益海嘉里生产库数据归档方案

|  |  |
| --- | --- |
| 编制日期 | 2018-10-09 |
| 编制人 | 马飞 |
| 版本 | V1.0 |

1. 概述

为了确保益海嘉里生产环境运行稳定快速，建议将线上不再频繁查询的表进行归档至历史库进行查询操作。生产环境只保留最近几个月的数据。

1. 迁移准备

## 2.1归档表统计

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 数据库名 | 表名 | 表大小(MB) | 索引大小(MB) | 保留期限 |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

说明：请将需要归档的表填至上表中。建议只保留3个月的数据，将3个月外的数据进行归档。

## 2.2 将迁移表改造为分区表

申请维护窗口进行操作，步骤如下：

* + - 申请操作窗口，时间长短与表大小有关。
    - 确定分区键，表中无created的需要增加该列，类型为datetime。
    - 建立分区表，结构同原表，主键采用（id+created）。
    - 分区键上创建索引。
    - 将待归档表数据插入分区表。
    - 将原来改名并将分区表改名为原表名。
    - 核对数据无误后，删除原表。

**说明：**以上过程每张表只初始化一次，将普通表改造为分区表后，可以通过人工或自动化方式进行数据归档。

## 2.3申请停止应用操作窗口

每一张待归档的表改为分区表均需要停止应用，时间长短与表大小有关，需要根据表大小具体评估。如果停止应用窗口较短，只能一张一张表地进行操作。

1. 数据迁移

## 3.1 历史库配置

（1）硬件配置

根据历史库访问方式来访问频率来制定，下面只是假设历史库只是执行一些简单的查询且访问不频繁的情况下给出的配置：

|  |  |
| --- | --- |
| CPU | 8核 |
| 内存 | 32G |
| 磁盘 | 1.2T，支持动态扩容 |

（2）软件配置

软件配置与生产环境一致，数据库版本与生产环境一致或高即可：

|  |  |
| --- | --- |
| 操作系统 | CentOS6或CentOS7云主机或RDS |
| 数据库 | MySQL5.7.17以上 |

## 3.2 归档检查

表在归档前需要核对该表是否已经迁移至历史库中，归档过哪些分区。

## 3.3 导出分区

根据表的保留期计出出要从生产环境中备份哪些分区还原至历史库中。

## 3.4 导入分区

将导出的表分区数据导入至历史库中。

## 3.5 数据核对

核对生产环境中归档分区与历史库是否一致。

1. 自动化

## 4.1 将上述步骤程序化

为了减少表归档过程中人工核对时间过长，容易出错的情况及问题排查，需要将上述过程做成程序，方便进行配置，提高效率。

## 4.2 每月定时进行迁移

设置定时任务，在指定时间运行数据归档任务。

1. 历史数据查询

## 5.1 历史库访问方式

（1）通过JDBC连接，使用程序方式查询。

提供IP、端口、数据库名、字符集、用户名口令信息。程序通过JDBC连接数据库并通过程序展示数据。

（2）通过图形工具连接查询数据

通过图形工具直接，编写SQL语句查询数据。通过图形化工具导出csv，excel等文件。