Appendix I: Prevalence of multi-cloud

Introduction

I.1 In this appendix, we present:

(a) our analysis of providers' data on the prevalence of multi-cloud among customers;

(b) cloud providers' submissions on the prevalence of multi-cloud;

(c) a detailed analysis of an AWS quantitative analysis on the prevalence of multi-cloud; and

(d) other customer evidence.

I.2 Our main findings can be found in Chapter 3.

I.3 The ability to multi-cloud and switch between cloud providers influences the nature of competition in the markets and at the extremes:

(a) If all customers are freely able to switch or use more integrated forms of multi-cloud, then cloud providers have a greater incentive to make their offerings competitive with their rivals. This is because if customers are able to switch or multi-cloud then they would be able to switch all or part of their workloads away to, or place new workloads with, rivals in response to their incumbent cloud provider becoming less competitive (eg due to higher prices or lower quality) relative to its rivals.

(b) Conversely if customers are unable to switch or use any kind of multi-cloud, then cloud providers have a lower incentive to make their offerings competitive with their rivals. This is because customers would not be able to switch existing workloads or place new workloads with a rival in response to their incumbent cloud provider becoming relatively less competitive than its rivals.

Our analysis of cloud providers' customer data

I.4 We have estimated multi-cloud prevalence using customer data from cloud providers.

I.5 We requested customer data sets from AWS, Microsoft and Google that identified customer names and annual spend on their respective clouds for 2020, 2021, 2022, and 2023.

I.6 The advantage of this analysis relative to surveys is that we avoid any potential issue of customers misunderstanding what multi-cloud is, as we define it for the purposes of the investigation. For example, customers using both private cloud and public cloud would not be counted as using multiple clouds in our analysis, but such customers may have responded in surveys that they use multiple clouds.

Methodology

I.7 We matched customers' names across the customer data sets from AWS, Microsoft and Google. We used two types of matching:

(a) Perfect matching: exact matches of customer names across data sets.

(b) Fuzzy matching: matches based on similar but non-identical strings in customer names. Fuzzy matching produces a similarity score based on how good the match is, with 0 meaning the two are not a match and 1 meaning a perfect match. We chose to use fuzzy matching to capture additional matches where customer names may have been recorded slightly differently across the providers' data sets (eg 'Company A' in one data set but 'Company A LTD' in another).

I.8 We excluded customers that spent less than $1,000 a year on a provider. In the first instance, this was because one provider provided their data set on the basis of customers spending at least $1,000 on their cloud. Further, we consider this approach to be appropriate because it eliminates customers that are spending relatively little on a cloud, and therefore more likely only to be trialling that provider, as opposed to using multiple clouds in a material way.

I.9 We note that this method counts customers as using multiple clouds in a binary manner: customers are counted as using multiple clouds if they spend over $1,000 on another cloud, irrespective of the size of that workload.

Choice of fuzzy matching threshold

I.10 As part of our fuzzy matching exercise, we needed to identify a suitable threshold for the similarity score for each match. The choice of threshold influences the accuracy and quality of the results, and so requires thorough sensitivity tests to make sure the choice of threshold most accurately reflects the reality of multi-cloud.

I.11 In our working paper, we tested a range of different thresholds to understand the extent to which the identified fuzzy matches identified true matches or false matches. In doing this we took a random sample of matches that were identified to be in certain ranges and manually inspected if they were true matches or not. For each data set, we drew a random sample of 50 observations (giving a total sample of 150

observations for each similarity score) and manually checked whether each 'match' was, based on our own knowledge, correct.

I.12 Each sample was then given a score based on the proportion of matches deemed correct, eg if 75/150 matches were accurate the sample was given a score of 50%. This was then repeated for several samples of different similarity scores, the results of which are listed in the table below.

Table I.1: Fuzzy matching sensitivity checks

Range	*Accuracy*
0.98 – 0.985	11.3%
0.985 – 0.99	32%
0.99 – 0.9925	33.3%
0.9925 – 0.995	46.7%
0.995 – 0.9975	62%
0.9975 – 0.9999	**78%**

Source: CMA analysis

I.13 0 table above shows that, as similarity score rises, the accuracy of the sample does too. Immediately below a similarity score of 0.99, the matches found by our fuzzy matching command are less than one-third accurate. Immediately over a similar score of 0.99, the accuracy of matches is over one third and grows to just under 50% as we approach 0.995.

