# Concentration Risk: Perspectives from Microsoft

September 2020

# Contents

# Introduction

**As adoption of cloud computing becomes more prevalent in the financial services industry, the topic of concentration risk has consistently been a source of interest, in discussions with regulators and customers concerning outsourcing, including use of cloud services. Due to a lack of clarity on these issues, financial institutions may conclude that a risk averse posture dictates a multi-cloud strategy must be adopted. Microsoft's global legal team has, in fact, found no regulatory guidance mandating a multi-cloud strategy. Rather, as with all forms of outsourcing, concentration risk is one of many risks that must be assessed, with a plan for avoidance, acceptance, or mitigation. While regulatory guidance exists, it deliberately leaves implementation details to individual institutions. The "Guidelines on outsourcing arrangements", published by the European Banking Authority,[1] as an example, raises concentration risk as a potential issue but leaves the corresponding response to its member banks. Accordingly, these institutions must develop governance and have assurance plans in place to manage such risks when using cloud services, as they do today in their use of existing legacy systems or other forms of outsourcing arrangements.**

Concentration risk is, at times, used as a reason not to use cloud services, or to mandate multi-source strategies, without sufficient consideration for the underlying risks that need to be addressed, including: service failure, cyber security, financial stability of the provider, and potential for vendor lock-in. Both regulators and decision makers in the banking and capital markets industries should understand these issues to effectively navigate the complexities involved. Indeed, much deeper analysis is required before broadly categorizing use of cloud itself as raising concentration risk, or in considering implementation of a multi-sourcing strategy. Cloud adoption, especially as used for material banking systems, is still nascent and concentration risk for cloud is nowhere near the level of concentration risk as it exists already today for other third-party arrangements such as use of mainframe and other legacy on-premises infrastructure. Further, competition among outsourcing providers is intense, and a mix of systems with institutions without an "all in the cloud approach" is here to stay for a long time. With this in mind, risk and procurement officers at financial institutions need to respond to regulation and ensure their decisions are optimized against meaningful risk without holding their individual institution back from the opportunity these technologies offer. This White Paper provides information on steps to assess and mitigate against such risk and, at the same time, implement approaches without the need to adopt a multi-sourcing strategy, which has its own drawbacks.

[1] See https://eba.europa.eu/sites/default/documents/files/documents/10180/2551996/38c80601-f5d7-4855-8ba3-702423665479/EBA%20revised%20Guidelines%20on%20outsourcing%20arrangements.pdf?retry=1. (The **"EBA Guidelines"**)

As a starting point, it is critical to understand the specific regulations governing an institution, as well as the real world factors to consider in a concentration risk assessment. From a global perspective , the Financial Stability Board acknowledged that *"there do not appear to be immediate financial stability risks stemming from the use of cloud services by financial institutions."*[2] There are regional guidelines to consider as well. The UK's Prudential Regulation Authority has stated[3] that it will continue to monitor "*discussions and potential future regulatory developments relating to systemic concentration risk*" without making any mandate or implementing requirements for multi-sourcing. As stated:

***the PRA will expect firms to assess the resilience requirements of the outsourced service and data and determine which of the available Cloud resiliency options is most appropriate. These may include multiple availability zones, regions or service providers.***[4]

So, while acknowledging that concentration risk is an area to monitor, the UK PRA acknowledges that *"if configured correctly, cloud services can significantly improve the operational resilience of individual financial firms."*[5]. Financial institutions should not lose sight of this when assessing such risks and strategies in adopting cloud services and other forms of outsourcing.

Microsoft has worked with many financial institutions that have determined cloud services like Azure, Microsoft 365 and Dynamics 365, offer net lower risk, in managing systems and banking operations in a safe, secure and resilient way, consistent with regulatory expectations wherever they do business around the globe. Many have concluded that a multi-cloud strategy carries greater risk than a primary cloud strategy. Their risk management plans conclude that selecting Microsoft as a primary cloud vendor is not only consistent with regulatory requirements, but meets the strict regulatory compliance needs aligned to their risk management strategy. As financial institutions know well, proper management of risk can be turned into a competitive advantage. In addition, concentration risk is not a net new area of concern, but one that has existed with other systems for decades. For instance, the historically high concentration of international payments happening through a single well secured global payments service provider may have led to positive outcomes because of concentration. Indeed, we believe outsourcing the risks associated with legacy on-premises systems to Microsoft also fits the criteria of being a highly secure and resilient

[2] See https://www.fsb.org/2019/12/third-party-dependencies-in-cloud-services-considerations-on-financial-stability-implications/

[3] See https://www.bankofengland.co.uk/-/media/boe/files/prudential-regulation/consultation-paper/2019/cp3019.pdf?la=en&hash=4766BFA4EA8C278BFBE77CADB37C8F34308C97D5 (The **"PRA Outsourcing Consultation"**)

[4] The PRA Outsourcing Consultation, paragraph 2.42.

