



TOP 5 BIG DATA & Analytics

USE CASES



SEARCH
TECHNOLOGIES

Turning Big Data into Accessible, Searchable Business Intelligence

Along with cloud computing, social media, mobile, digital channels, and large repositories of business data comes the challenge of extracting critical, actionable information. Scattered data in multiple systems, costly data warehouse expenses, and complicated and slow SQL queries can all impede business success – but that does not have to be the case.

Enabled by NoSQL databases, advanced search technologies, and powerful open source platforms, big data and predictive analytics applications are no longer a future outlook for organizations; they play an essential role and should be integrated into every business strategy.

How is your organization defining the path for collecting, storing, enriching, and transforming data from disparate sources into real-time business intelligence?

This white paper identifies our top five big data and analytics use cases and the impact they can have on various businesses.



1 Financial Services

Utilizing Log Analytics for Risk Management and Mitigation

Banks use regular financial risk model calculations across all of their lines of business to determine their overall risk profile. Bank IT staffs run these calculations continuously to ensure that they comply with rules and that liquidity and cash balances are adequate. In order to sufficiently analyze and monitor application performance in real-time to ensure a near 100% service level requirement for these critical applications, banks can depend on open source log analytics solutions.

CHALLENGES

If banks do not have an open source log analytics solution in place, an application job failure can trigger minimum capital reserve requirements (potentially impacting banking operations and profits), draw the scrutiny of regulators, and leave the bank blind to the level of risk exposure.

IT staff might also spend hours poring over log files to trace problems. Job failure detection in this sense is manual and reactive and it can take hours just to isolate the problem in the log entries. Also, IT staff in development and QA might find code development and improvement difficult, as many developers might not be able to access the servers where the logs reside.



SOLUTION

Banks can transition to an open source log analytics solution that allows them to aggregate log data from all of their servers for analysis and alerting. This would reduce the time required to identify root causes of problems and allow IT teams to monitor their systems proactively.

ROI metrics potentially include:

Efficiency: Reduce time to identify root cause of problems

Quality: Reduce failure rate of risk calculation jobs

- Help Dev/QA teams debug and improve code
- Improve confidence
- Improve support for Federal Reserve Stress-tests

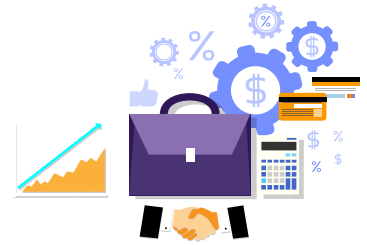
Financial: Reduce or eliminate “minimum capital reserve” requirements triggered when risk calculations fail

Regulatory: Meet Federal Reserve requirements

2

E-Commerce

Creating a Better Online Shopping Experience with Log Analytics



A system-wide logging infrastructure combined with log analytics is a powerful way to improve operational performance in IT and business. By analyzing event logs and system metrics that track performance of IT systems and user logs that capture behavior of users who interact with e-commerce sites, retailers can gather data and discover insights leading to more agile operations, improved competitive advantage, and increased site revenue.

CHALLENGES

The exponential growth in e-commerce activities and transactions for online retailers are generating terabytes or more of log data every day, making log management and analysis nearly impossible. The combined challenges of data volume and the associated storage and commercial software licensing costs have led many retailers to look for more cost-effective alternatives.

To put this into perspective, one of our retail customers had been paying hundreds of thousands per year for a commercial solution to analyze transaction data for just one of their e-commerce sites. They were quoted nearly three times this amount to quadruple the data for additional sites. They estimated that their data needs would continue to grow to terabytes of data which would cost them millions of dollars in the long run. Even at deep discounting they did not see how they could afford to keep a commercial solution.

SOLUTION

Retailers are moving toward open source technologies hosted on their own infrastructure. Open source log analytics solutions offer the necessary abilities to collect, process, and analyze massive volumes of log data without having to dump it into relational data warehouses for retrieval through SQL queries. With their customizability, speed, and affordability, open source technologies have emerged as solid competitors to commercial software. With open source technologies, retailers can better monitor their event logs, system metrics,

and user logs which contain valuable information that functional teams and executives can mine for performance-improving insight and decision-making.

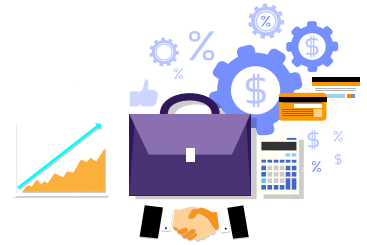
The analytics dashboard shown on the next page (Kibana 5), can be created for real-time data analyses and presentation. **Kibana** is the visualization component of the **Elastic Stack** – one of the most popular open source log analytics applications available today.

