Research report by APImetrics

October 2022

UK OPEN BANKING API PERFORMANCE 2021-2022

Dr Paul M Cray
Head of Artificial Intelligence and Machine Learning, APImetrics

01 Introduction

The UK continues to be at the forefront of the global Open Banking revolution thanks to the proactive attitude of the regulators. These regulators helped create an Open Banking ecosystem that encourages and facilitates smaller banks and new entries, including fintechs and neobanks, to participate in the Open Banking market. As the most advanced Open Banking market in the world, the UK provides an example of best practices in the implementation of API-based Open Banking.

We studied the performance of the large CMA9 UK banks, traditional High Street banks, credit card providers and building societies, and new entrant banks (neobanks). The endpoints were provided by the banks and measured using our patented APImetrics quality scoring system, CASC (Cloud API Service Consistency).

This report is generated from real API calls made using the FAPI compliant consent process with the partnership of tomato pay, a leading open banking provider in the UK.

All calls were made between July 1, 2021 and June 30, 2022

02
Key Findings

- Most, but by no means all, applications are hosted in the UK
- While many are hosted on AWS cloud, at least 25% of banks self-host
- There are significant differences in the performance of the three groups
    - Neobanks performed well
    - CMA9 provided good service overall
    - Performance issues are evident in the traditional banking providers
- Most providers were able to maintain a FAPI consent journey that achieved high APImetrics quality scores
    - There are significant differences in quality and latency among the groups
    - Neobanks are by far the fastest; traditional banks are the slowest with a 300ms difference
    - Some traditional banks suffered from periods of poor availability
- Where major TPPs are hosted impacts a bank's API performance
    - UK locations are fastest (under 400ms for Azure, Google, and IBM), but some non-UK locations can be faster
    - When every millisecond counts, choosing a cloud/location with superior DNS and TCP Connect Time can help maintain a performant API
- When benchmarked against general IT services using the same measurement system, the neobanks perform well, but CMA9 and traditional banks lag
- UK-hosted applications perform better than other locations; due to distance and other factors we would not recommend hosting in Nordic data centers

We find it remarkable that there is a significant difference between the performance of the different providers given that all types of providers had to implement this infrastructure in the last three years and none of the metrics we analyzed are overly dependent on historical technical debt.

Our analysis of Open Banking API services extends our established methodology and benchmarks for monitoring and ranking API performance and reinforces the importance of actively monitoring APIs.

2.1. Honorable Mentions

Neobanks

- Tide Bank has consistently topped the analysis with a quality score close to perfect – we do not run calls against a live account for this, so this is based purely on the consent access process
- Monzo – solid performance from a neobank where we are making full functional calls and a high CASC score over 9.0 for the year

CMA9

- NatWest Group has highly consistent good performance across multiple brands where we are making AISP and consent calls – this is even more remarkable when you consider the range of NatWest group's brands and technical debt

Traditional Banks

- Creation Cards have maintained high consent scores
- Virgin Money Group is among the most improved traditional banks and has locked in solidly good quality scores for the period

2.2. Hosting Patterns in UK Open Banking

In compiling the data for this report, we were able to analyze the hosting providers and locations for UK open banking where possible. Across all banking providers, AWS (Amazon Web Services) were the leading option, used by 30% of the providers; approximately 25% of the banks self-hosted and another 25% were unable to determine their core hosting service.

The remaining banks were split evening between Azure and Google services. Where we were able to determine a hosting location, 35% were hosted inside the United Kingdom and 25% in the Republic of Ireland. The remaining hosting was distributed evenly across Belgium, Denmark, France, Germany, and the Netherlands.

As will be shown in this report, the distance from the hosting solution can play a significant role in performance. Traditional banks solutions had more hosting outside of the United Kingdom (70% of providers, compared to 60% for neobanks and CMA9 banks).

If we combine numbers for UK and Ireland, we see that 70% of the CMA9 are hosted in the UK and Ireland, 60% of neobanks, and just 50% of traditional banks. This may further explain what we see with some of the performance differences.

2.3. Performance Differences

Financial-grade API (FAPI) consent journey endpoints are not reliant on legacy technology, technical infrastructure, or business processes, so we didn't expect such significant differences in total time. These differences can be attributed to several factors:

- Whether the bank uses an in-house or third-party solution for consent endpoint implementation
- Infrastructure used for implementation, such as
    - Length of time public keys are cached and stored for validation
    - Performance of the infrastructure serving the consent journey
- Human resources made available for API implementation, testing, monitoring and maintenance
    - Neobanks have a strong cultural focus on API quality and no legacy systems to maintain
    - CMA9 banks have significant human resources, but need to maintain legacy systems
    - Traditional banks tend to focus on off-the-shelf solutions that need more tuning and maintenance than they always have available to deliver the highest quality service

03
Scope of Report

We made ~17 million API calls to Open Banking endpoints and measured availability and latency components including DNS (name look up), connection, handshake, and backend processing times.

