

# What is the best way to store an actual document file?

 [mongodb.com/community/forums/t/what-is-the-best-way-to-store-an-actual-document-file/111714](https://mongodb.com/community/forums/t/what-is-the-best-way-to-store-an-actual-document-file/111714)

June 18, 2021



A

AbhiAbhijith Nair

Jun '21

Hello all,

I am trying to build a system which requires storing an actual document file in the database along with other details. For example, assume I need a table where I store the candidate information like name, email, phone etc, and resume file (doc, docx, pdf) who are applying for a job I posted online.

I don't want to use amazon S3 to store the document and use the link in the database. I'm building an application which requires low latency. So having the resume file content , alongside the candidate data helps.

In my use-case I don't actually require the resume file to be frank, rather I need the text inside the resume as a string. So, which is the most efficient way, should I extract the text from resume document file and store it in mangoDB as text (or) store the resume file as binary format. Which approach is more scalable?

The only reason I'm choosing to store text is, if I store the resume file itself I have to fetch and extract text from the document using some document parser which further increases my latency. On the other hand I'm thinking that storing such large texts from resume is scalable or not?

I'm new to mongoDB. Appreciate any help or suggestions. Note: The resumes document files are less than 1mb.

Solved by [Prasad Saya](#) in [post #6](#)

The data is stored in the document in its original size. The data and the indexes in the database are finally stored as data files. MongoDB uses a storage engine which manages this, and the presently it is the Wired Tiger Storage Engine; see [Wired Tiger Storage Engine - Compression](#). See [MongoDB Sto...](#)

P

[Prasad Saya](#) Database Rebel

[Jun '21](#)

Hello [@Abhi](#), welcome to the MongoDB Community forum!

MongoDB data in a document can be upto maximum size of 16 Megabytes. A resume's text is likely to be few (less than 100) kilobytes only, if stored as string (or text) data type. In case you are storing the resume document as a PDF or DOCX file, it is still going to be about few hundred kilobytes. So, you can store the resume as text with "String" data type and the PDF/DOCX as "Binary data" within the document itself, along with other data like name, address, phone, email, etc. See [MongoDB BSON Types](#) [814](#).

There is also an option store large document files, *files greater than 16 MB size*, in MongoDB GridFS (see this post for more details [Process of storing images in MongoDB](#) [524](#)).

A

[AbhiAbhijith Nair](#)

[Jun '21](#)

Hello [@Prasad Saya](#),

Thank you for the information. I just have one more doubt, Since there's not much different on how I store the document. I also need a way to lookup/search basic candidates details like name, phone, email, title etc.,

If I query the database directly from the UI/Frontend for these columns, will the latency be a problem, or do I need to index these columns in order to make the search faster. How fast is mongoDB? Or do I need to choose elasticsearch for these?

P

[Prasad Saya](#) Database Rebel

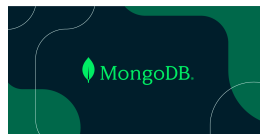
[Jun '21](#)

@Abhi , when you query from a browser - for example you enter some search criteria for name field - the search string is passed to the database query in the application server program (this is generally a Java, NodeJS, Python, etc., program and uses appropriate MongoDB Driver software). The program sends the query to the database server where it is executed and the results are returned to the application program which in turn is sent to the client (the browser, in this case).

The query on the database will run based upon the search criteria; for example, `db.usersCollection.find( { name: input_name } )`. For the query to run efficiently and in a performant way, an index can be defined on the collection's `name` field. Indexed searches are fast.

Reference:

 [mongodb.com](https://mongodb.com)



## FAQ: Indexes — MongoDB Manual 13

---

A

AbhiAbhijith Nair

Jun '21

@Prasad\_Saya, Since you mentioned the resumes is around few hundred kilobytes, when I store it in MongoDB is it compressed or stores it in original file size, We also want to calculate the cost and which server to choose, in order to do that we need to know how much disk space will it take when I store them.

Let's say I have around 100 million resumes or something, if each file takes around 100-200 kilobytes the amount of disk space is huge right?. I know that it converts JSON document to BSON, does it also compress my size? Is there a way to calculate that?

For example, If I store the resume text in Elasticsearch for each document, it's taking around 15-20 kilobytes only.

P

Prasad\_SayaDatabase Rebel

Jun '21

A Abhi:

| when I store it in MongoDB is it compressed or stores it in original file size,

The data is stored in the document in its original size. The data and the indexes in the database are finally stored as data files. MongoDB uses a storage engine which manages this, and the presently it is the Wired Tiger Storage Engine; see [Wired Tiger Storage Engine - Compression](#) [61]. See [MongoDB Storage - FAQs](#) [55] for how to find the size of your collection size, and other details (specifically see the section **Data Storage Diagnostics**).

Solution



Closed on Jun 23, '21

This topic was automatically closed 5 days after the last reply. New replies are no longer allowed.