CONTINUE READING

2

E-Commerce

Creating a Better Online Shopping Experience with Log Analytics



SOLUTION

The business benefits of real-time log indexing include:

- More uptime and fewer outages
- Less fraud and reduced security risk
- Reduced growth in infrastructure
- Faster problem resolution

By moving from a commercial software vendor to open source, retailers can save millions of dollars per year with no volume restrictions other than available hardware. Open source software also makes new features and functions possible such as real-time end-user personalization which will lead to happier shoppers and ultimately, increased conversion and revenue.



3

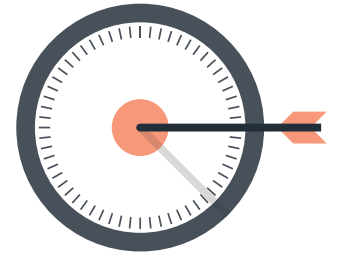
Market Research

Utilizing Big Data Analytics to Create Better Campaigns

Market research is a very important component of today's business strategy. It provides important information to identify and analyze the market size, market need, and competition. Market-research techniques typically encompass the analysis of data to gain insight or to support decision making. There are many market research firms worldwide that provide expert analysis around data and market research and oftentimes, the findings are accessible to subscribing online customers.

CHALLENGES

Many market research firms have not improved their market research collection and analysis processes for years. In addition, both the amount of data and the number of data sources have increased exponentially in recent years. Thus, without leveraging big data solutions and scalable data processing techniques, these firms are missing out on valuable takeaways and insights that could immensely benefit their subscribers.



SOLUTION

Combining big data processing techniques, search engines, and analytics applications, market research firms are able to get a complete view of their various media channels (i.e. email campaigns, print advertisements, banner advertisements, social media, mobile apps, video, etc.), analyze the efficacy of their various campaigns, and present the results to their online subscribers effectively through:

- Data acquisition from high volume feed
- Data processing and ingestion into search engines and big data platforms
- Data enrichment using machine learning, predictive analytics, manual evaluations, and other techniques
- Real-time analytics via custom visualization dashboards

As data accuracy and completeness enable insightful analyses and well-informed decisions, the ability to collect, process, analyze, and visualize vast amounts of data in real-time can give market research firms an unparalleled competitive advantage.



4

Precision Agriculture

Modernizing Agriculture with Big Data and Analytics

Agriculture has molded human history for centuries, since early human beings first domesticated plants and animals, and suddenly came across a new sustainable activity that encouraged them to settle, develop tools, and master techniques. With big data and analytics, we now have a method to tie agricultural data together and make agriculture an exact science, saving money, and increasing productivity.

CHALLENGES

Smart tractors, equipped with GPS, can collect a great amount of geo-coordination data and store information about product usage and covered areas. This data is often recorded in farm management software systems. Now, consider all tracked field operations applied to numerous crops every year, producing large amounts of valuable data stored in databases. Agricultural technology advancements have made collecting and storing data easy; but what about processing and analyzing this data to support productivity? How do we process and obtain information from all this data, which keeps growing in real-time?



SOLUTION

Raw data from farm management systems is extracted and fitted into a data domain model that is designed to facilitate data manipulation and analysis. This process is called ETL: Extract, Transform, and Load. This data model is populated by millions of records, creating a data lake that holds valuable information ready for being queried and discovered.

The data model contains different entities and collectively integrates to describe agricultural activities. It revolves around field operations, such as planting, pesticide and fertilizer applications, and harvesting.



[CONTINUE READING](#)

4

Precision Agriculture

Modernizing Agriculture with Big Data and Analytics

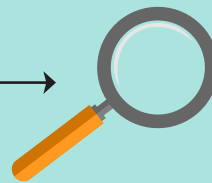


FARM MANAGEMENT SYSTEMS

(Relational Databases, XML Files, etc.)



BIG DATA FRAMEWORK



SEARCH ENGINE



USER

Enrichments

Inbound data is not only loaded, it is also enriched in various ways, for example:

- **Data normalization:** Farm management systems often allow the farmer to type free text, which generally create multiple representations of the same entity.
- **Organic soil composition:** Based on a particular field's geographical location, soil properties can be obtained from external sources.
- **Weather conditions:** Similarly, past weather conditions as well as forecasts for a particular location can be obtained from an external Application Programming Interface (API).

Technology Stack

A big data platform, such as Cloudera, Elastic, HPC, and many others, can be used to obtain, process, analyze, and massage the data in order to build the desired data model.

