**利用华为云完成综合案例的注意事项，以综合案例1为例：**

任务类型（应用场景）：离线（批处理），实时（流式计算），即席查询（例如Kylin）

**软件资源**：

华为云：MRS服务，DLV服务，1-2个弹性公网IP

本地：SSH客户端，Navicat

环境准备工作：

1. 购买华为云的MRS服务（注意选择批处理）和DLV服务（DLV服务可以免费试用1个月），。
2. 购买2个弹性公网IP，其中一个绑定到master1结点，另一个绑定到MRS的CloudManager，理论上用命令行使用Sqoop完成ETL可不使用CloudManager。
3. 使用SSH客户端测试HDFS和HIVE能否正常工作，通过SSH客户端将数据复制传到master1节点，并上传到HDFS。
4. 购买MySQL云数据库（建议5.X版本），并在华为主机的访问控制策略中的入站规则中打开3306端口（可能需要），此处配置好后可以利用本地的Navicat等工具远程测试连接华为的MySQL，此处可利用Navicat在MySQL中建库建表用于测试。
5. 关于Loader的测试，Loader底层封装了开源的Sqoop，用于完成数据ETL操作。因为华为云MRS版本的迭代较快，因此可参考华为离线实验指导书的Loader配置方法，更新Sqoop所需Mysql的JDBC驱动，注意此处要使用chown命令给驱动文件授权，例如，chown omm:wheel mysql-connector-java-5.0.27.jar，。注意master1和master2上都需要执行这个操作。
6. 可在HIVE中建库建表用于测试，注意建表时的字段分隔符，否则查询时会出现NULL值。
7. 在Hue中使用Loader，建立link和job用于测试。
8. 关于DLV的使用，建立数据连接时一定要注意连接类型，如果是自己在master节点上安装mysql，则只能选“mysql”，如果直接购买华与的Mysql服务，同样选择“mysql”，若选择“云数据库Mysql”需要另行购买华为的CDM服务（用于数据迁移）。

**购买华为云MRS服务时的必选项**

注意：默认只安装Hadoop，因此必须勾选Spark、Hive和Loader。如果不勾选Loader，无法使用MRS自带的Loader（sqoop）在hive和MySQL之间导数据。 如果勾选所有组件，第一次初始化集群会耗费大量时间，另外也会对集群性能造成一定影响。

**下述标注红圈的为必选项：**

