



CSC Expedites Big Data Application Development for Government

KEY HIGHLIGHTS

INDUSTRY

- > Public Sector

LOCATION

- > Washington, DC, US

BUSINESS APPLICATIONS SUPPORTED

- > Enterprise data hub empowering application development in the cloud

IMPACT

- > Applications developed in less than 30 days, vs. 8 months
- > Developers spend 80% of their time innovating and 20% administering infrastructure, vs. the reverse previously
- > Security comes as a service based on the framework, i.e. HIPAA, ITAR, etc.

TECHNOLOGIES IN USE

- > Hadoop Platform: Cloudera Enterprise
- > Hadoop Components: Apache Accumulo, Apache Spark, Apache Storm, Cloudera Search, Impala, Kerberos, MapReduce

Overview

[42six](#), a subsidiary of [Computer Sciences Corporation \(CSC\)](#), is a software development company and systems integrator for the U.S. Government Intelligence Community (IC) and the Department of Defense (DoD) that specializes in Big Data processing, analytics, and advanced applications support using distributed technologies.

The organization has developed an application engine called [EzBake \(EzBake.io\)](#), known as the [DoDIIS App Engine](#) within the government, which helps agencies including the [DIA](#) (Defense Intelligence Agency), [US Navy](#), [US Army](#), and other intelligence agencies efficiently migrate their analytics and data processing applications to enterprise data hubs running in the cloud, so they can take advantage of today's distributed technologies. EzBake, which runs on [Cloudera's big data platform](#), creates a faster, more secure, and scalable framework for applications to access data and to reduce the barrier for federal developers writing such applications. This is accomplished by providing the following features:

- > Classic platform-as-a-service (PaaS) utility to easily deploy apps without knowledge of DevOps tools
- > Technology agnostic approach, in that the framework supports any database or framework
- > Security throughout the framework with object-level security built into every database
- > A simple means to query for data using a distributed query ([Impala](#)) on top of all data stores, not just the [Hadoop Distributed File System](#) (HDFS) or [Apache HBase](#)

The solution helps analysts and end users gain access to more data in a very secure manner, allowing them to build better applications with greater productivity. With better applications, better tools, and more minds working together, agencies can drive increased innovation while focusing limited IT resources.

The Challenge

According to industry analyst firm [IDC](#), the 'Digital Universe' will multiply ten-fold between 2013 and 2020—from 4.4 trillion gigabytes (GB) to 44 trillion GB¹. This vast data presents challenges for any organization seeking to extract insights that will help them gain business advantage. For government agencies that exist to serve and protect our nation, the ability to look even more closely at data streaming in from new devices, and across many more data types, presents even greater challenges. Consider cost constraints relative to taxpayer-funded budgets, data security, and barriers to moving production applications from siloed operations to new distributed platforms in the cloud for greater cross agency analysis.

Most government agencies are limited in their technical ability to transition existing applications—operating in silos and with finite access to data sources—to become broadly available, cloud-based applications for expanded use.

¹ The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things. IDC, April 2014. <http://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm>

“The understanding of the ecosystem, and understanding of the evolution of Hadoop, is obviously a key factor of the business value of Cloudera. But it was also the understanding of the technologies and how they’re migrating and moving. So, for instance, the introduction of Impala was an important part of the decision, and the understanding of the need to insert new technology like Spark. It’s this knowledge base, not just the core product, that exists within the company.”

MATT CARROLL
CTO OF DEFENSE & INTELLIGENCE, CSC

As Matt Carroll, CSC’s chief technology officer of defense and intelligence explained, “The modern CIO really has a challenge: downward trending budgets. Yet they want to get more capability based on these technologies that are growing around the world. So they want more with less money. These customers are coming to us and asking, ‘How do we migrate our existing applications, within our existing portfolio, in an efficient manner to take advantage of these distributed technologies?’”

Carroll decided to address these challenges, and set out to build a scalable framework that would reduce the barrier to entry for federal developers and provide standardized security to meet their needs for Big Data applications, right out of the box.

