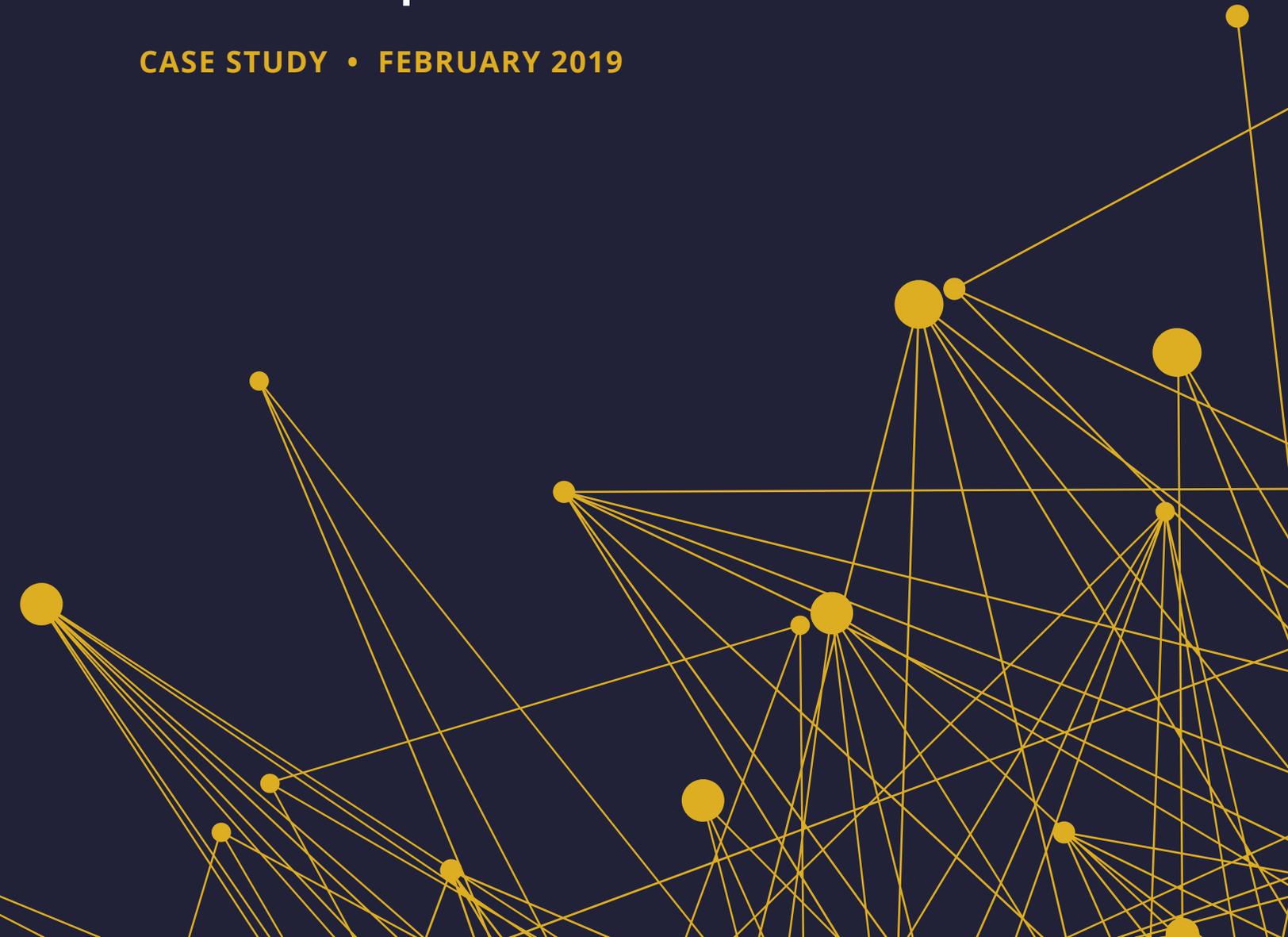




Io-Tahoe's Smart Data Discovery accelerates insights from Centrica's sensitive data landscape

CASE STUDY • FEBRUARY 2019



The Company

Centrica plc is a British multinational energy and services company with its headquarters in Windsor, Berkshire. Its principal activity is the supply of energy and services to over 25 million customer accounts mainly in the UK, Ireland and North America through strong brands such as British Gas, Direct Energy and Bord Gáis Energy, all supported by around 15,000 engineers and technicians.

The Challenge

“We knew that Centrica had an enormous amount of information about its customers globally,” says Mike Young, Group Chief Information Officer. “We found that we’d accumulated a significant amount of data. We also had confidential corporate information about profit and loss statements as well.

“In a GDPR-driven business environment, this meant we had to figure out precisely what we had and where it resided,” Young says. “In smaller organizations, that may be done on an anecdotal basis, with people responding and analyzing data based on certain questions. Many companies start and finish there, since they’re not particularly big. Centrica, however, is a global behemoth, with over 25 million customers worldwide. We needed to approach this effort far differently, as our sheer depth and breadth of services was a complicating factor; even though people think of us as a gas and electric service, we provide much more.”

Young says that implementing a data discovery strategy became a key focus, “so that we could prove to the Information Commissioner’s Office (ICO) that we applied science to the issue in an effort to resolve our challenge.” The ICO is the independent regulatory office tasked with dealing with a wide range of privacy regulations, including the UK Data Protection Act and the General Data Protection Regulation (GDPR), the Privacy and Electronic Communications Regulations across the UK; and the Freedom of Information Act and the Environmental Information Regulations.

