上启动 map

24、在大数据计算服务中，使用 MapReduce 开发一个分词程序，即输入为保存在表中的多篇文章（每篇文章一行记录），参考一个字典文件（该文件会经常更新）进行分词。简而言之，在执行该程序的时候，既需要输入待分词的文章（已经保存在表中了），又需要输入字典（为经常变动的本地文件），以下最合理的解决方案为？ D

- A. 将字典文件保存在本地，在 MR 中通过访问本地最新的字典文件实现
- B. 将这个字典与源代码一起进行打包，直接执行即可
- C. 将字典打成 jar 包，使用 jar 命令执行 MR 程序时，通过 resource 参数指定该 jar 包
- D. 每次都把字典文件作为资源上传到大数据计算服务上，然后 MR 中通过资源的方式来访问

25、Bob 使用 java 开发了一个大数据计算服务的 MapReduce，用来对海量交易数据做时序分析，结果该 MR 性能较差，某些 map 存在明显的长尾效应，___B___方法可以有效的消除长尾效应。

- D. 只关心处理的列，忽略其余的多余列

E. 合理使用 combiner，适当的减少数据交互量

- F. 通过 set 命令或者 Jobconf 修改 MR 参数，增大 map / reduce 的内存
- G. 通过 set 命令或者 Jobconf 修改 MR 参数，影响分片的大小和个数

26、传统的数据仓库通过加工、整理、分析企业的历史数据，提供有效的信息用于辅助企业决策，常见的应用类型有即席查询、固定报表、OLAP 分析以及数据挖掘等。阿里云提供的数据类产品可以非常便捷、高效的实现传统的数据仓库。上述的四项功能可以通过以下哪个产品或者产品组合实现？（C）

- A. 大数据计算服务
- B. 分析型数据库
- C. 分析型数据库、大数据计算服务
- D. MySQL 版云数据库、分析型数据库

27、某在线广告 APP 采用了阿里云的技术，通过对网上搜集的海量数据的分析，给企业主提供精准的广告投放人群。该 app 的产品经理介绍该系统能实现以下技术指标：1-能存储海量数据，单表大小超过 10TB 2-支持多张上亿的表进行 Join 3-支持多种筛选条件，包括连续范围筛选、取值筛选、and / or / not 条件组合等 4-支持快速（3 秒以内）对数十个维度分别进行透视分析 根据以上信息，请判断该 app 最有可能使用的是阿里云的哪一款产品？（C）

- A. 分布式关系型数据库
- B. 表格存储
- C. 分析型数据库
- D. 大数据计算服务

28、大数据计算服务中同步数据到分析型数据库中时，特别要注意用户权限问题。账户 [ALIYUN\\$Bob@aliyun.com](mailto:ALIYUN$Bob@aliyun.com) 要从大数据计算服务中将表 T_ODPS 导入到分析型数据库的表 T_ADS 中去，在大数据服务中，必须将表 T_ODPS 的 select 和 describe 权限赋给以下哪个账户？（A）

- A. [ALIYUN\\$garuda_build@aliyun.com](mailto:ALIYUN$garuda_build@aliyun.com)
- B. [ALIYUN\\$root@aliyun.com](mailto:ALIYUN$root@aliyun.com)
- C. [ALIYUN\\$Bob@aliyun.com](mailto:ALIYUN$Bob@aliyun.com)
- D. [ALIYUN\\$garuda_data@aliyun.com](mailto:ALIYUN$garuda_data@aliyun.com)

29、大数据计算服务中的日志表 log 是一张分区表，分区键是 dt，每天产生一个分区用于存储当天新增的数据，现在共有 dt='20160101'至 dt='20160131'共 31 个分区的数据，为了删除 20160101 当天新增的数据，可以通过___B___方式实现。

- E. delete from log where dt='20160101'
- F. alter table log drop partition (dt='20160101')
- G. truncate table where dt='20160101'
- H. drop partition log (dt='20160101')

30、某企业的数据仓库运行在大数据计算服务上，开发人员在加工数据时发现用户表 user 中的用户标识字段 user_id 有部分脏数据，正确的格式应该是 8 位的字符串。可以使用 SQL 语句___D___将脏数据过滤出来。

- E. select* from user where length(userid)<>8

- F. `select* from user where length(userid)>8`
- G. `select* from user where length(userid)<8`
- H. `select* from user where length(userid)<>8 or userid is null`

31、某企业需要一个能够从身份证号中分析得出所属省份的自定义函数，哪种自定义函数适用于这种场景？（B）

- E. User Defined Splitted Function（用户自定义分组函数）
- F. User Defined Scalar Function（用户自定义标量函数）
- G. User Defined Table Valued Function（用户自定义表值函数）
- H. User Defined Aggregation Function（用户自定义聚组函数）

32、大数据计算服务的数据仓库的管理员发现用户表 `user` 没有备注信息，为了方便其他人员的使用，可以通过____A____方式给这张表加上备注。

- E. `alter table user set comment' my comment';`
- F. 删除表 `user`，在重建时指定 `comment` 选项
- G. `touch table user set commnt' my comment';`
- H. `change table user set comment' my comment;`

33、某 BI 系统构建在大数据计算服务上，某数据分析员在分析订单的地域分布时，需要把订单表 `order`（占用存储空间约 10G）与地域维表 `region`（占用存储空间 100M）关联起来，两张表的结构如下：`create table order(order_id string, region_id string, order_amt bigint); create table region(region_id string, region_name string)`在运行时发现由于订单大量集中在上海地区，数据的倾斜导致整体运行时间较长，运行的 SQL 语句如下：`select region_name, sum(order_amt)amt from order t1 join region t2 on t1.region_id=t2.region_id;`此时可以考虑通过__A__优化 SQL 的运行效率。

- E. 使用 `mapjoin hint: select/*+mapjoin(t2)*/region_name,sum(order_amt)amt from order t1 join region t2 on t1.region_id=t2.region_id;`
- F. 使用 `mapjoin hint: select/*+mapjoin(t1)*/region_name,sum(order_amt)amt from order t1 join region t2 on t1.region_id=t2.region_id;`
- G. 调整两张表的位置：`select region_name, sum(order_amt) amt from region t1 join t2 on t1.region_id=t2.region_id;`
- H. 减小 SQL 的 `splitsize`，增加计算资源

34、大数据计算服务中的客户表 `customer` 是一张非分区表，现在想要清空表的内容然后重新上传数据，可以通过____B____将数据清空。

- E. `insert overwrite table customer select* from customer;`
- F. `truncate table customer;`
- G. `delete from customer;`
- H. `drop table customer;`

35、大数据计算服务中，临时数据表 `tmp_item` 是一张非分区表，开发人员在建表时指定了 `lifecycle` 属性为 30，且使用一次后未再进行任何操作和访问。30 天后这张表会____B_____。

- E. `tmp_item` 会被自动重命名为 `tmp_item.deleted`
- F. `tmp_item` 表会被自动删除掉

- G. 不会任何变更
- H. tmp_item 中的数据会被清空，表结构仍存在

36、大数据计算服务的开发人员创建了日志表，并且导入了历史数据后发现表的一个字段命名拼写有错误，可以采用___D___更正此问题成本最低。

- E. 将原表重命名，然后建张正确的表，再导入历史数据
- F. 在表的末尾新添加一个字段
- G. 删除表后重建，然后重新导入历史数据
- H. 通过 alter 命令重新命名该字段

37、在大数据计算服务的数据仓库中的订单表 fact_order，建表语句如下：create table fact_order(order_id string, order_amt double, order_dt string) partitioned by (dt string);此表中的数据是从 ods_order 加工而来，ods_order 建表语句如下：create table ods_order(order_id string, order_amt bigint, order_dt string); ods_order 中有一条记录数据值是 order_id order_amt order_dt 0001 100 20160301 运行 SQL 语句将数据从 ods_order 加载到 fact_order 中：insert overwrite table fact_order partition(dt='20160301') select* from ods_order; 对此语句的执行结果描述正确的是___B___。

- E. 提交 SQL 时会报语法错误
- F. 语句可以执行，order_amt 的值会被自动的转为 double 类型
- G. 语句可以执行，但是这条数据会被当作脏数据丢弃
- H. 目标表与源表中的数据类型不一致，执行出错

38、大数据计算服务中的 fact_order 表的生命周期属性设置成了 30 天，开发人员通过以下命令创建了一张新的 fact_lottery_order 表：create table fact_lottery_order like fact_order; 对于新表的描述正确的是___A___。

- E. fact_lottery_order 的生命周期属性是无穷大
- F. fact_lottery_order 的生命周期属性不会被设置
- G. fact_lottery_order 的生命周期属性是 30
- H. fact_lottery_order 的生命周期属性是 0

39、大数据计算服务中，日志表 ods_order 是一张非分区表，结构如下：create table ods_order(order_id string, order_date string, order_amt decimal); fact_order 是一张按日期分区的表，结构如下：create table fact_order(order_id string, order_amt decimal) partitioned by (order_dt string); 现在需要将数据从 ods_order 加载到 fact_order 表中，并且按照 order_date 进行分区，执行了以下 SQL: insert overwrite table fact_order partition(order_dt) select order_id, order_amt, order_date as order_dt from ods_order; ods_order 表中的 order_date 有部分记录取值为 NULL，对执行结果描述正确的是___B___。

- E. 语句可以正常执行，其中 order_date 为 NULL 的记录会被放入 order_dt='BAD'的分区
- F. 语句可以正常执行，在提交时会报语句错误
- G. 语句可以正常执行，其中的 order_date 为 NULL 值的记录会被丢弃
- H. 语句无法正常执行，系统会抛异常

40、大数据计算服务提供了 Tunnel SDK，支持高速上传下载数据。实际使用的时候有两种方式，第一种是通过 odpscmd 客户端，调用封装好的 tunnel 命令，第二种是直接使用 Tunnel

SDK 进行编程，通过调用程序进行个性化的上传下载。目前 SDK 支持以下哪种语言？（A）

- E. java
- F. C++
- G. Perl
- H. C#

41、使用大数据计算服务 tunnel 命令上传数据文件 data.csv 到表 t_data 成功后，不小心又之行了一次该上传命令，以下说法中正确的是？（A）

- E. 最后一次上传的数据会追加插入到表 t_data 中去
- F. 最后一次上传的数据会覆盖插入到表 t_data 中去
- G. 记录重复会报错退出
- H. 最后一次上传的数据会把 t_data 中已存在的相关记录更新

42、使用大数据计算服务 tunnel 命令上传 log.txt 文件到表 t_log 中去，t_log 为分区表，分区列为（p1 string, p2 string）。下列命令中正确的是？（C）

- E. tunnel upload log.txt t_log/p1="b1"/p2="b2"
- F. tunnel upload log.txt t_log/(p1="b1",p2="b2")
- G. tunnel upload log.txt t_log/p1="b1",p2="b2"
- H. tunnel upload log.txt t_log(p1="b1",p2="b2")

43、大数据计算服务提供了大数据的存储和计算服务，非常适合应用于大数据分析的领域。以下说法中错误的是哪一项？（C）

- E. 可以实现大型互联网企业的数据仓库和 BI 分析
- F. 提供了便捷的分析处理海量数据的手段，用户可以不必关心分布式计算细节，从而达到分析大数据的目的
- G. 可以支持实时 OLAP 分析
- H. 可以基于历史数据，进行用户特征和兴趣挖掘 5

44、大数据计算服务提供了客户端 odpscmd，支持 windows / linux 或者 OS X 等操作系统。在执行前需要对 odpscmd 的配置文件 odps_config.ini 中的某些参数进行设定，包括用于鉴权的 access_id/access_key,指定服务地址的 end_point,以及 tunnel 的服务地址 tunnel_endpoint 等。其中有一项参数为：project_name，关于这个参数的说法，正确的是？ C

- A. 如果不指定该值，且账户下就一个 project 的话，连接成功后会自动进入该 project
- B. 指定的项目必须存在，否则连接过程中会报错
- C. 如果给该值指定了恰当的 project 名字，连接成功后，会自动进入该 project
- D. 如果不指定该值，连接成功后会自动进入上次退出时所在的 project

45、大数据计算服务存储到文件系统上时，使用了_____ A _____的文件格式，降低了无效的磁盘读取操作。

- E. 列存储
- F. 文档存储
- G. 行存储
- H. 键值对存储

46、大数据计算服务支持基于标签的安全，它是项目空间级别的一种强制访问控制策略，它的引入是为了让项目空间管理员能更加灵活地控制用户对敏感数据的访问。LabelSecurity 可以对敏感数据的哪个粒度进行控制？ C

- E. 分区
- F. 表
- G. 列
- H. 行

47、大数据计算服务中，通常情况下，prj1 项目空间的成员若要访问 prj2 项目空间的数据，则需要 prj2 项目空间管理员执行繁琐的授权操作：首先需要将 prj1 项目空间中的用户添加到 prj2 项目空间中，再分别对这些新加入的用户进行普通授权。实际上，prj2 项目空间管理员并不期望对 prj1 项目空间中的每个用户都进行授权管理，而期望有另一种更高效的机制能使得 prj1 项目空间管理员能对许可的对象进行自主授权控制。以下哪种授权方式可以帮助 prj2 项目空间管理员实现他的预期目标？ C

- A. LabelSecurity
- B. Role
- C. Packet
- D. ProjectProtection

48、大数据计算服务项目空间 prj1 中存在表 t1，其中包含一列名字为 id，owner 打开 LabelSecurity 设置后，执行如下命令：SET LABEL 1 TO TABLE t1;SET LABEL 2 TO TABLE t1 (id); SET LABEL 3 TO TABLE t1; 此时，t1 表中的 id 列的敏感等级标签是什么？ C

- A. 1 级
- B. 3 级
- C. 2 级
- D. 0 级

多选

1、关于大数据计算服务中的 Policy 授权，说法正确的有？（3 个） B C D

- E. 授权或撤销授权时，对 Subject（如 User）没有要求，即授权的 User 存在或者不存在都可以
- F. 删除一个对象时，与该对象关联的 Policy 授权不会被删除
- G. 支持带限制条件的授权
- H. 授权或撤销授权时，要求 Object（如 Table）必须已经存在

2、随着社会数据收集手段的不断丰富及完善，越来越多的行业数据被积累下来，数据规模已经增长到了传统软件行业无法承载的海量数据级别。大数据计算服务可以很多的利用存储和计算能力去处理、分析这些海量数据，发现价值。以下说法中正确的有____ B C D ____（3个）

- E. 大数据计算服务提供了高并发低延迟（毫秒级）的处理能力，对海量数据进行灵活分析
- F. 大数据计算服务提供了高并发吞吐量上传下载工具，可以将外部数据源中的数据高效的同步到大数据计算服务中去
- G. 大数据计算服务支持使用 SQL 或者 MapReduce 编程对海量数据进行处理
- H. 大数据计算服务可以对海量数据进行基于复杂算法的分析，如聚类、分类、回归、主成分分析等。

3、大数据计算服务的整体架构包括了四个主要的部分，其中第一部分为客户端，即客户发起请求的方式。客户端包括__ A B ____。

E. Java SDK

F. ODPS CLT(Command line tools)

G. Nodejs SDK

H. RESTFul API

4、大数据计算服务中关于 tunnel 命令行工具的说法，正确的是？（2个） A C

- A. tunnel 上传到分区表时，会把数据直接上传到指定的分区中去，不保证业务逻辑
- B. tunnel 命令上传文件时，不能进行压缩
- C. tunnel 上传的数据是先写到临时目录，最后确定成功后才写到结果目录
- D. tunnel 上传过程中不加密，数据以明文方式在网络中传输

5、tunnel upload 命令可以上传数据到大数据计算服务中去，请判断以下哪些数据对象可以作为上传的目标？（2个） B D

- E. 资源
- F. 非分区表
- G. 视图
- H. 分区表

6、正则表达式可以用非常灵活的方式来处理字符串，在大数据计算服务 SQL 中的正则函数支持 POSIX 格式，以下对正则表达式描述正确的有__ A B C D ____（4个）

- A. [:blank::]匹配空格和 TAB 制表符
- B. \$匹配行尾
- C. [:alnum:]匹配字母字符和数字字符
- D. [:digit:]匹配数字字符
- E. \$^匹配空行

7、在大数据计算服务 SQL 中使用 distribute by 和 sort by 可以对数据进行局部排序，以下对 distribute by、sort by 的用法描述正确的有__ A B D ____。（3个）

- E. 数据会按照 distribute by 指定的列进行分片，hash 值相同的数据会在一个分片中
- F. 使用 sort by 子句时，必须有 distribute by 子句
- G. sort by 和 order by 的用法没有区别