[5] The PRA Outsourcing Consultation, paragraph 2.5.

service provider and can have net positive advantages to individual institutions, as well as financial consortia and markets.

## What is Concentration Risk

There are two elements to concentration risk: (i) macro risk at a system wide level within the financial ecosystem, and (ii) micro risk at the level of the financial institution. These comprise separate risks. The first constitutes an assessment regulators must address, by looking at the totality of dependence on any one provider, for core systems that run banking and insurance operations. The second involves an assessment at the customer level only, where an over-reliance on one provider could significantly impair operations of the financial institution.

### i. **Macro Risk**

Concentration risk at a macro-level concerns systemic risk at the financial ecosystem, as a collective, in relying on one vendor. The concern being that, if something catastrophic were to occur, the entire economy of that region would be affected, given the concentration of financial institutions affected. Note that this is primarily the concern of regulators, not their supervised institutions. But as they have not mandated that institutions multi-source vendors with legacy systems, so too they have not with cloud. They may use other tools at their disposal, such as examining such providers when warranted. (The Bank of England has such regulatory powers to request and obtain information from third party outsourcers, including rights of examination under the Financial Services Markets Act of 2000.)

When it comes to cloud computing, this risk remains mostly theoretical:

1. Cloud adoption of core banking systems is nascent.
2. The industry itself is fragmented.
3. Legacy and mainframe systems dominate IT functions of core banking systems.
4. Various traditional financial services are already systemically concentrated, such as market data and trading platform providers.
5. Use of systems is fragmented.
6. From a global perspective, as banks do more business regionally and globally, such assessments even become more disparate.
7. Distributed use of implementations across multiple data centers mitigates risk of single point of failure.

Conversely, cloud allows easier remote inspection, removes vendor lock-in for mainframe and similar legacy infrastructure, and provides greater resiliency because of its distributed architecture, which can provide more assurance, not less, if appropriate governance is in place. Regulators have at their disposal tools to examine services providers, and Microsoft commits in its contracts to the right of regulatory examination, as required under major guidelines (e.g., EBA). This enables regulators to take a macro view in assess risks, as appropriate, based on adoption of cloud services in the marketplace.

### ii. Micro Risk

This second category is the set of risks that apply to individual institutions, as opposed to collectively. The concern is that a single institution takes a dependency on a vendor such that a catastrophic event affecting that vendor would impact the institution and, depending on the systemic importance of that institution, the larger economy in that region.

As further explained below, such risks are addressed by the overall resiliency of cloud services and, customer configuration to mitigate risks of a single point of failure. Overall resiliency is addressed by the distributed architecture of systems, with geo-replication of such services with regional pairing to address potential catastrophic events. Further, customer configuration requires (depending upon criticality of systems), use of availability zones and configuration of regional pairing so if one full region were to be impacted, another would still be active.

Regulators themselves have acknowledged the benefits of cloud computing because of the resiliency it provides. Indeed, the PRA explained recently that, "reliance on outsourcing and third parties bring potential benefits and opportunities, including, in the case of Cloud, potentially enhanced resilience compared to firms' on-premise data centers (provided that firms oversee the provision of Cloud services effectively and take appropriate steps to protect their applications and data)"[6].

## Risk Management

In working with regulators, Microsoft has come to understand their desire to see that we treat risk management with as much diligence as they hold their supervised institutions accountable for. As important context financial institutions around the world have depended on Microsoft's operating systems, database, and server products as part of their infrastructure for more than two decades. In recognizing that the financial services industry and other critical infrastructure industries rely on our technologies (both legacy software and cloud computing), we have an enterprise risk management function (ERM) that operates across the entire company, documents and categorizes risks, and reports them to Microsoft's board of directors.  Each top-level risk gets assigned to a member of the senior leadership team – some even being assigned to the CEO himself – for management, monitoring and reporting back up to the board of directors. This flows down to each of the business units at Microsoft. This level of governance at the business unit level, with ultimate reporting to the Microsoft Board of

[6] The PRA Outsourcing Consultation, paragraph 1.9 on page 3.

