

The Guardian

the challenge:

How do you ensure that web content is properly presented and exposed to 5 million readers?

the solution:

By building an analytics solution on Elasticsearch, processing 40 million documents per day to deliver real-time visibility of site traffic across the organization



CASE STUDY HIGHLIGHTS

Leverage real-time analytics

- Easily query 360 million documents
- See traffic for all content as it happens
- Gain insight into how updates impact site traffic

Empower the organization

- Give the entire organization real-time insight into audience engagement
- Democratize analytics access for more than 500 users
- Encourage a culture of exploration and innovation for all employees

Keeping up with the ever changing news cycle

Starting out in 1821 as a UK-based newspaper, today The Guardian is a global provider of news content. The company site, theguardian.com, is one of the world's most popular websites with 5 million unique visitors per day – the third largest English-speaking newspaper website in the world.

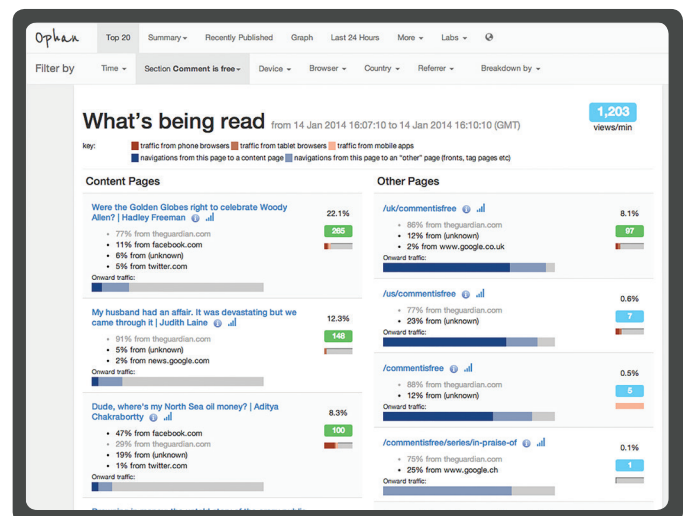
Ophan, the Guardian's in-house developed analytics system, enables users across the company – including editors, journalists, the search optimization team, and developers – to see in real-time exactly how users are interacting with the content. In the news environment, which changes every minute, real-time visibility is invaluable. The Guardian leverages the data generated by Ophan to ensure that content is given exposure at the right time, on the proper social media platforms, with the right headlines.

Processing 40 million documents per day in Elasticsearch

“Before Elasticsearch enabled The Guardian to develop Ophan, we used a traditional analytics package which had a four-hour lag,” recalls Graham Tackley, Director of Architecture at The Guardian. “Trying to get data out of it was horrendous. It was painfully slow. So the ability to see the results of what we did, to have any clue at all, just wasn’t there. We were shooting in the dark.”

Elasticsearch gave The Guardian the freedom to build a very powerful analytics system in-house, rather than relying on a generic, off-the-shelf analytics solution. Powered by Elasticsearch and processing 40 million documents per day and delivering real-time results, Ophan has grown to be an enterprise-wide analytics tool used throughout the organization, with over 500 active users. A large portion of The Guardian's business relies on Elasticsearch to understand how their content is being consumed.

The use cases for Elasticsearch at The Guardian are varied: the visibility afforded by the analytics system is used to see how many hits each content item receives; which headlines and content generate more traffic; where traffic is being referred from; which social media platforms to promote specific content on and when, to gain maximum exposure; and which links to provide the reader to click on next. Engineers are even using Elasticsearch to diagnose website performance issues by searching through events.



“Elasticsearch enables our team to focus on improving the content and headlines, and the promotion of content,” says Tackley. “It’s all about giving a great experience to the user, and showing them what they would be interested in next. Obviously it’s good for us as well because we get more clicks, but it’s also good for the reader because it is giving them content that interests them.”

Responding to change in real time

“We are a news organization,” Tackley explains. “We need to respond to the news agenda. A significant portion of our traffic will get a lot of traffic in a very short time. In that type of circumstance, we need to be able to respond at its peak, and so we need to have the information right away. If we wait until the end of the day to see what’s happening, it would be too late.”

Elasticsearch provides the real-time visibility The Guardian needs to ensure the right content is being promoted on the right social media venues at the right time.

“Elasticsearch improves our understanding of social media’s impact on our traffic, and has enabled us to use social media platforms better,” Tackley says.

Democratizing access to analytics

In addition to real-time improvement, minute by minute, The Guardian also drives overall improvement of the site because the entire organization is learning how to fine tune content and headlines to meet readers’ expectations.

“As part of the editorial process, understanding what content gets traffic and what doesn’t is very important,” Tackley explains. “One of the great accomplishments that we’ve been able to achieve using Elasticsearch is empowering journalists to investigate their content’s audience. We are democratizing access to data, so the editors and journalists can learn and explore themselves. Elasticsearch encourages a culture of self-exploration, which is very exciting.”

“We have seen a change in attitude within the organization,” he continues. “A couple of years ago only top management could look at traffic data. Among everyone else there was a fear that if we looked at traffic data, we are bound to turn into a tabloid paper. Now, people across the organization understand that being able to see what’s happening to their content helps them do their jobs.”

Scalability without sacrificing productivity

“Scaling of Elasticsearch has been fantastic for us,” Tackley says. **“When we introduce a new feature that stresses Elasticsearch more than we expected, we add capacity to our Elasticsearch cluster. Every time we do that it works perfectly. Being able to scale up fast has been invaluable to help our speed of innovation.”**

“The fact that we only have to do fairly light amounts of optimization to be able to do fairly complex faceting is a big advantage,” adds Phil Wills, Senior Software Architect at The Guardian. “We can query over 360 million documents without having to spend enormous amounts of time optimizing – and Elasticsearch has enabled us to do that with a small development team, not spending all of our time working on this aspect. Without Elasticsearch there is no way we would’ve been able to implement a number of features that we have, in the time we have.”

The Guardian’s benefits using Elasticsearch

✓ Driving more page views

Elasticsearch helps The Guardian improve content, headlines and promotion in a variety of ways, ultimately increasing the number of page views and the site’s success.

✓ Enhanced user experience

The Guardian utilizes Elasticsearch-powered analytics to provide readers with more content that meets their demands, which enhances the user experience on the company’s website.

✓ Empowering the team

Offering access to Elasticsearch across the organization has empowered editors and journalists to get more involved, and take a proactive approach to improving the site and its content.

✓ Improved site performance

The Guardian’s IT operations team utilizes Elasticsearch to track how any changes impact site performance, diagnose issues and keep the site up and running at peak performance.

Elasticsearch is on a mission to organize data and make it easily accessible. We deliver the world’s the most advanced open source search and analytics engine available and make real-time data exploration available to anyone. By having a laser focus on achieving the best user experience imaginable, Elasticsearch has become one of the most popular and rapidly growing open source solutions in the market. Today, Elasticsearch is used by thousands of enterprises in virtually every industry. We take good care of our customers and users, providing production support, development support and training worldwide.