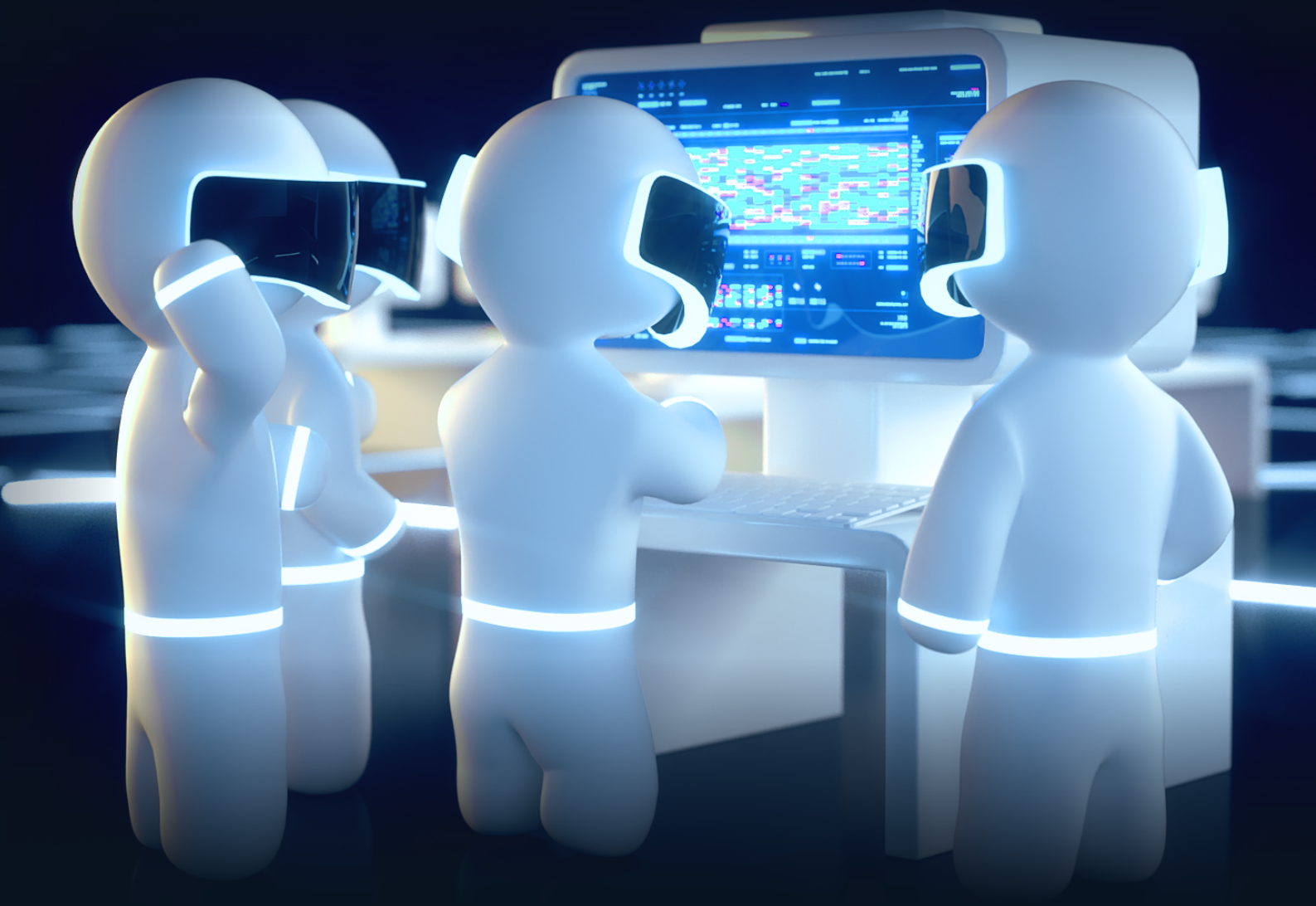
dotfive™



Department of Health & Social Care

NHS Test and Trace

Designing and implementing a cutting-edge COVID-19 data analytics platform and infrastructure

Providing critical technical knowledge and expertise to attain a high level of data quality, to support government decisions in response to a global pandemic.

