

Google Fusion Table

MILIND SURYAWANSHI^{1,2}

¹School of Informatics and Computing, Bloomington, IN 47408, U.S.A.

²Electronics and Telecommunication Engineer, Pune University, 2010

*Corresponding authors: laszewski@gmail.com

Paper-1 S17-IO-3020, April 12, 2017

Fusion Table is a web service provided by Google Inc. It is a data management tool, provides the feature like: data gathering, storing data in tables, visualizing stored data, merging tables, sharing table, view and download the tables.

© 2017 <https://creativecommons.org/licenses/>. The authors verify that the text is not plagiarized.

Keywords: Fusion Tables, data, query, I524

<https://github.com/cloudmesh/classes/blob/master/docs/source/format/report/report.pdf>

1. INTRODUCTION

Fusion Table is a “cloud Software as a Service (SaaS) application” [1]. It’s a data visualization web application, which gathers, stored, share and visualize data. User can store data in tables and can represent it in the form of different charts, maps and plots. This information can be share, merged, published by individual user or website.

Fusion table provides capacity to import tables/files up to 250MB [2] of the type .csv (Comma Separated Value), .tsv (text-delimited files), .kml (KML), spreadsheets (like: .xls, .xlsx, .ods) and google spreadsheet. Google provides quota of 1 GB per user to store tables. This 1 GB limit does not count the shared tables and tables in trash as a user quota [2]. Fusion table helps to visualize the stored data in unambiguous chart format, like: pie charts, bar charts, line plots, scatterplots, timelines and the geographical maps are provided. Fusion Tables service got launched in June 9 2009, and added to the Google Docs feature in 2011 [3]. Now they have released the v2 (version-2) of fusion table and v1 (version-1) has been deprecated from May 3rd, 2016 and be available till August 1st, 2017 [4].

Fusion tables uses HTTP request to perform the various tasks on tables. It allows users to [5]:

- create and delete tables
- read and modify the table column name and type
- insert, update and delete the rows in a table.
- Can change the access control of certain tables and its visualization.
- Fire queries on row to perform required tasks.

Data operation on table are performed through subset of SQL queries using HTTP request, also it used for retrieve the tables either in CSV or JSON file format. It uses the JSON format to make

the changes in default behavior of table structure, metadata and visualization settings [5].

2. ARCHITECTURE

Fig-1 shows the architecture of Fusion Table; we are discussing the architecture component of Fusion Table in brief. Which explains the storage stack, Bigtable and user table schema information.

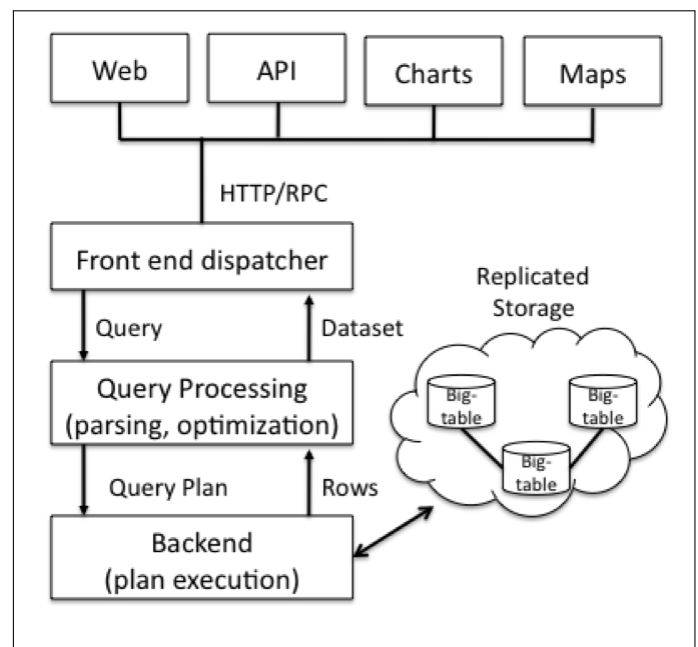


Fig. 1. Fusion Table Architecture

The service request in Fusion table is generated through Web, API, Charts and Maps. Web is a Fusion Table website, API will be from app which provides Fusion Table services, Charts to visualize the chart and to use spatial, the Maps service being used. These service request gets handled by the front end dispatcher. It will convert the above service request into required format and passes it to Query Processing. Front end dispatcher receives the dataset from Query Processing. The Query Processing convert the query into query plan. Here the query will be getting parsed and optimize to enhance the response of a query. Then Backend module's will execute the query plan received from Query Processing. Backend module will be responsible for fetching the diverse data of different table, sizes and queries from Storage. Also provides the requested (through optimize queries) data in rows [6].

3. STORAGE SYSTEM

Fusion Table provide large data storage capacity, it internally uses two-layer data storage, Bigtable and Magastore. It known as Google storage stack.

Bigtable: Bigtable stores data in tuples (it's a row, in a context of database), in key-value pair form. Its also stores the transaction timestamp history, the timestamp is a time when the tuples get written in table. Bigtable sorts the keys and stores the shard key in respective servers, according the key. Bigtable performs read and write operation. Write operation done automatically where read operation can be performed in 3 forms. Read by key prefix, read by key range or read key by exact key.

Megastore: It is "library on top of Bigtable" [6]. Megastore provide higher level operation like consistent secondary indexes, multi-row transaction and consistent replication.

