1. 参数解析，封装命令行参数，覆盖同名默认配置项。
2. distcp默认基于快照进行（可以在启动时追加-snapshot false指定不基于快照的方式进行distcp）。提取本次所有需要同步的数据表，解析数据表location信息，形成源集群目录集合A，根据每张表的内外部属性和源集群目标集群版本信息将对A集合进行转换形成B集合。
3. 获得源集群和目标集群中所有可快照目录集合，将上一步获得的A集合和B集合进行判断，如果所有目录都在可快照目录中则提取本次需要打快照的目录，如果存在不可快照目录则给出提示并退出。
4. 对源和目标集群分别建立快照。
5. 启动线程池并发处理处理表，每个线程获得一张数据表后获得表的属性信息和是否为分区表，判断目标集群是否存在该表，不存在的话会在目标集群创建该表并替换对应的表级location信息，如果是分区表会继续创建对应的分区并替换分区location；目标集群存在该表的话则会进行元数据比对，针对表级别和分区级别元数据会逐一比对，并alter到目标集群。
6. CDH5和CDH6的内部表同步到CDP7后依然是内部表，存储在/warehouse/tablespace/managed/hive，外部表同步到CDP7后依然是外部表，没指定非默认路径的则存储在/warehouse/tablespace/external/hive，如果指定了路径则保持源目录路径。
7. 元数据同步完成后会进行主数据的同步，默认采用的是基于快照的distcp，相当于采用的是源集群中的快照文件同步到目标集群的落地目录，防止在同步过程中源文件发生变更导致同步失败。
8. 所有表处理完成后线程池销毁，回到主线程，统计本次数据同步匹配到的数据表总数、元数据发生变化的表数量、元数据未发生变化的表数量、数据同步成功的表数量、数据同步失败的表数量、主任务执行结果。
9. 对所有同步成功的数据表进行数据一致性校验，检测同步到目标集群的数据文件和源集群是否一致。
10. 清理源集群中本次产生的快照文件，对其进行删除，目标集群快照则继续保留。