H. 在一个分片中，数据会按照 sort by 指定的列进行局部排序

8、当大数据计算服务的内置函数无法满足业务需求时，可以使用 Java 开发自定义函数，对于普通的自定义标量函数的描述正确的有__ B C D ____。（3 个）

E. 可以实现多个参数类型不同的 evaluate 方法，系统调用函数时会根据参数的类型自动匹配到正确的方法

F. 在函数类中，函数的逻辑通过 evaluate 方法实现

G. 每个用户只能使用自己创建的 UDF

H. 在 UDF 中可以使用 MaxCompute 的资源（Resource），比如通过 add file 上传的文件资源

9、在大数据计算服务 SQL 中可以用逻辑运算连接多个条件，以下逻辑运算结果正确的有__ A B C E ____。（4 个）

A. True AND True=True

B. False OR False=False

C. True AND False=False

D. NULL and False=NULL

E. True or NULL=True

10、大数据计算服务 SQL 中支持 join、left join、right join、full outer join 等，对于连接条件描述正确的有__ A C D ____。（3 个）

E. 使用 mapjoin hint 时，语法上可以允许没有连接条件

F. 只能有一个连接条件

G. 没有 mapjoin hint 时，多个连接条件之间必须是 AND 关系

H. 没有 mapjoin hint 时，必须是等值连接

11、大数据计算服务中，检查数据何时被修改时需要参考 lastmodifiedtime 属性，对于分区表，在分区级别和表级别都有 lastmodified 属性。对此属性的影响，以下描述正确的有？（2 个） B C

E. 更新一个分区中的数据时，只会影响该分区的 lastmodifiedtime 属性

F. 更新一个分区中的数据时，表的 lastmodifiedtime 属性会被更新

G. 更新一个分区中的数据时，分区的 lastmodifiedtime 属性会被更新

H. 更新一个分区中的数据时，不会更新表和分区的 lastmodifiedtime

!!!!!! 12、大数据计算服务表 fact_sale_detail 的建表语句如下：create table fact_sale_detail(sale_date datetime, region string, amt decimal) partitioned by (dt string); 包括 dt='20160229', dt='20160301', dt='20160302', dt='20160303' 共四天的数据，每个分区中只包含当天的销售数据。Udf_substr 为自定义的数据（UDF），功能和 substr 相同但是性能更好。现在想要统计 3 月份的按天销售汇总金额，以下哪种方式经过分区剪裁后只需要付 3 个分区的数据久可以完成计算？ A B C D （4 个）

A. select dt, sum(amt) from fact_sale_detail where dt like '201603%' group by dt;

B. select dt, sum(amt) from fact_sale_detail where substr (dt,1,6) = '201603' group by dt;

C. select dt, sum(amt) from fact_sale_detail where dt in('20160301','20160302','20160303') group by dt;

D. select dt, sum(amt) from _sale_detaol where udf_substr(dt,1,6)='201603' group by dt;

E. `select dt, sum(amt) from fact_sale_detail where dt>='20160301' group by dt;`

13、用大数据计算服务构建海量的数据仓库时，分区表是一种很常见的做法，对于分区表的描述正确的有____ B_C D ____。（3个）

- E. 支持 hash 分区、范围分区、列表分区及组合分区
- F. 在通过 SQL 读取分区表中的数据时，可以通过指定分区的方法只读取一部分分组，减少 IO 开销
- G. 可以单独处理指定分区中的数据，不会对其他分区数据产生影响
- H. 对于过期的数据可以将对应的分区 drop 掉，不会影响其他分区中的数据

14、大数据计算服务中，表 `ods_sale_detail` 存放的是销售明细记录，建表语句如下：`create table ods_sale_detail(sale_date datetime, region string, amt decimal);`其中 `sale_date` 字段表示销售的日期，`amt` 表示销售金额，想要找出所有汇总销售额大于 10000 的日期以及当日的销售额，以下哪几个 SQL 可以达到这个目的？（2个） C D

- E. `select sale_date, sum(amt) from ods_sale_detail group by sale_date where amt>10000;`
- F. `select sale_date, sum(amt) from ods_sale_detail group by sale_date where sum(amt)>10000;`
- G. `select sale_date, sum(amt) from ods_sale_detail group by sale_date having sum(amt)>10000;`
- H. `select* from (select sale_date, sum(amt)s_amt from sale_date group by sale_date) sub where s_amt>10000;`

15、大数据常见的计算方式有三种：离线计算、流计算以及在线计算，关于这几种计算方式的说法正确的有？（3个） B C D

- E. 流计算是消息驱动的，数据更新一般是定时更新
- F. 离线计算的数据更新以定期的批量更新为主
- G. 在线计算一般在用户发起请求时（比如查询）发生
- H. 流计算可以应用于实时数值统计、实时事件预警等领域

16、数据仓库是一个面向主题的、集成的、相对稳定的、反映历史变化的数据集合，主要用于支持决策、面向分析型数据处理。大数据计算服务除了能很好的处理大数据语境下的分析功能外，也是传统数据仓库上云的解决方案之一。下列哪些传统数据仓库的功能可以通过大数据计算服务实现？（3个） B C D

- E. 即席查询
- F. 定时的 ETL 任务（数据抽取、转换和加载）
- G. 固定报表分析
- H. 数据挖掘

!!!! 17、使用大数据计算服务中的 MR 时，MR 中的哪些操作会引发 Java 安全沙箱的限制而报错？（3个） A B D

- E. 起子进程执行 Linux 命令
- F. 获取本地 IP
- G. 本地文件以资源方式上传至大数据计算服务，在 MR 中以资源方式访问该文件
- H. Java 反射限制

18、在开发大数据计算服务的 MapReduce 时，可以使用 JobClient 类中的接口提交 MapReduce

作业。目前提交方式支持 ____ C D ____ （2个）

- E. 远程提交
- F. 本地提交
- G. 阻塞方式
- H. 非阻塞方式

!!!! 19、MapReduce 最早是由 Google 提出的分布式数据处理模型，随后受到了业内的广泛关注，并被大量应用到各种商业场景中。以下场景中，_ B C D ____ 适合用 MapReduce 来实现。（3个）

- E. 交互式查询：基于海量数据的透视分析，用户可以通过上卷、下钻、切片等交互操作，了解数据集细节
- F. 机器学习：监督学习、无监督学习、分类算法如决策树、SVM 等
- G. 文本统计分析：比如词频 TFIDF 分析；学术论文、专利文献的引用分析和统计；维基百科数据分析等
- H. Web 访问日志分析：分析和挖掘用户在 web 上的访问、购物行为特征，分析用户访问行为

20、大数据计算服务图计算框架在计算过程中通过对图进行编辑、演化最终求解出结果，在计算过程中用户可以对图进行哪些编辑操作？（4个） A C D E

- A. 增加点
- B. 读取本地数据修改点的权值
- C. 增加边
- D. 删除点
- E. 修改边的权值

21、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。周期性调度任务每天自动调度的实例都是在当天 23:30 开始生成的，以下说法正确的是？（3个） A B D

- E. 任务 task1 是按天调度的周期性调度任务，2016 年 4 月 9 日 23 点 00 分，把 task1 的定时时间从原来零点改成 8 点 30 分，4 月 10 日，task1 的实例定时时间为 8 点 30 分
- F. 任务 task1 是任务 task2 的上游任务，2016 年 4 月 9 日 11 点 30 分，把 task2 的上游任务换成 task3，当天 task2 的自动调度实例依赖的还是 task1 的实例
- G. 2016 年 4 月 9 日 11 点 30 分新增的按天调度的周期性调度任务，当天可以自动调度运行
- H. 2016 年 4 月 9 日 11 点 30 分新增的按天调度的周期性调度任务，4 月 10 日可以自动调度运行

22、Data IDE 中，通过运维中心的任务管理列表可以对任务进行一些批量修改，从而提高项目管理员管理调度任务的效率，其中批量操作功能包括？（3个） B C D

- E. 添加报警
- F. 修改资源组
- G. 修改调度周期
- H. 修改负责人

!!!! 23、Data IDE 中，一次性工作流任务不会被自动调度运行，需要运行时，可以到运维中

心对一次性任务操作测试的方式手法运行，以下关于一次行 workflow 说法正确的是（）（2 个）

A C

- E. 一次性 workflow 创建好后，无法修改成周期调度 workflow
- F. 一次行 workflow 测试的时候，业务日期只能选择前一天的日期（昨天之前）
- G. 周期调度任务的依赖属性的上游任务不能配置为一次性 workflow
- H. 一次行 workflow 的依赖属性的上游任务可以配置为周期调度任务

24、Data IDE 中，任务 task1 是按天调度的周期任务，task1 生成分区表 table1，该表每次执行生成的分区（分区名 ds）取前一天日期的年月日（格式为 yyyyymmdd），要使用调度系统提供的时间参数定义 table1 的分区，使得 task1 每次调度运行时自动替换时间值，task1 应该如何配置？（2 个） A B

- E. task1 的代码中 table1 的分区 ds=\${bdp.system.bizdate}
- F. task1 的代码中 table1 的分区 ds=\${var}，task1 的参数配置：var=\${yyyyymmdd-1}
- G. task1 代码中 table1 的分区 ds=\${bizdate}
- H. task1 的代码中 table1 的分区 ds=\${var}，task1 的参数配置：var=\${yyyyymmdd}

25、Data IDE 中，可以通过脚本开发创建 SQL 脚本文件，用于执行一些临时性的代码。以下哪些情况下可以使用脚本文件？（3 个） B C D

- E. 数据科学家在脚本文件里调度机器学习任务实现复杂分析
- F. 数据分析师使用 SQL 探查表数据
- G. 数据开发人员使用建表语句建表
- H. 数据开发人员发调度任务调试成功后，使用 SQL 检查表中的数据是否正常

26、Data IDE 中，某个数据同步任务的源表为云数据库 RDS（MySQL）中的表，目标表为大数据计算服务中的表，运行的时候任务报错，执行日志中提示“ODPS-0130131:Table not found - 'xxx_prj.xxx_table”，以下说法正确的是（A C D）（3 个）

- E. 表不存在有可能是配置完任务之后被删除了，所以执行就报错了
- F. 该日志说明目标表不存在
- G. 该表不存在有可能是配置的时候选错了表名
- H. 该表不存在有可能是当前所属项目为生产项目，任务在开发项目配置测试好了，发布到生产项目后，没有到生产项目创建该表

27、在 Data IDE 中使用数据同步任务将云数据库 RDS（MySQL）加载到大数据计算服务时，为了提高单个数据同步任务的速度及效率，可以采取（B D）方式？（2 个）

- E. 数据过滤条件
- F. 讲云数据库 RDS 表的主键配置为切分键
- G. 容错记录条数
- H. 提高作业速率上限的配置

28、Data IDE 中，通过数据管理可以查看表的血缘关系以及表元数据信息，以下（A B C）信息变更后通过数据管理可以实时刷新查看（正确答案 3 个）

- E. 表新增的字段结构信息
- F. 表修改的描述信息
- G. 表修改了某个字段信息

H. 表中新增的数据内容

!!!! 29、Data IDE 中组织管理员是指组织的管理者，拥有组织层面的管理权限，以下操作中，（ B C D ）只能由组织管理员才能完成（正确答案 3 个）

E. 添加数据源（项目管理员添加）

F. 添加调度资源组

G. 添加组织成员

H. 创建 Data IDE 项目空间

30、Data IDE 中的数据同步支持离线批量数据同步模式，离线数据同步是指数据周期性、成批量地从源端系统传输到目标端系统，关于数据同步周期说法正确的是（ ）（正确答案 4 个）

A B D E

F. 支持天级调度

G. 支持周级调度

H. 支持实时调度

I. 支持小时级调度

J. 支持月级调度

判断题

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数据同步。在使用 DataX 进行数据同步时，源端数据的读取支持断点续传。 对

2、Data IDE 中，通过数据管理可以查看表的元数据，但无法查看某个表使用哪些表使用 错

3、在 Data IDE 中使用数据同步任务将云数据库 RDS 数据导入大数据计算服务（MaxCompute，原 ODPS）时，将 MaxCompute 侧数据清理规则配置为导入前保留已有数据，可以有效避免测试引入脏数据带来的主键冲突问题 对

4、在 Data IDE 中配置数据同步任务，将主键冲突选项配置为“视为脏数据，保留原有数据”时，数据同步任务进行时，脏数据还是会写入目标表中。 错

5、Data IDE 中，调度系统提供一些常用的时间参数，使周期任务自动调度运行时参数会自动替换具体的时间值。其中 \${bdp.system.cyctime} 参数值是取任务运行时的定时时间（年月日时分秒），若按天调度的周期任务配置定时时间为 20:00，则 2016-01-01 当天自动调度时该参数赋值结果是 20160101200000。 对

6、Data IDE 中，虚拟节点类型是空跑节点，即调度时直接返回成功状态不会真正执行。两个周期调度任务 task1、task2 之间没有依赖关系，业务上也没有共同上游任务，但是希望可以有一个共同的父人物进行统筹管理，那么可以创建一个虚拟节点作为 task1 和 task2 的上游任务。 对

7、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。 对

8、图是一种数据结构，由点和边组成，点和边包括权值，用点表示要研究的对象，边可以用来反映之间的各种关系。对

9、大数据计算服务在正式执行 Map 前，需要将输入数据进行“分片”。所谓分片，就是将输入数据切分为大小差不多相等的数据块，每一块作为单个 Map Worker 的输入被处理，以便于多个 Map Worker 同时工作。对

10、大数据计算服务的 MapReduce 的输入可以为空，也可以为视图。错 （不可以为视图）

11、通过阿里云的数据集成服务，可以将云数据库、分布式关系型数据库、分析型数据库中存储的结构化数据同步到大数据计算服务中去。 错 （ads 只进不出）

12、在一个大数据计算服务的 SQL 语句中，可以读多张表中的数据，但是只能更新一张目标表中的数据。错 （多路输出？）

13、大数据计算服务中，在命令行工具 odpscmd 中使用 select 语句查看表中的记录时，可以结合 limit 子句来限制输出结果的行数。 对

14、大数据计算服务的函数分为标量函数、聚合函数及表值函数几种类型，各种函数都可以嵌套使用。对

15、大数据计算服务中，SQL 的 order by 子句对结果数据进行排序，默认是升序，NULL 值会被认为是最小值。对

16、大数据计算服务中，任何有创建函数权限开发者都可以创建与系统内置函数同名的函数，系统在调用时会优先调用用户自定义的函数。错 （自定义函数不能与内置函数同名）

17、大数据计算服务的函数 getdate 可以用来获取系统当前时间，某 SQL 语句中使用了 getdate 函数，某次执行时，返回多条结果记录，这些返回记录中 getdate 的值是一样的。 对

18、使用大数据计算服务的 Tunnel SDK 上传下载数据时，任务会被提交至 Tunnel 服务器（tunnel endpoint），由其进行响应和处理。 对

19、表是大数据计算服务的数据存储单元，它在逻辑上是由行和列组成的二维结构。所有的数据都被存储在表中，各种不同类型的计算任务的操作对象都是表，同时，用户可以创建表、删除表、上传数据到表中，同时也可以更新表中符合条件的数据。 对

20、当大数据计算服务项目空间的 Owner Alice 决定对另一个用户授权时，Alice 虚线将该用户添加到自己的项目空间中来，只有添加到项目空间中的用户才能够被授权。 否 可以通过资源授权的方式

Odps 复习注意事项：

一、 基本概念

1、分区

分区使用限制：

- 单表分区层级最多 6 级；
- 单表分区数最多允许 60000 个分区；
- 一次查询最多查询分区数为 10000 个分区。

目前，MaxCompute 分区仅支持 string 类型，不承诺其他分区类型的正确性，不支持其他任意类型的隐式类型转换。

2、生命周期

- 分区表若指定生命周期，则根据各个分区的 LastDataModifiedTime 判断该分区是否该被回收。不同于非分区表，分区表的最后一个分区被回收后，该表不会被删除。
- 生命周期只能设定到表级别，不能再分区级设置生命周期。

3、资源

资源 (Resource) 是 MaxCompute 的特有概念 。

MaxCompute 资源的类型包括：

- File 类型；
- Table 类型：MaxCompute 中的表；
- Jar 类型：编译好的 Java Jar 包；
- Archive 类型：通过资源名称中的后缀识别压缩类型，支持的压缩文件类型包括：.zip/.tgz/.tar.gz/.tar/jar 。

