UK Data Service Achieves Secure Modern Big Data and Analytics on AWS

Challenge
The UK Data Service (UKDS) needed to modernize its data and analytics platform to deliver and provide intelligence insights to researchers in an efficient and cost-effective manner. A solid technology framework was necessary to cost-effectively support large data sets, provide strong data governance, and deliver robust analytics.

Solution
UKDS chose open source technologies to avoid vendor lock-in, including Apache Nifi, Tomcat and Hadoop, and worked with Cloudwick, AWS and Hortonworks, to successfully implement a hybrid solution that provides strong security and data governance, enabling researchers get more out of big data.

Benefits
• Strong security and data governance ensures confidential personal data is protected
• Heavy processing and graphics are pushed to the cloud, resulting in cost savings, scalability, elasticity and flexibility
• The hybrid solution presents data in a way that is useful to researchers

Leveraging the Data Lake with Hadoop and AWS

This dilemma led the UKDS to select a hybrid model, with a data lake that would link their data with an on-premise and cloud-based solution. “The hybrid model enables tight security but also provides a seamless way of delivering information to those who need it,” said Cunningham. “In addition to meeting our vast data storage needs, the framework would need to be secure but not so much so that nobody uses it.”

UKDS wanted to use open source technologies to avoid vendor lock-in, which is important to the public sector. It would also need to be a repeatable and sharable solution, a design other organizations could eventually leverage. UKDS chose open source Hadoop to support large data set processing and storage, specifically Hortonworks Data Platform (HDP®), along with Amazon Web Services (AWS) as its cloud solution.

While academic organizations prefer to build systems themselves, it is not always optimal and efficient, especially when budget is limited and technologies are still maturing, like Hadoop. Cunningham put together a comprehensive team for the project, assigning graphic artists, IT staff, and engineers, but it was missing Hadoop experts. Hortonworks solved this by introducing UKDS to Cloudwick, the leading provider of open source services with extensive cloud and big data experience.

“I knew partnering with an experienced organization like Cloudwick, which had successfully implemented a solution like this one many times, was a must.”
- Nathan Cunningham, UKDS

Cloudwick Delivers AWS Solution with Strict Security and Data Governance for Appropriate User Accessibility

Founded in 1967, UK Data Archive (UKDA) in Colchester, is an internationally renowned organisation that acquires and curates the UK’s largest collection of social and economic data. Funded by the Economic and Social Research Council (ESRC), it has been designated a Place of Deposit by the National Archives, allowing it to curate public records. It is also the lead organisation of the UK Data Service (UKDS), which provides access to the UK’s largest collection of regional, national and international social, economic and population data.

Nathan Cunningham, UKDS Associate Director, leads the organisation’s Big Data Network Support (BDNS), which provides users with access, support and guidance on big data. “Big data generated by the public and commercial sectors represents a significant resource for society and it can provide a researchers with valuable insight,” said Cunningham. “The BDNS helps researchers make the most of big data for knowledge exchange and impact. Thus, in order to present data in a way that is useful to researchers, we needed the right framework.”

The UKDS big data is multi-dimensional because it draws on how things have changed over time - rather than being long and thin, this data is extremely wide. The organization needed a technology stack that would enable researchers to see data together and derive future insights, which was a contextual, economic and governance challenge. “We know that citizens want researchers to have all the information technologically possible at their fingertips, but they also want it to be kept secure,” said Cunningham.

Cloudwick
sales@cloudwick.com
www.cloudwick.com
this one many times, was a must,” said Cunningham. “Cloudwick proved to be invaluable, coming to the table with years of Hadoop and AWS experience and expertise, and they managed the rollout much more cost-effectively than an internal team with limited experience could have.”

**Expertise in Building the Data Lake**

Cloudwick migrated data from the organization’s SQL database to a cloud and on-premises Hadoop-based data lake using Apache Tomcat, an open-source Java Servlet Container that provides a web server environment for Java code. Heavy processing and graphics are pushed to the cloud yet the data owner can still trace it in real-time, which is crucial to obtaining accurate analysis and results.

HDP® delivers consistent operations, security, data governance and robust analytics across the UKDS clusters, while Hortonworks DataFlow (HDF), built entirely on open source software, including Apache NiFi, accelerates the flow of data in motion into HDP. NiFi, with its enterprise-grade security and management at scale, is used to ingest data, assign access levels and enable users to develop data flows with a user-friendly drag-and-drop graphical user interface.

The UKDS has worked closely with data owners to structure license agreements assigning data governance, decided by data owners, in one of three ways: open (accessible to anyone), safeguarded (available only to those with specific access rights, such as a paid subscription), and secure (accessible to nobody except the data owner, such as social security numbers). NiFi, Hadoop and AWS deliver elasticity, security, scalability and flexibility for the UKDS, paving the way for other organizations to review the solution and use the referenceable architecture for their own implementations.

“Cloudwick understood our vision and never tried to give me a sugar coated presentation, telling me everything I needed to know,” said Cunningham. “They understood our requirements and our budget constraints. It’s not always common to find a partner with the desire and dedication to making something work and providing insight on how it could evolve in the future. Both Cloudwick and AWS were dedicated to getting the solution to work.”

The hybrid solution provides UKDS with an ISO 27001 compliant platform that is secure, encrypted and auditable, delivering appropriate data access to everybody who needs it. “Cloudwick and AWS have delivered a solution that enables researchers to get more out of big data,” said Cunningham. “In turn, we will share the design, plan and cost details with other organizations and further maximize the return. We’re very happy with the outcome and appreciate the expertise, experience and support provided by Cloudwick, Hortonworks and AWS.”