

USING DATA ANALYTICS TO BUILD A STABLE FUTURE

>500X

Improvement on the speed of reporting

Impact

- 'One Bank Data Architecture' A centralised data platform to break down data silos and facilitate research
- More than 500x improvement on the speed of reporting, from 2 days to under 5 minutes
- Data is reliable and compliant

The Bank of England is the central bank of the United Kingdom formed in 1694. The Bank of England's mission is "to promote the good of UK citizens by maintaining monetary and financial stability". As well as managing the UK's currency, supply of money and interest rates, the institute has a diverse range of responsibilities including gathering and analyzing data from banks, building societies, credit unions, insurers and mortgage companies to inform policy decisions and guide UK government departments and international organizations.

To maintain financial stability the Bank of England uses a number of key mechanisms, policies and frameworks. As the 'Lender of Last Resort' - where if commercial banks suffer a shortfall they can borrow money from the Bank of England - it supervises and stress tests banks, insurers, and building societies for the potential impact of hypothetical adverse scenarios. This is a complex data intensive modeling exercise, enabling the Bank to assess the resiliency of each organisation to risks and to ensure they have enough capital to withstand these events.

The Bank of England has more than 400 economists, hundreds of analysts, data scientists and modellers, all of whom support the Financial Policy Committee (FPC) and Monetary Policy Committee (MPC) in making vital data-driven decisions.

Removing data silos

For over a decade the Bank of England has been successful in building business intelligence systems, serving operational analysts and economists to view datasets via pre-developed reports and OLAP cubes. However, these systems take a considerable amount of time to build and are usually funded locally by a team with a specific remit. The result is that datasets have become silo'd making it difficult to combine with other contextual datasets at speed or scale.

In addition, there was a growing demand from the analytical community for access to an increasingly diverse range of datasets including those self-obtained, and the ambition to apply sophisticated statistical data modelling techniques to these.

"Using our traditional stack of tools to satisfy the demands of our analysts and economists was becoming more of a challenge" said Nick Vaughan, Domain SME for Data Analytics & Modelling at the Bank of England. "In 2014 the Bank launched its strategic plan, 'One Mission, One Bank' and a key pillar of that strategy was the 'One Bank Data Architecture'. This plan called for, amongst other things, a centralised data platform to break down data silos and facilitate bank-wide research. So we needed a system that would allow us to collect, store and process larger datasets and share these with our researchers and data science community, as well as support operational systems. The EMIR (European Market Infrastructure Regulation) data presented our biggest and most immediate challenge in this respect."

Transforming to connected, fast and open data analytics

The Bank of England set about transforming its use of traditional data tools, with its first forays into Hadoop technologies occurring within the Business Intelligence teams in 2013. For ease of use, developers began experimenting with Microsoft Azure's HDInsight to learn more about HDFS/Hive and associated data processing technologies.

“Instead of processing and aggregating smaller slices of EMIR data sequentially and threading back together for final analysis, using HDP our analysts are now able to take their code to the granular dataset in its entirety. HDP has also enabled us to successfully and efficiently collect other datasets for research and analysis alongside our established EMIR Hive based warehouse.”

Adrian Waddy, Data Platform Delivery Lead, Bank of England

By then deploying a small Hortonworks Data Platform (HDP) instance on premise, developers were able to further progress their learnings with more relevant datasets and use cases to the Bank of England, and subsequently put these into practice with the creation of a social media monitoring system in the run up to the Scottish Referendum vote in 2014.

In 2019 the Bank of England now has multiple production environments supporting a number of systems, including the collection of EMIR (European Market Infrastructure Regulation) data, relating to derivative trades that involve any UK regulated entity. The Bank collects 40 million rows of derivative transactions every day, from six external trade repositories across 80 individual files and with varying schemas over time. Utilizing HDP, the analysts are able to significantly reduce the time it takes to process a dataset containing over 15 billion (and growing) derivative transactions. Report packs can be produced in five minutes where previously they took two days for instance.

“Instead of processing and aggregating smaller slices of EMIR data sequentially and threading back together for final analysis, using HDP our analysts are now able to take their code to the granular dataset in its entirety,” said Adrian Waddy Data Platform Delivery Lead at the Bank of England. “HDP has also enabled us to successfully and efficiently collect other datasets for research and analysis alongside our established EMIR Hive based warehouse.”

The Bank has recently introduced an additional HDP cluster tuned specifically for Spark to support EMIR data specialists in applying more sophisticated data science techniques for big data using a variety of open source tools – such as SparkR, PySpark, SparkSQL and distributed machine learning through MLlib.

Delivering a future-proofed data platform

The Bank of England is now in the process of building a second data platform, five times larger than the existing platform and spanning multiple data centres. This will allow it to analyze many more large datasets, such as those relating to securities transactions which are larger than the EMIR datasets, while also supporting multiple operational reporting systems and hundreds of researchers, analysts and econometricians. It also plans to bring more data from the legacy data warehouse estate onto the data platform overtime.

Ultimately, by delivering a key component of the ‘One Mission, One Bank’ Strategic Plan, the Bank of England is starting to leverage the benefits of a central data platform. The Bank of England can continue to fulfil its duty of providing a stable and reliable foundation for the UK economy and broader banking ecosystem. In a fiercely competitive environment where innovation is transforming the way services are delivered by banks, change is inevitable, but the data sitting behind needs to remain reliable and compliant to regulation both in the UK and beyond.

“Economics and finance needs to make an on-going investment in Big Data and data analytics if it is to rebalance the methodological scales. And early research, including at the Bank, suggests the returns to such activity could be high, deepening our understanding of the economy and financial system.

These returns will best be harvested if there is strong collaboration between statistical authorities, policymakers, the commercial sector, research centres and academia. The Bank of England can play a catalytic role in bringing this expertise together.” - Andrew Haldane, Chief Economist, 19th April 2018