I.14 Based on the above results, we have chosen 0.99 as the lower bound for the similarity score of each match. To use a higher threshold would exclude several accurate matches, potentially exerting undue influence on the results of the exercise. Using a lower threshold, while including some matches that may otherwise have been missed, would include several false positives, artificially increasing our calculated rate of multi-cloud.

I.15 On this basis and given that the 0.99 threshold already captures some false positives, we consider 0.99 the most appropriate threshold even if it excludes some true matches.

Limitations

I.16 Our analysis is subject to the following limitations and should be interpreted in light of these caveats.

(a) The analysis is sensitive to the threshold chosen for matching customer names. As discussed above, fuzzy matches are assigned a similarity score based on how good the match is. We conducted sensitivity checks of the quality of matches at different thresholds and therefore chose a cut-off of 0.99 similarity score for the purposes of the analysis. If this threshold is too high, it would mean we miss ‘true’ matches, leading to an underestimate of

multi-cloud prevalence. Conversely, if the threshold is too low it would mean we match 'false' matches, leading to an overestimate of multi-cloud prevalence.

(b) Customers may have been recorded under different names in different data sets. If so, the fuzzy matching would not identify these customers, even though they are using multiple clouds, resulting in an underestimate of multi-cloud prevalence.

(c) New smaller customers may not be paying much to a cloud provider if the initial cloud credits cover most of their needs. These customers will not be identified in the matching exercise if their recorded spend is less than $1,000 even if in subsequent years the same activity would lead to a spend over $1,000.

(d) Our analysis is based on the data sets from AWS, Microsoft and Google. This means that customers that use other cloud providers, such as Oracle or IBM, as an alternative cloud will not be identified in the matching exercise. We adjust our estimates by assuming that 50% of all customers of all other providers multi-cloud [1] (an assumption we believe to lead, if anything, to an overestimate of the prevalence of multi-cloud, given our results below).

(e) Given the data available, we cannot tell where customers that use multiple clouds lie along the spectrum of siloed multi-cloud to integrated multi-cloud. For example, if firms have different subsidiaries that use different clouds, but the clouds do not communicate, we will record them here as customers that use multiple clouds in the same way we would a customer that has highly integrated clouds.

Results

I.17 In this section, we set out the:

(a) prevalence of multi-cloud, unweighted and weighted, by total annual cloud spend [2] across all customers;

(b) average spend split across clouds across all customers that multi-cloud;

(c) prevalence of multi-cloud by spend band; and

[1] We make this assumption as we consider customers of smaller cloud providers are more likely to be multi-clouding than those of larger providers. For example, the Jigsaw report describes IBM and Oracle as 'secondary' cloud providers (see Cloud Services Market Investigation Qualitative Customer Research conducted by Jigsaw (2024), paragraph 1.3.9, page 31 and page 32).
[2] This weighting gives those with higher total annual spends on cloud services a greater 'weight' to reflect their increased importance relative to those with lower spends.

(d) average revenue split by spend band.

Prevalence of multi-cloud

I.18 The table below shows the results of our analysis on the prevalence of multi-cloud, both unweighted and weighted by spend.

Table I.2: prevalence of multi-cloud, unweighted and weighted by spend, 2020-2023

Prevalence of multi-cloud	*2020*	*2021*	*2022*	*2023*
Unweighted (%)	7.04%	7.50%	7.48%	7.50%
Weighted by revenue (%)	**34.80%**	**37.41%**	**38.07%**	**38.47%**

Source: CMA analysis of customer data provided by AWS, Microsoft and Google

I.19 Table above shows that based on our analysis:

(a) approximately 7% - 7.5% of customers in the data set use at least two of AWS, Google and Microsoft (unweighted); and

(b) about 35-40% of all spend is by customers that multi-cloud.

Average spend split of customers that multi-cloud

I.20 Based on our analysis, the average spend split across clouds when operating a two-cloud architecture is around 80/20 – that is 80% on the primary cloud and 20% of spend on the secondary cloud.