Designing and implementing a cutting-edge COVID-19 data analytics platform and infrastructure

In 2020, the world changed.

Out of nowhere, every country on the planet was affected by the emerging COVID-19 pandemic.

Department of Health & Social Care

NHS Test and Trace

BACKGROUND

Initial actions and decisions varied widely between governments. The situation very quickly developed to the point of the UK going into lockdown, at which point there was a high level of demand upon GDS and other technical departments in government.

We were involved from the start, contributing advice as part of the COVID-19 response team along with other key technologists and service providers. In those early days, we implemented the Shielded Vulnerable People programme in collaboration with GDS, the NHS, and the Cabinet Office.

The on-going pandemic led to the creation of a new government department — the Joint Biosecurity Centre (JBC), as a collaboration between the Department of Health and Social Care (DHSC), the NHS, and the Cabinet Office. Reporting directly to Number 10, JBC immediately became responsible for COVID data analytics, as well as the Test & Trace programme.

We were involved from inception, even before JBC had an official name, and before its existence was even announced to the public. One of the most critical requirements was for the data produced by JBC to be as reliable as possible — and for that, Excel was not sufficient. It had been widely reported in the media that some COVID analytics data had been erroneous due to Excel having been used, due to limitations not being recognised until errors were found in published data.

JBC needed a proper data analytics platform, and we were asked to design, architect, and ultimately build this platform, along with all the gold-plated processes needed in software development, data confidence, and infrastructure resilience.

It has been our immense privilege and honour to lead on this project and assist our government — and fellow citizens — during this period. It is a "we were there" story to tell the grandkids, that every involved member of the Dotfive team feels extremely proud of, and which leaves a lasting legacy to help protect the nation against future events as well as having a small but important role in the events of this time.

CHALLENGES

One of the main problems was time. Infection rates were soaring, lockdown decisions needed to be made, and the only way to make those with any hope of being correct or effective was to have accurate. We therefore had to decide what was needed when, and how to deal with immediate concerns whilst simultaneously starting the longer-term solution design and development process.

We also had to establish the full set of requirements, which was no easy task, as JBC was growing daily with new departments, teams, and task forces being created. We needed to understand and identify the data analytics requirements of all teams, even though the people on those teams were themselves brand-new and still figuring things out themselves.

Additionally, a committee had been formed — the Test & Trace Data Analytics Platform group (TTDAP) — which we were a part of, but it quickly became bogged down in red tape. The timeframe of TTDAP establishing outcomes at its level soared, to the extent where it became clear that it would only satisfy and validate the longer-term plan, and would not produce anything in time to help address the immediate needs.

We needed to establish a governance procedure to ensure that there were the appropriate aspects of control and review in place for data analytics. This was tricky as JBC was filling up with data analysts — who are not software engineers. We needed to bring a proper software development process to the data analytics code, along with full automated test coverage, without getting in the way of the analysts and slowing them down.

A variety of data sources were in use, much of which came in as CSV files from Public Health England (PHE) and were being processed using AWS Athena. This was fine as a short-term stop-gap solution, but would not scale, so a better approach would be needed for the full platform. We needed to design an appropriate platform, using AWS, to handle processing of the data, along with all the best possible approaches to automation so that deployments, tests, and scalability would all be automatic.

SOLUTION

We quickly evaluated the possible approaches, and advised use of the open-source edition of Plotly Dash, a Python-based framework for charting. We successfully convinced JBC leadership that this was the correct solution, and not the commercial version of Plotly Dash, which would have cost £50k per server and introduced a number of technical hurdles. With the open-source version we had more control, which was vital for integration with other government systems.

Once this had been agreed and approved, we started development on a platform prototype, finishing this on time for immediate use and adoption, whilst simultaneously producing extensive technical planning documentation describing two main phases of development. These phases would achieve the critical functionality as soon as possible, and also describe how to satisfy the longer-term requirements. This documentation was fed back into TTDAP, and formed the basis of their own, much longer assessment and write-up.

