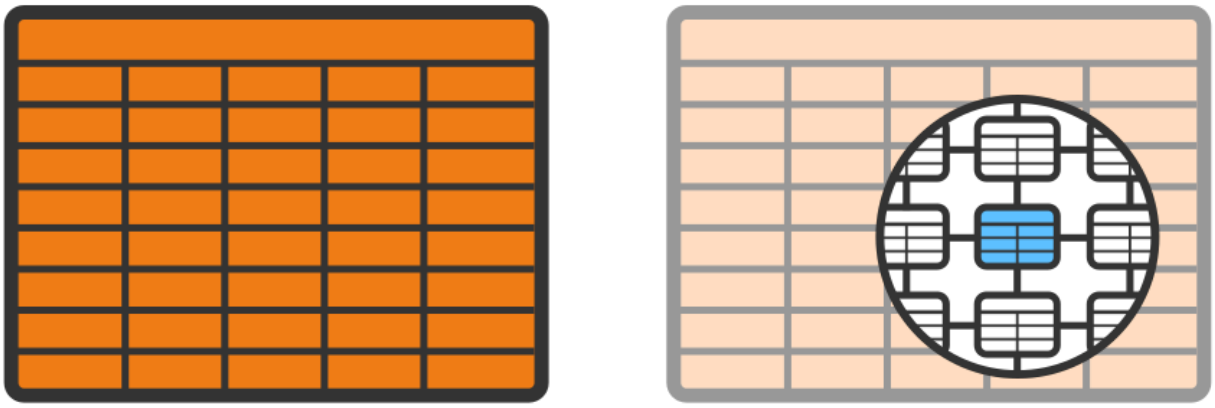
Overview

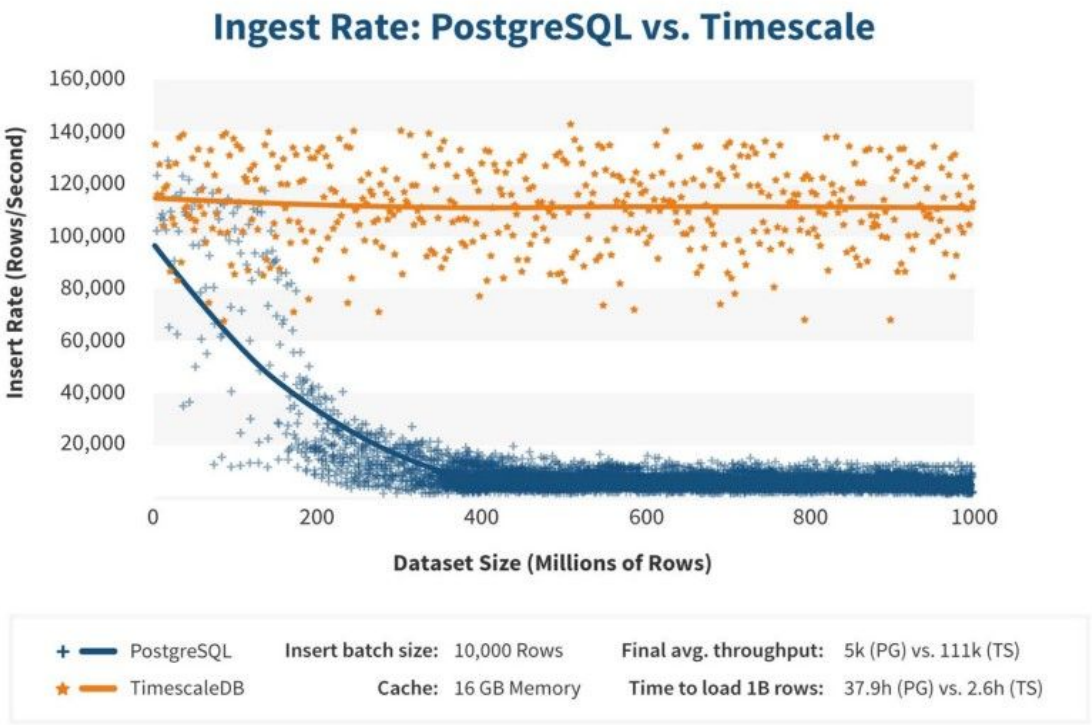
ADCIRC is a numerical ocean circulation model used to simulate hurricane storm surge, tides, and coastal circulation. The ADCIRC Region III Simulation was an effort to "develop and apply a complete end-to-end modeling system, with all required forcing inputs, for updating the floodplain levels for coastal and inland watershed communities" in FEMA Region III. This region includes the Delaware Bay, Chesapeake Bay, Delaware-Maryland-Virginia Eastern Shore, Virginia Beach, and all tidal tributaries and waterways connected to these systems. The simulation used 159 synthetic tropical storms that were simulated using the ADCIRC storm surge and SWAN wave models.

