**Contents Created for the  Chunk\_File Project**

***Team\_46***

The following contents were created for the different sections of the Bytsfy web app and the names of each individual who contributed and their contributions is provided below.

**Slack Username: Red**

**Github Username: Red\_Yinka**

**Name: IDOWU Olayinka**

CONTENTS IN THE FAQ SECTION

* Can I save the processed file on the platform for download later?

Yes. We keep your processed files in our servers except you delete them yourselves. Your files are saved on our servers so you can come back at any time to download and re-download them.

* Can I delete my files after processing?

Yes. You can prefer to delete your documents by yourself at the end of each splitting, conversion, or compression. Right after, they are completely removed forever from our servers. Or you can decide to save it for a while then come back for deletion when the file is no longer needed.

* Are my files probed by bytsfy?

Utterly not. Your files are yours and yours alone! Not in any way would we analyse, check or copy them. No one can access your files except you. Your files in our servers are strictly secured.

* Can bytsfy be accessed through any web browser?

Bytsfy can be accessed through Chrome, Firefox, Explorer +10, UC and Safari. To run our tools smoothly, we recommend that you work with the browsers listed. our system requirements are basic and if at any point you experience issues, we recommend you contact us.

* How do I upload my files?

You can click on the “Upload file” button or drag and drop your file from your computer.

* Are my files safe and secure with Bytsfy?

Yes. Your data is processed as safely as possible. For more privacy, our service includes end-to-end encryption for all file uploads. Our compliance is in line with one of the most non indulgent safety standards in the world(GDPR), EU Data Protection Regulation. These steps ensure security, safety and satisfies most of the corporate data privacy policies.

* Is there a limit to the size of files I upload?

200MB. For better performance minimum data size should be up to 100bytes. And maximum 200,000kB

* How can I have access to the tool?

You can have access to all the tools of Bytsfy with just a click on register.

* How do I leave a comment or feedback?

Here it is: Kindly send us your valuable suggestions, feedback, or questions to us.

* How do I contact you when I have a question?

You can click on Contact here. You can also find the Contact link on the Navigation bar at the top of the screen.

* Why does my conversion take so long?

Unfortunately, this is not entirely in our hands. Processing speed depends on many factors, so the time that will take you to get your files converted will depend as well on your own Internet connection speed, the size of the selected files and, mostly, on how busy our servers are.

**Slack Username: Eric\_254**

**Github Username:**

**Name: GITONGA Eric Gituma**

A CONTENT IN THE RESOURCE SECTION

**CSV AND JSON FILE COMPARISON**

CSV stands for Comma Separated Value. It is a delimiting text in which a comma is used to separate the values. It allows us to store data in a tabular format and is greatly used to transfer database data between programs that work on different data formats. Every line of the file forms a record and all the lines have fields that are equal in number. The filename of a csv file has a *.csv* file extension. An example of data stored in a csv file is shown below:

*Id,name,gender,age*

*1,barack,male,13*

JSON is a short form of JavaScript Object Notation which was pioneered by Douglas Crock Ford. It is a lightweight and compact data exchange text format in which data is stored in key/value pairs.  Curly braces are used to hold an object. Each object is represented in a Key/Value pair form where the key and value are separated by a colon. The pairs are separated by a comma. The data is organized in ordered lists called arrays indicated by square brackets. Its valid data types are string, object, number, array, Boolean and Null.

The filename of a JSON file has a *.json* file extension. It is similar in syntax to the code used to create JavaScript object. This makes it possible for programs to convert JSON data to JavaScript objects and vice versa. It is easily understood by both human and machine and is used for storing and transfer of data.

An example of data stored in a JSON file containing two arrays each with two objects is:

*{“students”: [{“Id”: “1”, “name”: “barack”, “gender”: “male”, “age”: “13”},*

*{“Id”: “2”, “name”: “erica”, “gender”: “female”, “age”: “14”}]},*

*{“teachers”: [{“Id”: “1”, “name”: “mark”, “gender”: “male”, “age”: “28”},*

*{“Id”: “2”, “name”: “esther”, “gender”: “female”, “age”: “24”}]},*

JSON is more lightweight, scalable and most preferred for storing and transfer of files of big sizes compared to CSV.

An object can have several different fields in JSON. The order of these fields is no significant. However, all records in CSV file should have the same fields with the same order.

A header is not required in JSON since every object has its field name. CSV file with the same fields and layout will require the first record to be a header with the field.

JSON objects are unordered set of name/value pair. CSV file might have records of different kinds and dimensions.

JavaScript data types are used in JSON while there are no data types in CVS. Data types in CSV are identified by descriptions in other external data.

JSON can have nested structures. JSON is used on data that has log records maintained in repetitive sections. CSV doesn’t have nested structures and its core structure is two dimensional and data presented in tabular format.

**Slack Username: Timi**

**Github Username:**

**Name: Oluwamayowa Olaoye**

SOME CONTENTS IN THE RESOURCE SECTION

**WORKING WITH CSV AND JSON FILES FOR DATA SOLUTIONS**.

Modern industries today are flooded with data and as a result , there is great value in processing and analyzing the data to derive insights from it. Discovering actionable insights can birth great value in any business by stimulating creative ideas.