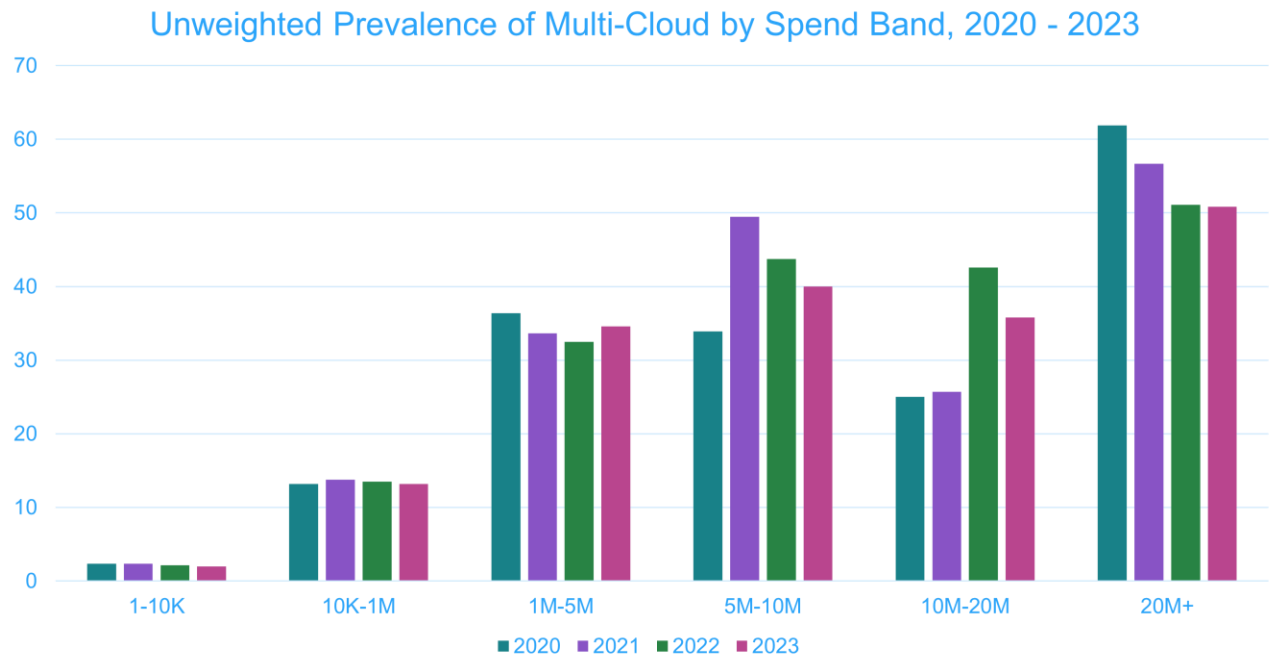
I.21 In a three-cloud architecture, the average spend split is approximately 75/20/5.

Prevalence of multi-cloud by spend band

I.22 We have also considered how some of these metrics differ when splitting customers into different spend bands (eg $1,000 to $10,000, $10,000 to $1 million). This is to better understand what is driving the differences between unweighted and weighted estimates of the prevalence of multi-cloud (see 0 above) and the extent to which the average spend split of customers that multi-cloud differs based on the size of the customer.

I.23 The figure below presents the unweighted prevalence of multi-cloud by spend band.