We categorized the UK Open Banking landscape into groups of similar types of banks:

- CMA9 banks – the largest incumbent banks
- Traditional banks – established smaller banks and building societies
- Neobanks – innovative entrants to the UK banking services market

APImetrics monitored FAPI-compliant journey endpoints for 28 Open Banking brands, up from 16 from the previous year, classified as follows:

CMA9

Allied Irish Bank (GB)
Bank of Ireland
Bank of Scotland
Barclays Business
Barclays Personal
Danske Bank
Halifax
HSBC Bank (Business)
HSBC Bank (Personal)
Lloyds Bank
Nationwide Building Society
NatWest
RBS
Santander

Traditional

B
Capital One
Creation Cards
Cumberland Building Society
Sainsbury's Bank
Tesco Bank
TSB
Virgin Money
Yorkshire Building Society

Neobank

Cashplus Bank
Monzo Bank
NewDay - Amazon Mastercard
Tide
Vanquis Bank

04 Methodology

APImetrics used our active API performance and quality monitoring system with a Software Statement from our partner Tomato Pay, a UK third-party provider (TPP), to make standardized end-to-end consent calls for the financial-grade API (FAPI) journey at 28 bank brands approximately every 5 minutes from APImetrics software agents hosted at 22 cloud locations in Europe.

From the raw metrics we obtained, we used our patented Cloud API Service Consistency (CASC) technology to generate a quality score each month for each brand.

Each API call consists of a sequence of steps that take place before data is exchanged with the target server. These are Name Lookup (DNS), TCP Connect (the digital connection with the remote server), and TLS Handshake (the secure handshake between the two services).

As a rule of thumb, the TLS handshake is, at a minimum, double the TCP connect time. In practice, it can be even longer as there might be server analysis required to ensure that the incoming connection is valid.

Pulling these together shows why it is important to consider both DNS and TCP connect times as part of your overall expected latency.

AWS UK has an average DNS lookup time for all UK banks of ~10ms, and a similar TCP connect time. Allowing another 20ms for the handshake process, we can see that there would be, on average, a 40ms overhead for using this server.

In contrast, AWS Sweden has a DNS lookup of ~30ms with connection and handshake times of at least 90ms – 5x slower than connection times from the UK data center.

Google Finland, both the most remote and secure data center, increases that overhead to over 200ms before any call-related data has been transferred.

Once the connection is established, the call request is uploaded to the target server. The target server then handles the query and a response is sent. The total time elapsed is then logged.

No APImetrics customer data was used in the generation of the reporting data; all data points were developed independently.

05
Detailed Results

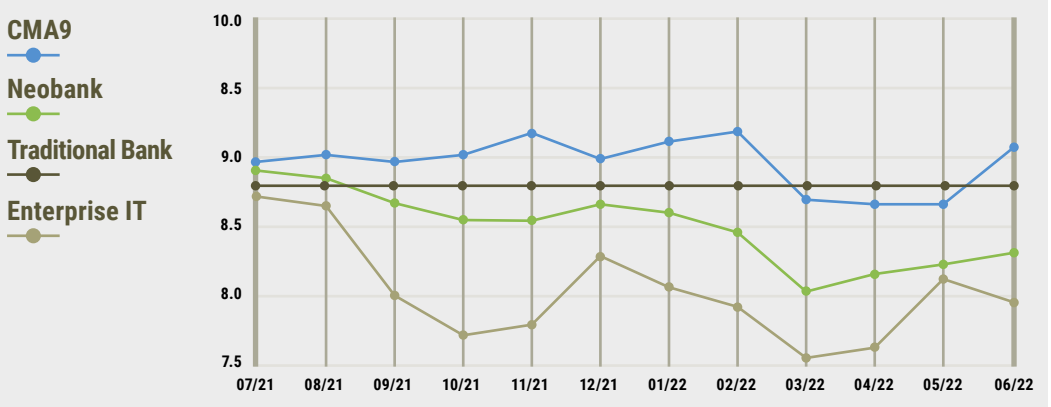
In interpreting the meaning of our results, we have considered context and compared the performance of UK Open Banking consent endpoints to a collection of the top Enterprise IT APIs we monitor. The value derived from this exercise is that these are highly stable, high-quality APIs from some of the largest IT providers in the world that run focused, API-first engineering teams. This gives us an insight into how well, or not, the banking providers are doing.