The ADCIRC Region III Simulation database contains 156 of these synthetic tropical storms. The database contains 32+ TB of data. TimescaleDB, an extension to PostgreSQL, enabled the storage of this large amount of data. TimescaleDB uses hypertables which look like and act like regular PostgreSQL tables but are "actually an abstraction or a virtual view of many individual tables (see image below) holding the data, called chunks".



These chunks sub-divide the data based on a time interval. The chunks allow for more efficient memory management, than a single large table, enabling the scaling of hypertables to very large sizes. TimescaleDB claims to "regularly test TimescaleDB to 10+ billion rows, while sustaining insert rates of 100-200k rows / second (1-2 million metric inserts / second)". Users "with more powerful hardware, have scaled TimescaleDB to 500 billion rows of data while sustaining 400k row inserts / second".

Below is a graph comparing the ingest rates for a regular PostgreSQL table vs. a TimecaleDB hypertable. As the number or rows increase the insert rate for the PostgreSQL table drops precipitately, while the hypertable holds steady.



In the case of this database, the ADCIRC data has a 30 minute time interval, and the hypertable chunks have a time interval of 2 hours, so each chunk contains four timesteps. There are three tables per storm containing different ADCIRC model variables:

- fort63 - zeta (water surface elevation above geoid, units: m)
- swan63 - hs (significant wave height, units: m), tps (Relative Peak Period, units: s), and dir (wave direction, units: degrees_CW_from_East)
- fort64 - u_vel (water column vertically averaged east/west velocity, units: m s-1) and v_vel (water column vertically averaged north/south velocity, units: m s-1)

Each storm table contains a column for the variables listed about, and column for the node, and timestamp of the data.

fort63 :

```
  Column   |            Type             | Collation | Nullable | Defaulta
-----------+-----------------------------+-----------+----------+---------
 node      | integer                     |           | not null |
 zeta      | numeric                     |           |          |
 timestamp | timestamp without time zone |           | not null |
Indexes:
    "bp1_dp1r3b1c2h1l1_fort63_pkey" PRIMARY KEY, btree ("timestamp", node)
    "bp1_dp1r3b1c2h1l1_fort63_node_timestamp_idx" btree (node, "timestamp" DESC)
    "bp1_dp1r3b1c2h1l1_fort63_timestamp_node_idx" btree ("timestamp" DESC, node)
```

swan63:

```
  Column   |            Type             | Collation | Nullable | Default
-----------+-----------------------------+-----------+----------+---------
 node      | integer                     |           | not null |
 hs        | numeric                     |           |          |
 tps       | numeric                     |           |          |
 dir       | numeric                     |           |          |
 timestamp | timestamp without time zone |           | not null |
Indexes:
    "bp1_dp1r3b1c2h1l1_swan63_pkey" PRIMARY KEY, btree ("timestamp", node)
    "bp1_dp1r3b1c2h1l1_swan63_node_timestamp_idx" btree (node, "timestamp" DESC)
    "bp1_dp1r3b1c2h1l1_swan63_timestamp_node_idx" btree ("timestamp" DESC, node)
```

fort64:

```
  Column   |            Type             | Collation | Nullable | Default
-----------+-----------------------------+-----------+----------+---------
 node      | integer                     |           | not null |
 u_vel     | numeric                     |           |          |
 v_vel     | numeric                     |           |          |
 timestamp | timestamp without time zone |           | not null |
Indexes:
    "bp1_dp1r3b1c2h1l1_fort64_pkey" PRIMARY KEY, btree ("timestamp", node)
    "bp1_dp1r3b1c2h1l1_fort64_node_timestamp_idx" btree (node, "timestamp" DESC)
    "bp1_dp1r3b1c2h1l1_fort64_timestamp_node_idx" btree ("timestamp" DESC, node)
```

The naming convention for the storm tables are:

To save space the lat/lon coordinates, for each node, and bathymetry data have been stored in a separate table named r3sim_fort_geom:

```
      Column      |         Type         | Collation | Nullable | Default
------------------+----------------------+-----------+----------+---------
 node             | integer              |           | not null |
 lon              | numeric              |           |          |
 lat              | numeric              |           |          |
 bathymetry       | numeric              |           |          |
 geom             | geometry(Point,4326) |           |          |
Indexes:
    "r3sim_fort_geom_pkey" PRIMARY KEY, btree (node)
    "r3sim_fort_geom_index" spgist (geom)
```

This table also contains the PostGIS geometry (geom). This geometry was indexed using SP-GiST which creates partitions that need not be of equal size, and therefore works well with ADCIRC's unstructured triangular grid.