Case Study
Intel® Xeon® Processor E5 Family
Big Data Analytics
Communications/Media

More engaged readers with customized news content

Intel® big data professional services helped Next Media Limited develop a big data article recommendation engine based on CDH*, a distribution of Apache Hadoop* from Cloudera, to bring more relevant content to its readers

NEXTmedia

Next Media Limited is Hong Kong's largest publicly-listed Chinese-language print media company, publishing newspapers and magazines in both Hong Kong and Taiwan, with over 30GB of view logs (approximately 30 million records) generated per day from its website and mobile applications.

Looking to explore international business opportunities and expand its presence and operations globally, Next Media collaborated with Intel to broaden its content reach and readership through a timely and content-based filtering article recommendation engine.

Challenges

- **Adapt to changes in news distribution.** Keep pace with the shift in news distribution as it evolves in the digital age by adopting digital channels such as portals, search engines, social media, mobile, and video.
- **Reach out to a bigger audience.** Grow readership by serving news, videos, and interactive games based on readers' personal interests.
- **Keep readers engaged.** Provide fresh content that meets individual readers' interests without having editors spend a substantial amount of time manually updating the news platforms with interesting topics.

Solution

- **Build an article recommendation engine based on big data technology.** Work with data scientists from Intel to architect, design, and develop a proof-of-concept (POC) big data-based article recommendation engine that will cater to readers' personal interests using CDH*, a distribution of Apache Hadoop* from Cloudera.

Technology Results

- **Provided high-performance content processing and results analysis.** Utilizing the highly distributed processing architecture of Apache Hadoop allowed real-time news recommendations based on readers' historical behavior models.
- **Enabled support for local language in the analytic engine.** Successfully built a big-data-based news recommendation engine that analyzes Chinese data from news feeds.

Business Value

- **Improved user experience and user engagement.** Engaging users with more relevant content at the right moment allows them to spend more time on the digital platforms.
- **Lessened editorial workloads.** Editors are more responsive in choosing the articles readers prefer and providing them in real time, as well as putting more value on content that can be utilized in other business processes.

"By working closely with Intel's data science team, we were able to develop an article recommendation engine that enabled us to get a deeper understanding of our readers' preferences. Through this new big data cloud platform based on the Intel® Xeon® processor E5 family and CDH*, a distribution of Apache Hadoop* from Cloudera, we can now provide personalized content that caters to our readers' interests, expanding our article recommendations beyond news content."

– Timothy Yiu
Chief Operating Officer, Digital Business
Next Media Limited

With the advent of digital technology, publishers in Hong Kong and Taiwan are challenged to keep pace with the shift in news distribution – from the traditional approach to adopting digital channels such as search engines, social media, mobile, and video. This change in the consumption habits of readers prompted Next Media to aggressively adopt leading technological trends to take its business forward and cater to the needs of its readers.

By coming up with the Apple Daily* news website and mobile app, Next Media became one of the very few early adopters of the digital news platform in the industry. Since its launch, Apple Daily has become the leading digital news platform in terms of page views and unique visitors from Hong Kong and Taiwan.

Growing and retaining audience in the face of rapid competition for mindshare remained a major challenge on new digital platforms. For example, to keep readers interested, news articles on the top page of each news platform need to be refreshed frequently based on editorial picks, number of views, and timeliness. Editorial recommendations were made manually, so editors would spend a considerable amount of time choosing which content should be provided to readers.

Despite the editors' efforts, all readers were presented with the same articles on the news platforms, regardless of their interests. As a result, it was very difficult to keep the readers' attention, even with the use of social media. Traditional data analytics tools were not suitable for creating customized recommendations to keep readers engaged and encourage them to stay on the news platforms.

In collaboration with Intel, Next Media* builds big data article recommendation engine that brings relevant content to readers in real time

To increase the time readers spend on the news platforms and attract more page views and unique readers, Next Media collaborated with data scientists and Hadoop engineers from Intel® big data professional services to come up with possible solutions. Intel big data professional services was chosen for its data scientists' vast experience in working with large enterprises on diverse algorithms, such as content and collaborative filtering.

Attract more loyal readers with personalized content

Next Media and Intel developed a POC article recommendation engine using CDH, a distribution of Apache Hadoop from Cloudera, and Intel® Xeon® processor E5 family-based servers. Data scientists from Intel helped architect, design and build the analytics engine that meets the unique requirements of Next Media.

"Data scientists from Intel big data professional services have demonstrated their unique position in the big data industry – that it can work with the ecosystem agnostically yet provide business value in data analysis. Their vision in big data aligns closely with us, enabling a sustainable and scalable Next Media enterprise that generates business value from our massive data growth," said Timothy Yiu, chief operating officer for digital business at Next Media Limited.