5.1. Endpoints

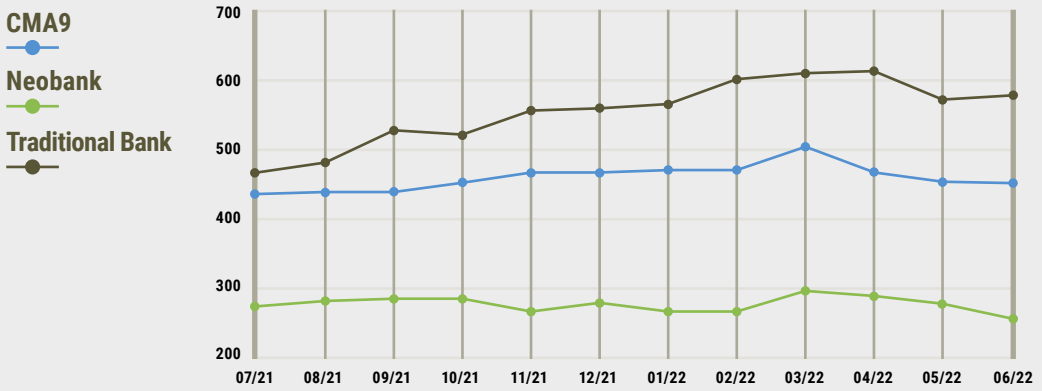
Neobanks tend to be technology-first organizations, so it's not surprising that they perform at a higher level than business-first organizations such as the CMA9 and traditional banks. Both types performed consistently in the CASC Green Zone.

Traditional banks are always the least performant, spending just as much time in the Green Zone as in the Yellow Zone – meaning that some additional work is needed for better consistency.

QUALITY SCORE OF BANK BRAND APIS OVER TIME (BASED ON APIMETRICS CASC SCORE)

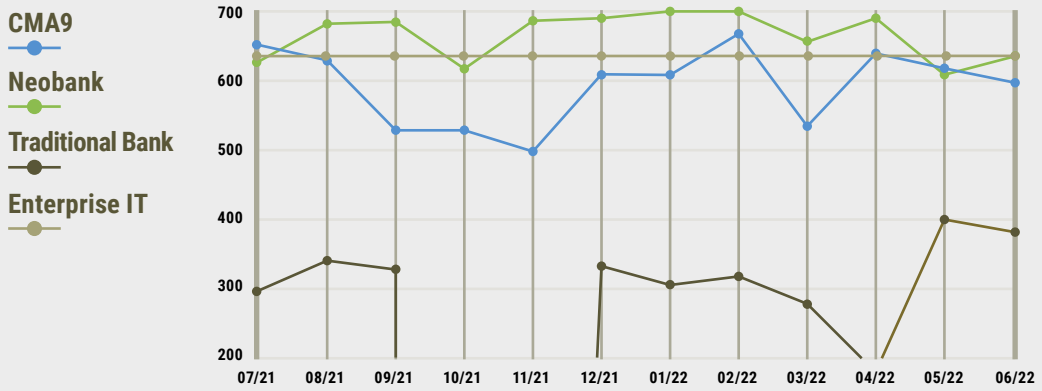


TOTAL LATENCY (IN MILLISECONDS) OF API CALLS MADE BY BANK TYPE



There is an established latency metric of ~450ms for a delay to become perceptible to a human. Neobanks all perform better, the CMA9 are close, and traditional banks are significantly outside this interval of human latency perception.

AVAILABILITY BY BANK BRAND OVER TIME



The overall availability of traditional banks was impacted by the performance of two brands, one of which suffered major issues in October and November 2021.

The availability of general Enterprise IT APIs monitored by APImetrics in the same period is shown. We would generally consider Enterprise IT APIs from vendors such as Microsoft, Google, Dropbox, Box, and so on as being the target quality and performance standard for all API providers.

Type	Average availability, July 2021-June 2022
Neobanks	99.98%
Top Enterprise IT APIs	99.96%
CMA9 Banks	99.93%
All Brands	99.74%
Traditional Banks	99.34%

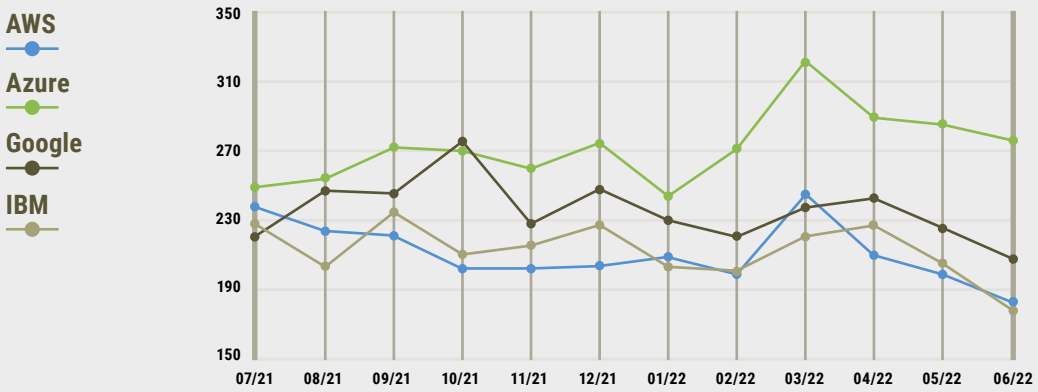
5.2. Cloud Datacenter Performance

Most applications built to call Open Banking APIs are hosted in a cloud data center from one of the major providers. With APImetrics we have made the same calls, using the same infrastructure and configuration from AWS, Azure, Google, and IBM Cloud data centers.

