EMS Annual Meeting Abstracts
Vol. 16, EMS2019-769, 2019

A Geo-Temporal NoSQL Database for Atmospheric Observations Derived from Mode-S Data

Conor Lally
Met Éireann, Climate & Observations Division, Dublin, Ireland (conor.lally@met.ie)

New clustered time-series specific databases based on distributed processing frameworks, such as Hadoop, have made huge improvements in the processing and accessing of time-series data. NoSQL columnar type databases have proved particularly advantageous. These innovations are providing substantial benefits in both performance and scalability and offer innovative approaches to handling high-volume high-velocity data. However, time series data also presents some design challenges for columnar databases. In particular, regional Hot-spotting caused by monotonically incrementing values such as time-stamps which direct database requests to a single node in the cluster. The solutions to these issues make time-series specific databases unsuitable for multidimensional spatio-temporal data that contains a location as well as a time aspect. For this type of data, a geo-temporal database is required.

Atmospheric observations derived from aircraft transponder signals are an interesting example of such spatio-temporal data. Commonly known as Mode-S, these constitute an abundant source of low cost but high-quality atmospheric measurements that can be exploited for meteorological purposes. For clustered columnar databases Mode-S data is particularly problematic due to the extremely uneven distribution of its observations in both time and space.

This research examines the feasibility of using a columnar database as a geo-temporal database for Mode-S atmospheric observations. This is addressed by developing and implementing a prototype system. The design process for this system identified several strategies for storing and accessing spatial-temporal data in columnar databases. It also assessed the different columnar databases available identifying those with features that make them unsuitable for the task.

A series of experiments examining the prototypes capabilities were conducted which reveal substantial improvements for data ingestion and recover operations, far in excess of those provided by traditional relational database systems. These experiments and the knowledge gained during the research and design processes proves definitively that a distributed columnar approach to managing and storing multidimensional spatio-temporal data is possible.

Given the success of this research it is believed that a suitable columnar architecture would provide an efficient platform for the distribution of large spatio-temporal datasets for Cloud Computing, such as the outputs of Numerical Weather Prediction model runs.