**BigQuery - Storage Organisation**

Traditional relational databases like MySQL stores data row by row (record oriented storage). BigQuery supports several input formats — CSV, JSON, Datastore backups, AVRO and when the file imports the data it converts into internal representation.

On the other hand BigQuery uses columnar storage, where each column in the table is stored in a separate file block and the columns are stored in a single **capacitor** file.

**Capacitor**

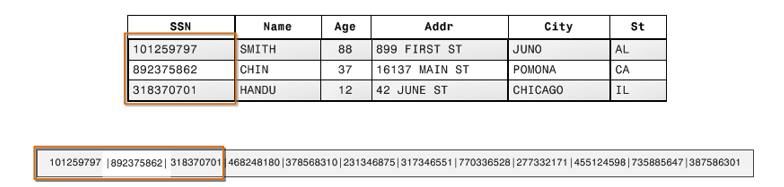
Capacitor is Google's next-generation columnar storage format. BigQuery stores data in a proprietary columnar format called Capacitor. Capacitor compresses data and allows BigQuery to operate on the compressed data on the fly without decompressing it. **Fig.1** shows a columnar storage format.

**Advantages of having columnar storage:**

* **Minimization of Traffic** - Whenever you fire a query, required column values on each query alone are scanned and passed for query execution.

Example: a query ‘select address from project\_test.test\_db.employee’ would only access the address column values.

* **Better compression ratio** - Since a similar type of data is organized together in columnar storage, it can achieve a higher compression ratio as compared to general row-based storage.



**Fig.1**

<https://cloud.google.com/bigquery/docs/saving-sharing-queries>

https://www.google.com/search?channel=trow5&client=firefox-b-d&q=table+location+%2C+where+does+the+bigQuery+stores+the+data

<https://stackoverflow.com/questions/43222120/where-the-data-will-be-stored-by-bigquery>

https://stackoverflow.com/questions/46587586/how-bigquery-use-data-stored-in-google-cloud

## **Location considerations**

We can choose a location for your data, consider the following:

1. [Cloud Storage](https://cloud.google.com/bigquery/external-data-cloud-storage): When you query data in [Cloud Storage](https://cloud.google.com/storage/docs) through a BigQuery [external table](https://cloud.google.com/bigquery/docs/external-tables), the data you query must be in the same location as your BigQuery dataset, in either a [regional](https://cloud.google.com/storage/docs/locations#location-r) or [dual-region](https://cloud.google.com/storage/docs/locations#location-dr) bucket. For example:

* Single region: If your BigQuery dataset is in the Warsaw (europe-central2) regional location, the corresponding Cloud Storage bucket must also be in the Warsaw region because there is currently no Cloud Storage dual-region location that includes Warsaw.
* Dual-region: If your BigQuery dataset is in the Tokyo (asia-northeast1) region, the corresponding Cloud Storage bucket must be a bucket in the Tokyo region or the ASIA1 dual-region (which includes Tokyo).
* Multi-region: Because external query performance depends on minimal latency and optimal network bandwidth, using multi-region dataset locations with multi-region Cloud Storage buckets is **not** recommended for external tables.

1. [Cloud Bigtable](https://cloud.google.com/bigquery/external-data-bigtable): When you query data in [Cloud Bigtable](https://cloud.google.com/bigtable/docs) through a BigQuery [external table](https://cloud.google.com/bigquery/docs/external-tables), your Cloud Bigtable instance must be in the same location as your BigQuery dataset.

* Single region: If your BigQuery dataset is in the Belgium (europe-west1) regional location, the corresponding Cloud Bigtable instance must be in the Belgium region.
* Multi-region: Because external query performance depends on minimal latency and optimal network bandwidth, using multi-region dataset locations is **not** recommended for external tables on Cloud Bigtable.

1. [Google Drive](https://cloud.google.com/bigquery/external-data-drive): Location considerations do not apply to Google Drive external data sources.