Latency is known to become perceptible to end users as call durations exceed 450ms – a data point established by IBM with Network based computing.

With this in mind, we can see that cloud datacenter choice for an application can have a significant effect.

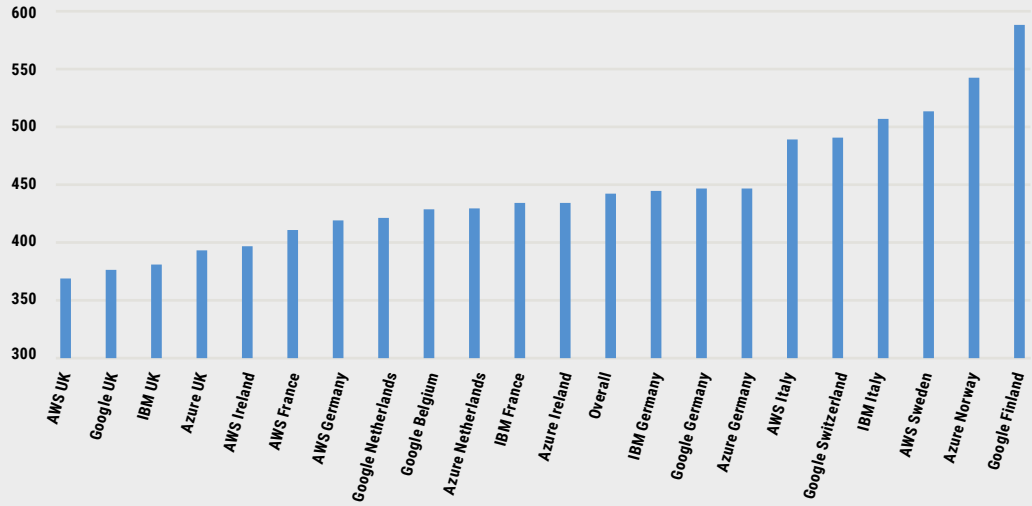
TOTAL CALL TIME IN MILLISECONDS BY CLOUD DATACENTER PROVIDER OVER TIME



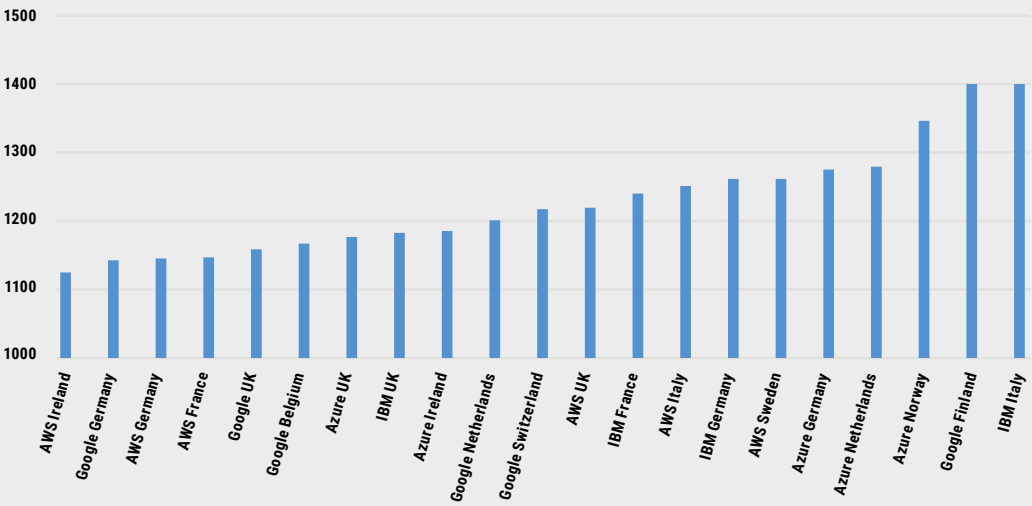
In terms of the total time taken for the entire API call sequence to complete including networking factors, the difference between the CMA9, traditional, and neobanks is stark. Neobanks are typically ~150ms faster than CMA9 banks and 350ms faster than traditional banks. Such a large difference in total time is likely to have a noticeable impact on user perception of service quality.

