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; and

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

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 market 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 cloud 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.

Limitations

I.10 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 entirely 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.11 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

(d) average revenue split by spend band.

[1] We make this assumption as we consider customers of smaller cloud providers are more likely to be multi-clouding than those of the three main 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 Research (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.

I.12 We note that this analysis is not intended to be read as a precise estimate of the prevalence of multi-cloud – rather, the general magnitude of the results will be what informs any conclusions based on these results.

Prevalence of multi-cloud

I.13 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 (%)	6.8%	7.2%	7.4%	7.5%
Weighted by revenue (%)	37.6%	38.5%	40.0%	40.8%

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

I.14 The 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 37-40% of all spend is by customers that multi-cloud.

Average spend split of customers that multi-cloud

I.15 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.

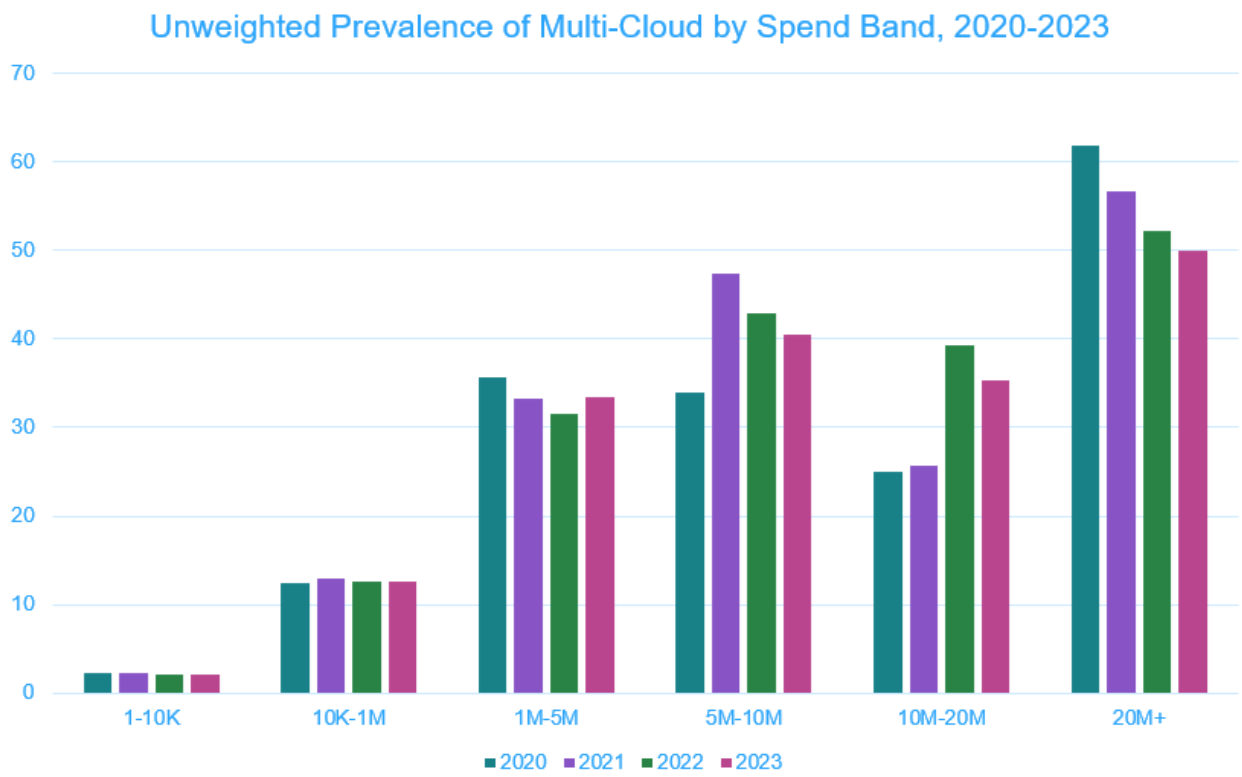
I.16 In a three-cloud architecture, the average spend split is approximately 75/20/5.

Prevalence of multi-cloud by spend band

I.17 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 Table I.2 above) and the extent to which the average spend split of customers that multi-cloud differs based on the size of the customer.

