



Business challenge

Today, the most advanced, exciting data science technologies come from the open source community. How could Fifth Third Bank safely integrate these tools into its highly secure IT environment?

Transformation

Fifth Third Bank used IBM® Data Science Experience to build a powerful, scalable environment that provides access to Jupyter Notebook, R and Python libraries, while meeting IT security requirements.

Business benefits:

Unlocks

the ability to harness cutting-edge machine learning and deep learning techniques

Maintains

security with a single point of control for open-source tools and data

Scales

seamlessly and cost-effectively as demand for analytics increases

Fifth Third Bank Builds a security-rich environment for state-of-the-art data science

Fifth Third Bancorp is a diversified financial services company headquartered in Cincinnati, Ohio. As of March 31, 2018, the Company had \$142 billion in assets and operated 1,153 full-service Banking Centers and 2,459 ATMs with Fifth Third branding in Ohio, Kentucky, Indiana, Michigan, Illinois, Florida, Tennessee, West Virginia, Georgia and North Carolina.

“Data Science Experience allows us to explore many more models and run many more iterations in the same amount of time, helping us get the results the business needs, fast.”

Jeffrey Allard

Lead Data Scientist, Decision Sciences Group
Fifth Third Bank

Share this



Embracing a new generation of data science tools

In the past few years, data science has emerged from the shadow of statistics and become a central pillar of analytics strategy for many businesses. Advances in hardware, such as the availability of efficient GPU clusters, now make it practical for companies to train complex neural networks cost-effectively, while new software frameworks are helping to simplify model design, development and deployment.

The real surprise, from an enterprise perspective, is that much of the innovation in data science tooling is coming from open source communities. In many cases, the primary expertise of data scientists graduating from university today is in open source languages like R and Python, frameworks like TensorFlow and Keras, and development tools like Jupyter Notebook, rather than legacy statistics software.

The Decision Sciences Group (DSG) at Fifth Third Bank quickly recognized the potential of open source software to help it gain a competitive advantage—both to optimize marketing analytics and to help it attract top data science talent.



When the bank decided to retool its data science practice, the DSG team realized that it had a golden opportunity to embrace open source. Instead of rewriting thousands of lines of code to pipe data into its legacy models, the team decided to move away from the legacy modeling platform altogether, and rebuild its models from scratch in Python and R.

Security, however, posed a challenge. In many industries, data scientists might be free to download the latest frameworks, and start building models right away—but in a financial services environment, where data governance is subject to strict regulations, it was vital to impose much tighter safeguards for the use of customer data.

Brian Robinson, Principal Application Developer at the bank, comments: “The question was, as a bank, how do we leverage open source software safely to build state-of-the-art models? We needed to create a controlled environment that would empower our data scientists to access these tools, while still passing muster with our IT security standards.”

Building a safe environment for data science

Fifth Third Bank decided to build a Hadoop cluster to support big data analytics and train models efficiently. By deploying the cluster on-premises, the bank would ensure that all its sensitive data remained within its own firewall at all times—helping to maintain compliance with IT security policies.

Jeffrey Allard, Lead Data Scientist for the Decision Sciences Group, says: “We looked at various Hadoop offerings from different vendors, but we were agnostic about the big data infrastructure itself. Our key requirements were around data science tooling and security.”

When the team saw a demo of [IBM Data Science Experience](#), they quickly recognized its potential to meet these requirements. Data Science Experience is a platform that gives data science teams access to a broad range of open source and enterprise tools covering the whole data science lifecycle from data preparation to model development. As a hybrid solution, it can be deployed on-premises, or accessed via public or private clouds.

“At first, we just set up personal accounts in the cloud version of Data Science Experience, and saw how we could use Jupyter Notebook for model development,” says Jeffrey Allard. “If we could deploy the same solution on-premises with Data Science Experience Local, it would give us the controlled environment we needed to build models with real banking data.”

The bank worked with IBM to set up a small IBM Data Science Experience Local cluster with three nodes. The environment quickly grew—first to a nine-node, and then to an 11-node cluster—as the DSG team began moving more of its model training workloads onto the platform.

Brian Robinson says: “From an IT perspective, Data Science Experience Local makes growth seamless, because it scales horizontally. When you need more power, you just add more nodes.”

Data Science Experience enables DSG and other groups within Fifth Third Bank to access both the data and the open-source tools and frameworks they need, within a controlled environment. Instead of data scientists downloading random packages (which could be infected with malware) to their own PCs, each package is downloaded centrally and scanned by the bank’s security team before it is made available to users.