Due to our participation and involvement in TTDAP along with supporting the Waste Water Programme (WWP) and Test & Trace/NIHP, we were uniquely positioned in JBC, and were asked to act as a conduit to relay requirements and ensure that all departments would have their needs met by the long-term solution that TTDAP were planning. We achieved this by maintaining regular communication with key stakeholders, and helping to steer and validate the direction being taken centrally.

Once the prototype platform was in place, the data analysts could start using it, and from that point on it was used for the generation of reports going to Number 10 for decision-making. We then introduced best-practice procedures to establish how code should be reviewed and unit tests created before declaring a new piece of analytics code as verified to be accurate, and the results safe to use.

We wrote full unit test coverage for the reports, and restructured the code we were given, identifying areas of commonality and introducing libraries for abstracting it, reducing the amount of report-specific code. We also created data architecture designs for using databases, allowing the platform to scale and support more users. We then assisted JBC's data engineers as they implemented this approach for the specific data they needed.

With the platform in daily use and data generation fully tested, we then worked on the infrastructure side, using Infrastructure as Code (IaC) approaches so that the entire container-based platform could be created with full automation. We also implemented automated pipelines for the testing and deployment of new updates, and created ETL (Export Transform Load) processes to deal with data ingest, processing, and storage.

RESULTS

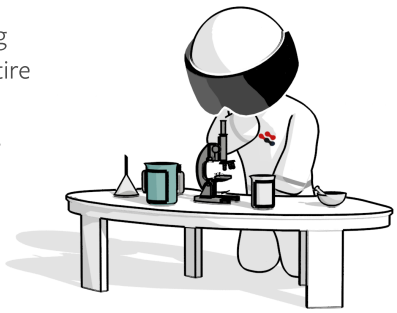
The introduction of the data analytics platform saved time in data processing and report production, as this had previously been dealt with manually. We customised it as needed for each department, notably Test & Trace and WWP. The WWP deals with, for example, detection of COVID in sewer samples, so each department had their own specific focus and set of data and reports.

Our involvement and contribution to TTDAP planning allowed the critical objectives of the longer-term plan to be validated, and TTDAP were then able to build upon the technical foundation we produced in the form of the prototype and Phase One development — especially as we finished it ahead of schedule. Phase Two has not yet started at the time of writing, and is a larger, longer-term objective, which will bring about automated test containers for individual data analysts, and other major features.

We were able to ensure that the Prime Minister made his decisions and announcements with the assurance of possessing verified, accurate data.

In the process of reviewing code and implementing tests we found several bugs, some of which would have led to significant errors in the reports. Because we caught and fixed those issues, we were able to ensure that the Prime Minister made his decisions and announcements with the assurance of possessing verified, accurate data — something that is of huge importance to the nation when the nature and timing of lockdowns is being evaluated.

The work on the DevOps and infrastructure side has achieved a robust, scalable platform, with documented, tested, automated processes and clear service layers. This now supports scaling up by adding more server nodes as required — and while avoiding paying any license fees. As such, we actually represented a net budget gain to JBC, as the total cost of our work was lower than the license costs originally budgeted.



dotfive™

www.dotfive.co.uk
hello@dotfive.co.uk
+44 (0) 8456 808 805
+1 (307) 222 4842

dotfive limited | frontell house, west coker hill, west coker, yeovil, somerset, ba22 9dg
dotfive inc. | 1908 thomas avenue, cheyenne, wyoming, 82001-3527

Technologies used:


PYTHON


FLASK


PLOTLY DASH

JUPYTER


PANDAS

Python | Flask | Plotly Dash | Jupyter | Pandas | JavaScript | React | PostgreSQL | Docker | Kubernetes | Minikube | Apache Spark | Apache Atlas | AWS | AWS EC2 | AWS EKS | AWS Athena | AWS CDK | AWS Sagemaker | Skaffold | Terraform | CI/CD | IaC |TDD

dotfive™

www.dotfive.co.uk
hello@dotfive.co.uk
+44 (0) 8456 808 805
+1 (307) 222 4842

dotfive limited | frontell house, west coker hill, west coker, yeovil, somerset, ba22 9dg
dotfive inc. | 1908 thomas avenue, cheyenne, wyoming, 82001-3527