Figure I.1: unweighted prevalence of multi-cloud by spend band (%), 2020-23



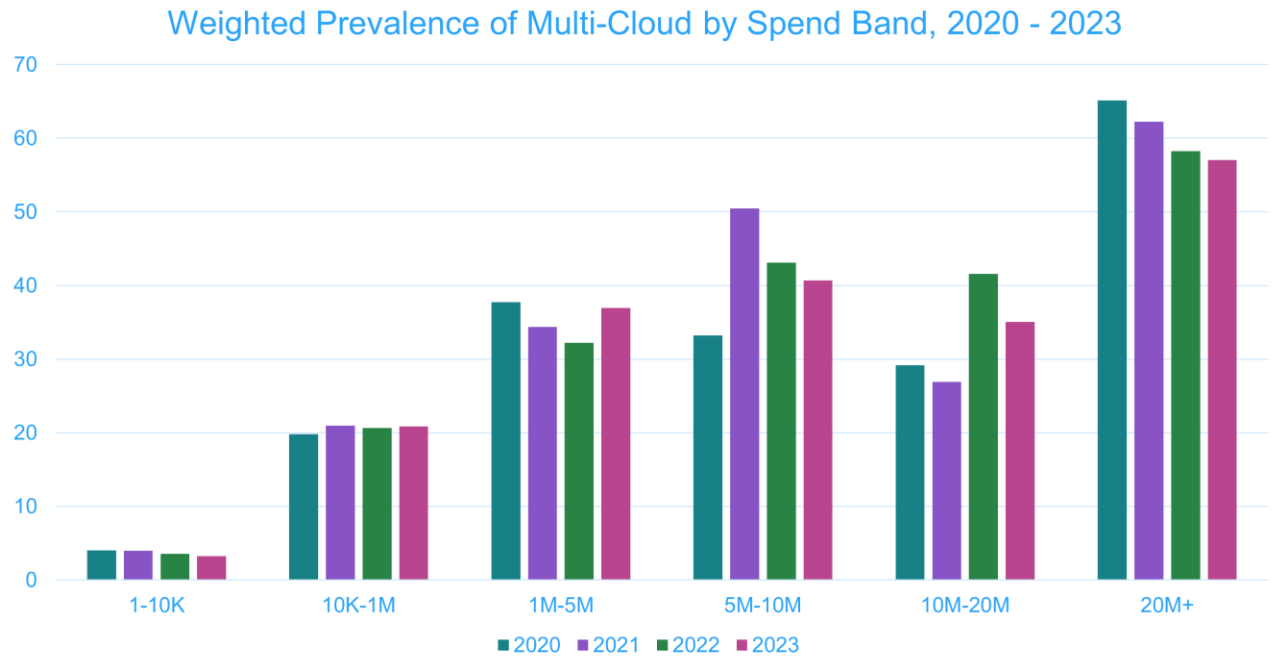
Source: CMA analysis of customer data provided by AWS, Microsoft and Google

I.24 The figure above indicates that the prevalence of multi-cloud tends to increase with the amount of total spend on cloud.

I.25 In particular, our analysis indicates that only around 2% of customers that spend less than $10,000 on cloud use multiple clouds, compared to 50-60% of customers that spend over $20 million.

I.26 The figure below presents the weighted prevalence of multi-cloud by spend band.

Figure I.2: weighted prevalence of multi-cloud by spend band (%), 2020-23



Source: CMA analysis of customer data provided by AWS, Microsoft and Google

I.27 The figure above indicates that, also when weighting customers by their cloud spend, the prevalence of multi-cloud tends to increase with the amount of total spend on cloud.

I.28 In particular, our analysis indicates that only around 3.5% - 4% of spend from customers that spend less than $10,000 on cloud use multiple clouds, compared to 55-65% of spend from customers that spend over $20 million.

Average spend split of customers that multi-cloud by spend band

I.29 0 table below presents the average percentage of spend that customers in each spend band allocated to their primary cloud (ie the cloud with the highest spend) in the years 2020-2023.

Table I.3: Average spend split of customers that multi-cloud by spend band, 2020-2023

Spend band	*2020*	*2021*	*2022*	*2023*
Less than 10k	66.9	67.1	66.8	66.5
10K – 1M	82.3	82.6	82.7	83.0
1M – 5M	88.9	90.9	91.3	90.5
5M – 10M	86.7	89.9	91.4	92.4
10M – 20M	94.3	82.0	82.7	83.7
Over 20M	79.3	85.4	87.1	86.1

Source: CMA analysis of customer data provided by AWS, Microsoft and Google

I.30 The table above shows that, in general, customers in higher spend bands who use multi-cloud concentrate more of their spend on their primary cloud. In comparison, lower-spend customers who use multi-cloud have a more even split across clouds.

I.31 We note that the fluctuations in average spend split in the higher spend bands (ie $10 million+) is likely to be due to the low number of observations in those bands. As such, if those few customers change their behaviour year-on-year, it will be reflected in the overall spend band averages changing.

Other evidence on the prevalence of multi-cloud

I.32 This section sets out cloud providers and customers' views on the prevalence of multi-cloud.

