Craig Wentworth

Premium report

Adoption & Practices

Transport for London creates an open data ecosystem with Amazon Web Services

September 2015

This report examines Transport for London (TfL)'s use of Amazon Web Services (AWS) – both to improve the agility and responsiveness of its own digital services, and to establish a third-party developer ecosystem (primarily of tech start-ups) that can leverage its openly-licensed travel data stream to build their own travel-planning applications for Londoners.

MWD Advisors case study reports are designed to help organisations considering or actively working with collaboration technology understand how others have worked to obtain benefits from collaboration technology, and how they have worked to overcome challenges that have arisen along the way. All our case study reports follow a standard model, and are researched using a standard process which is driven by senior MWD Advisors analysts.

Find out how to access related research at http://www.mwdadvisors.com/ec/membership.php.

MWD Advisors is a specialist advisory firm which provides practical, independent industry insights to business leaders and technology professionals working to drive change with the help of digital technology. Our approach combines flexible, pragmatic mentoring and advisory services, built on a deep industry best practice and technology research foundation.

Case study key facts

Organisation	Transport for London (TfL)
Industry	Transport and travel
Current goals	With its flagship tfl.gov.uk website now running on Amazon Web Services for over a year, TfL is now focusing on two main areas: enhancing the personalisation available to customers on its own site; and, in parallel, fostering a closer relationship with the third-party app and portal providers that contribute digital solutions of their own for London's travellers based on TfL's (openly licensed) transport data.
Current approach	Migrating TfL's digital projects to the cloud has enabled the organisation to become much more agile as it iterates new website and service designs, in response to changing customer requirements (in particular a rise in mobile and geo-location based queries for journey planning). The migration has also made it easier for development teams to access real-time streams of travel and transport data, bringing much more immediacy and higher granularity to its services. TfL has also chosen to release this data under an open data licence, which has helped to establish an ecosystem of third-party developers also working on digital travel-related projects for Londoners.
Outcome	TfL's website serves over 3 million page views to between 600,000 and 700,000 visitors a day, with 54% of visits coming from mobile devices. TfL has been able to scale interactive services to this level (its previous site was static), by leveraging AWS services as an elastic buffer between its back-office services and the 76% of London's 8.4 million population that uses the site regularly to plan their journeys. Some 6,000 developers are now engaged in digital projects using TfL's anonymised open data, spawning 360 mobile apps to date. The exposure that this transport data now enjoys has required TfL's various teams to take responsibility for raising its quality and format consistency – benefitting both internal and external data consumers alike.
Tools and suppliers used	Amazon Web Services – including Amazon Elastic Cloud (EC2), Amazon Simple Storage (S3), Amazon Route 53, Amazon Simple Queue Service (SQS), Amazon Relational Database Service (RDS), Amazon Simple Notification Service (SNS), and Amazon Glacier.

Organisation background

Transport for London (TfL) is a statutory body, created under a UK Act of Parliament, to implement the Mayor of London's Transport Strategy and manage London's integrated transport network (including surface and underground rail, buses and trams, cycling, river services, and taxis).

Its main digital asset is the tfl.gov.uk website, which serves between 600,000 and 700,000 visitors (3 million page views) per day. It provides status information on London's transport network, including live arrivals data for buses, trains and the underground Tube network, a 'jams cam' monitoring London's road traffic, and a journey planner. It also hosts press campaigns (such as those promoting cycling) and consultations on future transport network plans and regulations, etc.

TfL's Digital Team, under Head of Online Phil Young, is responsible for the organisation's website and related digital services, and also for the open data initiative upon which the ecosystem of third-party developers rely for live streamed transport data.

Project background

Cloud adoption at TfL started off as a way for the organisation to better embrace agile DevOps methodologies and deliver a much tighter integrated, more personalisable and efficiently scalable website. However, it also became a way of providing a scalable means to serve a burgeoning third-party developer ecosystem (some 6,000 strong, with 360 apps in production) that's grown up to develop independent solutions to London's travel problems, based on the anonymised open data that TfL releases concerning its transport network (e.g. timetable information, live running data, etc.).

