

MongoDB GridFS 文件系统

李焕贞

河北师范大学软件学院

<https://docs.mongodb.com/manual/core/gridfs/>



本章大纲

- 分布式文件系统简介
- GridFS 存储介绍
- GridFS 使用演示

分布式文件系统简介



文件系统 (File System) 是操作系统用来组织磁盘文件的方法和数据结构，计算机通过文件系统来存储管理数据。

典型的系统包括UFS (Unix File System), EXT (Extended file system) 以及NTFS (New Technology File System) 等。

分布式文件系统简介



分布式文件系统，将固定存储于单台机器上的文件系统，扩展到多台机器上，每个节点负责存储部分数据，众多的存储节点组成一个文件系统网络，各个节点之间通过网络进行通信和数据传输。

分布式文件系统简介

分布式系统	特 点
NFS (Network File System)	相当于RPC Service,使用者可以像使用本地文件系统一样来访问远程的文件
HDFS (Hadoop File System)	采用master/slave架构设计,可以运行于廉价的普通硬件之上,适合超大数据集的应用程序
TFS (Taobao File System)	适合海量非结构化的小文件存储,采用HA架构;文件名可以直接映射到物理地址
FastDFS	轻量级的分布式文件系统;采用对等架构设计;可以根据文件ID迅速定位文件



本章大纲

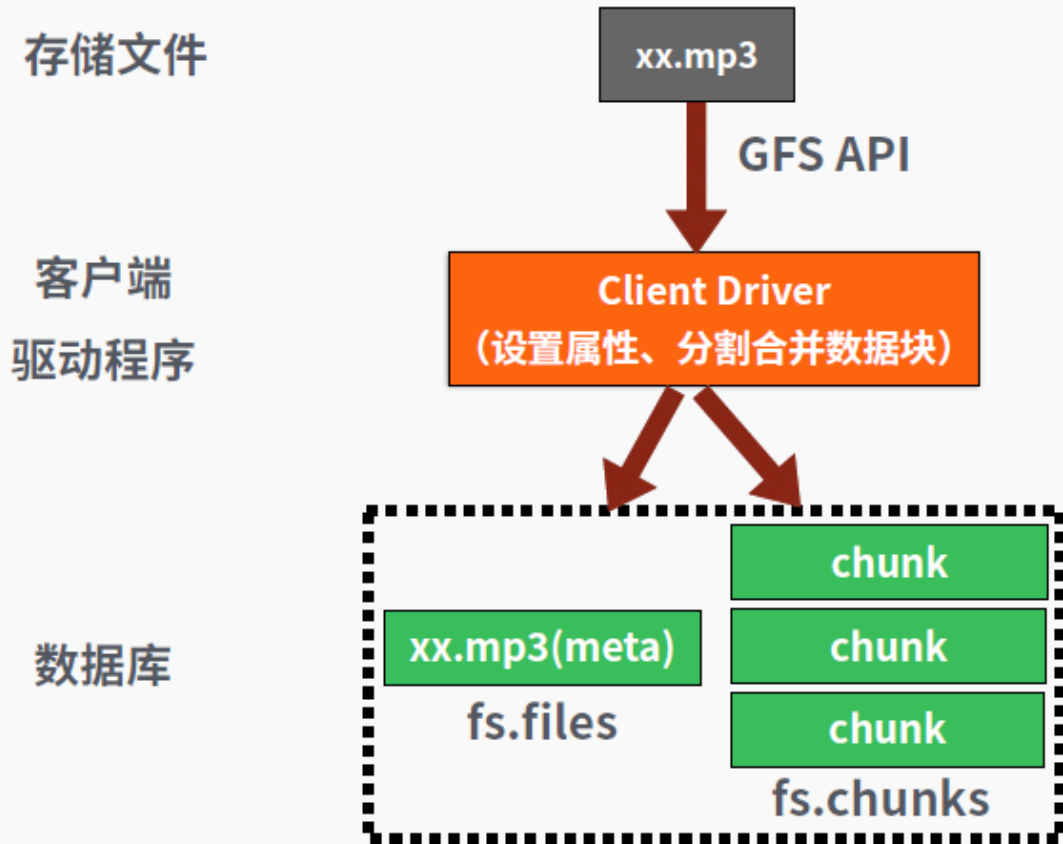
- 分布式文件系统简介
- GridFS 存储介绍
- GridFS 使用演示

GridFS存储介绍-简介

GridFS是MongoDB提供的**二进制数据**的存储解决方案，是专门为**大数据文件存储**提供的存储方案。

GridFS是MongoDB制定的，如何在数据库中存储大文件的（specification）规范。MongoDB数据库本身并没有实现GridFS的任何代码，具体代码实现交由客户端驱动程序来完成。