Cloud providers' views

I.33 Cloud providers generally submitted that they consider that multi-cloud architectures are common and that enabling customers to multi-cloud is part of their business strategy. Some cloud providers said that using multiple clouds is particularly prevalent among large customers. In particular:

(a) Microsoft said that its working assumption is that all customers multi-cloud at least to some degree,[3] and that multi-cloud is now the de-facto standard.[4]

(b) AWS submitted that use of multi-cloud is widespread.[5] As supporting evidence, AWS mentioned statistics from the Flexera 2023 'State of the Cloud' Report, Oracle's 'Multi-Cloud in the Mainstream' report, and Gartner.[6]

(c) Google said that it is a challenger to AWS and Azure, so in order to win customers it has to convince customers with workloads already in the cloud to move any future workloads to its cloud.[7]

(d) [✂] said that large customers are, or could be more likely to, multi-cloud.[8] Providers suggested this is due to the customers' internal resources, wider variety of use cases, skills and scale. [9]

I.34 We note that an internal document from one cloud provider is not consistent with its submission to us that multi-cloud is widespread. The document states, in the context of launching a new product, that [✂] was not at the beginning of a multi-cloud support strategy. The document said that key decision-makers in private and public companies tended to choose one single provider. In particular, the document states that, in relation to [✂].[10]

---

[3] Microsoft's response to the CMA's information request [✂].
[4] Microsoft's response to Ofcom's interim report dated 31 May 2023, paragraph 104.
[5] AWS' response to the CMA's information request [✂].
[6] AWS' response to the CMA's updated issues statement and working papers, paragraph 16.
[7] Google's submission to the CMA [✂].
[8] Responses to the CMA's information requests [✂].
[9] Responses to the CMA's information requests [✂].
[10] [✂] response to the CMA's information request [✂].

I.35 In relation to the type of multi-cloud used by customers, Microsoft said that the lack of use of 'integrated' multi-cloud is because it does not yield significant customer benefits and that there may be good reasons why customers concentrate their spend around a primary provider and/or to run different workloads separately in different clouds.[11] Microsoft also said that, of customers that multi-cloud, it estimated that there is roughly a [✂] split between 'single app, single cloud', 'same app, any cloud' and 'app spans multiple clouds' architectures, respectively.[12]

I.36 AWS and Google pointed to the use of specific cloud services as an indicator of customers adopting multi-cloud:

(a) AWS explained that customers can manage their users in another on-premises or cloud directory and then connect them into that provider's cloud through the provider's Identity Access Management (IAM) solution. It said that approximately [✂]% of AWS' customers with annual spending greater than $[✂] use third party identity sources. It added that, of these customers, approximately [✂] use AzureAD (now EntraID) and a further [✂] use Google IAM.[13]

(b) Google said that the fact that a customer is using BigQuery Omni can give an indication that the customer is deploying a multi-cloud strategy.[14] This product allows customers to query data on other cloud providers' storage services (for example, Amazon Simple Storage Server or Azure Blob Storage), and is therefore designed for multi-cloud data management.[15] Google said that it is a pioneer of tools designed to remove operational friction typically associated with using multiple clouds.[16]

I.37 We consider that both of the metrics submitted by AWS are helpful indicators of the presence of multi-cloud – although we acknowledge that they do not provide information on the level of integration of such multi-cloud.

I.38 Similarly, we consider that, since BigQuery Omni is used for querying data from other clouds to use on Google, this also provides an indicator of the prevalence of multi-cloud. While it is not possible to know the degree of integration of such multi-cloud architecture, the mere use of BigQuery Omni suggests some non-trivial level of integration is adopted by customers using it. Nonetheless, Google submitted that [✂] customers used BigQuery Omni in 2023 – we consider this to indicate a low level of multi-cloud under this metric.[17]

---

[11] Microsoft's response to the Issues Statement, paragraph 40.
[12] Note of meeting with Microsoft [✂].
[13] AWS' submission to CMA [✂].
[14] Google's response to the CMA's information request [✂].
[15] Introduction to BigQuery Omni - Google Cloud
[16] Google's submission to the CMA [✂].
[17] Google's response to the CMA's information request [✂].

I.39 Cloud providers generally submitted that there is a high prevalence of multi-cloud in the market, and one provider submitted that customers multi-cloud 'when it works for them'.[18] However, we note that in general these submissions provide limited evidence on the prevalence of multi-cloud. To this point, [✂] cloud providers highlighted surveys that sought to estimate prevalence of multi-cloud, and we consider these surveys below. One cloud provider submitted a quantitative analysis which we consider in the next sub-section.