Solution

After evaluating different options, Carroll led his organization to build an application engine called EzBake, comprised of [Cloudera Enterprise](#) at its core. “EzBake obfuscates the complexities of distributed technologies like [\[Apache\] Hadoop](#) to allow application engineers to rapidly take advantage of them,” explained Carroll. “While still maintaining the necessary elasticity and security to allow applications to rapidly come onto the engine within 30 days, they should be mission capable. And within one hour, they should be able to start development. That, as a whole, is EzBake.”

Underpinned by HDFS and open data formats, EzBake enables federal organizations to easily build centralized enterprise data hubs that support various user preferences, methods, workloads, and security requirements. Cloudera was uniquely suited to partner with CSC in this initiative, offering the most robust, stable, and versatile Hadoop platform that incorporates with modularized, leading-edge components of the stack including:

- > [Apache Accumulo](#): a derivative of HBase, which enforces a standard of access at the cell level
- > [Apache Spark \(incubating\)](#): making it easy to develop fast, unified Big Data applications combining batch, streaming, and interactive analytics
- > [Apache Storm](#) and [Apache Kafka](#): allowing data ingest at unparalleled rates
- > [Cloudera Search](#): enabling data discovery and exploration for non-SQL or Java users
- > [Impala](#): offering near real time analytics through interactive SQL for Hadoop

“Partnering with Cloudera was a natural fit,” Carroll explained. “[Cloudera’s Hadoop distribution](#) is, without a doubt, the most premier and stable one out there. The understanding of the ecosystem, and understanding of the evolution of Hadoop, is obviously a key factor of the business value of Cloudera. But it was also the understanding of the technologies and how they’re migrating and moving. So, for instance, the introduction of Impala was an important part of the decision, and the understanding of the need to insert new technology like Spark. It’s this knowledge base, not just the core product, that exists within the company.”

Impact: Higher Productivity, Faster Innovation

By deploying Cloudera within its EzBake application engine, CSC has built a scalable framework that reduces the barrier to entry for federal developers to migrate their Big Data applications to a cloud-based enterprise data hub environment, with standardized security right out of the box. The solution helps analysts and end users gain access to more data, in a more secure manner, so they can build more applications, faster, and with streamlined migration to the cloud.

“People want data at their fingertips,” noted Carroll. “We are a species that loves sensors. Data is always available, and the cloud provides a means to deliver that to the right person at the right place at the right time. That’s ultimately the goal of the enterprise data hub in the cloud.”

“With EzBake, developers focus 80% of their time building new applications and 20% of their time on infrastructure and integration tasks, versus the reverse which was the case previously.”

As a result of CSC's innovative use of Cloudera, U.S. government agencies are putting their annual IT budgets to better use, building new applications with greater efficiency and developing new features that add value. Whereas typical application development used to take six to eight months, developers can now create those same applications in under 30 days. Freed from time-consuming migration, infrastructure, and integration tasks, developers can now focus 80% of their time building new innovations, as opposed to 20%, which was the case prior to using EzBake.

Public agencies benefit from nearly 10X scalability of their data management infrastructures, as standardized security meeting federal mandates is automatically embedded across all data and all applications. In addition, with automated resource management, CSC can now support over 400 different projects, enabling simultaneous deployment and testing in unison, which is helpful for IT staff maintaining the system.

About Cloudera

Cloudera is revolutionizing enterprise data management by offering the first unified Platform for Big Data: The Enterprise Data Hub. Cloudera offers enterprises one place to store, process and analyze all their data, empowering them to extend the value of existing investments while enabling fundamental new ways to derive value from their data. Founded in 2008, Cloudera was the first and is still today the leading provider and supporter of Hadoop for the enterprise. Cloudera also offers software for business critical data challenges including storage, access, management, analysis, security and search. With over 20,000 individuals trained, Cloudera is a leading educator of data professionals, offering the industry's broadest array of Hadoop training and certification programs. Cloudera works with over 900 hardware, software and services partners to meet customers' big data goals. Leading organizations in every industry run Cloudera in production, including finance, telecommunications, retail, internet, utilities, oil and gas, healthcare, biopharmaceuticals, networking and media, plus top public sector organizations globally. www.cloudera.com