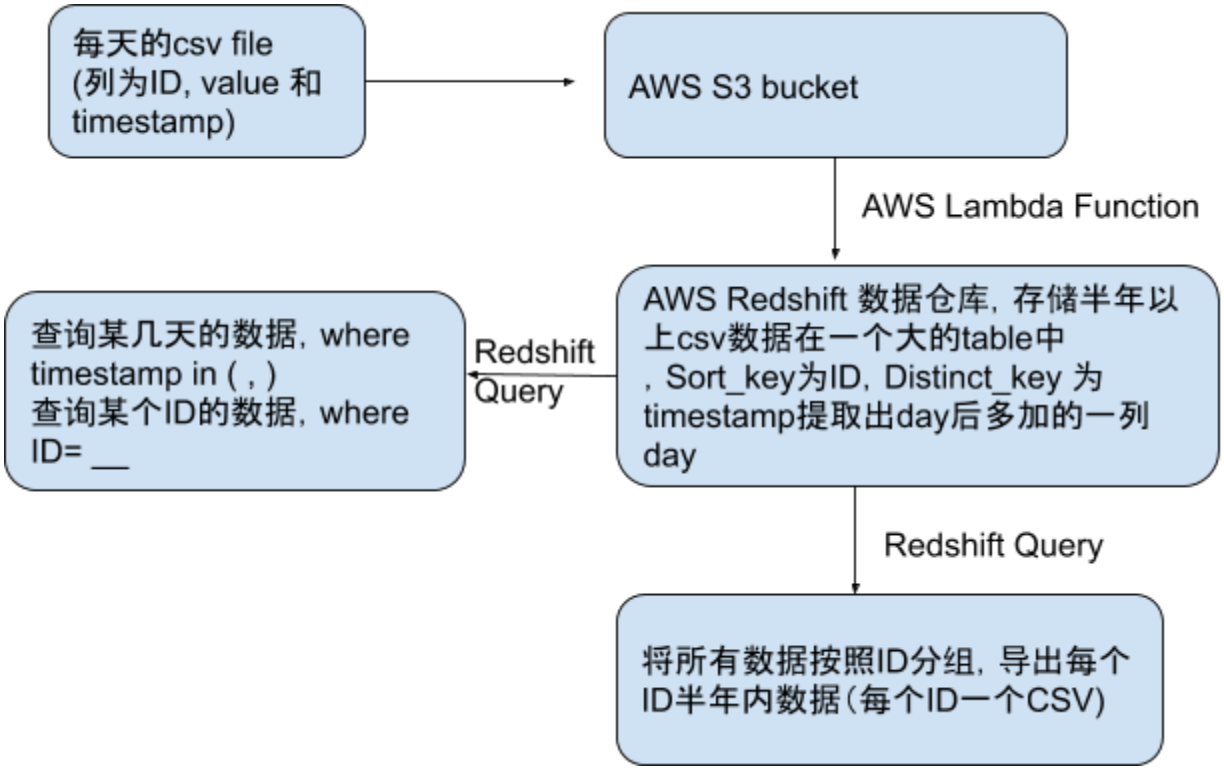
第一步：每天的数据存储到Redshift数据仓库中
Redshift数据仓库是AWS新研发出的一种数据仓库模式，可以同时有几个Work Node一起进行。有Query功能。



如果只用AWS Redshift query： select (*) from table where ID=__ 的方式，想要每天导出近30000个ID的半年数据（意味着做30000次where = query）几乎不可能实现。
所以想要将Redshift中的全部data根据ID partition后生成做30000个csv，只能采用另一个方式

第二步：把Redshift中半年的数据（40GB）根据ID进行分组。提取出每个ID半年内的数据并生成CSV。