- Calls made from AWS data centers were generally faster by the end of the year ending June 2022 compared to calls made from Azure data centers
- IBM Cloud has showed the most improvement

TOTAL CALL DURATION IN MILLISECONDS PER DATA CENTER

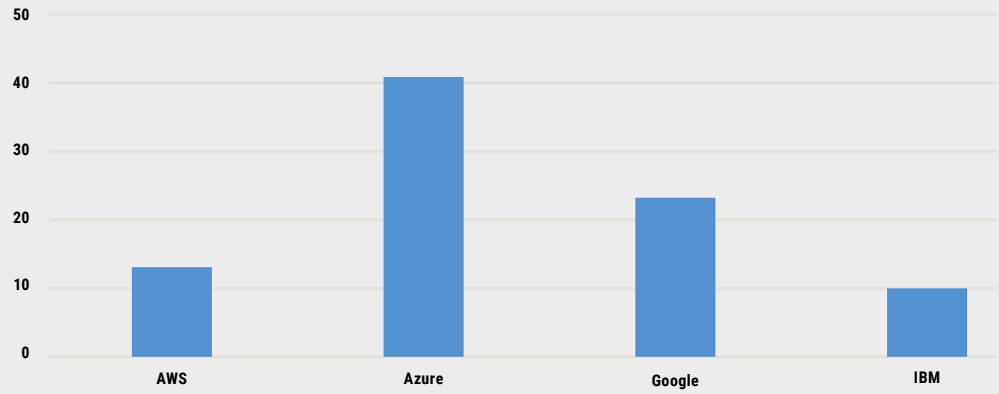


P99 TIME IN MILLISECONDS PER DATA CENTER



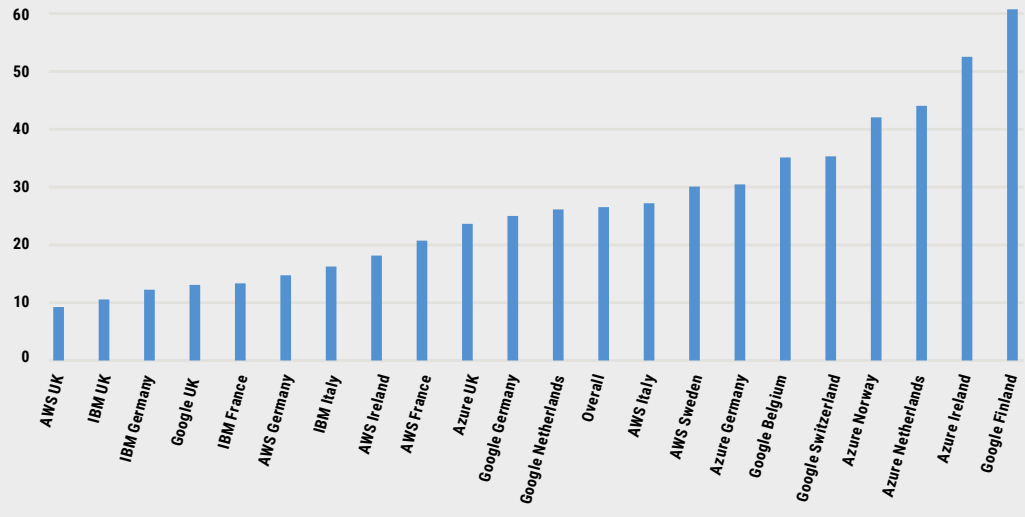
- Three cloud locations in the UK have an average total time less than 400ms
  - In 2020-21, no cloud locations had an average total time less than 400ms, indicating an overall improvement in network infrastructure
  - Azure UK at 411ms is on average 28ms slower than IBM UK and 10ms slower than AWS Ireland
- p99 times are >1100ms from all cloud locations, an improvement on 2020-21 (> 1200ms from all cloud locations)
  - In 2020-21, the four UK cloud locations had the four lowest p99 times; this was not the case in 2021-22
- Nordic data centers should be avoided for hosting UK Open Banking applications due to their distance from the primary data locations; calls made from Nordic locations being almost twice as slow as ones made from the UK or Ireland
  - Distance plus poor DNS resolution times make any data centers outside of proximity to the UK/Ireland a bad choice

