

White Paper

Web Application Acceleration with Couchbase

By Darrell Pratt, Architecture Leader, GBS

Introduction

The Answers on Demand application serves a large number of end users with a large amount of reporting data. Requests are taken from the users in form filtered query requests and report layout information and the system in turn returns large amounts of data that must be formatted into tabular and graphical data for the client to consume. This paper outlines the storage needs and challenges with this type of data and how a document storage solution such as Couchbase can increase the overall performance of this application.

Problem Overview

Nielsen Answers on Demand platform tracks and analyzes large amounts of consumer purchasing data from a variety of sources. These include point of sales data, loyalty card data and consumer panel data. This data is generally stored in Netezza backend databases and a large amount of metadata is generated and stored in a variety of sources as a result of this process.

Nielsen produces consumer insights by measuring what people buy and what people watch. At its core, these insights provide marketers with a view into the effectiveness of their marketing influence on consumers. Over the decades, Nielsen has evolved its measurement system to address the increasing complexity of this ecosystem and meet the increasing demand for accuracy of insights. Since a fundamental building block of our measurement system is the observation of influence in a sample or population, observer effect poses a principal challenge to the accuracy of our measurement.

“Observer effect” is the influence that an observer has on the outcome of an observed measurement. In a practical sense, this effect is found in many common measurements. Two examples are:

* Automobile tire pressure cannot be measured without letting some of the air out and, thus, influencing the pressure.
* The running score in a game could influence the performance of the team.

In marketing research, examples of the observer effect in action include:

* Online tags used by publishers or third-party measurement companies should be “sufficiently” fast, otherwise time spent processing the tag would affect the count of page views and skew the audience behavior metric.
* Web scraping is a method used for online pricing analysis, but that could contribute to count of page views.
* In some traditional trade markets, our capture methods for store sales could be influenced by tax implications and trading practices, and potentially impact our ability to analyze.

Given the ever-increasing demand for accuracy, these influences have significant impact and must be addressed. Nielsen uses a variety of methods to minimize the observer effect and ensure that the accuracy of our products stays within acceptable limits.

Nielsen Approach

At a macro level, Nielsen adopts a four-step process to produce information and insights for our clients:

1. Collection—acquisition of raw data
2. Classification—method to inform the data
3. Processing—aggregation
4. Delivery—information and insights.

Of the four steps above, processing and delivery are completely automated and observer effect has no impact. Collection and Classification are susceptible to observer effect and Nielsen has adopted a variety of guardrails to ensure that this does not have material impact on the overall accuracy of our insights.

Within our Buy business, we have three types of collection:

* Scan—electronic data feed from retailer Point of Sale (POS) systems. This is factual data, has no interpretations and, hence, is not subject to any observer effect.
* Audit—data captured through store visits performed by Nielsen field teams. Robust sample design and a resilient methodology created by our world class Measurement Science practice serve to offset observer effects introduced during these store visits. Additionally, the use of hand-held terminals for data capture and detailed audit instructions provided to auditors enable us to mitigate observer effect.
* Survey—data collected through interviews. The Survey method is necessary for various Nielsen businesses (e.g. BASES) as we provide forecasts to our clients on a concept (even before it becomes a product). Given the nature of this effort and the extensive human involvement, we have established several controls at Nielsen that include these:
* Leveraging CATI—Computer Assisted Telephone Interviews1
* Building controls into the processing step to compensate for any surveyor influence
* Specifying confidence intervals for estimates because results derived through the Survey method offer insights that are more subjective.

In the classification step, Nielsen does not perform any value judgments. Consumer items (products) are characterized and classified based on packaging information only. The characterization is based on a global dictionary and objective rules developed over time. The objective business rules are the Nielsen assets that drive quality and differentiate Nielsen in the marketplace.

On the Watch side of our business, we use four different methods for collection to accommodate various consumer devices and media:

* Metering—data streams directly from Nielsen panel homes. This is a fully automated process and is designed to collect data with very little intrusion to the panelist. While it is likely that panelist behavior may be influenced by the knowledge that he/she is being metered, we accommodate for it by not using the data within the first few days of installation and by leveraging a random probability design. We also have a very robust fault and flagging system to exclude homes for CATI crediting, if we find deviations from the norm.
* Set Top Box (STB)—return path data acquired by Nielsen from STB providers. Much like retailer POS systems, this data is an electronic feed sent to Nielsen without any observer influence.
* Tagging—data captured through ping backs when a user visits web pages. We address observer effect in this case by building appropriate compensation through our crediting rules. Some illustrative examples include:
* For Online Campaign Ratings, the Performance criterion for online tags is less than a specified threshold—the total time contributed by tag firing, referrer masking, and forwarding to data enrichment providers. Anything more than this specified threshold is likely to have a material impact on user viewing behavior.
* In-view metrics that determine if the ad was viewed by the audience need to meet Interactive Advertising Bureau (IAB) guidelines. Nielsen follows more stringent guidelines than those IAB recommends.
* Survey/Diary—the approach in Watch is similar to what we do within Buy to offset observer effect.

Conclusion

Observer effect is very real in any type of measurement activity. For a business such as Nielsen, whose focus is to provide insights based on consumer watch and buy behavior, observer effect is a critical element that needs to be addressed to ensure accuracy. Given the importance of quality to our brand value, Nielsen adopts principles and practices that demonstrate the highest degree of integrity, business ethics and commitment to the market. We actively research and deploy a variety of solutions to address observer effect as a part of our ongoing commitment to build quality into everything we do to enable success for our clients.

References

1. Resnick, Robert and David Haliday. Fundamentals of Physics. Wiley. 1960.
2. “Guidelines, Standards & Best Practices.” Interactive Advertising Bureau (IAB). Accessed 2012-09-20. <http://www.iab.net/>.