Row Store: Fusion table stores all users table in to a single table called Rows, this being possible by Bigtable. Users table row gets saved in a table with key of a row with id of user's table. If a new user adds a table, then Bigtable splits the Rows table into sub-tables. Now this sub-tables are handled by separate server, which provide good performance in querying millions of tables. Bitables stored the property value of table in string format. Please see table 1.1.

| Row Key (table Id, Row Id) | Index property | Non-Index property |
|----------------------------|---------------------------------------|----------------------|
| (123, 1) | model=328i, color=red, type=sedan | notes=sells quickly |
| (123, 2) | model=330i, color=red | |
| (124, 1) | price=20, location=warehouse, UPC=500 | |
| (124, 2) | price=32, location=shelf, UPC=430 | notes=reorder needed |
| ... | ... | ... |

Fig. 2. As an example of sub-table

Observe the above table, it's a subset of Rows table. It holds the values for two different table, table 123, and 124, with its row Id e.g.1, 2. If we observe the row of table 123, with row id-1 holds the property model, color, type and notes. Where row-2 holds the value of model and color.

4. QUERY PROCESSING

As mentioned above, Fusion Table uses Bigtable and MegaStore for storage, and it useless subset of SQL, for query processing.

In Fusion table we use different kind of queries to retrieve the results, following are the few common queries:

Prefix scan: This is the query type with prefix value. E.g. "select * from 111 limit 50". Here the result rows would be with prefix value = 111 up to the count 50;

Index prefix scan: Here the scan would be performing on index and the prefix value. E.g. "select * from 111 where color = 'blue'". This query will result the rows with prefix value 111 and the color is blue.

Index range scan: The queries which provides the result in range. E.g. "select * from 111 where price 10 and price 20". The query will result in rows, which have the prefix 111 and has the price in the range of 10 to 20.

Index join: Such queries particularly used in merging the table. It either merges the small table into big one or do the index merging. Parallel query will prefix the keys (in this example A.key and B.key) and compare the matching rows, such pairs then retrieved from the rows table. Query example: "select * from A.key = B.key".

5. VISUALIZATION FEATURE

Google provide powerful collection of visualization frameworks [7]. Which provides API to embed in website to show the real time data. Where the user can integrate the visualization in the website with data, which will show the refrection of data changes directly in the graph/maps. It provides data set interface to obtain data, for visualization [7].

Fusion tables facilitate the integration of geo location in website as well. Where user can upload a table with street address, points, lines or polygons. This data then rendered on server side and that will send to the user's website where it has been integrated, in the form of tiles (images). Following is the example of map service integration using Fusion Table.



Fig. 3. Map service integration using Fusion Table [8]

Like the maps, user can also integrate the pie charts, line charts, bar carts, scatter chart as per the requirement.

6. LATEST FEATURES

1. Controlled sharing of data, user can share the specific data which should be shared with other user.
2. Chart tooltip layout: Shows the information about the values after one would hover over them.

3. Filtering option: "NOT" queries and regular expression matching.
4. Custom table and column properties.
5. Extended quota limits up to 1GB per user, and user can store 250 MB data per table.
6. Media download of large tables: provide large tables download facility without having server timeout problem.

7. CONCLUSION

Google Fusion Tables is a web application used for sharing, visualizing and publishing of data on websites. It is also known as SaaS app, hosted and maintained by Google cloud services. This service can be accessible to user through Internet (using browser) or by using application. User can upload files (CSV, KML, ODS, XLS or Google Spreadsheet) to the Fusion Tables table. This data can be then visualize using different kind of charts, for private use or embed in website. Fusion Tables table can be used for real time data changes and its reflection on charts/map. As in collaboration of all the feature, Fusion table is helpful in data management tool and data analytics operation, with good scalability. Also it provides a real time data collaboration with add, delete, edit and comment feature to the users. Integration to Google will provide our tables, maps are searchable to Google's crawlers.

8. ACKNOWLEDGEMENT

Thanks to Professor Gregor von Laszewski for giving me an opportunity and encouragement to create a technical paper. Also the entire class and TA's for their suggestion for paper.

REFERENCES

- [1] Technopedia, "Google Fusion Tables," Web Page, Feb. 2017. [Online]. Available: <https://www.techopedia.com/definition/26624/google-fusion-tables>
- [2] Google Developers, "Fusion Table Help: Type and size of files to import," Web Page, Feb. 2017. [Online]. Available: <https://support.google.com/fusiontables/answer/171181?hl=en>
- [3] —, "Type and size of files to import," Web Page, Feb. 2017. [Online]. Available: https://en.wikipedia.org/wiki/Google_Fusion_Tables
- [4] —, "Fusion tables Rest API: Get Start," Web Page, Feb. 2017. [Online]. Available: https://developers.google.com/fusiontables/docs/v2/getting_started
- [5] —, "Fusion tables Rest API: Release Notes," Web Page, Feb. 2017. [Online]. Available: https://developers.google.com/fusiontables/docs/release_notes
- [6] W. S. HectorGonzalez, AlonHalevy, "Google fusion tables: Data management, integration and collaboration in the cloud," Tech. Rep., 2010.
- [7] Google Developers, "Google Charts," Web Page, 2017. [Online]. Available: <https://developers.google.com/chart/interactive/docs/?csw=1>
- [8] Gregor von Laszewski and Badi Abdul-Wahid, "Big Data Classes," Web Page, 2017. [Online]. Available: <https://support.google.com/fusiontables/answer/2527132?hl=en>