

ThinkData Works

ThinkData Works improves query and ingestion performance cost effectively with the Vertica Startup Accelerator Program.



Overview

ThinkData Works is a Toronto-based tech startup that specializes in the aggregation, standardization, and distribution of high value public data. Alternative data sources are a virtually untapped resource, but the difficulty of efficiently leveraging a variety of data in a variety of formats is a barrier to entry for most businesses.

To solve this problem, ThinkData Works developed Namara, a public-facing data management platform designed to standardize,

“We tried other data analytical platforms. AWS Redshift had a bunch of limitations and lacks geospatial support. Elastic Search has great query performance, but we ran into issues with data ingestion. Vertica provides us with everything we need and its flexibility, compression, and performance are superior, particularly on load time.”

BRENDAN STENNETT

Co-Founder & CTO
ThinkData Works, Inc.

reformat, and enrich a variety of data, effectively making it possible to take data from anywhere and turn it into anything. ThinkData Works has mapped to more than 1,000 data sources, and is connecting people to more than 100,000 data sets. Their customers include Altus Group, The Royal Bank of Canada, Thomson Reuters, and enterprise businesses and SMEs across the financial services, insurance, and other sectors.

Challenge

The company was building an ingestion pipeline that required an MPP analytical database due to the large number of reads and writes and need for analytics, high performance, and concurrency. AWS Redshift had many limitations, including a lack of geospatial support. Elasticsearch had better query performance, but ingestion times were slow and required heavy compute power. ThinkData Works required a columnar database to query a range of columns and attributes with a comprehensive set of built-in analytical functions and high performance for both loading and querying. Most importantly, because this startup was in the early stage of developing its ingestion data and search platform with just ten employees, the company required a cost-effective analytical database without a large initial investment.



At a Glance

- **Industry**
Information Technology
- **Location**
Toronto, Ontario, Canada
- **Challenge**
Solve the common access problem of open data with a cost-effective analytics platform for start-ups
- **Products and Services**
Vertica Analytics Platform
- **Results**
 - + Achieved sub-second query performance—
an order of magnitude faster than competing MPP databases
 - + Reduced ingestion time to build a more effective data pipeline
 - + Deployed Vertica on the Google Cloud Platform in minutes
 - + Implemented Vertica quickly with minimal learning curve and zero support tickets filed

“As a startup, it’s tough to justify a large initial software purchase. The Vertica Startup Accelerator Program enabled us to adopt Vertica at no cost for a year, helping us to invest more in our company as we grew and increased customer adoption of the Namara Platform.”

BRENDAN STENNETT

Co-Founder & CTO
ThinkData Works, Inc.

Contact us at:
www.vertica.com

Solution

“I have always wanted to use a column store for this type of challenge. Vertica proved to be the best choice for us. It is flexible, provides fast ad-hoc performance that is much faster than competing databases, and the compression of data and ingestion times were simply better with Vertica,” says Brendan Stennett, Co-Founder and CTO, ThinkData Works, Inc. “And, the Vertica Startup Accelerator Program afforded us the opportunity to get started for free for the first year, which enabled us to invest in our people and development as we were getting established.”

Providing access to public data is not sector specific and needed to appeal equally to professionals across the data use spectrum—from C-level executives to data scientists and academics. ThinkData Works realized that the common axis problem for every industry was gaining access to the data, first, and then having that data transformed to match their company’s internal standard.

To do this, ThinkData Works created techniques to search, index, and link to data via publicly accessible portals, and then focused on addressing varying data format issues. Ultimately, this led to the creation of Namara, which acts as a search and standardization engine for public data.

“Our data scientists eliminate all the data differences, so that we make it easier for our customers to consume clean data quickly,” Stennett explains. “We load all our data into Vertica and make it available via spreadsheet API access without having to export large,

cumbersome flat files. Our data is being constantly updated, so load time was critically important to us. Vertica excels at fast load time and top query performance.”

Results

ThinkData Works is now achieving sub-second query performance, regardless of the data set size. The company currently has 5 TB of data, but is growing both in data volume and as a company.

“We now have 20 employees and plan to double the size of our company in the next year,” says Stennett. “We are experiencing rapid growth in our customer and partner base, as well, helping them to solve real-world risk assessments and generating client lists for lead generation. They all require more external data to feed their models.”

ThinkData Works is also a Vertica commercial and OEM customer. This license flexibility has enabled them to meet the varying needs of its customers who prefer to spin up an entirely new deployment that includes the underlying Vertica database.

ThinkData Works runs Vertica on the Google Cloud Platform, deploying within minutes with a one-click install. Stennett reports that the learning curve with Vertica was minimal and the company has yet to file a single support ticket, so the customer satisfaction is high among his team.

“Vertica and the Vertica Startup Accelerator provides us with everything that we need—speed, flexibility, and value,” says Stennett.