4、任务

任务(Task)是 MaxCompute 的基本计算单元。SQL 及 MapReduce 功能都是通过任务(Task)完成的。

执行计划是由具有依赖关系的多个执行阶段(Stage)构成的。目前，执行计划逻辑上可以被看做一个有向图，图中的点是执行阶段，各个执行阶段的依赖关系是图的边。MaxCompute 会依照图(执行计划)中的依赖关系执行各个阶段。在同一个执行阶段内，会有多个进程，也称之为 Worker，共同完成该执行阶段的计算工作。同一个执行阶段的不同 Worker 只是处理的数据不同，执行逻辑完全相同。

备注:

- 在 MaxCompute 中，并不是所有的请求都会被转化为任务(Task)，例如：[项目空间\(Project\)](#)、[资源\(Resource\)](#)、[自定义函数\(UDF\)](#)及 [实例\(Instance\)](#)的操作就不需要通过 MaxCompute 的任务来完成。

二、 SQL

目前在 MaxCompute 中允许的最大 SQL 长度是 2MB。

1、join 操作

- 目前最多支持 16 个并发 Join 操作；
- 不支持笛卡尔积，即无 on 条件的链接。
- 在 mapjoin 中，最多支持 6 张小表的 mapjoin；

mapjoin

mapjoin 的基本原理是：在小数据量情况下，SQL 会将用户指定的小表全部加载到执行 join 操作的程序的内存中，从而加快 join 的执行速度。需要注意，使用 mapjoin 时：

left outer join 的左表必须是大表；
right outer join 的右表必须是大表；
inner join 左表或右表均可以作为大表；
full outer join 不能使用 mapjoin；
mapjoin 支持小表为子查询；
使用 mapjoin 时需要引用小表或是子查询时，需要引用别名；
在 mapjoin 中，可以使用不等值连接或者使用 or 连接多个条件；
目前 MaxCompute 在 mapjoin 中最多支持指定 8 张小表，否则报语法错误；
如果使用 mapjoin，则所有小表占用的内存总和不得超过 512MB。请注意由于 MaxCompute 是压缩存储，因此小表在被加载到内存后，数据大小会急剧膨胀。此处的 512MB 限制是加载到内存后的空间大小；
多个表 join 时，最左边的两个表不能同时是 mapjoin 的表。

2、其它

- ODPS SQL 目前最多支持 128 个并发 union 操作；
- 一般情况下，MaxCompute 最多允许 256 个表的 union all；
- 最多支持 128 个并发 insert overwrite/into 操作；
- 一般情况下，单个 SQL 里最多可以写 256 路输出；
- 在分布式环境下，单个进程最多只能输出 512 个动态分区
- 在现阶段，任意动态分区 SQL 不允许生成超过 2000 个动态分区；
- 动态生成的分区值不允许为 NULL；
- 当使用 Select 语句屏显时，目前最多只能显示 10000 行结果。
- 在 MaxCompute 中，共有四种 Task 类型：MapTask, ReduceTask, JoinTask 和 LocalWork；
- 输入表的数量不能超过 16 张。
- 一般情况下用户自建函数无法覆盖系统内建函数。只有项目空间的 Owner 才有权利覆盖内建函数

内建函数：

- (1) 所有日期函数，若某一个参数为 null 时，输出为 null；
- (2) to_date 某一参数为空时，报异常；sign 为空抛异常；space 为空抛异常；
- (3) 窗口函数：

窗口函数只能出现在 select 子句中，窗口函数中不要嵌套使用窗口函数和聚合函数，窗口函数不可以和同级别的聚合函数一起使用。目前在一个 MaxCompute SQL 语句中，可以使用至多 5 个窗口函数。现阶段，同一窗口内最多包含 1 亿行数据

3、视图

创建视图时，必须有对视图所引用表的读权限；
视图只能包含一个有效的 select 语句；
视图可以引用其它视图，但不能引用自己，也不能循环引用；
不允许向视图写入数据，例如使用 insert into 或者 insert overwrite 操作视图；
重建后视图本身的权限保持不变。

三、数据上传下载

1、数据库数据同步

- 离线批量的数据库数据同步，可以选择大数据开发套件；
- Oracle 数据库数据实时同步，可以选择 OGG 插件工具；
- RDS 数据库数据实时同步，可以选择 DTS 同步。

四、Udf

使用 UDTF 有如下使用限制：

同一个 SELECT 子句中不允许有其他表达式；
UDTF 不能嵌套使用；
不支持在同一个 select 子句中与 group by / distribute by / sort by 联用。

五、Mapreduce

不支持自定义比较器，支持自定义哈希函数，但优先级低于系统函数；

- 接受多表输入，且输入表的 Schema 可以不相同。在 map 函数中用户可以获取当前 Record 对应的 Table 信息；
- 输入可以为空，不支持视图(View)作为输入；
- Reduce 接受多路输出，可以输出到不同表，或者同一张表的不同分区。不同输出的 Schema 可以不同。不同输出间通过 label 进行区分，默认输出不必加 label，但目前不接受没有输出的情况。

六、用户及授权管理

1、用户管理

当一个用户被移除后，该用户不再拥有访问该项目空间资源的任何权限。

移除一个用户之前，如果该用户已被赋予某些角色，则需要先撤销该用户的所有角色。

当一个用户被移除后，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。MaxCompute 目前不支持在项目空间中彻底移除一个用户及其所有权限数据。

2、角色管理

每一个项目空间在创建时,会自动创建一个 admin 的角色,与项目空间 Owner 相比, admin 角色不能将 admin 权限指派给用户,不能设定项目空间的安全配置,不能修改项目空间的鉴权模型。Admin 角色所对应的权限不能被修改。

3、授权

所有的授权操作都必须由具有以下三种身份之一的用户来完成：

项目空间 Owner

项目空间中拥有 admin 角色的用户

项目空间中对象创建者

七、 1

DATA IDE

- 节点任务(node task): 一个数据执行的操作。可以与其他节点任务、工作流任务配置依赖关系, 组成 DAG 图。
- 工作流任务(flow task): 解决一个小业务的一组内部节点组成一个工作流任务。工作流任务数量建议小于 10 个。工作流任务内部节点, 无法被其他工作流任务、节点任务依赖。工作流任务可以与其他工作流任务、节点任务配置依赖关系, 组成 DAG 图。
- 内部节点(inner node): 工作流任务内部的节点, 与节点任务能力基本相同。其调度周期会继承工作流任务的调度周期, 无法进行单独配置, 依赖关系也按照拖拽关系。

实例 instance

在调度系统中的任务经过调度系统、手动触发运行后会生成一个实例。实例代表了某个任务在某时某刻执行的一个快照, 实例中会有任务的运行时间、运行状态、运行日志等信息。

例如: 设置每天 2:00 运行 task1 任务。调度系统会在每天 23: 30 会根据周期节点任务定义好的时间自动生成一个快照, 也就是 task1 的一个第二天 2:00 运行的实例, 第二天 2:00 到的时候, 同时判断上游任务已经完成, task1 实例就会如期启动运行。

只有开通大数据开发套件的阿里云账号(即主账号)可以创建项目, 其他子账号仅作为使用者在项目操作中

开发角色: 负责数据开发页面设计和维护工作流;

运维角色: 负责在运维中心页面管理全部任务的运行情况并做相应处理;

项目管理员角色: 除了拥有开发角色和运维角色全部权限外, 还可以进行添加 / 移出项目成

员并授予角色，创建自定义资源组等项目级别的操作；

部署角色：仅在多项目模式时审核任务代码并决定是否提交运维；

访客角色：仅有只读权限，可查看数据开发页面的工作流设计和代码内容。

*只有项目管理员角色才能够新建数据源，其他角色的成员仅能查看数据源。

一次性调度和周期调度——调度属性一旦选定，不可以再更改。

- 虚拟节点属于控制类型节点，在工作流运行过程中不对数据产生任何影响，仅用于实现对下游节点的运维控制。
- 虚节点在被其他节点依赖的情况下，如果被运维人员手动设置为运行失败，则下游未运行的节点将因此无法被触发运行，在运维过程中可以防止上游错误数据进一步蔓延。具体介绍请参见：[任务类型介绍](#)中的虚节点类型。综上所述，一般建议设计工作流时，默认创建一个虚节点作为根节点来控制整个工作流。具体操作如下：

同步任务保存后，点击右边的**提交**，将同步任务提交到调度系统中，调度系统会按照配置属性在从第二天开始自动定时执行。

不依赖上一调度周期

自依赖，等待上一调度周期结束，才能继续运行

等待下游任务的上一周期结束，才能继续运行

等待自定义任务的上一周期结束，才能继续运行

特别说明：如果是 23:30 以后提交的任务，则调度系统从第三天开始才会自动周期生成实例并定时运行。

大数据开发套件提供了测试运行、补数据和周期运行 3 种触发方式，详情如下：

- 测试运行：手动触发方式。如果您仅需确认单个任务的定时情况和运行，建议使用测试运行；
- 补数据运行：手动触发方式。如果您需要确认多个任务的定时情况和相互依赖关系，或者需要从某个根任务开始重新执行数据分析计算，可以考虑使用此方式；
- 周期运行：系统自动触发方式。提交成功的任务，调度系统在第二天 0 点起会自动生成当天不同时间点的运行实例，并在定时时间达到时检查各实例的上游实例是否运行成功，如

果定时时间已到并且上游实例全部运行成功，则当前实例会自动触发运行，无需人工干预。

说明：

手动触发和自动调度的调度系统根据周期生成实例的规则一致：无论周期选择天/小时/分钟/月/周，任务在每一个日期都会有对应实例生成，但仅在指定日期的对应实例会定时运行并生成运行日志；非指定日期的对应实例不会实际运行，而是在满足运行条件时将状态直接转换为成功，因此不会有运行日志生成。

- 补数据运行是手动触发方式，但生成的实例会与周期自动运行的实例存在上游依赖关系；若该任务有上游任务没有运行成功，定时时间到了不会触发运行。故补数据时建议从上游开始触发，或者当任务为“未运行”状态时，检查周期运行的下游任务是否已经运行成功不会单独运行。
- 根据前文所述的实例生成规则，配置为每周日凌晨 2 点运行的任务 `write_result`，补数据运行时选择的业务日期是周一(业务日期=运行日期-1)，则实例会在 2 点真正运行；如果不是周一，则实例在 2 点转换为成功状态并且没有日志生成。

什么是调度资源？

调度资源用于执行或分发调度系统下发的任务，大数据开发套件的调度资源分为以下两种模式：

1. 默认调度资源；
2. 自定义调度资源。

自定义调度资源指用户自购 ECS，配置成可以执行分发任务的调度服务器。组织管理员(主账号)可以新建自定义调度资源，调度资源包括若干台物理机或 ECS，用于执行数据同步任务、SHELL 任务、ODPS_SQL 任务、OPEN_MR 任务。

注意：

任务默认都是在默认调度资源组中执行/分发，用户可以通过 **运维中心 > 任务管理 > 修改资源组** 来修改任务的执行/分发的调度资源组，ODPS_SQL 任务只能在默认资源组上运行，如果需要在自定义的资源组上跑 SQL 任务，需要在项目管理下的调度资源修改默认资源组。

Mysql 切分键

MySQL Reader 进行数据抽取时，如果指定 `splitPk`，表示用户希望使用 `splitPk` 代表的字段进行数据分片，数据同步因此会启动并发任务进行数据同步，这样可以大大提供数据同步的效能。

- 推荐 `splitPk` 用户使用表主键，因为表主键通常情况下比较均匀，因此切分出来的分片也不容易出现数据热点；
- 目前 `splitPk` 仅支持整型数据切分，不支持字符串、浮点、日期等其他类型。如果用户指定其他非支持类型，忽略 `splitPk` 功能，使用单通道进行同步；

- 如果 **splitPk** 不填写，包括不提供 **splitPk** 或者 **splitPk** 值为空，数据同步视作使用单通道同步该表数据。

作业速率上限

作业速率上限是指数据同步作业可能达到的最高速率，其最终实际速率受网络环境、数据库配置等的影响。

作业并发数

从单同步作业来看：作业并发数*单并发的传输速率=作业传输总速率；

当作业速率上限已选定的情况下，应该如何选择作业并发数？

- ① 如果你的数据源是线上的业务库，建议您不要将并发数设置过大，以防对线上库造成影响；
- ② 如果您对数据同步速率特别在意，建议您选择最大作业速率上限和较大的作业并发数

错误记录数

表示脏数据的最大容忍条数，如果您配置 **0**，则表示严格不允许脏数据存在；如果不填则代表容忍脏数据，即如果出现脏数据，数据集成会记录并打印部分脏数据，方便用户排查。查看脏数据日志的方法：在运维中心-任务管理中，找到同步节点，右键点击查看节点运行日志。

在数据开发模块中除组织管理员无权限

调度配置：

- 任务名称：新建任务时填写的任务名称，不支持修改。
- 责任人：默认为当前登录人，也可以修改成本项目其他成员 (只有当前负责人以及管理员可以修改)。
- 类型：新建任务时选择的任务类型，不支持修改。
- 描述：新建任务时填写的描述，支持修改。

【说明】当创建一次性调度 workflow 任务时，workflow 不能配置调度属性；只有创建周期性调度 workflow 任务时，workflow 任务才能配置调度属性。

不支持每月月末调度，如果选择每月 **31** 日，则没有 **31** 日的那个月整个月都将不会调度。所以建议设置每月 **1** 日调度来计算上个月的数据

分钟调度：每天一个时间段里每隔 5*n 分钟调度一次，目前分钟调度只支持 5 分钟的倍数间隔。

- 不依赖上一调度周期：所有任务默认选择该选项，即不依赖任何任务的上周期实例。
- 自依赖，等待本任务上一调度周期结束，才能继续运行。使用场景：任务 A 当前周期数据来源依赖于任务 A 上周期执行的结果；或者小时/分钟调度任务 A 不允许实例并行。
- 等待下游任务的上一周期结束，才能继续运行：依赖第一层子任务的上周期。这种使用场景不多，选择此项，后续该任务一旦被其他任务直接依赖则实例都依赖所有第一层子任务的上周期实例。
- 等待自定义任务的上一周期结束，才能继续运行：使用场景：天任务 A 依赖一个数据是天任务 B 昨天产出的。

注意

1) 依赖属性配置的调度依赖是同周期依赖和跨周期依赖不冲突。任务 A 可以配置依赖属性依赖任务 B 也可以配置跨周期依赖任务 B，如此任务 A 既依赖任务 B 本周期也依赖任务 B 上周期。

2) 若任务 A 是小时/分钟调度，任务 B 是天调度，任务 B 配置依赖任务 A 的上周期，那么当天任务 B 的实例会依赖任务 A 昨天的所有实例。

3) 若任务 A 和 B 都是小时/任务调度，调度周期一样，任务 B 配置依赖任务 A 的上周期，则任务 B 每个实例都将依赖任务 A 昨天的所有实例和任务 A 与任务 B 同周期的前一个周期实例。

4) 如果天调度任务 A 配置跨周期依赖任务 B 的上周期，那么对任务 A 进行补数据的时候，补数据实例会去依赖任务 B 自动调度上周期实例，如果自动调度的上周期实例不存在则不依赖。

5) 长周期依赖短周期配置方式的详细内容请参见：[不同周期调度依赖实践](#)。

系统调度参数

https://help.aliyun.com/document_detail/30281.html?spm=5176.doc30280.6.585.mEi7EC

发布

【注意】工作流/资源/函数从开发项目发布到生产项目时，表、数据源配置等连接信息不随其一期发布至生产项目，需在生产项目中初始化。

数据管理模块是整个组织的数据管理，与组织项目的组织管理员/项目管理员两个角色有关联外，其他角色关联不大，是否能通过该模块创建修改表是看当前登录账号是否有对应 odps project 的权限。

通过数据管理模块创建的表可以进行业务类目划分。

运维中心仅对**开发、运维、项目管理员** 3 种角色开放。

ADS

数据的来源表在 MaxCompute 上，所以需要在 MaxCompute 上对云账号授予 **describe** 和 **select** 权限。

各个专有云授权的云账号名参照专有云的相关配置文档，一般为 **test1000000009@aliyun.com**

公共云上需要给 **garuda_build@aliyun.com** 以及 **garuda_data@aliyun.com** 账号授权。

- 实时更新的表刚刚建立后，会有一分钟左右的准备时间，这时候写入的数据需要在准备完成后才能查询，否则查询会报错。
- 在准备完成后，实时进行 **insert/delete** 数据变更，一般需要等待一分钟后才可查询。
- 分析型数据库进行实时插入和删除时，不支持事务，并且仅遵循最终一致性的设计，所以分析型数据库并不能作为 OLTP 系统使用。