Also, TfL was only previously able to supply historical travel data (e.g. the number of vehicles used from a given bike point over a year) piecemeal in response to Freedom of Information Act requests. Now the AWS platform helps TfL to deal with such things in an easier way, so the data can be used to enrich city planning with movement patterns and models (there are research use cases for the organisation's vast datasets – with university research groups extrapolating the impact of transport strategies and building scenarios for long-term planning).

Implementation characteristics and status

In its previous, Content Delivery Network (CDN) based incarnation launched in 2007, TfL's website was functional and informative but very flat and static. Now, with smartphones (rather than desktops) as the site's most common access method, and significantly raised user expectations of more real-time and personalised data-driven services, TfL had to re-think beyond CDNs. Phased development of the current site began in February 2012, culminating in a six-month period of parallel running in beta prior a full launch in March 2014.

The approach

TfL's technology stack is a mix of components (though fundamentally Windows-based). Google and Azure were amongst other vendors considered alongside AWS for cloud hosting (plus traditional in-house hosting), but TfL determined that AWS was much better placed for its needs in terms of service features and cost.

What followed was a re-development its software architecture and infrastructure which was closely related to its adoption of AWS services. To date TfL's cloud stack has been through nine iterations of design, each an improvement on the last (implementing learnings in terms of how customers actually used the services, and how best to architect them for cloud) to get it to where it is now.

Another driver was the effect of peak loads on the website due to snow days and strike days; these saw a 20x surge in visitor numbers compared with a regular day (most of which could come in a single hour at the start of the working day). In the past, TfL would have had to over-provision on capacity in order to account for the increased volumes of traffic, but now it can auto-scale on AWS. This responsiveness and flexibility (it can now scale down through the day as well, to get more efficiency) was a great incentive to go with the AWS' services and it's capabilities like these that got TfL, as an organisation, interested in cloud – and in AWS in particular.

Strategy

One of TfL's big strategic shifts over the last few years has been the release of its managed transport data. It's done this in order to encourage the growth of an ecosystem of tech start-ups focused on developing digital transport products and services for Londoners (following the principle that the more resources devoted to this cause, the better the overall service London's travellers receive – from whomever), and to support academic research groups looking to find longer-term solutions that optimise the transport network. Rather than this data being released only piecemeal (and in varying quality and formats), in response to individual *Freedom of Information Act* requests, the organisation has instead decided to release the data into the public domain proactively – and its goal is to continually improve the way it does this, and how it actively supports those who work with it.

AWS has essentially provided TfL with a channel to deliver travel data – to teams within TfL itself, and providing a gateway for third party developers and researchers to access it. Prior to TfL's Digital Team moving to a cloud-based infrastructure, the organisation was mainly limited to the supply of flat files of data – there were few real time feeds; plus it was difficult to cope with releasing travel data to multiple parties at web-scale. TfL's old infrastructure wasn't as rich as AWS' and couldn't be queried as easily. AWS' APIs for real-time access makes TfL's infrastructure much more interactive and responsive for developers to work with.

Phil Young, TfL's Head of Online, was the strategic lead for cloud adoption, and his Digital Team its main champions. TfL has a number of technical departments within it, each with its own requirements and priorities, depending on different centres of gravity (other parts of the organisation can be more traditional in their approach to development, operations, and infrastructure). However the Digital Team has more leeway to adopt an agile cloud approach. The challenge is when this agile cloud approach collides with a waterfall and on-premise approach. Many of the technology teams in TfL now 'get' cloud, partly because they've been able to see what the Digital Team has been able to demonstrate in terms of scale and complexity, showing that cloud-based systems can be robust and cost-effective, and illustrating what it can do now that it couldn't do before.

Across the organisation, as a result, many areas are looking at some form of cloud platform adoption. However, Young describes the process as like "turning a supertanker" since some parts of the business are committed to other infrastructure and so cloud in these contexts will inevitably take time to evaluate.