GridFS存储介绍-存储管理



GridFS存储介绍-存储原理

GridFS 用两个集合来存储一个文件：**fs.files**与**fs.chunks**。
默认情况下，使用fs.files存储文件的元数据，包含主要字段如下：

```
{  
  "_id": <ObjectId>,"length":<num>,  
  "chunkSize":<num>,"uploadDate":<timestamp>,  
  "md5": <hash>,"filename": <string>,  
  "contentType":<string>,"aliases": <string array>,  
  "metadata":<dataObject>  
}
```

GridFS存储介绍-存储原理

默认情况下，使用fs.chunks存储以255K进行分割的文件块，包含字段如下：

```
{  
  "_id": <ObjectId>,  
  "files_id": <ObjectId>,  
  "n": <num>,  
  "data": <binary>  
}
```

GridFS存储介绍-适用场景

GridFS主要适用于如下场景：

- 存储用户产生的文件且文件数量比较大
- 需要文件的自动备份和分布式存储
- 需要访问文件的部分内容
- 存储16MB以上的文件



本章大纲

- 分布式文件系统简介
- GridFS 存储介绍
- GridFS 使用演示

■ GridFS使用演示-mongofiles

`mongofiles`是MongoDB安装包中提供的一个工具，它专门用来存储文件。

`mongofiles`支持的操作包括查询所有文件、查看文件、上传一个文件、获取一个文件、按照`_id`来获取文件、按照文件名删除文件、按照`_id`来删除文件。

GridFS使用演示

上传文件: `mongofiles -d gfs put d:/demo/baidu.jpg`

```
C:\Users\think>mongofiles -d gfs put d:/demo/baidu.jpg
2019-03-07T11:11:20.689+0800    connected to: localhost
2019-03-07T11:11:20.840+0800    added file: d:/demo/baidu.jpg
```

显示文件列表: `mongofiles -d gfs list`

```
C:\Users\think>mongofiles -d gfs list
2019-03-13T10:05:48.910+0800    connected to: localhost
d:/demo/baidu.jpg             705
```

下载文件: `mongofiles -d gfs get`

```
C:\Users\think>mongofiles -d gfs get d:/demo/1122.png
2019-03-13T10:12:22.530+0800    connected to: localhost
2019-03-13T10:12:22.787+0800    finished writing to d:/demo/1122.png
```

删除 `delete` 搜索 `search`

GridFS使用演示

db.fs.files.find()

```
> use gfs
switched to db gfs
> db.fs.files.find()
{ "_id" : ObjectId("5c808bd8ff7c632fac9db59a"), "chunkSize" : 261120, "uploadDate" : ISODate("2019-03-07T03:11:21.049Z"), "length" : 705, "md5" : "803bb46a6acef395ed9353de2dcf26f5", "filename" : "d:/demo/baidu.jpg" }
>
```

db.fs.chunks.find()

```
> db.fs.chunks.find()
{ "_id" : ObjectId("5c808bd8ff7c632fac9db59b"), "files_id" : ObjectId("5c808bd8ff7c632fac9db59a"), "n" : 0, "data" : BinData(0,"R01G0D1hdQAMAKIAA0YyL+rU41lg6Jmd8e92dCky4eEgAv///yH5BAAAAAALAAAAAB1ACYAAAP/eLrc/jC2IEoZMATJu/9gyFUWlUyksIls677LUJbrEcXWDe98f+CWk4I0w/iOSNANKJQBC8mo9LEEDp8F3XR7r0IU2Cx3jHwKsUKyWqRhEEvGN3xN91BoCq819tTW/244Rk4m0kBGgI18UjF+d4U5A5KKf3IWiCCE0ZRraGxPnGqeIZpzoUyjdBptDpYmp1yumI9Bwq5QUQS6fn+1m31YmLdSBsW8xcjJBgAABLWfAMhXRQt60DWuJ8rb3AYK0d3h40HIBC7jC0TCYb/ZB+Th3/Dc4/PmLOgSukwK+xjMAAEmCwhQnrc8rFSxGkiwnoFnEnTp0mdq1Jw0DgIggwgu/08xABM+KkhGJdm9JJqMhGkyQIAATA3HxUwATu08jyQfECiXJyYzZwzy0SilcqWEmzkP1ES6LGmDcStshjspNJAgM0wwNmBa
```

■ GridFS使用演示-客户端驱动程序

MongoDB提供了丰富的各种语言的客户端驱动程序，官方支持的包括C、C++、C#、Java、Node.js、PHP以及Python等十几种。除此之外，还包括社区支持的Go（mgo）以及Erlang等。

<http://mongodb.github.io/mongo-java-driver/3.10/driver/tutorials/gridfs/>