Quantitative analysis from a cloud provider

I.40 A cloud provider submitted that the majority of the tenders it participated in between 2009 and 2021 were issued by its existing customers, indicating that customers do not view themselves as locked into a single cloud provider. This cloud provider said that its win rate in tenders for customers with existing workloads in its cloud is [✂] its win rate for other customers. It said that this shows it does not enjoy a significant advantage as an incumbent cloud provider.[19]

I.41 The cloud provider also submitted analysis on the distribution of revenue share of customers in its opportunity data[20] by the number of cloud providers the customers awarded tenders to between 2018 and 2022. This analysis showed that many of its customers [✂] used the cloud provider after awarding tenders to at least one other cloud provider between 2018 and 2022. It submitted that this is an indication of multi-clouding among customers.[21]

I.42 The cloud provider said that its analysis likely understates the prevalence of multi-cloud because: [22]

(a) many customers acquire IT services without a tender process. These customers would not be recorded in the opportunity data set;

(b) the provider did not participate in all tenders issued by customers;

(c) it is often not clear who won the tender. The analysis only flags customers as having awarded a tender elsewhere if the cloud provider knows the identity of the other competitor that won the tender; and

(d) some customers may have awarded tenders before or after the sample period.

[18] Microsoft's response to the CMA's Competitive landscape, Committed spend agreements and Egress fees working papers, paragraph 3
[19] [✂] submission to the CMA [✂].
[20] [✂]. [✂] response to Ofcom's information request [✂].
[21] [✂] response to the CMA's information request [✂].
[22] [✂] submission to the CMA [✂].

I.43 We consider that the analysis has several limitations, and therefore needs to be interpreted with care in light of these caveats and our view is that we should place less weight on it than our own assessment set out below:

(a) In the first instance, the analysis only includes customers that went through a tender process for their workload(s). While the provider said this may understate the prevalence of multi-cloud, we have not seen evidence to support a position for there to be lower prevalence among customers that tender relative to those that do not.

(b) Moreover, one potential explanation for demand being placed without a tender is that the customer may consider its current provider the only credible option. Therefore, the exclusion of this demand could bias the analysis in favour of suggesting that the use of multiple clouds is prevalent (by restricting the analysis to customers that have indicated a willingness to consider multiple clouds by running a tender).

(c) We understand that the majority of customers do not use a competitive tender process to acquire cloud services (see Chapter 3).

I.44 Further limitations that may affect the analysis are:[23]

(a) Weighting customers by their spend on the provider does not take into account how much customers may have spent on other cloud providers. To the extent that weighting by spend is informative, the results should be weighted by customers' total spend on public cloud.[24]

(b) Customers are counted as using multiple clouds if they put one workload on another cloud, irrespective of the size of the workload.[25] This is a limitation of our own analysis too, as discussed above. This implicitly assumes that all of these workloads face competition from rival providers.

(c) Data sets used by the provider may not be reliable. The main data set used consists of data where 'opportunities are manually made by members of [the provider's] sales team'. As a result, the data 'is not always comprehensive and may contain errors and inconsistencies'.[26] The provider noted that the data is 'often comprised of anecdotal feedback from the customer'. The

[23] These had already been raised by Ofcom during its market study: Ofcom's Final Report, Annex 3, paragraph A3.45-A3.48.
[24] We note that our analysis only includes spend on AWS, Microsoft and Google, and therefore will not capture spend on smaller cloud providers.
[25] Our analysis is also subject to the same limitation. We have estimated the average revenue split across providers (by customers' spend band) as one way to account for this factor.
[26] [✂] response to Ofcom's information request [✂].

provider is therefore ‘not able to accurately assess how representative and comprehensive the data set is in percentage terms’.[27]

I.45 Due to these limitations, we place limited evidential weight on the results coming out of this analysis. Overall, we place more weight on our analysis than that submitted by this cloud provider.

Quantitative surveys

I.46 AWS and Microsoft submitted that independent surveys and industry reports show that using multiple clouds is common.[28] We assess these surveys in Chapter 3 and Appendix C.

[27] [✂] response to the CMA’s information request [✂].
[28] Submissions to the CMA [✂]; [✂] response to the CMA’s information request [✂].