在分析型数据库中，创建一个数据库的账号为这个数据库的 **owner** 账号，拥有最大化的权限。

数据库

- 在分析型数据库中，数据库是用户和系统管理员的管理职权的分界点。
 - 系统管理员：可管理的最小范围即数据库粒度的参数，但未经授权无法查看和管理数据库内部的结构和信息。
 - 用户：对于数据库级别的参数，默认只能查看而不能修改。
- 在分析型数据库中，一个数据库对应一个用于访问的域名和端口号，同时有且只有一个 **owner** 即数据库的创建者。
- 分析型数据库是以数据库为粒度对用户的宏观资源进行配置，因此创建数据库时用户需要输入业务预估的 **QPS**、数据量、**Query** 类型等信息 用于智能的判断初始的资源分配。
- 分析型数据库的用户不能直接通过 **CREATE DATABASE** 的 **DDL** 语句创建数据库，只能通过 **DMS** 控制台界面来创建需要的业务数据库。

表组

表组是一系列可发生关联的数据表的集合。分析型数据库采用关系模型存储数据，也就是使用二维表来进行数据的组织和存储。分析型数据库为了管理相关联的数据表，引入了表组的概念。

- 表组是数据库的下一级实体，也是表的上一级。
- 在分析型数据库中一个表必须从属于一个表组。

分析型数据库中表组分为两类：维度表组和普通表组。

- 维度表组，用于存放维度表，目前有且仅有一个，在数据库建立时会自动创建（表组名称：“数据库名_dimension_group”），用户不可修改和删除。（维度表特征上是一种数据量较小但是需要和任何表进行关联的表。）
- 普通表组，有如下特征：
 - 表组是数据物理分配的最小单元，数据的物理分布情况通常无需用户关心，但是数据的副本数必须在表组上进行设定，一个表组所有表副本数一致。注：副本数指数据在分析型数据库中同时存在的份数。
 - 同一个表组内的普通表才能进行快速的 **hash join**。（0.9 版本之前分析型数据库对普通表仅支持同表组内 **join**。0.9.5 版本后支持不同普通表组内的普通表进行 **join**）
 - 一个表组的表可以共享一些配置，例如查询超时时间，如果表组中的单表也进行了这些配置的个性化，那么进行表关联时会通表组级别的配置进行覆盖。
 - 一个表组中的所有表的一级 **Hash** 分区的分区数建议一致，但非强制性要求。

表

分析型数据库中表分为维度表和普通表，普通表也称实时表或分区表。

- 维度表可以和任意表组的任意表进行关联，并且创建时不需要配置分区信息，但是对单表数据量大小有所限制，并且需要消耗更多的存储资源。（一般要求维度表单表不超过 1000 万）
- 普通表创建时至少要指定一级分区列和相关分区信息，并且指定存放在一个表组中。

分区

在分析型数据库中，维度表无分区，普通表的分区目前最多为两级。

分区种类有 **Hash** 分区和 **List** 分区两种。目前分析型数据库支持将普通表的一级分区设置为 **Hash** 分区，二级分区设置为 **List** 分区，即可支持 **Hash join** 又可支持增量数据导入。

- **Hash** 分区是根据导入数据时已有的一列的内容进行散列后进行分区的，目前多张事实表进行快速的 **Hash Join** 时 **JoinKey** 必须有分区列参与，同时这些表的 **Hash** 分区数必须一致。仅采用 **Hash** 分区的数据表，在数据装载时，将进行全量覆盖历史数据。

注：0.9 版本之前，不支持事实表在没有 **Hash** 分区键参与关联条件的情况下进行关联，0.9.5 版本后，通过 **Full MPP Mode** 或小表广播进行计算时无此限制，关于 **join** 的详细内容请看《多计算引擎和 Hint》章节。

- **List** 分区是根据导入操作时所填写的分区列值来进行分区的，即一次导入的数据会进入同一个 **List** 分区中，因此 **List** 分区是支持增量的数据导入的。

注：无论采用何种分区形态，分析型数据库均不需要在用户查询时指定分区列，但是指定分区列或分区列的范围进行查询可能会提高查询性能。

根据表的更新方式不同，分析型数据库的表分为离线批量更新的表和实时更新的表：

- 离线批量更新的表，适合从离线系统如 **MaxCompute**（原 **ODPS**）产出的数据结果导入到分析型数据库供在线系统使用。
- 实时更新的表，可以直接 **insert/delete** 单条数据，适合业务系统直接写入数据。同时实时更新的表不提供二级分区功能（因为默认支持增量，无需二级分区）。

注意，分析型数据库不支持读写事务，并且数据实时更新时一分钟左右才可查询，另外在一致性方面分析型数据库遵循最终一致性。

ODPS

MaxCompute 资源的类型包括：

- **File** 类型；
- **Table** 类型：**MaxCompute** 中的表；
- **Jar** 类型：编译好的 **Java Jar** 包；
- **Archive** 类型：通过资源名称中的后缀识别压缩类型，支持的压缩文件类型包括：**.zip/.tgz/.tar.gz/.tar/jar** 。

当一个用户被移除后，该用户不再拥有访问该项目空间资源的任何权限。

移除一个用户之前，如果该用户已被赋予某些角色，则需要先撤销该用户的所有角色。关于角色的介绍请参考 [项目空间的角色管理](#)。

当一个用户被移除后，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。

MaxCompute 目前不支持在项目空间中彻底移除一个用户及其所有权限数据。

删除一个角色时，**MaxCompute** 会检查该角色内是否还存在其他用户。若存在，则删除该角色失败。只有在该角色的所有用户都被撤销时，删除角色才会成功。

沙箱：

MaxCompute MapReduce 及 **UDF** 程序在分布式环境中运行时受到 **Java** 沙箱的限制 (**MapReduce** 作业的主程序则不受此限制)，具体限制如下：

- 不允许直接访问本地文件，只能通过 **MaxCompute MapReduce/Graph** 提供的接口间接访问：**读取 - -resources 选项指定的资源，包括文件、Jar 包和资源表等**；通过 **System.out** 和 **System.err** 输出日志信息，可以通过 **MaxCompute** 客户端的 **Log** 命令查看日志信息；
- 不允许直接访问分布式文件系统，只能透过 **MaxCompute MapReduce/Graph** 访问到表的记录；
- 不允许 **JNI** 调用限制；
- 不允许创建 **Java** 线程，不允许起子进程执行 **Linux** 命令；
- 不允许访问网络，包括获取本地 **IP** 地址等等都会被禁止；
- **Java** 反射限制，“**suppressAccessChecks**”权限被禁止，无法 **setAccessible** 某个 **private** 的属性或方法，以达到读取 **private** 属性或调用 **private** 方法的目的。

具体的说，在用户代码中直接使用下面这些方法会抛“**access denied**”异常：

- 切分键：
只支持类型为整型的字段。读取数据时，根据配置的字段进行数据分片，实现并发读取，可提升数据同步效率。只有同步任务是 **RDS/Oracle** 数据导入至 **MaxCompute** 时，才显示切分键配置项。

ODPS mr

MapReduce 处理数据过程主要分成 2 个阶段：**Map** 阶段和 **Reduce** 阶段。首先执行 **Map** 阶段，再执行 **Reduce** 阶段。**Map** 和 **Reduce** 的处理逻辑由用户自定义实现，但要符合 **MapReduce** 框架的约定。

- 在正式执行 **Map** 前，需要将输入数据进行“分片”。所谓分片，就是将输入数据切分为大小相等的数据块，每一块作为单个 **Map Worker** 的输入被处理，以便于多个 **Map Worker** 同时工作。

- 分片完毕后，多个 Map Worker 就可以同时工作了。每个 Map Worker 在读入各自的数据后，进行计算处理，最终输出给 Reduce。Map Worker 在输出数据时，需要为每一条输出数据指定一个 Key。这个 Key 值决定了这条数据将会被发送给哪一个 Reduce Worker。Key 值和 Reduce Worker 是多对一的关系，具有相同 Key 的数据会被发送给同一个 Reduce Worker，单个 Reduce Worker 有可能会接收到多个 Key 值的数据。
- 在进入 Reduce 阶段之前，MapReduce 框架会对数据按照 Key 值排序，使得具有相同 Key 的数据彼此相邻。如果用户指定了“合并操作”(Combiner)，框架会调用 Combiner，将具有相同 Key 的数据进行聚合。Combiner 的逻辑可以由用户自定义实现。与经典的 MapReduce 框架协议不同，在 MaxCompute 中，Combiner 的输入、输出的参数必须与 Reduce 保持一致。这部分的处理通常也叫做“洗牌”(Shuffle)。
- 接下来进入 Reduce 阶段。相同的 Key 的数据会到达同一个 Reduce Worker。同一个 Reduce Worker 会接收来自多个 Map Worker 的数据。每个 Reduce Worker 会对 Key 相同的多个数据进行 Reduce 操作。最后，一个 Key 的多条数据经过 Reduce 的作用后，将变成了一个值。

ACL 授权：

所有的授权操作都必须由具有以下三种身份之一的用户来完成：

- 项目空间 Owner
- 项目空间中拥有 admin 角色的用户
- 项目空间中对象创建者

上述权限描述中 Project 类型对象的 CreateTable 操作，Table 类型的 Select、Alter、Update、Drop 操作需要与 Project 对象的 CreateInstance 操作权限配合使用。单独使用上述几种权限而没有指派 CreateInstance 权限是无法完成对应操作的。这与 MaxCompute 的内部实现相关。同样，Table 的 Select 权限也要与 CreateInstance 权限配合使用。

使用时请注意：在添加用户或创建角色之后，需要对用户或角色进行授权。MaxCompute 授权是一种基于对象的授权。通过授权的权限数据(即访问控制列表, Access Control List)被看做是该对象的一种子资源。**只有当对象已经存在时，才能进行授权操作；当对象被删除时，通过授权的权限数据会被自动删除。**MaxCompute 授权支持类似于 SQL92 定义的 GRANT/REVOKE 语法，它通过简单的授权语句来完成对已存在的项目空间对象的授权或撤销授权。

当一个用户被移除后，该用户不再拥有访问该项目空间资源的任何权限。

移除一个用户之前，如果该用户已被赋予某些角色，则需要先撤销该用户的所有角色。关于角色的介绍请参考 项目空间的角色管理。

当一个用户被移除后，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。

MaxCompute 目前不支持在项目空间中彻底移除一个用户及其所有权限数据。

MaxCompute 只允许主账号将自身的 RAM 子账号加入到项目空间中，不允许加入其它云账号的 RAM 子账号，因此在“add user”时，无需再 RAM 子账号前指定主账号名称，

MaxCompute 默认判定命令的执行者即是子账号对应的主账号。

删除一个角色时，MaxCompute 会检查该角色内是否还存在其他用户。若存在，则删除该角色失败。只有在该角色的所有用户都被撤销时，删除角色才会成功。

大数据开发套件中成员角色类型对应的 MaxCompute 角色，以及各角色的平台权限详情

上述权限描述中 Project 类型对象的 CreateTable 操作, Table 类型的 Select、Alter、Update、Drop 操作需要与 Project 对象的 CreateInstance 操作权限配合使用。单独使用上述几种权限而没有指派 CreateInstance 权限是无法完成对应操作的。这与 MaxCompute 的内部实现相关。同样，Table 的 Select 权限也要与 CreateInstance 权限配合使用。

使用时请注意：在添加用户或创建角色之后，需要对用户或角色进行授权。MaxCompute 授权是一种基于对象的授权。通过授权的权限数据(即访问控制列表, Access Control List)被看做是该对象的一种子资源。只有当对象已经存在时，才能进行授权操作；当对象被删除时，通过授权的权限数据会被自动删除。MaxCompute 授权支持类似于 SQL92 定义的 GRANT/REVOKE 语法，它通过简单的授权语句来完成对已存在的项目空间对象的授权或撤销授权。

Tunnel

主要参数说明：

- **-bs, --block-size**: 每次上传至 Tunnel 的数据块大小，默认值：100MiB (MiB=1024*1024B)。
- **-c, --charset**: 指定本地数据文件编码，默认为'UTF-8'；不设定，默认下载源数据。
- **-cp, --compress**: 指定是否在本地上压缩后再上传，减少网络流量，默认开启。
- **-dft**: DateTime 类型数据格式，默认为'yyyy-MM-dd HH:mm:ss'；
- **-dbr**: 是否忽略脏数据(多列，少列，列数据类型不匹配等情况)。值为'true'时，将全部不符合表定义的数据忽略。值为'false'，若遇到脏数据，则给出错误提示信息，目标表内的原始数据不会被污染。
- **-e**: 指定 odps 的 endpoint。
- **-fd**: 本地数据文件的列分割符，默认为逗号','；
- **-h**: 数据文件是否包括表头，如果为 true，则 dship 会跳过表头从第二行开始上传数据。
- **-mbr, --max-bad-records**: 默认情况下，当上传的脏数据超过 1000 条时，上传动作终止。通过此参数，可以调整可容忍的脏数据量；
- **-ni**: NULL 数据标志符，默认为" "(空字符串)；
- **-rd**: 本地数据文件的行分割符，默认为'\r\n'；
- **-s**: 是否扫描本地数据文件，默认值为'false'。值为'true'时，先扫描数据，若数据格式正确，再导入数据。值为'false'，不扫描数据，直接进行数据导入。值为'only'时，仅进行扫描本地数据，扫描结束后不继续导入数据。
- **-te**: 指定 tunnel 的 Endpoint。
- **-tz**: 指定时区。默认为本地时区：Asia/Shanghai。

DataIDE

● 概念与杂项

组织是 Data IDE 的特有概念，组织的**基本对象为项目空间**；组织普通成员无权创建项目空间。

切分键：只支持类型为整型的字段，只有同步任务是 **RDS/Oracle 数据导入至 MaxCompute** 时，才显示切分键配置项。**推荐使用表的主键作为切分键**；读取数据时，根据配置的字段进行数据分片，实现**并发读取**，可**提升数据同步效率**。

DataIDE 支持**全量**与**增量**数据同步，利用 where 条件可进行全量同步和增量同步。

Data IDE 中，调度任务每次运行前都先将任务实例化，即生成一个实例，调度运行时实际上在执行相应的实例。在调度的不同阶段，实例会处于不同的状态，包括**未运行、运行中、运行成功、运行失败、等待时间、等待资源**等。当实例处于“未运行”状态时不能手工停止。

● 调试参数

➤ `${bdp.system.bizdate}`

`${bdp.system.bizdate}` 格式为 **yyyymmdd**，日常调度实例定时时间的**前一天**（年月日）。节点每天自动调度实例定时时间年月日减 1 天。

正常调度时，比如：任务定时在 14 号凌晨执行，那么 `${bdp.system.bizdate}` 替换的结果是 13 号。

测试，补数据时，比如：业务日期选择为 12 号，那么 `${bdp.system.bizdate}` 替换的结果是 12 号。

➤ `${bdp.system.cyctime}`

`${bdp.system.cyctime}` 格式为 **yyyymmddhh24miss**，日常调度**实例定时时间**（年月日时分秒）。yyyy 表示 4 位数年份，mm 表示 2 位数月份，dd 表示 2 位数天，hh24 表示 24 小时制的时，mi 表示 2 位数分钟，ss 表示 2 位数秒。

正常调度时，比如：任务定时在 14 号凌晨 1 点整执行，那么 `${bdp.system.cyctime}` 替换的结果是 当天 **yyyymm140100**。

测试，补数据时，比如：业务日期选择为 12 号（任务调度时间配置为每天凌晨 1 点整执行），那么 `${bdp.system.cyctime}` 替换的结果是 所选业务日期的 **yyyymm130100**。

`${bdp.system.cyctime}=${yyyymmddhh24miss}` 可以任意分解组合，以及指定格式，如定时时间年月日 `${yyyymmdd}`、`${yyyy-mm-dd}` 等格式可自定义，定时时间时分秒 `${hh24miss}`、`${hh24:mi:ss}` 等格式可自定义。

【注意】：所有参数都不允许包含空格！以 `cyctime` 为基准参与运行。**补数据**的时候，选中什么业务日期，参数替换结果是**业务日期+1**天，如果补数据选 20140510 这个业务日期，执行时 `cyctime` 替换结果是 20140511。