Directors, provides a level of overall assurance, structure and accountability to maintain the security and continuity of our cloud services.

For example, risk of outages in Azure is managed by a compliance team in the Cloud and AI engineering group, who drive compliance down to individual engineers and flow information all the way back up to the board via the ERM team. This is a mature program, with an internal investigations team, risk modeling (impact, likelihood) and practices (improve, monitor, tolerate, operate) that have been audited by many 3rd parties as part of our commercial and financial services audits.

### Vendor Stability

An underlying concern related to concentration risk is that an otherwise financially stable institution is taking a dependency on an unstable vendor. The concern is that under financial stress, a vendor could impact the institution’s stability. We counter this concern by pointing out that Microsoft is as well or better placed to mitigate risks arising due to adverse micro and macroeconomic events than even most banks. We have a robust hedging strategy both from financial as well as operational risk mitigation standpoint.  We hold a AAA credit rating from Standard & Poors, one of only two companies with this rating today. We’re the only company to be consistently placed in the top 10 by market cap evaluation for the past twenty years. The stability of Microsoft, relative to other vendors, and even to banks themselves, is hard to question.

### Vendor Lock-in

There is industry concern that once on the cloud, customers cannot move off the cloud in porting their data to either on-prem systems or to other cloud service providers. This is a risk that exists when selecting vendors for insourced and outsourced functions and must be carefully managed. The key for mitigation of this risk in the cloud is architecting application patterns with transportability of workloads as a consideration. Modern infrastructure patterns and standardization, including virtualization, containers, and open source standards, are entirely compatible with this goal. These patterns should be paired with an exit plan, to show how these applications could be moved to a different environment if needed.  Such compatibility and standardization are likely also to reduce vulnerabilities for the purposes of institutions' operational resilience mapping.

For these considerations, it is not expected that applications would need to be moved immediately. Exit plans should serve as a baseline for both well planned migrations as well as those under stress. Migrating from on-premises to cloud, in fact, provides an opportunity for institutions to move their application architectures to be more supportive of transportability than they are often on-premises. We have had GSIFI banks migrate from the highly proprietary Oracle database to PostgreSQL on Azure, an open-source database that can be run in most environments. Given economies of scale, and as witnessed over decades in traditional computing environments, the pattern of continued decreasing prices is expected to continue. These observed trends coupled with contractual mitigations, and proper exit planning, yield negligible residual risk.

### Outages

There is concern that an outage in a cloud will cause significant application downtime. While no infrastructure runs with zero downtime, there may be perception that on-premises availability exceeds that of cloud. This is often a consequence of news cycles that amplify outages in the cloud as being catastrophic, where more consequential on-premises outages affecting a single institution often go unreported. But uptime cannot be judged in absolute terms. It must be judged relative to current state, and it must be based on real risk, rather than perceived. It is possible that a risk averse posture slows migrations to the cloud and keeps an institution in an on-premises environment where they are in fact running *higher* risk of downtime.

Additionally, cloud availability has been demonstrated to consistently improve over time. Azure launched with options for 99.9% VM availability SLA. We added the availability sets feature, which brought the SLA up to 99.95%. We then added availability zones, which brought it up to 99.99%. And our modern services are pushing this even higher. Cosmos DB, a PaaS database service, offers 99.999% SLA for read and write availability. Given this history and the innovation that continues in the cloud, it is expected that availability will only improve over time.

The potential for outages has generated a desire to create "active/active" configurations for cloud applications to span multiple cloud providers. In real world implementations, we have seen these efforts as being counterproductive to the ends sought. Active/active configurations are designed seeking resiliency via diversity, referring to the fact that the more diverse systems are, the less likely they are to simultaneously encounter issues.

Efforts to span workloads between clouds increases risk beyond risk reduction, yielding higher net risk impact. Adding to this complexity is the need to implement a first party (built by the institution) or third party (built by yet another vendor) solution for managing active/active status between clouds. This takes the risk associated with Azure and transfers it to another party. If first party, it will be a one-off solution requiring development and maintenance – the antithesis of what most institutions seek in moving to the cloud. If third party, the risk associated with the cloud provider is now transferred to a typically smaller, less reliable vendor.

