

[Try It](#)[View Demo](#)

PREDICTIVE ANALYTICS

# How to Choose a Database for Your Predictive Project

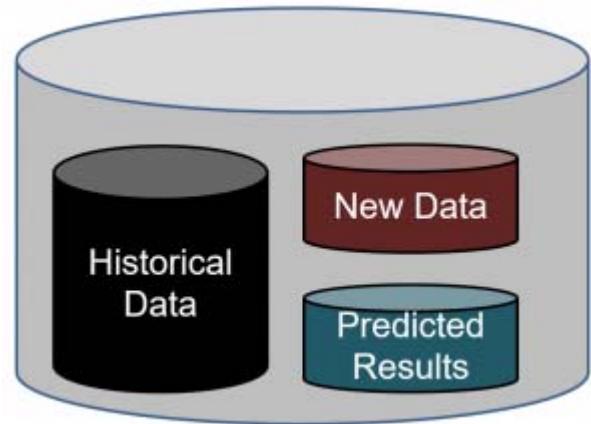
BY SRIRAM PARTHASARATHY

Predictive analytics has set a new bar for business intelligence. It makes applications infinitely more valuable, separates software from competitors, and offers new revenue streams. But predictive analytics platforms and implementations are also incredibly complex. Embedding effective predictive solutions in your application requires choosing the right talent and resources—and chief among them is a database (or databases) to house your datasets.

**>> Related: [What Is Predictive Analytics?](#) <<**

In predictive analytics implementations, there are two different datasets that come into question: Historical data, which is used to train your algorithm in how to predict an outcome. And new data, where future predictions need to be done.

Though most application teams tend to use separate databases to store their historical and new data, it's not uncommon to see teams wanting to have one database for all datasets. In such cases, one would use a six-month filter for historical data and a last day/week filter for predicting on new data.



As for choosing the specific database systems for your predictive analytics platform to pull from, consider these three factors: volume, support, and scalability.

## Data Volume

Data volume typically affects the historical data where transactions from prior months/years are saved. However, in some industries it also affects new data predictions, such as in cases

[Try It](#)[View Demo](#)

If your dataset spans thousands of terabytes of petabytes, on the other hand, look at non-relational databases that can scale by adding multiple nodes. These include Amazon Redshift, Google BigQuery, Snowflake, and Hadoop-based solutions. Note that these databases will require data experts to work with them.

## Support Resources

Many customers ignore the aspect of support and maintenance. But often, these companies find they need to hire talent to properly install, maintain, and tune their database servers for optimum performance. There needs to be processes in place for backups, dealing with occasional failures/data losses, etc. Always consider the investment you have to make for infrastructure/human resources and compare it to what you would pay for a dedicated service in the cloud.

Unless you have hundreds of different applications deployed, it's easier and cheaper to deploy with the cloud. And even in the former case, I would still recommend letting the experienced data professionals in the cloud manage the database for you—unless you find that there is a big financial gain by doing it in house.

## Data Scalability

Most companies start with vertical data scaling by adding more resources, such as CPU and RAM, to an existing machine. Cloud databases like Redshift scale nicely just by adding more nodes. You can also partition the data so different clusters handle data for different sets of customers.

The next line of attack is to do horizontal scaling by creating and clustering replicas of the database. You'll need to buy licenses for different machines where the database has to be installed. NoSQL databases like Cassandra and Mongo DB are good candidates to scale horizontally by distributing where the data is saved.

In summary:



---

[Try It](#)[View Demo](#)

- Consider the investment needed to install and maintain your databases in house versus in the cloud.

*Originally published February 22, 2019; updated on July 31st, 2020*

## About the Author

Sriram Parthasarathy is the Senior Director of Predictive Analytics at Logi Analytics. Prior to working at Logi, Sriram was a practicing data scientist, implementing and advising companies in healthcare and financial services for their use of Predictive Analytics. Prior to that, Sriram was with MicroStrategy for over a decade, where he led and launched several product modules/offerings to the market.