I.18 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



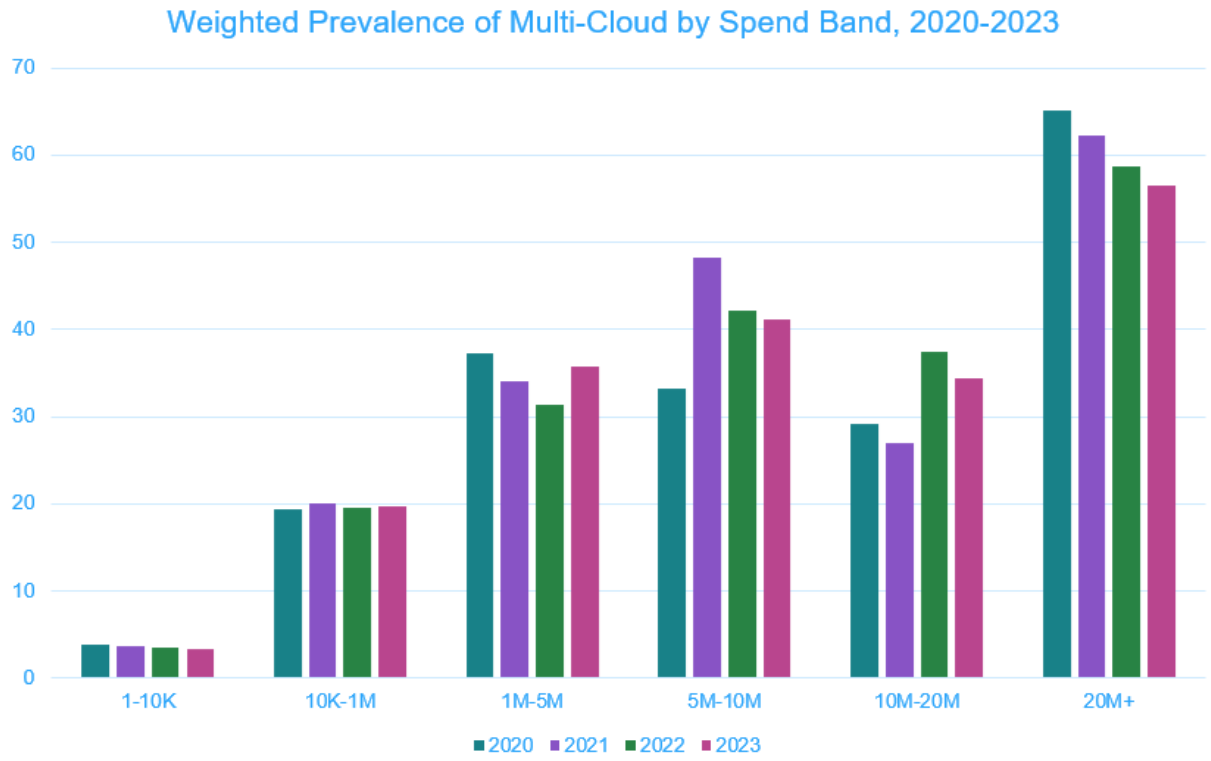
Source: CMA analysis of customer data provided by AWS, Microsoft and Google

I.19 The figure above indicates that the prevalence of multi-cloud tends to increase with the amount of total spend on cloud.

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

I.21 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.22 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.23 In particular, our analysis indicates that only around 3%-4% of spend from customers that spend less than $10,000 on cloud use multiple clouds, compared to about 55-65% of spend from customers that spend over $20 million.

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

I.24 The 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	67.1	67.1	66.6	66.4
10K – 1M	82.4	82.7	82.7	83.1
1M – 5M	88.8	90.7	91.3	90.3
5M – 10M	86.7	89.8	91.6	92.5
10M – 20M	94.3	82.0	83.4	84.6
Over 20M	79.3	85.4	86.3	85.2

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

I.25 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.26 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 a few customers in these spend bands 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

Quantitative analysis from a cloud provider

I.27 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.[3]

I.28 The cloud provider also submitted analysis on the distribution of revenue share of customers in its opportunity data[4] 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.[5]

I.29 The cloud provider said that its analysis likely understates the prevalence of multi-cloud because: [6]

(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.

---

[3] [✂] submission to CMA [✂].
[4] [✂]. [✂] response to Ofcom's information request [✂].
[5] [✂] response to the CMA's information request [✂].
[6] [✂] submission to CMA [✂].

I.30 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 above:

(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 of multi-cloud among customers that tender relative to those that do not.

(b) In addition, 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.31 Further limitations that may affect the analysis are:[7]

(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.

(b) Customers are counted as using multiple clouds if they put one workload on another cloud, irrespective of the size of the workload.[8] 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'.[9] The provider noted that the data is 'often comprised of anecdotal feedback from the customer'. The provider is therefore 'not able to accurately assess how representative and comprehensive the data set is in percentage terms'.[10]

[7] These had already been raised by Ofcom during its market study: Ofcom's Final Report, Annex 3, paragraph A3.45-A3.48.
[8] 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.
[9] [✂] response to Ofcom's information request [✂].
[10] [✂] response to Ofcom's information request [✂].

I.32 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.33 AWS and Microsoft submitted that independent surveys and industry reports show that using multiple clouds is common.[11] We assess these surveys in chapter 3 and appendix C.

[11] Submissions to the CMA [✂]. [✂] response to the CMA's information request [✂].