Some technical departments may choose to retain on-premises infrastructure for perfectly valid business reasons, and of course those which ultimately do opt for cloud may not all necessarily use AWS (TfL's business predictions are already on Microsoft Azure, for example). In such a heterogeneous hybrid environment, therefore, it's imperative that services can be effectively integrated and aggregated so that the end-to-end customer experience remains seamless (and the flow of anonymised open data to TfL's third-party developers isn't compromised). Young has described his approach to infrastructure decisions at TfL as "cloud when it's best; on-premise for the rest".

In terms of future milestones, TfL launched its current site with just a measure of personalisation and geo-based querying – so there's more personalisation it's intending to add. It needed the platform (on AWS) in place to be able to do this effectively at scale, and now that's bedded in it can add account based-access to provide a much richer and personalised experience across all its products, recycling information for individual customer travel planning, macro level planning, and proactive recommendations based on travel information. This is only feasible at scale because of how the cloud platform infrastructure can flex.

Organisation and people

Adoption

Rolling out TfL's website and access to transport data on AWS has demonstrated that customer requirements can be handled without breaking back-office systems. The API handles 10 million calls a day, though some back office elements can only get through 2 queries a minute – so the cloud platform acts as a buffer, providing many connections to customers through just one secure connection to the back end.

In 2010 (in line with the Mayor of London's and UK Government policies) TfL made the strategic commitment to release its travel data under an open licence. Since that time, TfL has invested time in marketing its open travel data to developers to drive adoption (and foster an ecosystem of travel apps), resulting in a high demand because of its quality and utility. In fact, because this data is open and exposed, those responsible for generating it within TfL have found themselves responding to outside pressure to improve its quality from the original multi-format datasets that first greeted developers. July 2015 sees TfL launching an enhanced proposition for its digital ecosystem with a single unified API superseding the legacy feeds, with members of its Digital Team spending more time with key developers to explore potential uses for its data.

Roles and resources

TfL's move to cloud has unified all its transport data under one roof (compared with the previous situation, where it was all on different systems, in different formats, in different parts of TfL). This has had the effect of simplifying developers' access to data and facilitating access to real-time streams in consistent formats they can work on.

Data is supplied by TfL's own teams (those responsible for managing the services that collect or generate the data). They also become internal data consumers themselves (for TfL's projects based on this data) alongside external developer / researcher consumers inhabiting TfL's ecosystem. In some cases vendors also become involved, where APIs needed to be commissioned to achieve connections with existing back office systems.

TfL regards its relationship with AWS as being "as close as [it] wants it to be", with a Solution Architect and an Account Manager on hand to liaise with so it's aware of new service developments on the AWS roadmap.

Governance

At the beginning of 2012, TfL's Digital Team numbered 40 people. By 2015, the original team had almost doubled in size, but this is seen as just an interim step (with further expansion likely) as the organisation sees new opportunities to exploit the cloud platform to develop more of its own apps and services, and to more fully support third-party developers, in order to meet its customers' growing expectations for digital channel engagement.

Technology and infrastructure

TfL has invested lots of time in DevOps automation. It has an agile infrastructure now, to go with its agile methodology (supporting 20 development environments, plus two for its website, one for the TfL Journey Planner, and three production environments – although TfL's Oyster payment card system[1] isn't on AWS). To maintain a 24-hour service it's been able to deploy to replica test environments, and also swap over to new in versions of the site (keeping the load balancing) using DNS changeovers.

The environments are all run by two DevOps staff day-to-day, with an external company providing a managed service on AWS to provide 24/7 infrastructure support. TfL itself doesn't have the resources to do this, and AWS doesn't provide support that far up the stack into TfL's applications.

TfL Digital's main components on AWS are EC2 (hosting MongoDB, SQL Server and Oracle instances that TfL manages itself, via a third-party support provider), S3, Route53, SQS, RDS for some data, SNS, and Glacier. The organisation now regularly runs 40- 50-node instances to serve current demand – up from 11 or 12 when it first started with AWS.

