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

Name	Data type	Description
Geography	`GEOGRAPHY`	A pointset on the Earth's surface (a set of points, lines and polygons on the WGS84 reference spheroid, with geodesic edges)
JSON	`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 2022-12-16 16:43:03 UTC by James
Updated 2024-01-21 14:51:44 UTC by James