

Travel Agency Masters Big Data with Google BigQuery



At a Glance

What they wanted to do

- Implement a powerful data analysis tool to solve technical problems quickly
- Gain insights into bus service booking patterns to serve customers better
- Avoid setting up and maintaining a complex infrastructure in-house

What they did

- Chose Google BigQuery to take advantage of Google's vast data processing infrastructure
- Analyzed large data sets in near real time using a simple, SQL-like language
- Saved time analyzing technical problems and customer booking trends

What they accomplished

- Analyzed data sets as large as 2 terabytes in less than 30 seconds
- Spent 80% less than they would have on a Hadoop infrastructure
- Strengthened the company by improving customer service and engineering quality

Organization

In 2006, online travel agency redBus introduced internet bus ticketing in India, unifying tens of thousands of bus schedules into a single booking operation. (Think of it as Expedia for bus booking.) Using BigQuery, redBus crunches terabytes of booking and inventory data in mere seconds and at a fraction of the cost of other big-data services. BigQuery also helps engineers fix glitches quickly, minimize lost sales and improve customer service.

Challenge

Executives at the Bangalore-based redBus needed a powerful tool to analyze booking and inventory data across their system of hundreds of bus operators serving more than 10,000 routes. They considered using clusters of Hadoop servers to process the data but decided the system would take too much time to set up and would require a specialized staff to maintain it. It also would not provide the lightning-fast analysis they needed.

"It would have taken at least a couple of hours to analyze anything," says Pradeep Kumar, a technical architect at redBus. "Crunching very large data sets would have been a day's job. We needed something more powerful to get the real-time analysis we were looking for."

Solution

Kumar and his colleagues learned about Google BigQuery and realized it was the right match for their data processing needs. The web-based service, which enables companies to analyze massive datasets using Google's data processing infrastructure, is easy to set up and manage since its simple, SQL-like query language doesn't require complex technology or specialized personnel. It also has low overhead costs.

"We explored several data analytics solutions. Nothing comes remotely close to the sheer power of Google BigQuery. It made large-scale data collection and crunching possible with little effort, which has translated to a significant business advantage."
—Pradeep Kumar, technical architect, redBus

The redBus team uses Google BigQuery as part of an intricate data collection and analysis process. Applications hosted on a range of servers continually pump information related to customer searches, seat inventory and bookings into a centralized data collection system. Engineers upload the data to BigQuery, which provides answers to complex queries within seconds.

About Google BigQuery

Google BigQuery is a web service that enables companies to analyze massive datasets – up to billions of rows in seconds – using Google’s infrastructure. Scalable and easy to use, BigQuery lets developers and businesses tap into powerful data analytics on demand using the familiar SQL query language.

For more information, visit
<http://cloud.google.com/products/big-query.html>

“Getting to the root of problems used to be really time-consuming. By the time we figured it out, customers might have given up. Now if there are booking problems, BigQuery helps us understand the reason right away. Choosing Google BigQuery was the right decision for our company.”
— Pradeep Kumar, technical architect, redBus

For example, BigQuery helps redBus staff:

- Learn how many times customers searched for seats and found none or very few available, indicating more seats should be added to a route
- Investigate decreases in bookings and notify engineers if a technical problem is the cause
- Identify server problems by quickly analyzing data related to server activity

Results

Google BigQuery provides near real-time data analysis capabilities at 20% of the cost of maintaining a complex Hadoop infrastructure. Queries that would have required a day to analyze on a Hadoop framework take less than 30 seconds using Google’s web-based service.

“We explored several data analytics solutions. Nothing comes remotely close to the sheer power of Google BigQuery,” Kumar says. “It made large-scale data collection and crunching possible with little effort, which has translated to a significant business advantage.”

The fast insights gained through BigQuery are also making redBus a stronger company. By minimizing the time it takes staff members to solve technical problems, BigQuery has helped improve customer service and reduce lost sales.

“Getting to the root of problems used to be really time-consuming,” Kumar says. “By the time we figured it out, customers might have given up. Now if there are booking problems, BigQuery helps us understand the reason right away. Choosing Google BigQuery was the right decision for our company.”