获取+/-周期的方法：

后 N 年： <code>\${add_months(yyyymmdd, 12*N)}</code>	前 N 年： <code>\${add_months(yyyymmdd, -12*N)}</code>
后 N 月： <code>\${add_months(yyyymmdd, N)}</code>	前 N 月： <code>\${add_months(yyyymmdd, -N)}</code>
后 N 周： <code>\${yyyymmdd+7*N}</code>	前 N 周： <code>\${yyyymmdd-7*N}</code>

后 N 天: \${yyyyymmdd+N}前 N 天: \${yyyyymmdd-N}

后 N 小时: \${hh24miss+N/24}前 N 小时: \${hh24miss-N/24}

后 N 分钟: \${hh24miss+N/24/60}前 N 分钟: \${hh24miss-N/24/60}

➤ **自定义参数：**
在大数据开发套件中以**\${变量名}**方式来声明变量，在参数中以**变量名=调度参数或常量**的方式来赋值。

● **角色说明**

Data IDE 角色	平台权限特征	MaxCompute 角色
组织管理员	指组织的管理者，可新建计算引擎、新建项目空间、新建调度资源、添加组织成员、为组织成员赋予组织管理员角色、配置数据类目等。	无
项目管理员	指项目空间的管理者，可针对项目空间基本属性、数据源、当前项目空间计算引擎配置和项目成员进行管理，并为组织成员赋予项目管理员、开发、运维、部署、访客角色。	role_project_admin
开发	开发角色用户能够创建工作流、脚本文件、资源和UDF，新建/删除表，同时可以创建发布包，但不能执行发布操作。	role_project_dev
运维	项目空间的运维人员，由项目管理员/项目所有者分配运维权限；拥有发布及线上运维的操作权限，无数据开发的操作权限。	role_project_pe
部署	部署角色与运维角色相似，但是其没有线上运维操作权限。	role_project_deploy
访客	访客角色的用户只具备查看权限，而无权限进行编辑工作流和代码等。	role_project_guest

ODPS

● **工具说明**

dship: 最初 ODPS 版本提供的数据上传及下载工具（基于 tunnel 实现）。

tunnel: 命令，用来上传，下载 odps 数据。最新 ODPS 版本鼓励用户直接使用 tunnel 命令实现原有 dship 工具的功能。

dpc: 基于 ODPS 的 DW/BI 工具解决方案平台。

dpc 的后台数据同步基于 cdp 完成，而 cdp 是基于 datax 实现的数据互通。

● 项目空间

项目空间 (Project) 是 MaxCompute 的基本组织单元，它类似于传统数据库的 Database 或 Schema 的概念，是进行多用户隔离和访问控制的主要边界。一个用户可以同时拥有多个项目空间的权限。通过安全授权，可以在一个项目空间中访问另一个项目空间中的对象，例如：[表 \(Table\)](#)，[资源 \(Resource\)](#)，[函数 \(Function\)](#)，[实例 \(Instance\)](#)

● 表

MaxCompute 的数据**存储单元**。MaxCompute 的表格分两种类型：**外部表及内部表**。

目前 MaxCompute 仅支持读取外部表数据，即**读取 OSS 数据**，不支持向外部表写入数据；

ODPS 的表不能修改字段属性，如要更改删除原表重新创建。

数据类型只能是：bigint, double, boolean, datetime, decimal 及 string。

目前 MaxCompute 单表支持的**分区数量上限为 6 万**。

表中建的**分区层次不能超过 6 级**。

表名列名的长度不超过 128 字节。

● 函数 (UDF)

用户可以使用 MaxCompute 提供的 Java 编程接口开发自定义函数 (User Defined Function，以下简称 UDF)。自定义函数 (UDF) 又可以进一步分为标量值函数 (UDF)，自定义聚合函数 (UDAF) 和自定义表值函数 (UDTF) 三种。

UDF (user defined function) (可以用在查询的 where 子句中)	(一个输入一个输出)
UDTF (user defined table function)	(一个输入多个输出)
UDAF (user defined aggregation function)	(多个输入一个输出)

- UDF 目前**只支持 Java 语言接口**，用户如果想编写 UDF 程序，可以通过 **添加资源 的方式**将 UDF 代码上传到项目空间中，**使用 注册函数 语句**创建 UDF。
- 如果用户需要使用 UDF 功能，**需要在工单系统上提交申请**，提供 odps project 名称，简单描述使用场景。只有**申请通过，开通好权限后才可以创建 UDF**。

ODPS 的客户端工具 odpscmd 中可以使用 **list functions** 查看当前 ODPS 项目空间中所有的自定义函数。

```
List functions;           --查看当前项目空间中的所有的自定义函数
List functions -p my_project; --查看指定项目空间 my_project 下的所有自定义函数
```

```
Create function           --创建函数
Drop function             --注销函数
```

与资源文件一样，**同名函数只能注册一次**。

一般情况下用户**自建函数无法覆盖系统内建函数**。只有项目空间的 **Owner 才有权利覆盖内建函数**。

如果用户使用了覆盖内建函数的自定义函数，在 SQL 执行结束后，会在 Summary 中打印出 **warning 信息**。

● 分区

分区表指的是在创建表时指定分区空间，即指定表内的某几个字段作为分区间。在大多数情况下，用户可以将分区类比为文件系统下的目录。MaxCompute 将分区列的每个值作为一个分区(目录)。用户可以指定多级分区，即将表的多个字段作为表的分区，分区之间正如多级目录的关系。在使用数据时如果指定了需要访问的分区名称，则只会读取相应的分区，**避免全表扫描，提高处理效率，降低费用。**

```
1. create table src (key string, value bigint) partitioned by (pt string); -- 目前，
   MaxCompute 仅承诺 String 类型分区
2. select * from src where pt='20151201'; -- 正确使用方式。MaxCompute 在生成查询计
   划时只会将'20151201'分区的数据纳入输入中
3. select * from src where pt = 20151201; -- 错误的使用方式。在这样的使用方式下，
   MaxCompute 并不能保障分区过滤机制的有效性。pt 是 String 类型，当 String 类型与
   Bigint(20151201)比较时，MaxCompute 会将二者转换为 Double 类型，此时有可能会精度损
   失。
```

● 任务

任务(Task)是 MaxCompute 的基本计算单元。SQL 及 MapReduce 功能都是通过任务(Task)完成的。在 MaxCompute 中，并不是所有的请求都会被转化为任务(Task)，例如：项目空间(Project)、资源(Resource)、自定义函数(UDF)及实例(Instance)的操作就不需要通过 MaxCompute 的任务来完成。

● 任务实例

在 MaxCompute 中，部分 任务(Task) 在执行时会被实例化，以 MaxCompute 实例(下文简称实例或 Instance)的形式存在。实例会经历**运行(Running)**及**结束(Terminated)**两个阶段。运行阶段的状态为 Running(运行中)，而结束阶段的状态将会是 Success(成功)，Failed(失败)或 Canceled(被取消)。用户可以根据运行任务时 MaxCompute 给出的实例 ID 查询、改变任务的状态，例如：

```
1. status <instance_id>; --查看某实例的状态
2. kill <instance_id>; --停止某实例，将其状态设置为 Canceled
```

● 视图

- **创建视图**时，必须对视图所引用表的**读权限**。(查询时不需要对引用表有读权限，有视图读权限就行)
- 视图只能包含一个有效的 select 语句。
- 视图可以引用其它视图，但**不能引用自己，也不能循环引用**。
- 不可以向视图写入数据，例如使用 insert into 或者 insert overwrite 操作视图。
- 当视图建好以后，如果视图的引用表发生了变更，有可能导致视图无法访问，

● 资源

MaxCompute 资源的类型包括：

- **File 类型**；
- **Table 类型**：MaxCompute 中的表；
- **Jar 类型**：编译好的 Java Jar 包；

- **Archive 类型**：通过资源名称中的后缀识别压缩类型，支持的压缩文件类型包括：.zip/.tgz/.tar.gz/.tar/jar；

● 连接

Odps

开头：公网 service.odps 内网或 VPC：odps-ext

结尾：公网 .aliyun.com 内网或 VPC：.aliyun-inc.com

Tunnel

开头：公网 dt.odps 内网或 VPC：dt-ext.odps dt-ext.xxx.odps.

结尾：公网 .aliyun.com 内网或 VPC：.aliyun-inc.com

● 多路输出

- 一般情况下，单个 SQL 里最多可以写 **128 路输出**，超过 128 路报语法错误。
- 在一个 multi insert 中，对于分区表，同一个**目标分区不可以出现多次**；对于未分区表，该**表不能出现多次**。（即不能有交叉相同目标）
- 对于同一张分区表的不同分区，不能同时有 insert overwrite 和 insert into 操作，否则报错返回。

● 鉴权模型

Security.CheckPermissionUsingACL(激活/交结 ACL 授权机制，默认：true)

Security.CheckPermissionUsingPolicy(激活/冻结 Policy 授权机制，默认：true)

Security.ObjectCreatorHasAccessPermission(允许/禁止对象创建者默认拥有访问权限，默认：true)

Security.ObjectCreatorHasGrantPermission(允许/禁止对象创建者默认拥有授权权限，默认：true)

Security.LabelSecurity(开启/关闭 LabelSecurity 安全策略，默认为 false)

ProjectProtection(开启/关闭项目空间的数据保护机制，默认为 false)

查看鉴权模型：Show SecurityConfiguration

两种访问策略：强制访问控制策略（MAC）、自主访问控制策略（DAC）

用户安全许可标签和文件敏感等级取值：**0 级最低，默认 0 级**。两者对应

显式设置的**列的敏感等级**优先级**高于表的敏感等级**，**和顺序、等级高低无关**

设置包时，若省略[using label<number>],则**默认为 0 级**，即只可以访问非敏感数据。（package）跨项目访问敏感数据时，包安装者的项目空间中的所有用户都将使用此许可访问级别。

LabelSecurity 对敏感数据的粒度可以支持**列级别**，管理员可以对表的任何列设置敏感标记（Label），一张表可以由**不同敏感等级**的数据列构成。

打开 LabelSecurity 安全机制开关：（Owner 设置,admin 角色没有此权限）

Set security.LabelSecurity=<true|false>

设置用户安全许可标签：

Set label <number> to user <username>

设置数据敏感等级标签：

Set label <number> to table tablename[(column_list)]

显示授权低级别用户访问高敏数据：

Grant label <number> on table <tablename>[(column_list)] to user <username> [sith exp <days>]

撤销显示授权：

Revoke label on table <tablename>[(columns_list)] from user <username>

清理过期的显示授权：

Clear expired grants;

查看一个用户能访问哪些敏感数据集：

Show label[<level>] grants [for user <username>]

查看一个敏感数据表能被哪些用户访问：

Show label[<level>] grants on table <tablename>

用户对指定表上列级别的 Label 授权：

Show label[<level>] grants on talbe <tablename> for user <username>

包安装者对包中敏感资源许可访问级别

Allow project <prjname>to install package <paname> [using label<n>]

● Package

被安装的 package 是独立的 ODPS 对象类型,若要访问 package 里的资源(即其他项目空间分享的资源),必须拥有对此 package 的 Read 权限。如果请求者没有 Read 权限,则需要向 ProjectOwner 或 Admin 申请。ProjectOwner 或 Admin 可以通过 ACL 授权或 Policy 授权机制来完成。

创建: create package <package_name>

删除: drop package <package_name>

添加要分享的资源到 package:

Add <project_object> to package <package_name>[with privileges privileges]

从 package 中删除分享对象:

Remove <project_object> from package <package_name>

允许其他项目空间使用此 package

Allow project <project_name> to install package <package_name> [using label<n>]

撤销项目空间对此 package 的使用许可:

Disallow project <project_name> to install package <package_name>

卸载 package:

Uninstall package <project_name.package_name>

查看已安装的 package:

Show packages;

查看 package 的详细信息:

Describe package <package_name>

● 数据流出（ODOS 规定：资源分享优先于数据保护）

- SQL: create table ownprj.t1 as select * from secretprj.t1
- MR: 能过 MR 将表读出后写到其他的项目空间去
- 导出: 通过数据导出工具, 将数据导出
- PAI: 将数据间接导出

通过以下对数据保护的项目也可以进行数据流出

- Trustedproject

- Exception policy
- Package 数据分享

● 权限

当一个用户被移除后，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。

ODPS 目前**不支持**在项目空间中**彻底移除一个用户及其所有权限数据**。

每一个项目空间在创建时，会**自动创建一个 admin** 的角色，并且为该角色授予了确定的权限：能访问项目空间内的所有对象，能进行用户与角色管理，能对用户或角色进行授权。与项目空间 Owner 相比，admin 角色**不能**将 admin 权限指派给用户，**不能设定**项目空间的安全配置，**不能修改**项目空间的鉴权模型。**Admin 角色所对应的权限不能被修改**。

ODPS 的 ACL 授权操作一般涉及三个要素：**主体（subject）、客体（object）、操作（action）**

Package 授权（项目空间之间的授权）

目前支持的对象类型不包括 Project 类型,也就是不允许通过 Package 在其他 Project 中创对象。**添加到 Package 中的不仅仅是对象本身,还包括相应的操作权限。**

当没有通过[with privileges privileges]来指定操作权限时,**默认为只读权限**,即 Read/Describe/Select。

“对象及其权限”被看作一个整体,添加后**不可被更新**。若有需要,只能**删除和重新添加**。

● Tunnel 命令支持的操作

upload: 帮助用户上传数据到 MaxCompute 的表中;

download: 帮助用户从 MaxCompute 的表中下载数据;

resume: 如果上传数据失败,通过 resume 命令进行断点续传,目前仅支持上传数据的续传。每次上传、下载数据被称为一个 session。在 resume 命令后指定 session id 完成续传。

show: 查看历史运行信息;

purge: 清理 session 目录;

help: 输出 tunnel 帮助信息;

参数说明

-bs,--block-size:每次上传至 Tunnel 的数据块大小,默认值: 100MiB (MiB=1024*1024B)。

-c,--charset: 指定本地数据文件编码,默认为'UTF-8';不设定,默认下载源数据。

-cp,--compress: 指定是否在本地压缩后再上传,减少网络流量,默认开启。

-dfp: DateTime 类型数据格式,默认为'yyyy-MM-dd HH:mm:ss';

-dbr: 是否忽略脏数据(多列,少列,列数据类型不匹配等情况)。值为'true'时,将全部不符合表定义的数据忽略。值为'false',若遇到脏数据,则给出错误提示信息,目标表内的原始数据不会被污染。

-e: 指定 odps 的 endpoint。

-fd: 本地数据文件的列分割符,默认为逗号',';

-h: 数据文件是否包括表头,如果为 true,则 dship 会跳过表头从第二行开始上传数据。

-mbr,--max-bad-records: 默认情况下,当上传的脏数据超过 1000 条时,上传动作终止。通过此参数,可以调整可容忍的脏数据量;

-ni: NULL 数据标志符, 默认为" "(空字符串);

-rd: 本地数据文件的行分割符, 默认为' \r\n';

-s: 是否扫描本地数据文件, 默认值为' false'。值为' true' 时, 先扫描数据, 若数据格式正确, 再导入数据。值为' false', 不扫描数据, 直接进行数据导入。值为' only' 时, 仅进行扫描本地数据, 扫描结束后不继续导入数据。

-te:指定 tunnel 的 Endpoint.

