

WHITE PAPER | JUNE 2017

A Blueprint for Advanced Analytics

Leveraging open source and advanced analytics to gain new customer insights

Bryan Whitmarsh
Director, Product Management, CA Technologies



Table of Contents

Introduction	3
Understanding the analytics agenda	3
Challenge	3
Unlocking insights can be costly and complex	3
Solution	4
Taking an open source approach	4
Flexible and reliable architecture	4
Adopting and developing best practices	5
Benefit	8
Seizing the analytics advantage	8
Conclusion	8
About the Author	9
Bryan Whitmarsh, Director, Product Management, CA Technologies	9

Introduction

Understanding the analytics agenda

In today's always-on world, customers expect a flawless experience—whether they're using a website, mobile app, smart watch or any other technology or channel. They expect an engaging and positive interaction that is tailored to their specific needs, at that specific moment.

To meet these demands, organizations need real-time insights into users' requirements, behaviors and actions—as well as visibility into the performance of the IT systems that support the customer experience. Although these insights are attainable, they often get lost in the growing volume and variety of data sets that exist today.

As a result, many organizations find themselves amassing plentiful data without maximizing its value. To unlock the insights needed to deliver richer customer experiences and better business outcomes, organizations need advanced analytics that yield maximum insights from data.

By capturing, correlating and analyzing different data sets and sources in real time, organizations can gain on-demand access to the meaningful insights they need. Only then can they deliver highly personal, “in-the-moment” customer experiences and ensure services are routinely available and secure.

Why analytics matter

75 percent of respondents believe IT analytics will help them gain a single view of the customer experience.

63 percent of respondents expect that IT analytics will enable them to predict problems before they have an impact on the business.

Source: **CA Technologies IT Monitoring and Analytics Tools Customer Research**, TechValidate, April 2017.

Challenge

Unlocking insights can be costly and complex

To uncover the insights that are now so critical to business success, organizations have invested billions in acquiring and implementing generic big data analytics tools, and training their employees to use them. However, the cost and complexity of this approach can often mean that it takes a long time to receive a return on this investment.

Off-the-shelf solutions can also come with limitations; their scalability might be inadequate or their flexibility restricted, which can hamper organizations' efforts to unlock the full business value of their data.

Building a custom analytics platform that is highly efficient, flexible and scalable will not only help organizations overcome these limitations but also enable them to probe more data sets. More compelling insights can be gleaned when data from different sources is correlated and analyzed together instead of in isolation.

CA Technologies embarked on a journey to build its own advanced analytics platform, and this experience can help provide a blueprint for organizations looking to establish advanced analytics.

Designing and deploying a custom analytics platform can be a massive challenge, however. Organizations need to ensure numerous open source products work together in harmony, and they also need to keep pace with changes to each of these components before, during and after implementation to prevent any performance issues.

Solution

Taking an open source approach

CA Technologies embarked on a journey to build its own advanced analytics platform, and this experience can help provide a blueprint for organizations looking to establish advanced analytics. Leveraging a combination of industry-leading open source and custom solutions, the CA platform offers advanced analytics across all the company's products.

By encouraging collaboration between multiple vendors and individual developers, open source software (OSS) can provide the core building blocks needed for common processes and systems, and so help reduce duplication and reinvention.

According to results from a CA survey, a third of organizations believe that the most important feature of an analytics platform is that it offer extensibility by leveraging open source technologies.¹

Built with a specific task in mind, OSS generally offers excellent functionality and performance. OSS projects are often among the first to feature new standards and approaches, meaning the solutions based on these projects often deliver enhancements in interoperability.

Flexible and reliable architecture

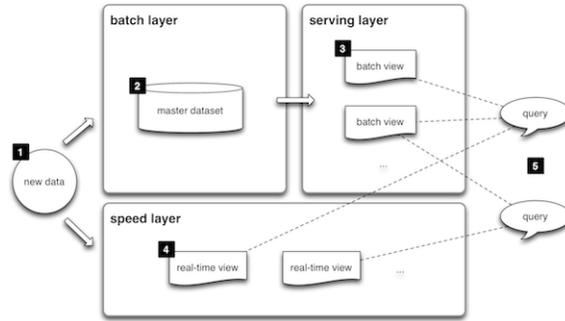
The goal of the new platform was to deliver real-time insights from behavioral, performance, operational and network data captured by a diverse range of CA solutions, including CA App Experience Analytics, CA Application Performance Management (CA APM) and CA Unified Infrastructure Management (CA UIM).

Beyond being flexible, the new analytics platform needed to be robust, fault-tolerant and scalable. The industry-proven Lambda Architecture, which is designed to handle massive quantities of data by taking advantage of both batch and stream methods, was an obvious choice. This architecture brings together three fundamental elements (see Figure A):

- A batch layer that aggregates data for the master data set and computes batch views ahead of time.
- A servicing layer that indexes the batch layer, so business insight gained can be presented without any latency.
- A speed layer that compensates for any latency in updates to the servicing layer by processing the most recent data.