Subsets of the data lake are incrementally pushed from the big data storage platform into search engine indices which can vary in content and structure, depending on the search need of each case. Coupled with visualization tools, this solution can provide search power and facilitating data

analytics. Stakeholders can further examine the data lake in order to research more complex business enquiries, such as "Why 2016 was an awful year in terms of Oil Seed Rape Yield? What correlates to higher yield?"

By analyzing the data using the technology stack described above and statistical software, farming companies can create a statistical method to narrow down which specific variables amongst a discrete set (for example radiation, soil content, geographical location, fertilizer treatment, temperatures, etc.) correlate to higher yield.

The platform is in a favorable position to serve data to different actors in the industry, always keeping in mind a single goal: boosting farming productivity. Farmers, manufacturers, buyers, retailers, consumers, and pretty much anyone involved in the agricultural chain will ultimately benefit from applying big data and analytics to farming.

5

Precision Medicine

Discovering Cures with Big Data and Analytics

Healthcare organizations and research institutes struggle today with processing and extracting value out of the increasing deluge of genomics data triggered by the decreasing costs of genomics data sequencing. Recent efforts in the bioinformatics field have focused around porting and refactoring existing tools to work in Hadoop-based environments (such as **Cloudera CDH**) and loading genomics data to tabular-based storage systems (such as Hive, Impala, and HDFS/Parquet) for analysis and visualizations. Another similar approach for bringing genomics data into the big data world emphasizes the usage of search engines as the primary storage mechanism.

Search engines are particularly suitable to help solve these challenges as they allow for substantial performance improvements as well as query facilities not supported by SQL-based engines, like the ability to provide faceting, full-text search, and the execution of multi-dimensional studies across many data sets.

CHALLENGES

Several research institutions for large hospitals we've recently worked with have created data lakes where they ingest clinical data from their EMR systems, flow cytometry applications, DNA sequencing data for their patients, DNA mutations/ variations aggregated from multiple public databases, and medical literature content. But they've needed help linking all of these separate data elements together to provide a mechanism that allows researchers and physicians to explore, visualize, and easily corroborate data sets from all of these sources. It's often common that the data in data lakes is not in a format that is suitable for easy aggregation and fast access from UI applications. Much of the raw data is in unstructured form expressed as file formats specific to and accessible only from bioinformatics tools.



To put this into perspective, the scale of genomics information being collected is enormous. It projects that as we do sequence our genomes, it's going to be significantly greater than any amount of data, including the amount of data flowing through social media channels. So a solution needs to be built to ingest all of this information that healthcare organizations and research institutions possess in order to derive answers on how to improve current health and alleviate future risks.

5

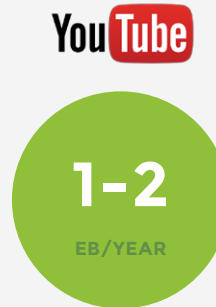
Precision Medicine

Discovering Cures with Big Data and Analytics

Big Data in 2015

Per Year Growth Estimates

Ref: Big Data Astronomical or Genomical
© 2015 Stephens et al
PB: Petabyte (1024 Terabytes)
EB: Exabyte (1024 Petabytes)



SOLUTION

By leveraging search engines, data can be parsed and ingested in a format easily accessible from web applications. The design for the search engine indexes is based on intelligent chromosome data sharding techniques which leads to substantial performance improvements for user queries.

On top of these search engine indexes, a research dashboard/application UI can be built that integrates clinical data, genomics data, and medical literature into a unified web-based interface that allows research institute users to perform cross-domain research studies and corroborate phenotypical with genomic data as well as visualize and analyze the data.

The research/dashboard application helps principal investigators to:

- Analyze and visualize the structured data
- Search over genome annotations data containing full-text
- Speed time to discovery of cures for children's diseases
- Ensure that research institutes can obtain funding more easily to pursue these cures

TOP 5 BIG DATA & Analytics USE CASES

TO LEARN MORE

Please contact Search Technologies to see how big data applications can be integrated into your systems for search, knowledge sharing, and analytics. If you are interested in how Search Technologies can help assess your system's big data readiness, please visit our [Big Data Application Assessment](#) resource.

ABOUT SEARCH TECHNOLOGIES

Search Technologies is the leading trusted and independent technology services firm specializing in the design, implementation, and management of search and big data analytics applications. We have delivered results for over 800 customers including industry leaders in e-commerce, publishing, media, financial services, professional staffing, manufacturing, as well as the government sector. Our experienced consultants and unique technical assets help us deliver customized search and big data analytics solutions that are easier to use, less expensive, more powerful, more reliable, and most importantly, aligned with your business objectives.

© 2017 Search Technologies. All Rights Reserved.

Find **Better** Answers.