# B3log 数据备份与恢复

版本：0.2.1, Dec 16, 2012

作者：[88250](http://88250.b3log.org)

[B3log 数据备份与恢复](#h.4w7w4kenk9dm)

[客户端](#h.nu72txqcn4fx)

[服务端](#h.ug3iyoy1drcm)

[关键点](#h.6jj06r12dtkv)

[备份中断](#h.5cv5638vcznf)

[恢复中断](#h.eg0l0s2ou20d)

[数据写入禁用](#h.jnwga3cjo2vk)

[使用指南](#h.bhhux081ssc4)

[参数说明](#h.iyw7pw7bxzgr)

[注意事项](#h.ss7ls0ahovt8)

[备份（导出）](#h.3f9ocljfn6ra)

[恢复（导入）](#h.v272t85139ai)

[其他](#h.5oxlqx58eel0)

数据备份/恢复主要用于数据导出/导入，典型的应用场景是应用平台切换（比如需要从 GAE 切换到 BAE）。

在 Latke 框架层面提供 HTTP 接口用于数据备份/恢复，供客户端（桌面程序）调用。

## 客户端

**备份**：调用服务端接口获取数据，并按照库名，分页参数（页号、页大小）写文件。例如：

/backup/

- article/

- 1\_10\_20120409103031.json

- 2\_10\_20120409103032.json

- ....

- archive/

- 1\_10\_20120409103033.json

- ....

- 20\_10\_20120409103040.json

/backup/ 是备份根目录，article/ 目录下面保存所有 article 库的数据，1\_10\_20120409103031.json 保存第 1 页（每页 10 条）数据，保存时间为 2012 年 4 月 9 日 10 时 30 分 31 秒。

对于同一个库，只进行单线程请求。

**恢复**：调用服务端接口保存数据，按照备份的分页文件提交。提交后修改文件名（加入提交时间），标识该文件已经恢复成功。例如：

/backup/

- article/

- 1\_10\_20120409103031\_20120410223000.json

- 2\_10\_20120409103032.json

- ....

- archive/

- 1\_10\_20120409103033\_20120410223130.json

- ....

- 20\_10\_20120409103040.json

## 服务端

默认情况下 Latke 数据访问接口为*禁用状态*。当类路径（WEB-INF/classes）下存在配置文件 remote.properties 时，数据访问接口才生效。

**remote.properties**

示例如下：

repositoryAccessor.userName=xxx

repositoryAccessor.password=xxx

配置分别描述了数据存取接口调用的用户名与密码。

处于禁用状态时，调用数据访问接口将返回 HTTP 状态码 501（NOT\_IMPLEMENTED）。

**获取存储写入状态接口**：getRepositoriesWritable()，获取所有存储是否可写入。

HTTP GET: /latke/remote/repositories/writable?***userName****=xxx&****password****=xxx*

返回值 JSONObject：

{

“sc”: 200,

“writable”: true,

“msg”: “”

}

**存储写入禁用/启用接口**：setRepositoriesWritable(boolean)，设置所有存储是否可写入。存储写入禁止只是针对其他对数据操作的功能，即使在不可写状态，备份/恢复接口仍然可以调用数据存储操作接口操作数据。

HTTP GET: /latke/remote/repositories/writable?***userName****=xxx&****password****=xxx&****writable****=true*

返回值 JSONObject：

{

“sc”: 200,

“msg”: “”

}

**存储名获取接口**：JSONObject getRepositoryNames()，获取存储名（即表名）。

HTTP GET：/latke/remote/repository/names

返回值 JSONObject：

{

“repositoryNames”: [

“”, “”, ….

]

}

**数据获取接口**：Result getData(userName, password, repositoryName, pageNum, pageSize)，提供分页查询。

HTTP GET：/latke/remote/repository/data?***userName****=xxx&****password****=xxx&****repositoryName****=xxx&****pageNum****=xxx&****pageSize****=xxx*

Result：

{

"sc":200,

"msg":"Got data",

"pagination":{

"paginationPageCount":11

},

"rslts":[{}, {}, ….]

}

**数据保存接口**：putData(userName, password, repositoryName, data)，提供数据写入。

HTTP POST：/latke/remote/repository/data?***userName****=xxx&****password****=xxx&****repositoryName****=xxx*

Data Body：

{

"data": [{}, {}, ....]

}

## 关键点

### 备份中断

每次请求获取数据前需要检查是否和上次请求的页连续（页号是否连续、页大小是否一致），这样如果某次请求失败的话，可以在下次请求时继续进行备份。

备份中断检查：检查当前页号、页大小是否存在前叙。比如当前准备调用获取接口获取第 3 页数据，页大小为 20。需要检查本地对应的库目录是否有名为 2\_20.json 的前叙。

### 恢复中断

选取没有提交时间段文件名的文件进行恢复提交，提交成功后加上时间标识。

### 数据写入禁用

在进行备份和恢复时必须禁止数据写入，避免备份数据不一致。

因为备份/恢复存在中断可能，所以禁止/恢复写入控制由用户指定。

在数据操作前查看方法调用栈，如果发现是恢复程序在操作数据，则忽略写入禁用标识。

## 使用指南

客户端备份/恢复工具是一个 Java 命令行程序，使用前请仔细阅读本说明。

### 参数说明

|  |  |
| --- | --- |
| -u <username> | 必须，用户名（参考 remote.properties） |
| -p <password> | 必须，密码（参考 remote.properties） |
| -s <server> | 必须，服务器地址，例如：localhost:8080 |
| -backup\_dir <backup\_dir> | 必须，数据备份（导出）目录 |
| -create\_tables | 数据库建表（非 GAE 环境首次恢复数据时必须设置） |
| -backup | 备份（导出）数据 |
| -restore | 恢复（导入）数据 |
| -repository\_names | 列出服务器上所有的数据存储名，并在指定的备份目录（backup\_dir）下分别创建对应存储名的目录 |
| -w <writable> | 配置数据存储是否可写，可选实参 true/false |
| -h | 打印帮助信息 |
| -v | 打印版本信息 |
| -verbose | 打印执行过程 |

### 注意事项

#### 备份（导出）

1. 在备份（指定 -backup 参数）之前，请禁用数据存储写入：-w false。整个备份结束后再打开写入开关：-w true
2. 备份过程中如果遇到非客户端序问题引起的中断（例如网络异常、服务端异常），可继续运行 -backup 进行备份，这样可以继续上次的备份
3. 如果要重新开始备份，请清空备份目录
4. 在备份期间（包括中断），不能进行恢复操作

#### 恢复（导入）

1. 在恢复（指定 -restore 参数）之前，请*禁用*数据存储写入：-w false。禁用操作针对的是应用调用，避免在恢复期间应用写入数据
2. 恢复过程中如果遇到非客户端序问题引起的中断（例如网络异常、服务端异常），可继续运行 -restore 进行恢复，这样可以继续上次的备份
3. 在恢复期间（包括中断）不能进行备份操作

#### 其他

1. GAE 等云环境请注意配额使用。如果一天的免费配额不能够完成整个备份/恢复，则可以等待配额重置后继续备份/恢复
2. 非 GAE 环境首次恢复数据时必须设置 -create\_tables 选项，用于数据库建表
3. 备份/恢复只能单线程进行，不能同时运行多个备份/恢复程序