AVERAGE DNS NAME LOOKUP TIME IN MILLISECONDS BY CLOUD PROVIDER



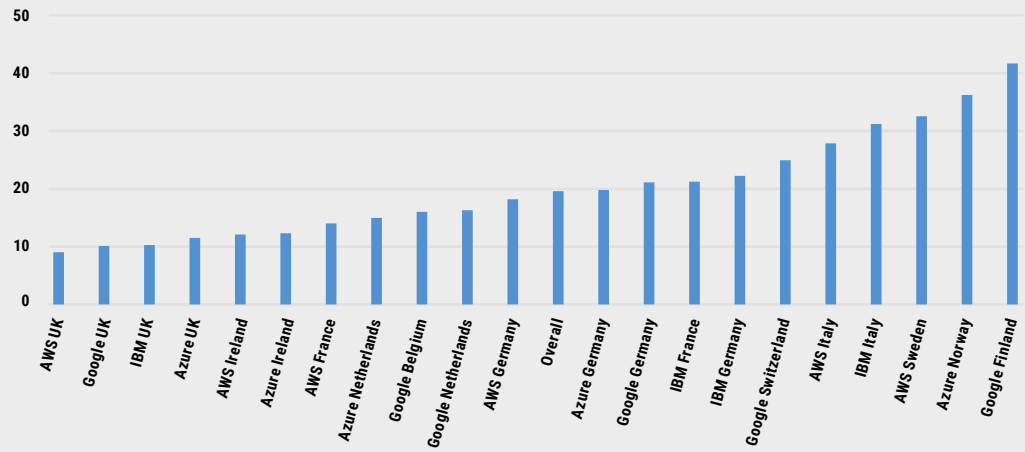
- There is a significant variation in DNS Lookup Time between clouds, with Azure and IBM Cloud one-third faster than AWS
- From our global monitoring, we've found that a well-configured setup should have a DNS Lookup Time of ~14ms or less
- AWS and IBM make good choices for DNS resolution from the cloud data center
- Azure is negatively impacted by slow DNS lookup times

AVERAGE DNS LOOKUP TIME BY CLOUD DATACENTER



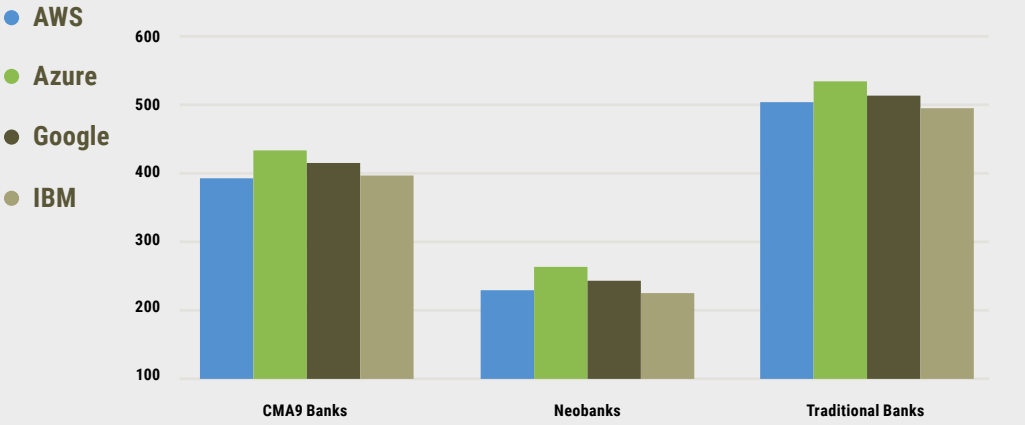
- Google Finland is known to implement strong security on inbound and outbound traffic, which explains the higher DNS resolution
- Azure Ireland is interesting as Ireland is a common hosting location, and this represents a significant latency overhead for that data center
- IBM Cloud provides best overall DNS performance across all data centers

AVERAGE TCP CONNECTION TIME IN MILLISECONDS BY CLOUD DATACENTER

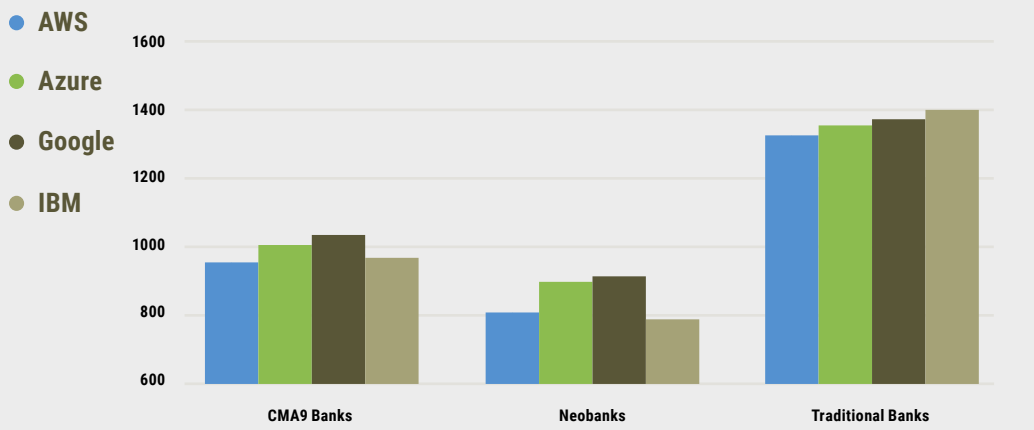


Pulling the data together and removing the impact of Nordic datacenters, we can see that the best performant data centers for Open Banking applications in the period 2021-22 were those provided by IBM and Azure in the UK and Ireland. The significant performance differences seen between the different types of providers were consistent across all cloud providers.