**Choosing Different Formats**

The data used in a model can be collected from different (external or internal) sources and saved in different file formats for processing. Your choice of data

format can greatly affect the space requirements, cost, and performance of project. There are several factors that need to be taken into account when trying to decide which data format to be used. In this article, our focus will be on the two most popular text-based file formats:CSV and JSON.

**What is a CSV format?**

A CSV (comma separated value) file is a matrix of data. Each row is an array

that represents a record, and each column represents a specific field within

that record. Each field (or element) is separated by a comma.

CSV files store data in a simple and easy to read manner. These can be opened

using pretty much any piece of software, even plain text editors like the Notepad.

**What is a JSON format?**

JSON (Java Script Object Notation) files store data based on JavaScript object index. Each object can have multiple keys/value pairs or other objects within it. Data can be saved in various data types including strings, arrays, integers, e.t.c

Unlike CSV, JSON allows for creating hierarchical structure of data. It is used for transmitting data in web/mobile applications projects as it is easy to integrate with APIs.

**When to use each format**

When deciding upon which format to use, consider asking the following questions:

USE CASE:

Who is going to use the data? Or how is it going to be used?  Some file formats are meant for general use, while others are for more specific use cases.

YOUR SYSTEM SPECIFICATIONS:

What ETL processes are you working with? What tools are you going to use for analyzing your data? Are you constrained on storage or memory?

YOUR DATA CHARACTERISTICS:

What is the type of your data? How large or small is it? How complex is your data? Is it structured or unstructured? How many columns are relevant to your use case, if not all?

Is your data evolving over time?

Now let's compare CSV and JSON formats on the considerations mentioned above.

CSV vs JSON

PROJECT TYPE:

JSON is a highly efficient data exchange format. In application projects,  JSON is used in the interaction between server and client (HTTP requests). It is lightweight and hence, does not burden the network. Due to its low memory demand, it can transfer great amounts of data pretty Quickly.

CSV files on the other hand are convenient for storing and analyzing problems

involving small datasets but are nowhere near a good option when your data is

huge and complex.

FORMAT TYPE:

Both CSV and JSON are text-based formats, hence are easier to use. They can

be read by the end-user who can also modify the file content.

These formats are usually compressed to reduce their storage footprints.

But their performance is subpar to binary file formats.

STORAGE:

CSV file format is so widely used because of its ability to store data in a simple tabular manner. However, CSVs are slow to query and difficult to store efficiently. But, they also have an excellent compression ratio.

JSON files are like a mini database for textual data JSON is a partially structured format and can be used to efficiently store tons of data that you may need to use within your project.

PARSING:

CSVs are simple in nature, hence simple to parse and split (just break on comma to

convert to column). Although there is a significant potential for data loss or data corruption if the application receiving CSV input isn't the same application that created it.

A key point to take note of is that CSV is not a standardized format. This means that a

file may be in the C SV format. This means that a file may be in CSV format, but not necessarily will be read by the CSV parser used in your project.

A JSON file is larger by default and more flexible. So, it is more technical to parse and split. JSON is loaded and parsed into memory and hence depending on the size, can take up a lot of disk space.

CSV is easier to parse than JSON and is therefore potentially faster to write.

FORMAT ORIENTATION:

Both CSV and JSON formats are row-oriented, meaning they are used when all fields

in a row need to be accessed.

Columnar formats are typically used when several columns are needed to work with but

not all. Parquet and ORC are examples of such formats.

SCHEMA EVOLUTION:

Schema evolution deals with the need to retain current data when your database structure changes with time. Unless your data is guaranteed to never change or is immutable in nature, your database design must involve handling schema updates.

You can choose CSV to create your tables if your schema evolution requires only

renaming of the columns but not reordering them. CSV does not support removing columns. Nor does it allow adding columns at the beginning or in the middle of the table.

For such operations, you can use other formats, preferably columnar.

JSON allows you to do all schema manipulations except the renaming of columns.

Takeaways

If you want to retrieve simple data as lists or a table with rows and some columns, CSVs are a good option. However, you need to remember that you lose type when exporting

data to CSV.

Use JSON format for communication. Although JSON is referred to as comparatively

better than CSV when dealing with huge data sets and in terms of scalability of files

for applications, you should avoid this format when working with big data. There are

more efficient alternatives.

More optimized data formats can be utilized to meet the needs of your data science

project in terms of splitability, compression, support, and the ability to support complex data structures. But for easy readability and faster read/write time, JSON and CSV are hands down preferable choices.

The biggest of business achievements in recent times have had data-driven approach

in some way or the other. Hence, it is important that you choose the right data storage format for your business needs.

**SHOULD I STORE DATA IN CSV OR JSON FORMAT?**

JSON is the preferred choice while working with a huge volume of data and in terms of scalability of files and application. CSV is great for working with smaller files and fewer data. JSON is said to be the current best data exchange format. It is lightweight,

compact and versatile. CSV is used mainly when a large volume of data is required to be sent and there is a problem with bandwidth.