-tz:指定时区。默认为本地时区: Asia/Shanghai。

● MR

由于 MaxCompute 的所有数据都被存放在表中, 因此 MaxCompute MapReduce 的**输入、输出只能是表, 不允许用户自定义输出格式, 不提供类似文件系统的接口。**

MR 模型中 **Map 阶段**是必须要有的。

MR 任务中如何在程序中设置 mapper 和 reducer 的个数?

mapper 数是根据 **split size** 来确定的, 默认是 **256M**, 可以使用 **setSplitSize** 调节;

reducer 数可以通过 **setNumReduceTasks** 设置, 默认为 mapper 数的 1/4。具体可以参考相关 SDK 文档。

(在 JobConfig 中的方法中通过 SetNumReduceTasks 设置)

ODPS 提供了扩展的 MapReduce 模型, 即可以支持 Map 后连接任意多个 Reduce 操作

在开发大数据计算服务 (MaxCompute, 原 ODPS) 的 MapReduce 时, 可以使用 JobClient 类中的接口提交 MapReduce 作业。目前支持**非阻塞方式**与**阻塞方式**提交方式。

MaxCompute MapReduce(MR)及 UDF 程序在分布式环境中运行时受到 Java 沙箱的限制(**MapReduce 作业的主程序则不受此限制**), 具体限制如下:

- 不允许直接访问本地文件, 只能通过 ODPS MapReduce/Graph 提供的接口间接访问: 读取 --resources 选项指定的资源, 包括文件、Jar 包和资源表等; 通过 System.out 和 System.err 输出日志信息, 可以通过 ODPS 客户端的 Log 命令查看日志信息;
- 不允许直接访问分布式文件系统, 只能透过 ODPS MapReduce/Graph 访问到表的记录;
- 不允许 JNI 调用限制;
- 不允许创建 Java 线程, 不允许起子进程执行 Linux 命令;
- 不允许访问网络, 包括获取本地 IP 地址等等都会被禁止;
- Java 反射限制, "suppressAccessChecks"权限被禁止, 无法 setAccessible 某个 private 的属性或方法, 以达到读取 private 属性或调用 private 方法的目的。

MaxCompute 客户端提供一个 jar 命令用于运行 MapReduce 作业, 具体语法:

- -conf <configuration file> : 指定 JobConf 配置文件;
- -classpath <local_file_list>: **本地执行时的 classpath**, 主要用于指定 main 函数所在的 jar 包的本地路径 (包含相对路径和绝对路径)。大多数情况下, 用户更习惯于将 main 函数与 Map/Reduce 函数编写在一个包中, 例如: WordCount 代码示例, 因此, 在执行示例程序时, -resources 及 -classpath 的参数中都出现了 mapreduce-examples.jar, 但二者意义不同, -resources 引用的是 Map/Reduce 函数, 运行于分布式环境中, 而 -classpath 引用的是 main 函数, 运行于本地, 指定的 jar 包路径也是本地文件路径。包名之间使用系统默认的文件分割符作分割(通常情况下, windows 系统是分号 ";" ,linux 系统是逗号 "," ,如果用户是在云端使用 mr 任务, jar 包之间的分隔是逗号 "," ,");

- `-D <prop_name>=<prop_value>`: 本地执行时, `<mainClass>` 的 java 属性, 可以定义多个;
- `-l`: 以本地模式执行 MapReduce 作业, 主要用于程序调试;
- `-resources <resource_name_list>`: MapReduce 作业运行时使用的资源声明。一般情况下, `resource_name_list` 中需要指定 Map/Reduce 函数所在的资源名称。请特别注意, 如果用户在 Map/Reduce 函数中读取了其他 ODPS 资源, 那么, 这些资源名称也需要被添加到 `resource_name_list` 中。资源之间使用逗号分隔, 使用跨项目空间使用资源时, 需要前面加上: `PROJECT/resources/`, 示例: `-resources otherproject/resources/resfile`。有关于如何在 Map/Reduce 函数中读取资源的示例请查看资源使用示例; 用户可以通过 `-conf` 选项指定 JobConf 配置文件, 该文件可以影响 SDK 中 JobConf 的设置。下面将给出一个 JobConf 配置文件的示例:

归并(Combiner)函数将 Shuffle 阶段**相邻的 Record**进行归并。用户可以根据不同的业务逻辑**选择是否使用归并函数**。归并函数是 MapReduce 计算框架的**一种优化**, 通常情况下 Combiner 的逻辑与 reduce 相同。当 map 输出数据后, 框架会在 **map 端**对相同 key 值的数据进行**本地的归并操作**。

应用场景

- **搜索**: 网页爬取、倒排索引、PageRank。
- **Web 访问日志分析**: 分析和挖掘用户在 web 上的访问、购物行为特征, 实现个性化推荐; 分析用户访问行为。
- **文本统计分析**: 比如莫言小说的 WordCount、词频 TFIDF 分析; 学术论文、专利文献的引用分析和统计; 维基百科数据分析等。
- **海量数据挖掘**: 非结构化数据、时空数据、图像数据的挖掘。
- **机器学习**: 监督学习、无监督学习、分类算法如决策树、SVM 等。
- **自然语言处理**: 基于大数据的训练和预测; 基于语料库构建单词同现矩阵, 频繁项集数据挖掘、重复文档检测等。
- **广告推荐**: 用户点击 (CTR)和购买行为 (CVR) 预测。

● 图计算

MaxCompute Graph 是一套面向迭代的图计算处理框架。图计算作业使用图进行建模, 图由点(Vertex)和边(Edge)组成, 点和边包含权值(Value), MaxCompute Graph 支持下述图编辑操作:

修改点或边的权值; 增加/删除点; 增加/删除边: 编辑点和边时, **点与边的关系需要用户维护**。

通过迭代对图进行编辑、演化, 最终求解出结果, 典型应用: **PageRank, 单源最短距离算法, K-均值聚类算法** 等等。用户可以使用 MaxCompute GRAPH 提供的接口 Java SDK 编写图计算程序。

图加载: 框架调用用户自定义的 GraphLoader 将输入表的记录解析为点或边; 分布式化: 框架调用用户自定义的 Partitioner 对**点进行分片**(默认分片逻辑: 点 ID 哈希值然后对 Worker 数取模), 分配到相应的 Worker;

● 执行计划

执行计划是由具有依赖关系的多个**执行阶段(Stage)**构成的。目前, 执行计划逻辑上可以被看做一个有向图, 图中的**点是执行阶段**, 各个**执行阶段的依赖关系是图的边**。

ODPS 会依照图(执行计划)中的依赖关系执行各个阶段。在**同一个执行阶段内, 会有多个进程, 也称之为 Worker**, 共同完成该执行阶段的计算工作。同一个执行阶段的不同 Worker 只是处理的数据不同, 执行逻辑

辑完全相同。

计算型任务在执行时,会被实例化,用户可以操作这个 [实例(Instance)] 的信息,例如:获取实例状态(Status Instance), 终止实例运行(Kill Instance)等。

- **ODPS 限制**
 - ODPS 表 string 列内容长度不允许超过 2MB。
 - 单个任务引用的资源数量不超过 512 个，分区表按照一个单位计算。
 - 单个任务引用的资源总计字节数大小不超过 64MB。
 - 单个任务的输入路数不能超过 128，单个任务的输出路数不能超过 128 路。
 - 单个任务中自定义 Counter 的数量不能**超过 64**。
 - 单个 Map 或 Reduce Worker 占用 memory 默认为 2048MB，范围[256MB, 12GB]。
 - 单个 Map 或 Reduce Worker 重复读一个资源次数限制 <=64 次。
 - 本地运行模式下，Map Worker 个数不能超过 100；Reduce Worker 个数不能超过 100；默认一路输入下载记录数 100。

- **数据类型**

类型	描述
STRING	字符串类型，长度不能超过 8MB。
BOOLEAN	上传值只支持” true”，” false”，” 0”，” 1”。下载值为 true/false 且不区分大小写。
BIGINT	取值范围[-9223372036854775807， 9223372036854775807]。
DOUBLE	1.有效位数 16 位
	2.上传支持科学计数法表示
	3.下载仅使用数字表示
	4.最大值： 1.7976931348623157E308
	5.最小值： 4.9E-324
	6.无穷大: Infinity
	7.无穷小: -Infinity
DATETIME	Datetime 类型默认支持时区为 GMT+8 的数据上传，可以通过命令行指定用户数据日期格式的 format pattern。

- **授权**

客体(Object)	操作(Action)	说明
Project	Read	查看项目空间自身(不包括项目空间的任何对象)的信息，如 CreateTime 等

Project	Write	更新项目空间自身(不包括项目空间的任何对象)的信息，如 Comments
Project	List	查看项目空间所有类型的对象列表
Project	CreateTable	在项目空间中创建 Table
Project	CreateInstance	在项目空间中创建 Instance
Project	CreateFunction	在项目空间中创建 Function
Project	CreateResource	在项目空间中创建 Resource
Project	CreateJob	在项目空间中创建 Job
Project	CreateVolume	在项目空间中创建 Volume
Project	CreateOfflineModel	在项目空间中创建 OfflineModel
Project	CreateXflow	在项目空间中创建 Xflow
Project	All	具备上述所有权限
Table	Describe	读取 Table 的元信息
Table	Select	读取 Table 的数据
Table	Alter	修改 Table 的元信息
Table	Update	覆盖或添加 Table 的数据
Table	Drop	删除 Table
Table	All	具备上述所有权限
Function	Read	读取
Function	Write	更新
Function	Execute	执行
Function	Delete	删除
Function	All	具备上述所有权限
Resource,Instance, Job, Volume	Read	读取
Resource,Instance, Job, Volume	Write	更新
Resource,Instance, Job, Volume	Delete	删除

Resource,Instance, Job, Volume	All	具备上述所有权限
OfflineModel	Read	读取
OfflineModel	Write	更新
OfflineModel	Delete	删除
OfflineModel	All	具备上述所有权限
Xflow	Read	读取
Xflow	Write	更新
Xflow	Execute	执行
Xflow	Delete	删除
Xflow	All	具备上述所有权限

上述权限描述中 **Project 类型对象的 CreateTable 操作，Table 类型的 Select、Alter、Update、Drop 操作需要与 Project 对象的 CreateInstance 操作权限配合使用。单独使用上述几种权限而没有指派 CreateInstance 权限是无法完成对应操作的。这与 ODPS 的内部实现相关。类似的，Table 的 Select 权限也要与 CreateInstance 权限配合使用。使用时请注意。**

● JOIN

- ODPS SQL 支持的 Join 操作类型包括：{LEFT OUTER|RIGHT OUTER|FULL OUTER|INNER} JOIN；
- ODPS 的 JOIN 支持多路间接，但不支持笛卡尔积，即无 on 条件的链接。
- 连接条件，只允许 and 连接的等值条件，并且**最多支持 16 路 join 操作。**
- 只有在 MAPJOIN 中，可以使用不等值连接或者使用 or 连接多个条件。

● MAPJOIN HINT

- left outer join 的左表必须是大表；
- right outer join 的右表必须是大表；
- inner join 左表或右表均可以作为大表；
- full outer join 不能使用 mapjoin；
- mapjoin 支持小表为子查询；
- 使用 mapjoin 时需要引用小表或是子查询时，需要引用别名；
- 在 mapjoin 中，可以使用不等值连接或者使用 or 连接多个条件；
- 目前 MaxCompute 在 mapjoin 中最多支持**指定 6 张小表**，否则报语法错误；
- 如果使用 mapjoin，则所有小表占用的内存总和不得**超过 512MB**。请注意由于 MaxCompute 是压缩存储，因此小表在被加载到内存后，数据大小会急剧膨胀。此处的 512MB 限制是加载到内存后的空间大小；
- 多个表 join 时，最左边的两个表不能同时是 mapjoin 的表。

● 数据倾斜

Join 造成的数据倾斜优化

- 小表并且没有超过 512MB,我们将上面的语句优化成 mapjoin 语句再执行
- 将倾斜的 key 用单独的逻辑来处理

GroupBY 造成的数据倾斜优化

- 在执行 SQL 前设置防倾斜的参数: `set odps.sql.groupby.skewindata=true。`

错误使用动态分区造成的数据倾斜

- 减少 MaxCompute 系统产生的小文件,使后续处理更快;
- 避免一个 Worker 输出文件很多时占用内存过大。

对 SQL 进行复习 https://help.aliyun.com/document_detail/48950.html

MaxCompute SQL 采用的是类似于 SQL 的语法,可以看作是标准 SQL 的子集,但不能因此简单的把 MaxCompute 等价成一个数据库,它在很多方面并不具备数据库的特征,如**事务、主键约束、索引**等。目前在 MaxCompute 中允许的最大 SQL 长度是 **2MB**。

MaxCompute SQL 支持在一个语句中插入不同的结果表或者分区(语法: FROM from_statement INSERT OVERWRITE | INTO…… [INSERT OVERWRITE | INTO……])

- 一般情况下,单个 SQL 里最多可以写 128 路输出,超过 **128 路**报语法错误。
- 在一个 multi insert 中,对于分区表,同一个目标分区不允许出现多次;对于未分区表,该表不能出现多次。
- 对于同一张分区表的不同分区,不能同时有 insert overwrite 和 insert into 操作,否则报错返回。

ODPS 对 **SQL DML 语句**进行解析后,ODPS 会对其进行解析,得出任务的执行计划。执行计划是由具有依赖关系的多个**执行阶段(Stage)**构成的

create table like 语句不会复制源表的**生命周期属性**。除生命周期属性外,列名、列注释以及表注释等均相同。

limit number 的 number 是常数,限制输出行数。当使用**无 limit 的 select 语句**直接从屏幕输出查看结果时,**最多只输出 5000 行**。

LIKE 和 RLIKE 的 source 和 pattern 参数均仅接受 string 类型;

In 的隐式转换规则:

In 右侧的 value 值列表中的数据类型必须一致;

当 key 与 values 之间比较时,若 **bigint, double, string 之间比较,统一转 double**,若 **datetime 和 string 之间比较,统一转 datetime**。除此之外不允许其它类型之间的转换。

Boolean 类型与其他类型都不能转换

wm_concat 与 concat

wm_concat 聚合函数，用途：用指定的 separator 做分隔符，链接 str 中的值。

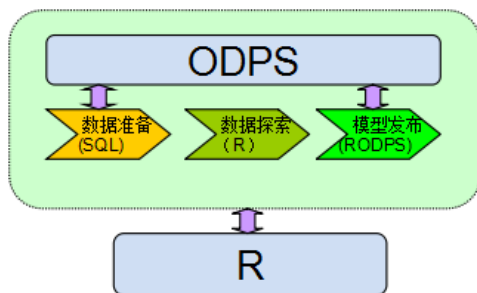
其输入与输出是多对一的关系，即将多条输入记录聚合成一条输出值。可以与 SQL 中的 group by 语句联用。

concat 用途：返回值是将参数中的所有字符串连接在一起的结果。

To_date

1. to_date('阿里巴巴 2010-12*03', '阿里巴巴 yyyy-mm*dd') = 2010-12-03 00:00:00
2. to_date('20080718', 'yyyymmdd') = 2008-07-18 00:00:00
3. to_date('2008718', 'yyyymmdd') -- 格式不符合，引发异常
4. to_date('阿里巴巴 2010-12*3', '阿里巴巴 yyyy-mm*dd') -- 格式不符合，引发异常，若是 '阿里巴巴 2010-12*03' 就能转换
5. to_date('2010-24-01', 'yyyy') -- 格式不符合，引发异常
6. to_date('2010-24-01', 'yyyy-24-01') 返回 2010-01-01 00:00:00
7. to_date('1234-2234', 'yyyy-yyyy') 会返回 1234-01-01 00:00:00。format 参数至少包含 "yyyy"，否则引发异常，如果 format 中出现多余的格式串，则只取第一个格式串对应的日期数值，其余的会被视为分隔符。

RODPS 能够提供一个桥接的方式，使得可以在 **R 语言环境** 中无缝的使用 MaxCompute (原 ODPS) 里面的数据、计算资源，类似于开源社区的 R Hive 和 Rhadoop 的功能。



数据导入命令发送后，数据并不会立刻导入到分析型数据库中，而是会在后台进行数据的导入工作。用户可以通过使用 SQL 命令或在 iDB Cloud 中进行查询数据导入状态。高级用户也可以直接在 [information_schema](#) 中查询全部数据导入的信息（具体见附录）。导入方法有三种：

- 1、通过 SQL 语句查询；
- 2、通过通过 iDB Cloud 查询数据导入状态 在 iDB Cloud 中；
- 3、通过 Meta DB 进行查询。

计费：MaxCompute 以 **项目 (Project)** 为基本单元，从 **存储、计算和数据下载** 三个方面分别计费。计费周期为：天

ODPS 的云账号认证使用**消息签名**机制

平滑扩容

当目前的 DRDS 无法支撑现有业务时，我们可以通过用户管理控制台进行 DRDS 的平滑扩容，以满足业务需求。平滑扩容的流程依次分别为（1）**配置**、（2）**迁移**、（3）**切换**、（4）**清理**四个步骤。

ADS

在分析型数据库中，**数据库**是用户所关心的**最大单元**，也是用户和分析型数据库系统管理员的管理权限的**分界点**。分析型数据库系统管理员最小可管理数据库粒度的参数，而无法未经用户授权来查看和管理数据库内部的结构和信息。

而数据库级别的参数对于用户来讲大多只能查看而不能修改。在分析型数据库中，一个数据库对应着一个用于访问的域名和端口号，并且**有且只有一个 owner 即为数据库的创建者**。

另外对用户的宏观的资源配置是以数据库为粒度进行的，因此创建数据库时用户需要输入业务预估的 QPS、数据量、Query 类型等信息用于智能的判断初始的资源分配。

分析型数据库以 **ECU (弹性计算单元)** 作为资源计量的最小单位。ECU (弹性计算单元) 拥有多种型号，每种型号的 ECU，标识着不同的 **vCPU 核数、内存大小、磁盘空间大小**。