RELATIVE LATENCY OF THE SAME CALLS MADE FROM DIFFERENT DATACENTERS IN MILLISECONDS BY BANK TYPE



AVERAGE DNS NAME LOOKUP TIME IN MILLISECONDS BY CLOUD PROVIDER



Traditional banks all have p99 times longer than 1200ms regardless of cloud, whereas neobanks have p99 times under a 1000ms. Slow calls can impact user experience and neobanks can provide not only faster endpoints than traditional banks, but more reliable ones with fewer outliers.

06
Recommendations

Only by monitoring API quality from the end-user perspective can an organization understand how its APIs behave. Internal monitoring will not expose issues for rapid escalation and resolution.

Be aware of the location you are planning to call your application from. UK Banking providers are generally hosted in the United Kingdom, Ireland, and the Netherlands. The farther you are from those locations, the slower your call latency will be.

Traditional banking providers are significantly slower than neobanks. If you are providing service to an audience using or integrating with traditional banks, design this overhead into your applications.

Best performance currently comes from applications built on AWS or IBM datacenters. Azure currently has DNS resolution issues in conventional use and may need additional work to bring performance up to acceptable levels.

07
About

APImetrics

APImetrics provides run-time API governance solutions for organizations offering API services across the Financial Services, Open Banking, Telecoms, Software, and IoT sectors. By enabling a holistic, end-to-end view of performance, quality, and functional issues across the API surface, we allow organizations to better serve their customers and end users.

Our patented technology automates the process of producing regulator-ready reports for financial services providers around the world.

Our active monitoring platform integrates with many of the leading developer operations suites and provides an API-centric view of:

- Real-time API performance from more than 80 locations worldwide on four clouds and six continents
- Fully integrated security monitoring designed and built for the needs of the financial services industry
- Machine learning based analysis driven by a database of more than a billion real API calls
- Integrated reporting, analysis, and alerting
- 360-degree visibility with Cloud API Service Consistency scoring (CASC), allowing for at-a-glance service and competitor comparisons

For more information, visit **apimetrics.io**, email **info@apimetrics.com**, or call +1 833 274 6389.

Tomato pay

The tomato pay API platform powers partners propositions focused on supporting small, medium enterprises, including its own simple, QR-code based payments and invoice app used by businesses and sole traders who want to receive payments in a fairer, cheaper and more ethical way.

Businesses and sole traders can benefit from a low-cost solution with no hidden fees, which can save money and time compared to current payment systems. It offers instant access to money as cash settlement happens almost immediately. It also offers access to all bank accounts and transactions in one place.

Businesses and sole traders will also be able to benefit from a quick and easy invoice solution later in the year. Invoices can be created within the app, with the option to give automated discounts, late penalties and send nudges to remind customers and clients to pay.

Plus, businesses can connect their bank account as digital payments are embedded within the invoice from the app - so no need to send bank details, and businesses will receive money owed instantly into their account.

Everyone can support their local communities and help them thrive by paying their neighbourhood businesses in a cashless, faster, cheaper, hassle-free way.

For more information,
visit **www.tomatopay.co.uk** or email **info@tomatopay.co.uk**

08 Glossary

4xx HTTP status codes
4xx HTTP status codes are generated when a request is made to an endpoint that does not exist or for which the user lacks the appropriate authorization. Because these types of issues indicate that the web server receiving the request is behaving as expected, 4xx client-side warnings should not generally be included when determining the performance and quality of an API endpoint.

5xx Server Error
A 5xx server error is an actual reported error from the application server hosting the APIs.

Agent
The APImetrics software agent run at various cloud locations around the world enabling synthetic calls to be generated as if they were being made by an end user or partner.

API
Although an Application Programming Interface (API) is a general concept in computer systems, in the current context we are concerned only with web APIs. A user makes an HTTP request to a published API endpoint. The request causes the web server that receives the request to return a payload containing information in a specified format or cause the state of some remote resource to be changed. APIs can thus be used to exchange useful data and information between systems.

API call
An API call is a single HTTP request made to a particular endpoint. Details of the request and the response are stored by APImetrics for further analysis to determine the performance and quality of the endpoint.

Authentication and Authorization
Access to a particular API endpoint may depend on validating the identity of the requesting party and that it has been granted the appropriate authorization. This might involve encrypted passwords or tokens generated and managed through a protocol such as OAuth 2.0 supplemented by specification such as FAPI (Financial-grade API).

Availability
Closely linked to pass rate. Strictly, the availability should always be higher than the pass rate. Calls to an endpoint may not pass because of authentication and authorization issues or because the request is malformed, but the endpoint is still available. APImetrics analyzes the results and calculates pass rate and estimates availability.

Cloud provider
An organization that provides a commercial service, hosting applications at a server. Well-known cloud providers include Google Cloud Platform (Google), Amazon Web Services (AWS), Microsoft Azure (Azure), IBM Cloud (IBM), all of which have many locations around the world.

