Voldemort – Review

Yixing Chen, yc3094

The paper *Serving Large-scale Batch Computed Data with Project Voldemort* introduced Voldemort – a general-purpose distributed storage and serving system. Voldemort is a key-value storage system, which is inspired by Amazon Dynamo. It is a bulk loading system capable of serving multiple terabytes of data. Because many social networking websites need to process hundreds of terabytes of offline data every day to make predictions, the data changes extremely frequently and the most data is only for read. In this case, the primary key only, non relational database is preferred. The novelty of Voldemort compared to Dynamo is that, the custom storage engine for bulk-loaded data sets.

A Voldemort cluster can contain multiple nodes , each with a unique identifier. A physical host can run multiple nodes, though at LinkedIn we maintain a one-to-one mapping. All nodes in the cluster have the same number of stores , which correspond to database tables. General usage patterns have shown that a site-facing feature can map to one or more stores. The storage engine includes Hadoop, HDFS, Voldemort and a driver program. The driver program triggers the build. Hadoop jobs are responsible or constructing the chunk sets on a per-node basis, which are the storage format and each contains an index file and a data file. The chunk sets are stored in HDFS. When the Hadoop job is completed, Voldemort nodes fetch the data from HDFS by a request sent from the driver. Checksums are also used in the build phrase and retrieve phrase. Over time as new stores get added to the cluster, the disk to memory ratio increases beyond initial capacity planning, resulting in increased read latency. The rebalancing feature allows the system to add new nodes to a live cluster without downtime. The rebalancing process is run by a tool that coordinates the full process.

In the evaluation part, the writers used a simulated data set where the key is a long integer between 0 and a varying number and the value is a fixed size 1024 byte random string. The evaluation is from three aspects - build times, read latency and production workload. They presented many graphs and results, and demonstrated that Voldemort provides faster data deployment and more read latency as the database increases and various client throughput.

In conclusion, the writers presented a low-latency bulk loading system capable of serving multiple terabytes of data, which provides more stable and reliable system performance. And in the future they will try to add many other interesting features to the read-only storage pipeline, and they are investigating additional index structures that could improve lookup speed and that can easily be built in Hadoop.