The DSG team has already used Jupyter Notebook within Data Science Experience to build new models to solve a wide range of marketing and customer analytics problems. Examples include helping direct marketing teams target the right prospects in a database of potential customers, predicting the balance and revenue from new customers’ accounts, and assessing the likelihood of customer attrition over the next five years.

“In addition to more traditional statistical models, we’re also using machine learning and deep learning techniques,” says Jeffrey Allard. “For example, we’re using neural networks to help with forecasting and sequence modeling for customer value analysis and marketing optimization.

“With access to leading-edge natural language processing frameworks in Data Science Experience, we can analyze unstructured text in ways that weren’t even feasible with our legacy platform.”

Faster, safer, and more cost-effective

Data Science Experience is helping Fifth Third Bank achieve greater accuracy and much faster training times for its models, compared to its legacy platform.

Jeffrey Allard comments: “The gain in performance that we get from the new open source tools is significant, even when we are just training a model on a laptop. When we run a big training job on the Data Science Experience cluster, there’s no contest—the new platform allows us to explore many more models and run many more iterations in the same amount of time, helping us get the results the business needs, fast.”

More importantly, the DSG team is able to harness new open source tools and frameworks confidently, without worrying about data governance and security issues. Every package available in Data Science Experience has been thoroughly checked and audited by the bank’s security experts, helping to reduce risk and facilitate compliance with information security policies.

The bank also expects to be able to grow the environment without significantly increasing its infrastructure costs. Brian Robinson says: “Each node in the cluster is just a virtual Linux server, and those servers can run on commodity hardware. As a result, adding new nodes is much less expensive than upgrading the enterprise-class servers that support our legacy platform.”

In the future, the DSG team hopes to take advantage of more features of Data Science Experience, such as automated model deployment. The team also plans to offer training to other groups within the business, to help them move their models to the new platform and share skills and data more effectively.

Jeffrey Allard concludes: “Data Science Experience gives us a powerful platform that we can use to build the future of data science at Fifth Third Bank. By giving us access to state-of-the-art open-source tools in a safe, fast and scalable environment, Data Science Experience is helping us transform customer analytics and attract a new generation of data scientists to an exciting career with Fifth Third.”

“With access to leading-edge natural language processing frameworks in Data Science Experience, we can analyze unstructured text in ways that weren’t even feasible with our legacy platform.”

Jeffrey Allard, Lead Data Scientist, Decision Sciences Group, Fifth Third Bank

Solution components

- IBM® Data Science Experience

Take the next step

IBM Data Science Experience is the enterprise data science platform that allows teams to explore, build and put their data science practice into production faster. For more information about how IBM Data Science Experience can transform industries and professions with data, visit ibm.com/products/data-science-experience. Follow us on Twitter at @IBMDataScience, on our blog at ibmbigdatahub.com and join the conversation #DSX.

Connect with us



© Copyright IBM Corporation 2018, IBM Corporation, 1 New Orchard Road, Armonk, NY 10504 U.S.A. Produced in the United States of America, August 2018.

IBM, the IBM logo, and ibm.com are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at “Copyright and trademark information” at ibm.com/legal/copytrade.shtml.

Not all offerings are available in every country in which IBM operates.

The performance data and client examples cited are presented for illustrative purposes only. Actual performance results may vary depending on specific configurations and operating conditions.

All client examples cited or described are presented as illustrations of the manner in which some clients have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual client configurations and conditions. Contact IBM to see what we can do for you.

It is the user’s responsibility to evaluate and verify the operation of any other products or programs with IBM products and programs.

The client is responsible for ensuring compliance with laws and regulations applicable to it. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the client is in compliance with any law or regulation.

Statement of Good Security Practices: IT system security involves protecting systems and information through prevention, detection and response to improper access from within and outside your enterprise. Improper access can result in information being altered, destroyed, misappropriated or misused or can result in damage to or misuse of your systems, including for use in attacks on others. No IT system or product should be considered completely secure and no single product, service or security measure can be completely effective in preventing improper use or access. IBM systems, products and services are designed to be part of a lawful, comprehensive security approach, which will necessarily involve additional operational procedures, and may require other systems, products or services to be most effective. IBM DOES NOT WARRANT THAT ANY SYSTEMS, PRODUCTS OR SERVICES ARE IMMUNE FROM, OR WILL MAKE YOUR ENTERPRISE IMMUNE FROM, THE MALICIOUS OR ILLEGAL CONDUCT OF ANY PARTY.



13017713-USEN-00