The results

Since the start of the project to roll out AWS in October 2012, TfL has seen a rise from 51% to 70% of customers using its website for real-time information. In the same period the number of customers using third-party apps powered by TfL data for the same purpose has risen from 27% to 40%.

TfL sees this as being due, to a large extent, to its ability to iterate quickly on the platform in a DevOps environment within AWS in order to improve its own services; and also to the fact that it can now satisfy real-time demands for live traffic data from the third parties whose services augment TfL's own in the London digital travel assistance marketplace.

TfL reports that it has found AWS' cloud platform to be even more flexible than it had thought it would be – not necessarily *simple*, but certainly flexible. There are things it's been able to do now which it didn't have the infrastructure to be able to do before – for example, it's now able to rapidly prototype ideas on its 20 development environments (where it previously only had 4), and replicate the production environment for parallel work in order to meet the dynamic demands of the business. This has dramatically cut down the time-to-market for new services, enabling TfL to respond effectively to customers' demands (for example, being able to move to an entirely new cloud infrastructure for its Journey Planner engine – which is now used by in excess of 5 million travellers – in under two months).

The organisation has also experienced a cost benefit because AWS' services are metered and paid for by the hour.

[1] Oyster is a plastic smartcard which can hold pay-as-you-go credit, and passes for travel on TfL's transport network (Oyster isn't used for any roads related payments) – see https://tfl.gov.uk/fares-and-payments/oyster/what-is-oyster.

Recommendations for adopters

In our conversations with TfL's Head of Online, Phil Young, and his Digital Team for this case study, they offered several recommendations for organisations embarking on a similar initiative:

- Have a good strategy for integrating with legacy IT. TfL had a couple of false starts and learnt from its experiences – achieving success with cloud platforms requires a different approach than when working exclusively in an on-premises environment. In any organisation for which 'hybrid' is likely to be a pragmatic way forward for the time being, integrating the old with the new is crucial in order to maintain and raise the quality of existing services whilst new ones are developed and rolled out.
- Think about security and who's going to provide it at which points in your technology stack. Cloud platform providers such as AWS can block ports but organisations typically need much more when it comes to protecting the apps they expose to their customers, and they need to be able to do this for themselves.
- Keep your eyes open for potential opportunities as well as potential pitfalls. You'll be learning about what a cloud platform can do to transform your business as you go along, so be prepared to capitalise on where you're able to do things you couldn't do before; but make sure you can still do what you used to do too (if you still need to do it); and keep focused on the end game when roll-out and adoption feels harder than you expected.
- Make sure your organisation's financial processes are able to handle variable bills each month. It can be a challenge to get the correct budgeting model (forecasting costs can be tricky), but without this you'll not be able to realise all the cost benefits. Also look into reserved instance pricing to achieve further savings.

Best practice insights

Transport for London's cloud adoption with Amazon Web Services is a story of an organisation that identified a clear need – being able to respond to rapidly changing customer dynamics (in terms of travellers' mobile needs) – and then got to a point where it is iterating its own flagship online services much faster; learning how it can exploit the features of cloud-hosted environments to make its own data much more accessible for its own teams, which is leading to more personalised services… and then also making this newly-visible live data work for third parties too.

From its growing ecosystem of mobile app developers helping today's travellers get from A to B, to academic research teams helping TfL improve London's transport network in the longer term, TfL's open data policy has enabled third party organisations to benefit from real-time, highly granular, high quality transport data; and the keen interest these parties have shown has created a quality feedback loop that's encouraged all data contributors within TfL to improve their data – where previously an internal-only market hadn't been able to achieve such a data quality push.

The success of TfL's third-party app ecosystem has also encouraged the organisation to make a strategic decision to get much closer to key developers than had originally been envisaged. Whether this morphs into more overt partner relationships remains to be seen, but the open data policy has certainly shown that there's a ready market for digital services targeting Londoners' transport problems – whoever solves them.