

## Setting the Right Price: Yahoo! Relies on Infobright for Fast Analysis of Advertising Impression Data

### Overview

Yahoo! is the premier digital media company. Their business focus is on creating a content, communications, and community platform that delivers rich consumer experiences and advertising solutions across all digital screens. Yahoo attracts more than half a billion consumers every month in more than 30 languages—making Yahoo! one of the most visited and trusted Internet destinations.

Yahoo! is headquartered in Sunnyvale, Calif., and has more than 13,000 employees in 25 countries, provinces, and territories.

### Business Challenge

Yahoo! reaches more than half a billion people worldwide—that’s one out of every two Internet users—and offers advertisers creative ways to put their messages in front of the right set of customers in the right contexts at the right time in order to drive their businesses’ growth.

Setting the prices for display ads is the responsibility of Yahoo!’s Pricing and Yield Management (PYM) team. They are also responsible for pricing analytics and pricing-yield business operations for all of Yahoo!’s display advertising business. This involves sophisticated analysis of very large volumes of impression-level Web data by highly skilled analysts within the PYM team. An “impression” in this context is an ad served to a user on a particular web page. Advertisers measure the reach (and therefore the value) of their ads by tracking ad impressions, and publishers sell ad space based on impressions.

However, this involves far more than just counting impressions. Also factored in are attributes such as geography (where the ad displays) and demographic information about visitors to the specific Web site and their browsing history. With the move toward finer and finer ad targeting (using more attributes to determine which ad to display), the process of pricing has become more complex as the team must price each specific set of attributes. In addition to setting the right price in the guaranteed display ad market, the PYM team must also determine how impressions are monetizing in their secondary marketplace—this is required for establishing appropriate pricing and deal evaluation.

Not surprisingly, as more attributes are captured for analysis, the volume of data collected every day continues to increase. For the PYM pricing analysis application, that equals 20 to 30-million records per day.



### INDUSTRY Online Advertising

#### SOLUTION

High-performance platform for the analysis of terabytes of detailed online advertising data to set display-ad prices.

“Using Infobright allows us to do pricing analyses that would not have been possible before. We now have access to all of our detailed Web impression data, and we can keep 6x the amount of data history we could previously. This enables us to quickly determine the value of certain combinations of attributes and price accordingly.”

**Arvind Hariharan**  
Sr. Director, Pricing and Yield Management  
Yahoo!

## Technology Challenge

In addition to their analysis and pricing functions, the PYM team also has responsibility for implementing and supporting their analytic infrastructure. The team does their analysis using a Web-based application developed with Ruby on Rails, the open-source web application framework for the Ruby programming language. The application lets the analysts run standard queries and display the results graphically. In addition the analysts need to be able to create and run complex, ad-hoc queries and get very fast responses.

Previously Yahoo! used a traditional row-based database to capture and store the data. However the high cost and high administrative effort it took was a barrier to storing all of the data the PYM team needed. Only summary data of user behavior could be stored, and the data history was limited to 30 days. This severely limited the number of attributes that could be considered in the pricing analysis, thereby limiting the analysis the PYM team could do and making the analysis less accurate.

The PYM team decided they needed a new database to meet their needs:

- Allow the storage of all of the detailed impression data instead of only summary data
- Extend data history to 6 months versus the 30 days they could keep previously
- Provide fast queries and flexible ad-hoc analytics
- Reduce the amount of hardware required
- Reduce the administrative effort involved in managing and tuning the database

The PYM team decided to investigate using an open-source analytic database to meet these requirements. After researching Infobright and Calpont's InfiniDB, the team chose Infobright, as it was a better match to their requirements to analyze large volumes of machine-generated data such as Web logs.

## Results with Infobright

Now in production, the Infobright database has allowed the Yahoo! PYM team to meet the requirements listed above:

Store all detailed data: Yahoo! is currently loading close to 30-millions records every day. This now includes all of the detailed data that the team wanted to have access to, rather than just the summary information they were able to store previously.

Keep much more history, in much less space: Yahoo! now has the ability to store the six months of data they need for the most accurate analysis. This means that there are about 6-billion records in the database. What's more, as Infobright provides outstanding levels of data compression, the approximately 6TB of raw data that has been loaded only uses 600GB on disk (10:1 compression).

How does the Infobright database enable such deep compression? Infobright's industry-leading compression rates are a result of the product's architecture. As data is loaded into Infobright, it is stored in Data Packs (65K fixed-sized segments within each column). Infobright compresses each Data Pack individually using the optimal compression algorithm for that data, resulting in very deep data compression, typically 10:1 to 40:1.

## CASE STUDY

The rate of compression depends on the specific type of data. In addition, as Infobright does not use indexes to achieve fast query performance, no additional space is needed to accommodate indexing. With traditional row-based databases, the size of the database on disk often exceeds the raw data volume by a significant factor, sometime as high as 2x when many indexes are required.

**Faster queries, flexible ad-hoc analytics:** The PYM team has also seen faster queries with Infobright, even though these queries are now running against much more data than had been the case before. Queries that took several minutes previously now run in seconds with Infobright. As Infobright eliminates the need to create indexes for performance, PYM analysts can also create and run new ad-hoc queries and get fast responses—without the need to have a database administrator do any tuning.

**Easier to maintain and support:** Most databases depend on expensive, skilled resources like database administrators (DBAs) to maintain and tune the database. Infobright was specifically designed to eliminate the need for DBAs and other personnel to administer and tune the database. Instead, built-in intelligence within the Infobright Knowledge Grid architecture automatically delivers high performance and fast query speed without manual effort. This allows the PYM team to focus on their primary responsibilities rather than devoting effort to database maintenance.

Moving forward, the PYM team will be adding supply data to the database as they extend the usage of Infobright to support other important activities. Infobright has given the PYM team the ability to do more in-depth analysis than they could previously, enabling them to deliver a higher degree of accuracy in their pricing analysis.

### About Infobright

Infobright's high-performance database is the preferred choice for applications and data marts that analyze large volumes of "machine-generated data" such as Web data, network logs, telecom records, stock tick data and sensor data. Easy to implement and with unmatched data compression, operational simplicity and low cost, Infobright is being used by enterprises, SaaS and software companies in online businesses, telecommunications, financial services and other industries to provide rapid access to critical business data.

Try it yourself! Download a free trial of our Enterprise Edition at <http://www.infobright.com/Products/Product-Demo/>.

#### CONTACT

**Corporate Headquarters**  
 47 Colborne Street, Suite 403  
 Toronto, Ontario M5E1P8  
 Canada  
 Tel. 416 596 2483  
 Toll Free 877 596 2483

**Sales**  
 Email: [info@infobright.com](mailto:info@infobright.com)  
 Americas  
 Tel. 312-924-1695  
 EMEA:  
 Tel. +353 (0)87 743 7107  
 APAC:  
 Tel. +1 416 596 2483 x 225

**OEM Sales**  
 Email: [oem@infobright.com](mailto:oem@infobright.com)  
 Americas  
 Tel. 416 596 2483 x 225  
 EMEA:  
 Tel. +353 86 826 7639  
 APAC:  
 Tel. +1 416 596 2483 x 225

[www.infobright.com](http://www.infobright.com)  
[www.infobright.org](http://www.infobright.org)  
 Twitter: [Infobright](https://twitter.com/Infobright)