By combining batch processing's comprehensive and accurate views of the master data set with stream processing's views of real-time data, Lambda balances visibility with reliability and responsiveness.

Figure A.
The Lambda Architecture



Adopting and developing best practices

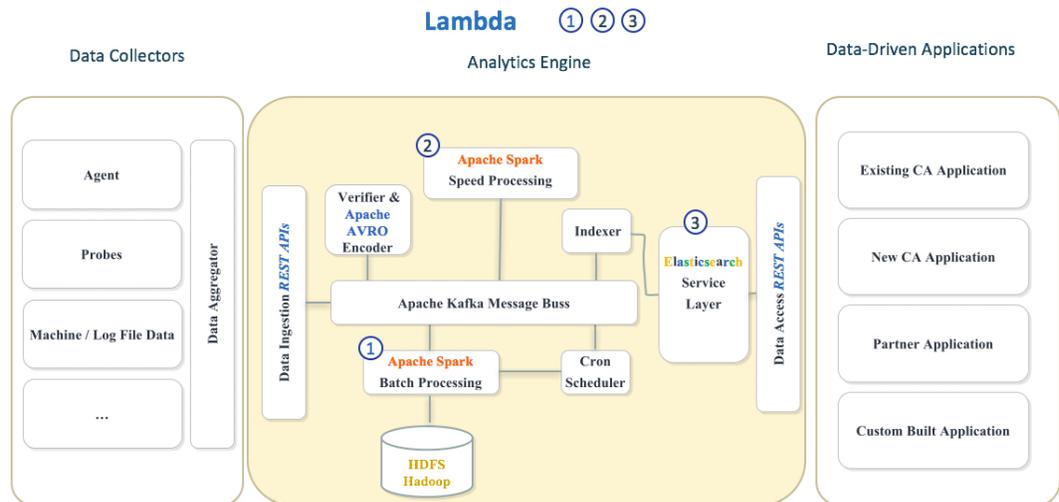
CA staff began building the new analytics platform within the company’s internal incubation program, CA Accelerator (see Figure B).

With more than 10 open source products to bring together, each moving at its own pace, the journey was complex and presented various challenges. We found that there is no single magic open source technology that can cover all the requirements for building an industry-leading

analytics platform. Different open source technologies were hand-picked based on their particular strengths. For example, Elasticsearch proved to be a great solution for the service layer of our Lambda-based architecture. However, before data can be delivered to the service layer, you first need an efficient message bus to load the data to. We used Kafka for this purpose. Then you need to process data on the Kafka bus for both your batch and speed layers. Apache Spark was selected for its capabilities in these areas. For batch processing you need a storage solution that is tuned for large amounts of historical data, where instant access isn’t a requirement. Apache Hadoop HDFS worked best for us in this area. Then you need a scheduler to kick off jobs and the list of requirements goes on (see Table A below).

With each open source technology, tuning needed to take place to best fit our use case. One example of this tuning is in regards to Elasticsearch shards. Shards are subcomponents (individual Lucene instances) of indexes in Elasticsearch environments. CA developers began to realize two key considerations: First, that the balance between the number of shards and indexes was very important and, second, that the optimal balance depends on the use case. The team discovered that the best balance depends on what kind of data you have, how your cluster is set up, how fast you want to ingest data and how rapid you need your search to be. Some additional Elasticsearch lessons learned can be found in the “Lessons learned” section below.

Figure B.
The architecture of the advanced analytics platform from CA



© 2017 CA. ALL RIGHTS RESERVED.

Table A. Individual Analytics Platform Components

Component	Role	Solution Used	Why It's a Good Fit
Data ingestion	Allows collectors to add data	REST APIs	REST APIs allow easy ingestion of data from a wide variety of sources.
Message bus	Provides a shared set of interfaces that allow individual components to share data	Kafka	Kafka offers a high-throughput, low-latency message bus, which is not only very efficient but also easy to use by both data owners and data consumers.
Verifier and encoder	Ensures data is valid, then encodes it to improve scalability and storage efficiency	Avro	Avro Encoding improves storage and processing efficiency.
Scheduler	Schedules batch jobs	CRON	CRON is a simple and efficient job scheduler.
Batch processing	Executes batch processing	Apache Spark	Apache Spark jobs are an efficient mechanism for accessing and manipulating large-scale data sets.
Data repository	Manages batch data	HDFS	HDFS offers a more efficient solution for longer term historical data storage.
Data science	Runs complex algorithms	Custom developed	A custom-built data science engine using machine learning to deliver intelligent, contextual insights.
Speed processing	Provides real-time speed layer processing	Apache Spark	Apache Spark jobs are an efficient mechanism for accessing and manipulating large-scale data sets.
Service layer	Provides access to queries	Elasticsearch	Elasticsearch is a distributed full-text search engine that is highly scalable and efficient and that allows searches across a wide variety of data types.
Indexer	Allows the data to flow into the service layer	Custom developed	After considering alternative open source tools for moving data off the Kafka bus and into Elasticsearch, we decided to build our own use-case specific indexer to mitigate the bottlenecks arising from high ingestion rates.