Configuration
Internal and external network configuration, such as load balancers at the API gateway that direct requests to specific IP addresses, can have a significant impact on API performance and quality. For instance, problems with external configuration such as routing tables can cause requests to be misdirected, and load balancers can direct requests to IP addresses that do not support a particular service.

CASC score
Cloud API Service Consistency (CASC) is an APImetrics patented technology that combines various measures of API performance such as availability, latency, reliability, consistency of response and number of outliers, benchmarked against our unrivaled collection of historical API call records, to generate a single blended metric, much like a credit rating. The CASC score lets you see at a glance the quality of an API endpoint, whether it is getting better or worse and how it compares to other endpoints.

Green *Score: 8.00 or greater*
API is performing consistently with no performance issues

Yellow *Score: 6.00-7.99*
API is generally performing adequately, but there are some performance issues in need of attention

Red *Score: Below 6.00*
API is performing poorly. There are serious performance issues and in need of urgent remedial attention

CMA9 banks
The CMA9 are the nine large UK banks that are mandated by the Competition and Markets Authority (CMA) to expose certain Open Banking APIs and regularly return certain reports on the performance of the APIs. The banks are Allied Irish Bank, Bank of Ireland, Barclays, Danske, HSBC, Lloyds Group, Nationwide, NatWest Group, and Santander.

Download time
The time taken for a request to be downloaded from the web server to the agent.

DNS latency
DNS (Domain Name Server) is the global service that identifies where a particular service is located on the internet. The look up time is the time taken for the cloud service making the API call to identify where the target server is and route the request. The different techniques used for the look up task will affect service quality. How, for example, AWS handles DNS lookup, is different to how Google does.

Endpoint
The Universal Resource Indicator (web address) that is called when you make an API call. For the API call to work you will have URI + parameters of the call + security. This task is different to simply looking up the URL of a website where it is often just the URI that is needed.

Failure Rate
The proportion of calls made to an API endpoint that return an unexpected response.

FAPI
Financial-grade API (FAPI) is a technical specification that Financial-grade API Working Group of OpenID Foundation has developed. It uses OAuth 2.0 and OpenID Connect (OIDC) as its base and defines additional technical requirements for the financial industry and other industries that require higher API security.

GET
The simplest HTTP verb (others are HEAD, POST, PUT PATCH and DELETE) that sends a request to an API endpoint that gets a resource, such as a list of account transactions. Parameters and headers allow complex requests to be made with a GET.

Handshake time
The time to complete the process that sets up an HTTP connection, which is called a handshake.

Latency
In general, latency is the same as total time. The latency consists of several latency components including name lookup (DNS) time, handshake time, upload time, processing time, and download time. In this report, latency is reported in milliseconds.

Metric
A measure of some aspect of API endpoint performance such as the availability or median length of a latency component.

Network Infrastructure
The totality of the physical network elements that make up the systems that together comprise the internet. Includes switches, routers, and connectors such as fiber and microwave links.

Non-conformance
An API endpoint that does not respond according to its published specification is non-conformant. Typically, this might mean that the return payload has missing fields or contains incorrect information or the endpoint is generating errors and warnings despite the call being made according to specification.

OBIE
Open Banking Implementation Entity (OBIE) is the UK entity managing standards for Open Banking within the United Kingdom.

Open Banking
A new global paradigm for banking, financial and payment services that enables innovative new products and user experiences powered by data and information exchange through APIs.

Performance
The set of metrics such as availability, latency, reliability, and number of outliers that define how an API endpoint has behaved over a period of time.

Processing Time
One of the components of latency, this is the time the server takes to process a received request before sending the response back to the end user.

PSD2
Payment Services Directive 2 (PSD2) is a pan-European agreement on payment services to open up payments and banking services that is applicable to all financial services providers doing business in the EU and United Kingdom. Responsibility for the implementation of regulations lies with each country.

Quality
How good an API endpoint is from the end-user perspective. Although this can be challenging to measure, blended metrics such as the APImetrics CASC score provide a quantitative benchmark, allowing organizations to compare the quality of an API endpoint over time or compare two API endpoints at a glance.

Reliability
A reliable API endpoint tends to respond within a narrow range of times. A reliable endpoint may not necessarily be fast, but the variance in its latency will be relatively small.

Speed
The rate at which data is passed along a connection such as an intercontinental undersea fiber link. The more traffic, the slower the speed of the connection.

Total Time
The time between a request being made to an API endpoint and the whole of the response being received, including the name lookup (DNS) time, TCP connection time and handshake time.

Upload Time
The time taken for a request to be uploaded from the agent to the web server.

Version
APIs are often updated to make changes to the way the endpoints are invoked, or the content of the payload returned. It is important to ensure that the endpoint for the correct version is invoked. Often the URI for the endpoint will contain the version.


APImetrics
API INTELLIGENCE


tomato
pay™