By relying on multiple clouds, an institution would need to train security staff on two clouds. The same controls need to be implemented twice. The monitoring and operations teams would need to learn two different clouds, all spreading resources and expertise thin, and adding more risk than was offset via the attempt to reduce risk associated with a single cloud. Synchronizing data between clouds requires complex architecture, given higher latency between clouds than within a single cloud. This compounds the risk as read/write collisions and conflict reconciliation algorithms increase complexity as latency increases. Compare this to services provided natively by Microsoft that offer low latency replication, span multiple regions for high availability, and come with conflict reconciliation as a service.

We recommend relying on Microsoft investments in building "diversity as a service". At the platform layer, we have two DNS infrastructures configured active/active, one Windows and one Linux. We have multiple petrol providers for our generators. In order to avoid overreliance on the connectivity between New York and London, we ran a second subsea line connecting North America to Europe via Virginia to Spain. We have 60+ regions around the globe, making it relatively simple for customers to spread their workloads across many data centers for high availability and resiliency. We are investing in infrastructure and resiliency improvements at a pace that greatly exceeds most financial institutions. Taking advantage of these ever-improving innovations in diversity is a better strategy than building them in-house or adding another vendor to solve them.

## Security

Security presents an additional set of risks associated with concentration. As concentration risk is primarily concerned with catastrophic events, it is helpful framing to focus on security events that could be deemed systemic to a cloud provider. The most common scenarios raised are insider threat, including zero-day attacks and side channel attacks; and ransomware attacks. The concern with zero-day attacks is the threat that an exploit could attack concentration points in the cloud, essentially bringing the entire cloud down with one attack. Again, relative risk is important to consider. The cloud is inherently no more susceptible to these attacks than on-premises systems are. In fact, the scale of cloud is deemed an advantage when it comes to mitigating these risks. For the same reason that physical security practices dictate choke points for entry — it is easier to control and monitor movement and anomalies — this applies to concentration in the cloud. These are similar to the principles that drive banks to shut down all entry points from the Internet except known web interfaces and email. These allow focus on specific ports and protocols which can be hardened and monitored closely. Microsoft also understands concentration points in our environment. Code being deployed to production, as an example, is deemed a concentration point which is carefully managed and monitored. Humans have no standing access to production systems and the workstations used to deploy production code are hardened and air gapped from internet connected devices. On-prem legacy solutions, especially when operated by outsourcing providers, are not inherently less concentrated from a cyber risk perspective, as the recent "Cloud Hopper" incident demonstrated, because these vendors have connectivity into many on-prem environments.

### a) Zero-day attacks

Being able to patch systems quickly is one of the primary ways to mitigate zero-day attacks. Because the millions of servers that power the cloud are built and deployed at scale with automation, the operating systems and configurations are much more consistent and uniform than traditional IT estates, which are often a diverse set of infrastructure and software versions. Heterogeneity of versions and configuration is what slows down emergency patching. Automation is more difficult and failure more common.  As we observed in response to the Spectre and Meltdown vulnerabilities in 2018, Microsoft was able to patch the entire cloud infrastructure much more quickly than our customers were able to their on-premises environments.

### b) Side-channel attacks

A related concern is the threat of side-channel attacks, which refer to the risk that, in a multi-tenant environment, hackers could execute code on the same physical machines as a banking system runs on. It has been shown in controlled lab conditions that there are attacks from this vector that make it more likely to be effective than would be possible in physically isolated machines. For those that want to fully mitigate this risk, Microsoft does offer dedicated servers in the cloud. But the risk associated with side-channel attacks in real world cloud scenarios, outside of lab conditions, approaches zero. The reasons for this are that a) side-channel attacks assume the adversary knows the physical location of the target workload, b) can physically place themselves on that same location, and c) can stay on the physical location with the target for an extended period of time. The probability of all of these conditions to occur simultaneously in the cloud is effectively nil. Azure relies on hyper optimized use of physical resources and therefore load balances virtual tenants constantly. Consequently, two tenants are collocated on the same physical host for very short periods of time. For these reasons, there are no real-world examples of this type of attack in the cloud. Compare this to spear phishing attacks, which are one of the most common exploits seen.  The same resources that are spent architecting and managing an air gapped solution, over engineered for the theoretical concern over side-channel attacks, could be repurposed to focus on mitigation of realistic exploits.