Data access	Accepts user requests for data via analytics applications, such as CA App Experience Analytics	REST APIs	REST APIs make it easy for a wide variety of applications to query the platform.
Data visualization	Offers an interface for creating queries and designing dashboards and provides a graphical view of query results	Kibana	Kibana slices and dices Elasticsearch data very easily—allowing near real-time ad-hoc queries.

Lessons learned from using Elasticsearch

One shard doesn't fit most: The right balance of shards per index is critical, but depends on your specific needs, such as whether you require high ingestion rates or rapid searching.

Know your nodes: Spend time understanding hot nodes, warm nodes, master nodes and client nodes and how they can improve speed and efficiency.

Don't split your brains: Don't allow replica nodes to promote themselves if a connection with the master node is lost.

Avoid shooting yourself in the foot: Elasticsearch enables you to delete your own indexes, so be careful not to accidentally delete yours.

Don't spin with disks: It is imperative to use solid-state drives (SSD). Floppy disks and tape drives don't cut it.

Remember the two Ps: Don't expect to get it right the first time. Planning and prototyping will save you time and money.

Swapping can be bad: For Elasticsearch, you need to avoid memory swapping at all costs, so disable it at the OS level.

Monitoring is important: Use multiple types of monitoring to detect and resolve issues—before your customers start calling you.

Multi-tenant best practice tips

Colorize the data: In multi-tenant environments, assign a tenant identifier to all incoming data so you can track it effectively.

Control access: Tenants shouldn't have direct access to Elasticsearch; build an authentication layer to protect data integrity.

Benefit

Seizing the analytics advantage

The new analytics platform helps CA unlock the full potential of its product offerings by equipping customers with compelling, real-time insights across their IT and business operations.

You can now combine correlated data from such products as CA App Experience Analytics, CA APM and CA UIM, so you can establish a unified view of the digital experience and gain deep application and infrastructure visibility. The analytics platform leverages this visibility to deliver true business insights, enabling organizations to retain and attract new customers and increase revenue.

CA expanded these capabilities even further by creating a solution to help IT operations teams enhance monitoring and accelerate troubleshooting. The solution helps teams rapidly visualize and analyze CA UIM log files coming from multiple sources across large-scale, distributed systems. The advanced analytics platform ingests the log files, displays them in easy-to-read dashboards and provides easy-to-use search facilities, accelerating the resolution of infrastructure performance issues.

The digital experience insights advantage

Improve user experience: CA App Experience Analytics delivers meaningful insights that help organizations continuously improve application quality, design, performance and user experience.

Prevent application outages: CA APM can detect unique data patterns, enabling organizations to identify and triage problems before users are affected.

Manage and monitor your infrastructure: CA UIM provides a single, analytics-driven solution for proactively and efficiently managing modern, cloud and hybrid IT infrastructures.

Conclusion

To thrive in today's always-on world, organizations need to consistently deliver the exceptional experiences customers demand. To succeed, organizations must fully harness their data so they gain the meaningful insights needed to tailor customer engagements and ensure systems and services are always available.

Generic big data analytics tools often lack the capabilities or flexibility organizations need. Combining multiple open source and custom components enables organizations to develop a highly efficient, flexible and scalable analytics platform that meets their current and future needs. However, this can be a complex and costly process, so many organizations will require the help of an experienced partner that can maximize the results and minimize the risks.

By drawing on experienced partners with proven integrations, organizations can seize the analytics advantage—so they can enrich customer experiences, increase agility, boost efficiency and realize long-term growth.

To learn more about the work CA has done in this area, visit our **advanced analytics page**. Also, be sure to visit the **digital experience insights page** and learn about more resources and upcoming events. For more information about CA, go to **ca.com**.

About the Author

Bryan Whitmarsh, Director, Product Management, CA Technologies

Bryan has been in the software industry for more than 20 years, starting out as an engineer, moving into engineering management, then into technical sales and product management. The mobile market is a specific area of focus for Bryan as he has been in the mobile industry since the late '90s. Bryan is currently a director in the agile operations business unit at CA, where he is primarily focused on external facing product management activities. As a CA Champion, he is very active in CA Communities as well as various public communities. He lives in Idaho and loves the outdoors.



Connect with CA Technologies at ca.com



CA Technologies (NASDAQ: CA) creates software that fuels transformation for companies and enables them to seize the opportunities of the application economy. Software is at the heart of every business, in every industry. From planning to development to management and security, CA is working with companies worldwide to change the way we live, transact and communicate—across mobile, private and public cloud, distributed and mainframe environments. Learn more at ca.com.