

# KuaiLeShiFu

HOME

CONTACT

GALLERY

## 数据同步：全量与增量

2018-03-07 15:52 by 乘着风去破浪, 3222 阅读, 0 评论, 收藏, 编辑

### 1. 背景

数据如果保留多份，就会存在一致性问题，就需要同步，同步分为两大类：全量和增量

### 2. 概述

数据如果要保留副本，要么同时写（就是多写），或者进行复制：异步写（即从主数据拷贝到副本）；

同时写（多写），引出一个问题，写多少节点算成功（场景：分布式系统）？全部写成功才算成功，还是写大多数成功算成功，还是写指定几个节点算成功？

异步写的话，如果采用异步复制，那么实时性需要考量的话，就需要采用性能优先的架构。

### 3. 同步方式

数据同步一般分为两种方式：全量和增量。

#### 3.1 全量

全量，这个很好理解。就是每天定时（避开业务高峰期）或者周期性全量把数据从一个地方拷贝到另外一个地方；

全量的话，可以采用直接全部覆盖（使用“新”数据覆盖“旧”数据）；或者走更新逻辑（覆盖前判断下，如果新旧不一致，就更新）；

这里面有一个隐藏的问题：如果采用异步写，主数据物理删除了，怎么直接通过全量数据同步？这就需要借助一些中间操作日志文件，或者其他手段，把这些“看不到”的数据记录起来。

#### 3.2 增量

增量的基础是全量，就是你要使用某种方式先把全量数据拷贝过来，然后再采用增量方式同步更新。

增量的话，就是指抓取某个时刻（更新时间）或者检查点（checkpoint）以后的数据来同步，不是无规律的全量同步。这里引入一个关键性的前提：副本一端要记录或者知道（通过查询更新日志或者订阅更新）哪些更新了。

##### 3.2.1 确定更新点

采用更新时间戳、有的采用checkpoint等来标识和记录更新点。

好文要顶

关注我

收藏该文

乘着风去破浪

关注 - 1

粉丝 - 0

+加关注

» 下一篇：建模：通过ES平铺关系型数据库多表的数据

About

昵称：[乘着风去破浪](#)

园龄：[1年7个月](#)

粉丝：[0](#)

关注：[1](#)

[+加关注](#)

SEARCH

#### 最新评论

Re:elasticsearch6.0.0源码导入到idea总结  
我在core目录下执行gradle idea时总是报错 -- 夜无痕星

日历							随笔档案						
< 2018年10月 >							2018年10月(12)						
日	一	二	三	四	五	六	2018年9月(22)						
30	1	2	3	4	5	6	2018年7月(4)						
7	8	<a href="#">9</a>	10	11	<a href="#">12</a>	<a href="#">13</a>	2018年6月(5)						
14	15	16	17	<a href="#">18</a>	<a href="#">19</a>	20	2018年5月(3)						
21	<a href="#">22</a>	23	<a href="#">24</a>	<a href="#">25</a>	26	27	2018年4月(1)						
28	29	<a href="#">30</a>	31	1	2	3	2018年3月(5)						
4	5	6	7	8	9	10							

我的标签
<a href="#">Elasticsearch(1)</a>
<a href="#">gradle(1)</a>
<a href="#">netty源码分析(1)</a>
<a href="#">开发经验(1)</a>
<a href="#">数据(1)</a>
<a href="#">源码(1)</a>

推荐排行榜
-------

阅读排行榜
-------

1. 数据同步：全量与增量(3222)
2. elasticsearch6.0.0源码导入到idea总结(340)
3. 完整后端开发流程(229)
4. 建模：通过ES平铺关系型数据库多表的数据(168)
5. 如何编写高效的代码和优化代码(57)