### c) Ransomware

The last subcategory of security risk is ransomware. U.S. regulators began contemplating this risk in 2014, which led to a set of recommendations the industry responded to first via the Sheltered Harbor initiative, and now expanded to additional recommendations for snapshots for full systems and their dependencies. Sheltered Harbor calls for establishing an air-gapped set of backups for critical data such as customer account balances and allowing restoration of this data, even if restored to a different financial institution as a last resort. The focus of this work has been on protecting on-premises systems; and using cloud as the target for backup, in fact, has been deemed a viable solution. Similarly, in the cloud, it is possible to back up to air-gapped locations. For example, snapshotting data in Azure public and storing it in one of the Azure regions designated for national critical infrastructure would fulfill this requirement. Again, decomposing security risk shows that proper understanding and planning can leverage the cloud for benefit rather than detriment.

## Strategies Optimized to Mitigate Risk and Reduce Complexity

As institutions are evaluated based on their systemic importance and materiality, so too are the workloads running within them. It is important to categorize workloads accordingly and include proportionality in mitigation investments. The employee vacation tracking system does not require the same considerations for availability as core banking systems. Thus, the same resources, planning, and testing are not required. Being explicit about the requirements aligned to application "tiers" is recommended.

Understanding a cloud provider's recommendations for high availability is essential to ensuring the relevant tiers of application are built on architecture that maximizes uptime and is consistent with portability objectives and exit strategy. For Azure, a combination of availability zones for high availability, coupled with a regional replication strategy for disaster recovery, would be foundational. Understanding a provider's options for hybrid is an additional way to balance needs for diversity with needs for consistency. When additional levels of confidence are required, a migration over time can rely on the existing on-premises environment as part of a cloud exit strategy.

As security, risk, compliance, and technical staff already understand the on-premises environment, it's less overhead than learning a third cloud. For the most systemically important systems, relying on diversity within a single cloud can mitigate risk of downtime without spreading resources to a second or third cloud. For example, a PaaS architecture could leverage an IaaS configuration for a disaster recovery environment, yielding sufficient diversity without adding significant resource burden, as would be required in implementing multiple clouds. Typical cloud implementations start with IaaS, then add PaaS; so these are capabilities the institution would already have built.

As a final consideration, compare the complexity of multiple vendors found in on-premises environments. Most institutions seek to reduce their vendor count, due to their resources being spread thin against a diverse set of technologies. Multi-cloud simply replicates these same challenges in the cloud and therefore should be avoided.

## Business Continuity and Exit Planning

Regulators expect financial institutions to have appropriate governance models to address business continuity and exit planning. Indeed, the PRA's Outsourcing Consultation Paper (as with the EBA Guidance) expects firms to develop, document, maintain and test:

- Business continuity plan

- Exit strategy, which should cover and differentiate situations where a firm exits an outsourcing agreement due to disruption
    - An outage or the failure, i.e. insolvency or liquidation of the service provider ('stressed exit')
- Commercial, performance or strategic reasons in a planned and managed way ('non-stressed exit')[7]

Starting from this position and developing an exit plan provides a baseline for a cloud strategy that mitigates concentration risk and maximizes a financial institution's ability to take advantage of cloud for both innovation and improved risk posture. Having a documented exit plan can align strategy against all subcategories of concentration risk across an entire financial institution, as well as educate internal and external stakeholders. Microsoft guidance on Exit Planning can be found here: http://aka.ms/MicrosoftExitPlan.We have depth of experience supporting customers in overall governance and business continuity planning and would be pleased to support you further on this.

[7] PRA Consultation Paper at 2.38.

# Conclusion

Microsoft believes that proper planning, with the subcategories of concentration risk accounted for, will yield reduced residual risk as an outcome of a cloud migration. This constitutes a more rational approach, relative to a multi-cloud strategy, which in most cases is considered over engineering and net negative to an overall risk posture. Further, it does not even account for the risk posture of financial institutions with their existing environments, and if concentration risk at the mainframe level and controls provide the same (or possibly less) assurance than in moving to the cloud. A more holistic view and approach, may result in different conclusions, including that a more robust approach to using cloud services, including for core systems, is not only fully consistent with regulations, but may further mitigate risks from existing environments.