With the new big data platform, Next Media can now process the large number of view histories as a basis for the news recommendation engine.

The recommendation engine analyzes readers' views on articles to understand user behavior, deriving the likes/affinity to subsets of news topics. Data is obtained from two sources for the POC:

- Daily and real-time articles from the content management system (CMS)
- Tracking the number of views for specific articles

Enabling more language-focused content

Since topics of interests are based on the article content (which is in Chinese), it was imperative for the POC recommendation engine to handle character-based language. Written Chinese sentences are composed of a continuous string of characters and a combination of different characters forms words with different meanings. The challenge is grouping Chinese characters into words, which could be considered a natural language processing (NLP) technique similar to phrasal grouping in languages like English.

"Content processing with the character-based Chinese language had to be effective using the new POC, as opposed to word-based languages, where the linguistic processing may be simpler. The POC is able to combine linguistic processing and user behavior model formulation," explained Yiu.

As part of the POC's linguistic processing of article content, n-gram approach was employed in textual processing and analysis of the article content. A combination of n-grams was validated sufficient and efficient in supporting the identification of key words that form the basis of clusters of topics of interest. The validated data solution also supports retraining of the clusters as news evolves over time. The clusters formed are then used to determine the recommended articles based on individual readers' past viewing behavior and can be applied generally (across all news categories) or specifically (for a particular category). This delivers a personalized article recommendation list. The formed clusters also help in identifying similar and related articles.

Lighter editorial workloads with an automated solution

Through this data solution, article recommendation is automated and personalized, reducing editors' workloads in handpicking articles for the front page of the news platforms. More importantly, Next Media can now better understand its readers' behavior and attract more loyal users by bringing relevant, personalized content.

Since the value of news decreases quickly with time, the ability to process and recommend news in real time was important for the success of the POC. At the conclusion of the POC, the recommendation engine proved capable of processing, analyzing, and providing real-time content to readers via different news platforms.

Increasing business opportunities with enhanced user experience

The user behavior model formed from the data analysis can be used not just to improve readership, but also to gather insight into new business opportunities and increase advertising revenue from targeted advertisements. Since the engine is extensible, it can be applied or scaled to suit Next Media's business in different regions or segments.

Next Media is continuing to work with Intel to use the benefits of the article recommendation engine as well as to explore other opportunities beyond providing traditional news content. With a better understanding of readers' preferences, Next Media can provide content that caters to their interests. By getting more relevant content, readers can spend more time on Next Media's digital platforms, which in turn gives the company more time and added opportunities to build one-on-one relationships with its readers. Through this increased engagement and understanding of its readers' preferences, Next Media can further enhance the user experience with more responsive content and innovative products and services that evolve along with readers' changing behaviors and preferences.

Lessons Learned

- Big data enables new solutions and new ways for publishers to provide personalized, relevant content to readers. ebis alignment quiassita sust, coctassit autat omnimait.
- Intel® Xeon® processor E5 family and CDH*, a distribution of Apache Hadoop* from Cloudera, provide an efficient and cost-effective platform for big data solutions.
- Data scientists are building sound predictive analytical models through data mining and producing business intelligence solutions.
- Linguistic processing of news article content using the n-gram model is a feasible way to process Chinese language in a big data analytics engine.

Find a solution that's right for your organization. Contact your Intel representative, visit Intel's Business Success Stories for IT Managers (www.intel.com/itcasestudies) or explore the Intel.com IT Center (www.intel.com/itcenter).

* The n-gram model is a probabilistic model used in predicting the next item in a sequence. It is widely used in statistical natural language processing.

About Cloudera

Cloudera is revolutionizing enterprise data management by offering the first unified platform for big data, an enterprise data hub built on Apache Hadoop*. Cloudera offers enterprises one place to store, process, and analyze all their data, empowering them to extend the value of existing investments while enabling fundamentally new ways to derive value from their data. Only Cloudera offers everything needed on a journey to an enterprise data hub, including software for business-critical data challenges such as storage, access, management, analysis, security, and search. As the leading educator of Hadoop* professionals, Cloudera has trained over 22,000 individuals worldwide. Over 1,000 partners and a seasoned professional services team help deliver greater time to value. Finally, only Cloudera provides proactive and predictive support to run an enterprise data hub with confidence. Leading organizations in every industry, plus top public sector organizations globally, run Cloudera in production. www.cloudera.com

Printed in USA 0914/EDH/PMG/XX/PDF Please Recycle 331240-001EN