

BigQuery

Google BigQuery is a data warehouse platform.

Schema

You can define table schemas via JSON documents which get ingested at the same time as your data using the `bq` tool.

Data Types

Data Types List from [google documentation](#)

Name	Data type	Description
<u>Integer</u>	INT64	Numeric values without fractional components
<u>Floating point</u>	FLOAT64	Approximate numeric values with fractional components
<u>Numeric</u>	NUMERIC	Exact numeric values with fractional components
<u>BigNumeric</u>	BIGNUMERIC	Exact numeric values with fractional components
<u>Boolean</u>	BOOL	TRUE or FALSE (case-insensitive)
<u>String</u>	STRING	Variable-length character (Unicode) data
<u>Bytes</u>	BYTES	Variable-length binary data
<u>Date</u>	DATE	A logical calendar date
<u>Date/Time</u>	DATETIME	A year, month, day, hour, minute, second, and subsecond
<u>Time</u>	TIME	A time, independent of a specific date
<u>Timestamp</u>	TIMESTAMP	An absolute point in time, with microsecond precision
<u>Struct (Record)</u>	STRUCT	Container of ordered fields each with a type (required) and field name (optional)

Name	Data type	Description
<u>Geography</u>	GEOGRAPHY	A pointset on the Earth's surface (a set of points, lines and polygons on the <u>WGS84</u> reference spheroid, with geodesic edges)
<u>JSON</u>	JSON	Represents JSON, a lightweight data-interchange format

Differences between JSON and Record/Struct

JSON type allows you to ingest JSON without pre-defining the schema whereas a record/struct must be pre-defined and all the fields must be known in advance.

JSON fields are more fiddly to query and work with in general. It seems like you can't do things like UNNEST them.

Nested/Repeated Columns

To allow a column (or object) to repeat (e.g. to have an array of values) you must use `mode: NESTED` in your schema.

[See Nested and repeated columns documentation](#)

Revision #1

Created 16 December 2022 16:43:03 by James

Updated 21 January 2024 14:51:44 by James