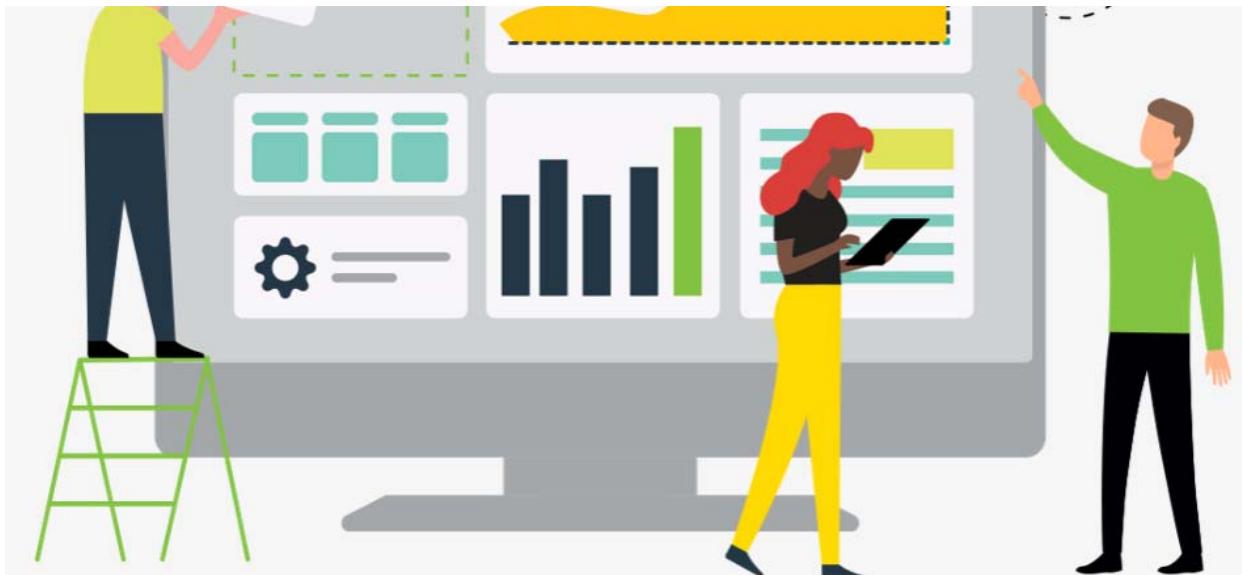
## You May Also Like...

[Try It](#)[View Demo](#)

#### PREDICTIVE ANALYTICS

### Predictive Analytics in Action: Manufacturing

By embedding predictive analytics in their applications, manufacturing managers can monitor the condition and performance of equipment and predict failures before they happen.

[Try It](#)[View Demo](#)

#### PREDICTIVE ANALYTICS

## How to Maintain and Improve Predictive Models Over Time

How do you make sure your predictive analytics features continue to perform as expected after launch? Follow these guidelines to maintain and enhance predictive analytics over time.

[Try It](#)[View Demo](#)

#### PREDICTIVE ANALYTICS

## Solving Common Data Challenges in Predictive Analytics

Once you know what predictive analytics solution you want to build, it's all about the data. Follow these guidelines to solve the most common data challenges and get the most predictive power from your data.

[Try It](#)[View Demo](#)

#### PREDICTIVE ANALYTICS

### Predictive Healthcare Analytics: Improving the Revenue Cycle

Efficiency in the revenue cycle is a critical component for healthcare providers. A failure in even one area can lead to critical revenue loss for the organization.

[Try It](#)[View Demo](#)

#### PREDICTIVE ANALYTICS

## Predictive Analytics in Action: 5 Industry Examples

Predictive analytics is transforming all kinds of industries. It can catch fraud before it happens, turn a small-fry enterprise into a titan, and even save lives.

[Try It](#)[View Demo](#)

#### PREDICTIVE ANALYTICS

### Predictive Analytics for Business Applications

Predictive analytics is the #1 feature on product roadmaps. Learn how application teams are adding value to their software by including this capability.

[Try It](#)[View Demo](#)

Call: **1-888-564-4965**

Email: [salesteam@logianalytics.com](mailto:salesteam@logianalytics.com)

---

Logi Analytics Confidential & Proprietary | Copyright 2020 Logi Analytics | [Legal](#) | [Privacy Policy](#) | [Site Map](#)