Young and his team, led by Daljit Rehal, Centrica’s Senior VP Digital and Data and Chief Data Officer, wanted to approach their task from a molecular level: “Not only would we be able to tell the ICO that we have customers’ names, but we can say where those names reside in a particular database...within a particular schema, field or more. We could tell them that we have names located in several different locations.

“We wanted to be fully above board in our efforts,” Rehal says. “We knew that would require a significant data discovery effort, given the amount of information with which we were dealing, and we knew it would require advanced technology.”

The Solution: Io-Tahoe

Centrica’s data discovery project began in September 2017. A decision was quickly made to use Io-Tahoe as the company’s smart data discovery platform.

“Going in, we realized there was not much information being shared between business units across the company,” Rehal says. “But we also realized that by combining that information in a data lake, and using Io-Tahoe, we could turn it into a strategic asset.”

The Centrica team initially deployed Io-Tahoe in a pilot project across four data sources, to demonstrate the software’s value. “We utilized Io-Tahoe’s sensitive data discovery capability, Rehal adds. “And it worked very well. We were able to process 30 billion records and 1.7 million columns in a fraction of the time we anticipated. We reported the results we received back to our architecture team, and they endorsed it for data discovery.” Given Centrica’s sheer size, even the initial project was significant in scope: Rehal and his team were able to deploy Io-Tahoe across 16 HP servers, each holding petabyte-sized loads. “One data source was 40TB alone,” he says. “We ended up working with more than 1200 databases and 1500 applications, covering millions of customers and associates.

“Using Io-Tahoe enabled us to connect to each data source and scan each source. It certainly validated a whole heap of assumptions, such as duplicated data across multiple sources. We knew that would be

the case, but Io-Tahoe allowed us to quickly articulate what and where it was,” Rehal notes. As a result, Centrica has a solid picture of what sensitive data resides where; “By knowing that, we can rank or classify our apps by potential risk. We know a lot more, far more than when we started, such as which applications contain the most amount of personal information, so that we can triage any remediation activities.”

Rehal says it would have been “absolutely impossible” to accomplish the project manually. “I’m not even sure this level insight could have been gained if it was done manually. It would have certainly diverted valuable resources away from other high priority business activities. To put things into perspective, we have a side project, taking personal information residing in non-production systems, and had a third party look at that. They spent eight months cataloging four data sources. Using Io-Tahoe, we did 22 data sources in one month.”

By implementing Io-Tahoe as Centrica’s smart data discovery platform, Rehal and his team are able to know where to look for any given data, as well as establishing who has access to it and for what purpose. They can also determine who has the responsibility for keeping the data valid and pertinent. “There are a number of steps we needed to take to fully implement a GDPR-compliant scenario, and Io-Tahoe’s discovery functionality has played a key role in helping us achieve our goal. Our customers have a right to know what data we hold about them. Our work has enabled that team to narrow the search. We know where that personal data resides exactly, so that when someone contacts us, we know exactly where to look. We’ve developed a more automated approach to subject access requests. Previously, we had to go, logging on from app to app, and it was very much a manual process. Io-Tahoe has enabled us to develop a streamlined solution that’s sharply reduced both our time and legal exposure.” It’s not just a one-off, either; Rehal says Centrica may receive several thousand requests concurrently. Each one, under the old manual system, could take hours or days to fulfill.

A key advantage, Rehal says, was Io-Tahoe’s ability to work with the data itself, without having to rely on metadata. This, he says, enabled his team to find hidden data elements that wouldn’t have been identified without a smart data discovery ML-based solution like Io-Tahoe. “Dynamic profiling proved essential; we found data which would not have been possible without absorbing vast amounts of Subject Matter Experts’ time. For example, we scanned SAP applications as many field names have Germanic origins or are not intuitive by just looking at the name. Io-Tahoe enabled us to swiftly complete the task,” he says.

The Future

“Doing change programs gets you thinking about privacy up front instead of an afterthought,” Young says. “You realize that privacy is not merely a bolt-on.” Using Io-Tahoe for smart data discovery has enabled Centrica to achieve results his team had not initially anticipated. “When my data privacy officer wants to know about his exposure on a given issue, we can quickly tell him, and he’s happy about that. We find ourselves working with him more closely. More and more of what we’re doing is embedded into our technology, so that we can address emerging issues like ePR, and others. We’ve even started up a program for data privacy by design.

“Io-Tahoe has proven itself repeatedly to be an effective platform. The sensitive data component we’ve used has certainly worked very well from a solution point of view. The interface is intuitive and the ability to understand the dashboards doesn’t take long at all to get. We’ve been using Io-Tahoe as it’s being developed,” Young says. He even says that when Centrica wanted to incorporate a feature which wasn’t part of the initial implementation, the Io-Tahoe team responded. “They were definitely receptive. We discussed this with the development team, and they were super quick to respond. They exceeded our expectations.”

Io-Tahoe and its smart data discovery capability is a key part of Centrica’s future. “Thanks to Io-Tahoe, we have a far better understanding of where all of our customers’ personal and sensitive data resides. There’s no need to stop what we’re doing; we’re making plans to use Io-Tahoe throughout all of Centrica,” Young concludes.



Confidentiality Notice

This file is private and may contain confidential and proprietary information. If you have received this file in error, please notify us and remove it from your system, and note that you must not copy, distribute or take any action in reliance on it. Any unauthorized use or disclosure of the contents of this file is not permitted and may be unlawful. Io-Tahoe LLC, 111 Broadway, Suite 601, New York, NY 10006. www.io-tahoe.com