

In this lecture, we will discuss...

- ✧ GridFS Introduction
- ✧ GridFS Architecture
- ✧ GridFS Collections
- ✧ Metadata

GridFS- Introduction

- ✧ **GridFS** - specification for **storing** and **retrieving** large files
 - images, audio files, video files, etc.
- ✧ File system to store the **chunks**
- ✧ Data is stored within MongoDB **collections**
- ✧ File are broken in to multiple chunks and a metadata file
 - stores each chunk of data in a **separate document**, each of maximum size **255KB**

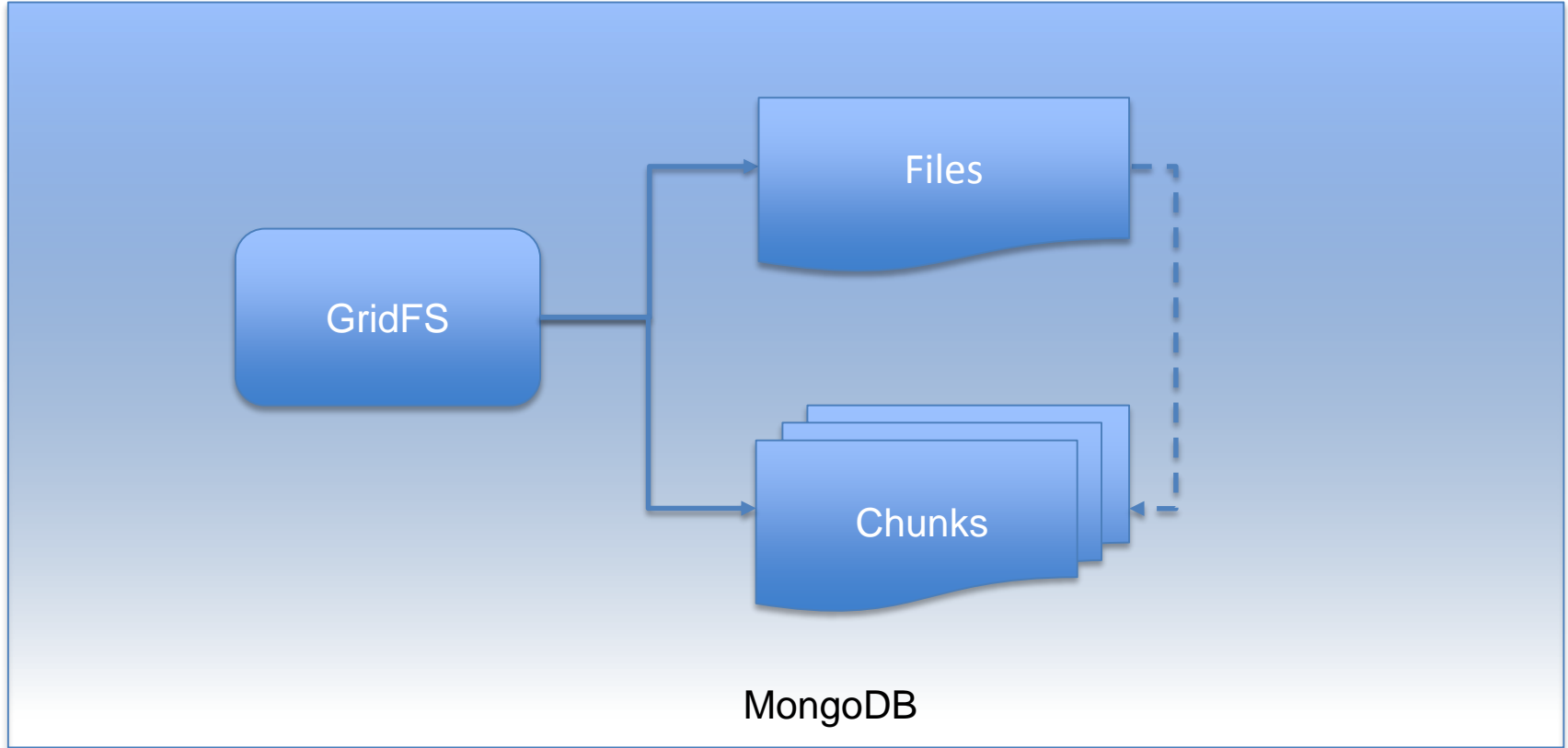


GridFS- Introduction

- ✧ 10 MB document of query-able data with normalized types as embedded – *embedded*
- ✧ 18 MB document of query-able data with normalized types as embedded – *embedded/linked*
- ✧ 10 MB document of image blob plus query-able data – store as *regular* document
- ✧ 18 MB document of mostly an image blob and a small amount of query-able data - *GridFS*



GridFS – Architecture



GridFS- Collections

- ✧ **GridFS** – uses two collections
 - fs.files - files's metadata
 - fs.chunks – file chunks
- ✧ GridFS will **reassemble** the contents when you access the chunks.



GridFS- chunks field

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

[source: mongodb.org](http://source.mongodb.org)

_id	The unique Object ID of the chunk.
files_id	The _id of the “parent” document, as specified in the files collection.
n	The sequence number of the chunk. GridFS numbers all chunks, starting with 0.
data	The chunk’s payload as a BSON binary type.



GridFS - files metadata

```
{
  "_id" : <ObjectId>,
  "length" : <num>,
  "chunkSize" : <num>,
  "uploadDate" : <timestamp>,
  "md5" : <hash>,

  "filename" : <string>,
  "contentType" : <string>,
  "aliases" : <string array>,
  "metadata" : <dataObject>,
}
```

source: mongodb.org



GridFS - files metadata

_id	The unique Object ID of the document.
length	Size of the documents in bytes
chunkSize	The size of each chunk. Default size is 255KB
uploadDate	The date the document was first stored by GridFS.
md5	MD5 hash
filename	A human-readable name for the document.
contentType	A valid MIME type for the document.
aliases	An array of alias string
metadata	Any additional information you want to store.



Summary

- ✧ GridFS - storing large binary files by breaking them into smaller “chunks”

What's Next?

- ✧ GridFS Demo
- ✧ Geospatial Indexes and Queries