分析型数据库会对通过 SELECT 语句进行查询的返回行数做全局的最大限制，公共云上目前为 **10000** 行，**若不加 LIMIT 或 LIMIT 行数超过 10000，则只会返回 10000 行**

连接方式与环境：

- 用户管理控制台中的 DMS
- 使用版本为 **5.4** 系列的 mysql-jdbc 驱动
- **mysql5.6** 的客户端
- 安装了 **php-mysql-5.1.x** 模块的 php 环境

复习 SQL :

https://help.aliyun.com/document_detail/26408.html

Explain:https://help.aliyun.com/document_detail/26420.html

ml

● 表组：

```
create tablegroup db_name.tablegroup_name options(minRedundancy=2 executeTimeout=30000);
```


minRedundancy: 副本数

executeTimeout: 查询超时时间

增加副本数可以一定程度的增加分析型数据库的最大承受的 QPS，但是数据存储费用也会相应增加。在绝大部分情况下，不推荐修改任何表组默认配置。

仅允许删除没有任何表的空表组，维度表组不允许删除。

修改表组: alter tablegroup db_name.tablegroup_name key=value; minRedundancy 修改后需要下次装载数据时才会生效。

ADS 普通表一级分区数不超过 256 个，单个分区数据量不超过 800W 为宜。

分析型数据库的用户不能直接通过 CREATE DATABASE 的 DDL 语句创建数据库，**只能通过 DMS 控制台界面来创建需要的业务数据库。**

扩容缩容: ALTER DATABASE SET ecu_count = N;

分析型数据库的表根据 updateType 分为批量更新表（仅能够离线批量更新数据）和实时更新表（能够通过 insert/delete 实时更新数据），用 **updateType (realtime|batch)** 以区分，如果 updateType 选项不填则默认为批量更新表。需要注意的是，updateType=realtime 暨为**实时更新表时，必须指定合法的主键并且不能有二级分区。**

事实表的创建上，默认有如下限制：

- （1）一张事实表至少有一级 Hash 分区并且分区数不能小于 8 个；
- （2）一个事实表组最多可以创建 256 个事实表；
- （3）一个事实表最多不能超过 1024 个列。

普通表的分区目前**最多为两级**，分区种类有 Hash 分区和 List 分区两种。目前分析型数据库支持将普通表的一级分区设置为 Hash 分区，二级分区设置为 List 分区，即可支持 Hash join 又可支持增量数据导入。

每个分区一般不超过 1500 万条记录为宜（亦可通过划分二级分区实现无限扩展）。

● 二级分区

二级 List 分区为**非动态分区**，暨分区值不是由数据本身决定的，而是**由每次导入数据时用户指定的**。

所以在进行分区信息定义时需要指定一个**和现有数据中的列不同的新列名**，以及这个列的类型（**目前仅支持 long**）。

二级分区有一个可选属性，available_partition_num，即为最大保留的二级分区数，当新的数据装载进来后，若线上存在的二级分区数大于这个值，那么会根据二级分区的值进行排序，下线最小的若干分区的数据。

实时更新表不支持二级分区。

● 索引

- 列属性上，一个列可以设置列属性 **disableIndex = true**，用于屏蔽分析型数据库的默认列索引，不过需要注意的是，要如此做，则这个列应该不在实际查询中所筛选和计算的。
- 分析型数据库中主键的概念是弱化的，分析型数据库不要求一个表有主键，有主键的表的性能和用法上和没有主键的表之间没有任何区别。
- 分析型数据库拥有高度智能的自动化索引机制，所以通常用户无需亲自为自己的数据表配置索引。

- 一种情况例外：暨用户需要对某列进行 Hash Join 时，无论是事实表之间的 Join 还是事实表和维度表的 Join，都需要为事实表的该列建立索引。

批量导入：data pipeline（如 SQL 命令 Load Data From XXX） HTTP Rest-API

● Explain

当用户发起一个 explain 查询到分析型数据库系统后，分析型数据库会**抽样一个数据分区**来分析执行计划，并以图形方式展现给用户。

当用户通过查询的方式，想要获取文本格式的 explain 语句后，将会得到如下的 json 串。

EXPLAIN 的 ResultSet 记录。其中第一行为（1）**逻辑计划**，第二行为（2）**物理计划**。

物理计划各个节点说明：

JoinExecutor 表示 Join 的节点；

TableExecutor 表示参与计算的表信息

● 权限

查询数据的权限： 查询表数据需要 SELECT 权限，最小级别是列；并非所有查询都需要该权限，例如 SELECT now()

导出数据权限： 导出数据同时需要 DUMP DATA 和 SELECT 权限；同时需要数据导出目的地的数据写入相关权限

聚合： Database -> Table[Group] -> Column，即每个权限级别能聚合其下面级别的所有权限

● ECU

ECU（弹性计算单元）拥有多种型号，每种型号的 ECU，标识着不同的 vCPU 核数、内存大小、磁盘空间大小

用户在创建数据库时需要根据自己的需求选择这个数据库的 **ECU 型号**，以及初始的 **ECU 数量**（必须是**偶数个，至少两个**），ECU 型号 DB 创建后**不可修改**，ECU 数量可以在使用中**随时调整（扩容/缩容）**缩容和扩容**都不是瞬时的同步操作**，可以使用元数据查询状态。

RDS

RDS 基本功能：**实例管理、账号管理、数据库管理、网络管理、安全管理**、标签管理（根据标题摘录）

MySQL 存储引擎：目前支持 InnoDB、TokuDB 两种引擎（MyISAM 引擎由于自身缺陷，存在数据丢失的风险，因此仅部分存量实例暂时支持，新创建实例的 MyISAM 引擎表会自动转换为 InnoDB 引擎表）。**不支持 Memory 引擎**。如果您创建 Memory 引擎的表，我们将自动为您转换成 InnoDB 引擎的表。

同一实例下的数据库共享该实例下的所有资源，其中：

MySQL 版实例最多可以**创建 500 个数据库**

SQL Server 2008 R2 版实例最多可以**创建 50 个数据库**

PostgreSQL 版实例和 PPAS 版实例**没有数据库个数限制**

RDS 采用的**主备 M-M 的高可用架构**，其主备之间的数据同步依靠日志的方式实现，MySQL 数据库为 **binlog**，

SQL server 为**事务日志**

RDS for SQLServer

- 为保障实例的稳定及安全，RDS for SQL Server 有部分使用上的限制，详情如下：
- 单个实例数据库个数限制为 50 个
- 数据库账号限制为 500 个
- 不提供在命令行创建数据库和账号的权限，无 grant 权限
- 出于安全考虑，某些 SQL Server 功能不能够使用，如：分布式事务、Windows 域账号登录、邮件功能、BI 分析功能、报表功能、数据库级别的 DDL 触发器（触发器中含有修改表结构、创建表、增删索引等 DDL 语句）、程序集、Service Broker、SQL Server Profiler、复制功能、策略管理、代理启动停止功能等

● DTS

DTS 提供的**数据订阅**功能意在帮助用户实时获取 RDS 增量数据，实现数据分析，消息转发及本地灾备等功能，但是目前 DTS 支持的 RDS MySQL 订阅功能存在一定的限制，具体限制如下：

- MySQL5.6 binlog 的 binlog_row_image 格式必须为 FULL
- MySQL 存储引擎只支持 myisam、innodb
- MySQL 字符集只支持 latin1、latin2、gbk、utf8、utf8mb4、binary

● RDS 购买方式

包年包月实例

在合同期内，实例**只能升级配置，不能降级配置或者释放**

实例到期 7 天后会被释放

按时付费实例

购买按时付费实例需要账号实名认证，并且账号余额不少于 100 元

实例可以随时释放

欠费 24 小时后会锁定，欠费 7 天后会被释放

RDS 实例将会被锁定，无法进行读写。

● 实例空间文件

除了您正常的**数据**外，还有数据库实例正常运行所需要的空间，比如系统数据库、数据库**回滚日志**、**重做日志**、**索引**等。同时 RDS for MySQL 数据库实例产生的一天内的 **Binlog 日志**也会占用这个空间。这些文件会保证 RDS 实例正常稳定运行。

● RDS 升级

云数据库 RDS 版支持用户将低版本数据库升级到高版本，但是无法降级版本。具体可升级版本请参见实际界面。

注意：目前该操作仅适用于数据库 MySQL 5.5 到 MySQL 5.6 的升级。

在数据库升级过程中，RDS 服务可能会出现 1 次 30 秒的闪断，请您尽量在业务低峰执行升级操作，或确保您的应用有自动重连机制，以避免闪断造成的影响。

● RDS 优势

高可靠性

双机热备 RDS 采用热备架构，物理服务器出现故障后服务秒级完成切换。整个切换过程对应用透明。

多副本冗余 RDS 服务器中的数据构建于 RAID 之上，数据备份存储在 OSS 上。

数据备份 RDS 提供自动备份的机制。用户可以自行选择备份周期，也可以根据自身业务特点随时发起临时备份。临时备份**每天可触发次数没有限制，不支持第三方工具**进行物理备份

数据恢复 支持按备份集和指定时间点的恢复。在大多数场景下，用户可以将 **7 天内任意一个时间点**的数据恢复到 RDS 临时实例上，数据验证无误后即可将数据迁回 RDS 主实例，从而完成数据回溯。

高安全性

防 DDoS 攻击 遭受 DDoS 攻击 -> 流量清洗 -> 攻击达到黑洞阈值进行黑洞处理

流量清洗：只针对外网流入流量进行清洗，处于流量清洗状态的 RDS 实例可正常访问。流量清洗的触发和结束由系统自动完成，单个 RDS 实例满足以下任一条件即触发流量清洗：

PPS (Package Per Second) 达到 **3 万**；

BPS (Bits Per Second) 达到 **180Mb**；

每秒新建并发连接达到 1 万；

激活并发连接数达到 1 万；

非激活并发连接数达到 10 万。

黑洞处理：只针对外网流入流量进行黑洞处理，处于黑洞状态的 RDS 实例不可被外网访问，此时应用程序通常也处于不可用状态。黑洞处理是保证 RDS 整体服务可用性的一种手段。黑洞触发条件如下：

BPS (Bits Per Second) 达到 2Gb；

流量清洗无效。

黑洞结束条件如下：

黑洞在 **2.5 小时**后自动解除。

建议用户通过内网访问 RDS 实例，可以使 RDS 实例免受 DDoS 攻击的风险。

访问控制策略

用户可定义允许访问 RDS 的 IP 地址，指定之外的 IP 地址将被拒绝访问。

每个账号只能看到、操作自己的数据库。

系统安全

RDS 处于多层防火墙的保护之下，可以有力地抗击各种恶意攻击，保证数据的安全。

RDS 服务器不允许直接登录，只开放特定的数据库服务所需要的端口。

RDS 服务器不允许主动向外发起连接，只能接受被动访问。

专业安全团队 阿里巴巴集团安全部门负责 RDS 的安全技术支持。

高性能：参数优化、SQL 优化建议、高端硬件投入

便宜易用：即开即用、按需升级、透明兼容、管理便捷

● RDS MySQL5.5 版本新特性

- 1、提高性能和可扩展性** 支持多个回滚段增加并发处理能力，插入缓冲的改进提升数据库写性能。
- 2、提升易管理性和效率** 支持 InnoDB 数据页压缩，数据页的压缩使数据文件体积变小，减少磁盘 I/O，提高吞吐量。
- 3、提高可用性** 分区功能的增强，可 drop 分区表。
- 4、改善检测和诊断** 引入 performancn_schema, 用于监控 mysql 监控服务器运行时的性能。

- **只读实例**

前提条件（只读实例只能应用于 MySQL 实例的 5.6 版本）

- 当前只读实例只支持数据库类型为 MySQL 的 RDS 实例
- 要创建只读实例，主实例要求 MySQL 5.6 及以上版本。在升级主实例版本前，请做好兼容性测试；或者新建一个 MySQL 5.6 的实例，将数据从主实例复制到新建实例，然后在新建实例上创建只读实例

功能特点

- **规格可以与主实例不一致，并可以随时更改规格（没有时间限制）**，可以更方便的弹性升降级
- 支持按小时计费，使用更灵活，费用更便宜，只读实例欠费 24 小时后会被锁定，欠费 7 天后会被释放，请确保账户有足够余额
- 不需要维护账号与数据库，全部通过主实例同步
- **独立的白名单配置**
- 提供系统性能监控：RDS 提供近 20 个系统性能的监控视图，如磁盘容量、IOPS、连接数、CPU 利用率、网络流量等，用户可以轻松查看实例的负载
- 提供优化建议：RDS 提供多种优化建议，如存储引擎检查、主键检查、大表检查、索引偏多、缺失索引等，用户可以根据优化建议并结合自身的应用来对数据库进行优化

功能限制

- 1 个主实例最多可以创建 **5 个只读实例**
- 备份设置：不支持备份设置以及临时备份
- 数据迁移：不支持数据迁移至只读实例
- 数据库管理：**不支持创建和删除数据库**
- 账号管理：**不支持创建和删除账号，不支持为账号授权以及修改账号密码功能**
- 实例恢复：不支持通过备份文件或任意时间点创建临时实例，不支持通过备份集覆盖实例
- **创建只读实例后，主实例将不支持通过备份集直接覆盖实例来恢复数据**

DRDS

- DRDS 的读写分离功能是一种对应用透明的读写分离实现，应用在不需修改任何代码的情况下，只需要在 DRDS 控制台中**调整读权重**，即可将流量按照需要的比例在主实例与多达 **5 个只读实例**之间调整。
- 落到只读实例上读所操作的数据都是从**主实例上异步**同步的，延迟在**毫秒级别**，所以个别要求实时性的 SQL 请通过 DRDS hint 指定主库执行。
- DRDS 默认提供了 3 种策略（**主库读、均衡、只读均衡、自定义**），每种策略都会有不同的读权重分配，用户也可以自定义
- **写权重则统一走主实例，不分流。**

DRDS 对于单机事务完整支持，也就是业务中一个事务中的各个 SQL 最终都落到同一个数据库即可保障强一致，对于跨数据库的分布式事务，DRDS 提供最终一致分布式事务给业务使用，drds 实例中每一个 RDS 的分库**限制是 8 个**，但每个分库里面**分表的数量没有限制**

OTS

OTS python CLI 工具需要 **2.7 版本** 的 python 支持。

操作：

单行数据操作：**GetRow / PutRow / UpdateRow / DeleteRow**

多行数据操作：**GetRange / BatchGetRow / BatchWriteRow**

表需要提供的参数主要有：

- TABLE_NAME：表的名称。
- PrimaryKeySchema：主键的名称及类型。
- timeToLive：数据表的数据过期时间。
- maxVersion：数据表属性列的最大版本号。
- ReservedThroughput：数据表的预留读写吞吐量，容量型实例不再支持预留吞吐量的设置。

Table Store 以**实例和表**的形式组织数据，通过**数据分片**和**负载均衡**技术，达到规模的无缝扩展

- 表格存储支持多个主键，**第一个主键为分区键**，分区键**不允许设置为自增列**。除分区键以外，其它主键都可以设置为自增列。每张表最多**只允许设置一个主键**为自增列。
- Table Store 会根据表中每一行分片键的值所属的范围自动将这一行数据分配到对应的分片和机器上，以达到负载均衡的目的。
- 具有相同分片键的行属于同一个分片，**一个分片可能包含多个分片键**。Table Store 服务会根据特定的规则对**分片进行分裂和合并**，以达到更好的负载均衡，这个**过程是自动的**，应用无需关心。
- 为了防止分片过大无法切分，**建议**单个分片键下的所有行**数据大小之和不要超过 1 GB**。

● 读/写吞吐量

为读服务能力单元和写服务能力单元，简称 CU（Capacity Unit），是对数据读写操作的最小计费单位。1 单位读能力表示从数据表中读一条 4 KB 数据，1 单位写能力表示向数据表写一条 4 KB 数据。操作数据大小不足 4 KB 的部分向上取整，如写入 7.6 KB 数据消耗 2 单位写能力，读出 0.1 KB 数据消耗 1 单位读能力。**单表的预留读写吞吐量 0-5000 如有需求提高上限，请联系客服。**

各操作消耗的写服务能力单元的计算规则：

PutRow - 本次消耗的写 CU 为修改的行主键数据大小与属性列数据大小之和除以 4KB 向上取整，若指定条件检查不为 IGNORE，还需消耗该行主键数据大小除以 4KB 向上取整的读 CU。如果操作不满足应用指定的行存在性检查条件，则操作失败并消耗 1 单位写服务能力单元和 1 单位读服务能力单元 PutRow 详解

UpdateRow - 本次消耗的写 CU 为修改的行主键数据大小与属性列数据大小之和除以 4KB 向上取整，UpdateRow 中包含的需要删除的属性列，只有其列名计入该属性列数据大小，若指定条件检查不为 IGNORE，还需消耗该行主键数据大小除以 4KB 向上取整的读 CU。如果操作不满足应用指定的行存在性检查条件，则操作失败并消耗 1 单位写服务能力单元和 1 单位读服务能力单元 UpdateRow 详解

DeleteRow - 被删除的行主键数据大小除以 4KB 向上取整。若指定条件检查不为 IGNORE，还需消耗该行主键数据大小除以 4KB 向上取整的读 CU。如果操作不满足应用程序指定的行存在性检查条件，则操作失败并消耗 1 单位写服务能力单元 DeleteRow 详解

● OTSClient

创建 OTSClient 对象，在构造函数中指定 **EndPoint**、**AccessKeyID**、**AccessKeySecret**、**实例名字**。

通过拼接多列主键列形成分片键的表减少了单个分片键下的总数据量大小。

● 限制项汇总

以下为表格存储的限制项说明，部分限制范围表明用户能够使用的最大值，并不为建议值。为保证更好的性能，请合理设计表结构和单行数据大小。

限制项	限制范围	说明
一个阿里云用户帐号下可以保有实例数	不超过 10	如有需求提高上限，请联系客服。
一个实例中表的个数	不超过 64	如有需求提高上限，请联系客服。
实例名字长度	3-16 bytes	字符集为 [a-zA-Z0-9] 和连词符 (-)，首字符必须是字母且末尾字符不能为连词符 (-)。
表名长度	1-255 bytes	字符集为 [a-zA-Z0-9_]，首字符必须是字母或 (_)。
列名长度限制	1-255 bytes	字符集为 [a-zA-Z0-9_]，首字符必须是字母或 (_)。
主键包含的列数	1-4	至少 1 列，至多 4 列。
String 类型主键列列值大小	不超过 1 KB	单一主键列 String 类型的列列值大小上限 1 KB。
String 类型属性列列值大小	不超过 2 MB	单一属性列 String 类型的列列值大小上限 2 MB。
Binary 类型主键列列值大小	不超过 1 KB	单一主键列 Binary 类型的列列值大小上限 1 KB。
Binary 类型属性列列值大小	不超过 2 MB	单一属性列 Binary 类型的列列值大小上限 2 MB。
一行中属性列的个数	不限制	不限制单一行拥有的属性列个数。
单行数据大小	不限制	不限制单一行中所有列名与列值总和大小。
单表的预留读写吞吐量	0-5000	如有需求提高上限，请联系客服。
读 请 求 中 columns_to_get 参数的列的个数	0-128	读请求一行数据中获取的列的最大个数。
表级别操作 QPS	10	一个实例的表级别操作每秒不超过 10 次，表级别操作见表操作。
单表 UpdateTable 的次数	上调：无限制， 下调：无限制	需要遵循单表的调整频率限制。
单表 UpdateTable 的频率	每 2 分钟 1 次	单表在 2 分钟之内，最多允许调整 1 次预留读/写能力值。
BatchGetRow 一次操作请求读取的行数	不超过 100	N/A
BatchWriteRow 一次操作请求写入行数	不超过 200	N/A

BatchWriteRow 一次操作的数据大小	不超过 4 MB	N/A
GetRange 一次返回的数据	5000 行或者 4 MB	一次返回数据的行数超过 5000 行，或者返回数据的数据大小大于 4 MB。满足以上任一条件时，超出上限的数据将会按行级别被截掉并返回下一行数据主键信息。
一次 HTTP 请求 Request Body 的数据大小	不超过 5 MB	N/A

通过**拼接多列主键列形成分片键**的表减少了单个分片键下的**总数据量大小**。

云盾

用户创建监控项后，可对探测点的**响应时间、状态码、丢包率等**监控项设置报警规则。报警规则的统计周期和创建监控项时的统计周期一致。

报警服务支持**短信、邮件、旺旺、事件订阅**四种方式。旺旺仅支持 PC 端报警消息推送。如果您安装了阿里云 APP，也可以通过阿里云 APP 接收报警通知

- **云盾先知**

先知计划是一个帮助企业建立**私有应急响应中心**的平台（帮助企业收集漏洞信息）。云盾先知计划包括以下哪些功能：**A、漏洞收集 B、审核漏洞 C、自主设定奖励计划 D、协助漏洞修复**

- **反欺诈服务**

功能：垃圾注册、暴力破解、撞库、营销作弊、盗卡支付、垃圾广告或内容、恶意登入。**云盾反欺诈服务**是阿里巴巴的**大数据风控服务**能力开发的？

使用：可支持 WEB 和移动平台。通过前端接入，不同的平台，方式不一样：

- Web 平台、移动平台 **HTML5 应用**，前端页面引入 **JS 脚本**。
- **移动平台 Native** 应用，引入 **SDK**。

- **安全网络**

会将攻击者所有的请求都会进入安全网络的虚拟防护节点上，通过每个节点自带的安全防护功能，进行流量清晰过滤，真实的服务器始终在安全网络的防护下，避免黑客对其发动攻击。

- **高防 DDos**

防护多种 DDos 类型攻击 包括但不限于以下攻击类型 ICMP Flood、UDP Flood、TCP Flood、SYN Flood、ACK Flood 等

随时更换防护 IP DDoS 防护服务可随时更换防护的 IP，让您更自由配置 高级别防护

弹性防护 DDoS 防护阈值弹性调整，您可以随时升级更高级别的防护，整个过程服务无中断

精准防护报表 提供实时精准的流量报表及攻击详情，让您及时准确获得当前服务详情

- **主机防护 安骑士（通过反入侵服务进行监测入侵并通知客户）**

（非阿里云服务器也可以进行防护）

内容 主机密码防爆破、异地登录报警、网站后门检测、高危漏洞修复、木马文件查杀

特点 轻量化资源占用、实时报警、云平台联动防御

- **数据库防护**

Proxy+DB FW（数据库防火墙）

白名单限制访问源 IP

内网访问模式

防暴力破解

SQL 注入防护

SQL 审计 （不规范 SQL 审计进行报警）

- **应用防火墙（WAF） 保护两类服务器：云服务器 ECS 、负载均衡 SLB**

跨站攻击

CRLF 攻击

SQL 注入攻击

FastCGI 攻击

WEBSHELL 攻击

SLB

用户可以为创建的 SLB 实例指定一个易于识别的名称，如果用户不指定，那么系统将以该 SLB 实例的 loadBalancerId 作为名称，LoadBalancerID 是识别用户 SLB 实例的**唯一标识**，无法修改。

机器学习

阿里云机器学习主要应用在数据挖掘场景下，即指从大量的数据中通过算法搜索隐藏于其中信息的过程。通过统计、在线分析处理、情报检索、机器学习、专家系统（依靠过去的经验法则）和模式识别等诸多方法来实现隐藏信息的探索 and 发现。其场景主要有：

分类：分类可以找出这些不同种类客户之间的特征，让用户了解不同行为类别客户的分布特征，从而进行商业决策和业务活动，如：在银行行业，可以通过阿里云机器学习对客户进行分类，以便进行风险评估和防控；在销售领域，可以通过对客户的细分，进行潜在客户挖掘、客户提升和交叉销售、客户挽留等

聚类：通常“人以群分，物以类聚”，通过对数据对象划分为若干类，同一类的对象具有较高的相似度，不同类的对象相似度较低，以便我们度量对象间的相似性，发现相关性。如在安全领域，通过异常点的检测，可以发现异常的安全行为。通过人与人之间的相似性，实现团伙犯罪的发掘

预测：通过对历史事件的学习来积累经验，得出事物间的相似性和关联性，从而对事物的未来状况做出预测。比如：预测销售收入和利润，预测用户下一个阶段的消费行为等

关联：分析各个物品或者商品之间同时出现的机率,典型的场景如: 购物篮分析。比如超市购物时，顾客购买记录常常隐含着很多关联规则，比如购买圆珠笔的顾客中有 65%也购买了笔记本，利用这些规则，商场人员可以很好的规划商品摆放问题。在电商网站中，利用关联规则可以发现哪些用户更喜欢哪类的商品，当发现有类似的客户的时候，可以将其它客户购买的商品推荐给相类似的客户，以提高网站的收入。

【多选】【2个正确项】阿里云 MaxCompute 中，下列对 lifecycle 描述正确的是（AD）

- A、使用 lifecycle 设置表的生命周期
- B、lifecycle 的单位是月
- C、可以使用浮点数设置 lifecycle
- D、lifecycle 可以不用设置

【解析】[lifecycle days] days 参数为生命周期时间，只接受正整数。单位：天

【判断题】流计算可以实现多种复杂规则引擎告警，针对单——条数据存在多条复杂规则告警，且规则在系统运行时变化（X）

【解析】这类应该有专门的规则引擎系统解决，单流计算面对不是该问题域。

【判断题】现有的 Spark 作业可以无缝迁移到流计算（X）

【解析】Spark 部分涉及流计算可以考虑改造并迁移到流计算，用户可以完全省去运维 Spark 和开发 Spark 的各类成本，但无法做到 Spark 作业无缝迁移到流计算

Oracle 存储过程可以使用阿里云流计算替换（X）

【解析】流计算无法从功能上完全替换掉 Oracle 存储过程，两者面向问题领域不一样

一、多选题

1. MaxCompute 提供了哪些访问接口可以使用？

- A. SQL; B. MapReduce; C. Graph; D. odpscmd

二、判断题

1. MaxCompute 只能支持对结构化数据的存储和处理

- A、√ B、×

2. 在 MaxCompute 中，对表的列数和列内容的大小没有限制

- A、√ B、×

3. 在 select 语句的 where 子句中可以指定查询范围，这样可以仅仅扫描表的 指定部分避免全表扫描

- A、√ B、×

4. 使用 mapjoin 的性能比普通的 join 要快很多，与表的大小无关。

- A、√ B、×

1. 下列关于 MaxCompute 的描述中，错误的是（）

- A. 一种快速、完全托管的 TB/PB 级数据仓库解决方案
- B. 向用户提供了完善的数据导入方案
- C. 多种经典的分布式计算模型，能够更快速的解决用户海量数据计算问题
- D. 提供了完善的 SQL 访问，有效支持事务处理

二、多选题

1. 下列产品中，能够满足用户使用 SQL 进行交互式查询需求的是（）

- A. MaxCompute
- B. AnalyticDB
- C. RDS

D. TS

三、判断题

1. 在 MaxCompute 中使用 DDL 建表, 如果一开始并没有分区字段, 在使用时 可以动态增加或更改分区

A、√ B、×

2. 在 MaxCompute 中, 通常使用 SQL DML 中的 UPDATE 更新 MaxCompute 表或者分区中的数据

A、√ B、×

3. 在 MaxCompute 中, 通常使用 SQL DML 中的 DELETE 删除 MaxCompute 表或者分区中的数据 ()

A、√ B、×

1. 下列关于阿里云流计算的描述中, 错误的是 ()

- A. 是运行在阿里云平台上的流式大数据分析平台
- B. 提供给用户在云上进行流式数据实时化分析工具
- C. 可以使用阿里云 StreamSQL 进行流式数据分析
- D. 流计算的数据存储是基于盘古分布式文件系统的。

2. 可以用于的行为或相关事件进行实时监测和分析的产品是 ()

- A. AnalyticDB
- B. 流计算
- C. MaxCompute
- D. TS

二、多选题

1. 下列关于流计算特点的描述中, 正确的是 ()

- A. 实时且流式的
- B. 数据是无界的
- C. 事件触发
- D. 用户触发

2. 下列对流式计算的链路描述正确的是 ()

- A. 流计算全链路整体上更加强调数据的实时性
- B. 链路包括数据实时采集、数据实时计算、数据实时集成
- C. 一个流计算作业必须至少使用一个流数据作为源
- D. 流计算将计算的结果数据直接写入目的数据存储

3. 流式计算的应用场景包括 ()

- A. 实时数据仓库
- B. 实时监控
- C. 实时报表
- D. 流式数据分析

一、多选题

1. 阿里云流计算可以用于下列哪些应用场景中? ()

- A. 实时数据仓库
- B. 实时报表

C. 实时监控

D. 离线数据仓库

2. 阿里云流计算可以直接读写下列哪些产品的数据? ()

A. DataHub

B. 日志服务(SLS)

C. RDS

D. AnalyticDB

二、判断题

1. 阿里云流计算目前可以支持 SQL 和 MapReduce () **StreamSQL/UDF**

A. ☒ B. ☒

2. 流计算对不同的项目进行了严格的项目权限区分, 不同用户/项目想之间是无法进行访问、操作 ()

A. ☒ B. ☒

3. 在阿里云流计算中, 一个用户作业在运行期间随着数据量的突增提升了其作业 CPU 使用率, 这将会影响到其他用户作业的 CPU 使用情况 ()

A: ☒ B: ☒

答案: 一、多选 (1、ABC; 2、ABCD;) 二、判断 (1、☒; 2、☒; 3、☒;)

一、多选题

1. DataWorks 中, 工作流任务支持的调度类型包括:

A. 一次性调度

B. 周期调度

C. 混合类型调度

D. 需要第三方工具用于支持调度

2. DataWorks 中工作流任务如果配置为周期性调度, 所支持的周期包括哪些?

A. 月调度

B. 周调度

C. 天调度

D. 季度调度

E. 小时调度

F. 1 分钟调度

二、判断题

1. DataWorks 中, 工作流任务支持一次性调度和周期调度两种类型。如果将某工作流任务设置为按天调度, 则该任务每天都会被执行。

A. ☒

B. ☒

2. DataWorks 中, 通过数据开发面板的脚本开发, 可以创建 SQL 脚本文件并直接编辑、执行。

A: ☒

B: ☒

3. DataWorks 内置有数据同步工具, 不需要第三方的工具就可以完成数据的导入导出。

A: ☒

B: ☒

一、多选题

1. 阿里云 MaxCompute 中，下列对 lifecycle 描述正确的是（）

- A、使用 lifecycle 设置表的生命周期
- B、lifecycle 的单位是月。
- C、可以使用浮点数设置 lifecycle
- D、lifecycle 可以不用设置。

二、判断题

1. 在阿里云 MaxCompute 中，使用 Create Table 建表时，可以使用中文作为表的名字（）

- A、√
- B、×

2. MaxCompute 的表的名称对大小写字母是有区别的，列的名称对大小写没区别（）

- A、√
- B、×

4. MaxCompute 支持外部表，外部表中的数据可以存储在 OSS 或者 TS 中（）

- A、√
- B、×

1. 【多选】当前阿里云流计算对外接口包括（）

- A. StreamSQL
- B. UDF
- C. MapReduce
- D. Graph

【答案】AB

【解析】当前流计算对外接口定义为 StreamSQL/UDF，提供服务于流式数据分析、统计、处理的一站式开发工具，面向客户更多是数仓开发人员、数据分析师，这类客户不希望更多参与底层代码开发，而希望简单编写流计算 SQL 即可完成自身流式数据分析业务。

2. 【单选】Use Project 是 MaxCompute 服务端提供的命令。

- A.正确
- B.错误

【答案】B

【解析】Use Project 是 MaxCompute 客户端提供的命令。

【多选】MaxCompute 目前支持的分区类型包括：（ABCD）

- A. STRING
- B. TINYINT
- C. BIGINT
- D. VARCHAR

【解析】Partitioned by 指定表的 分区 字段，目前支持 TINYINT、SMALLINT、INT、BIGINT、

VARCHAR 和 STRING 类型

一、多选题

1. 阿里云 MaxCompute 中，下列对分区描述正确的是（ABC）

- A、需要在创建表时指定分区空间
- B、以将分区类比为文件系统下的目录。
- C、MaxCompute 将分区列的每个值作为一个分区
- D、目前仅支持 2 级分区

【解析】最高支持 6 级

二、判断题

1. 在阿里云 MaxCompute 中,可以在插入数据之前为一张非分区表添加分区，并插入数据。

A √

B ×

2. 在 MaxCompute 中，表中的某一列可以同时作为普通列和分区列。

A √

B ×

3. 在 MaxCompute 中使用数据时，如果指定需要访问的分区名称，仍有可能进行全表扫描。

A √

B ×