Deep Science Publishing
https://doi.org/10.70593/978-93-49910-62-1 _5

DeepScience
Open Access Books

Chapter 5: Optimizing Infrastructure-as-a-Service (IaaS) solutions to achieve agility, scalability, and resilience in financial IT systems

5.1. Introduction

Cloud Computing is a technological revolution impacting a vast cross-section of the world's economy, especially forms of Information Technology outsourcing. Specifically, Infrastructure-as-a-Service in Financial IT Systems, is the providing virtual computing infrastructure dedicated to facilitate businesses dealing with financial transactions such as government banks, private banks, non-banking financial companies, and online transaction platforms. IaaS empowers financial companies with scalable and flexible choices of infrastructure/devices, lowering costs for organizations with unpredictable usage patterns and at the same time highly enhancing reliability by providing services in a centralized manner, focusing on uptime as well as security of computing resources. Consequently, this paper seeks to outline the optimization process for sizing IaaS infrastructure in Financial IT Systems while dealing with financial projects and business scenarios (Arogundade & Palla, 2023; Oye & Matthew, 2024; Chowdhary, 2025).

Financial IT Systems are applications managing and aiding in conducting financial activities revolving around transferring money. Some examples are executing transactions related to withdrawing or depositing cash from Automated Teller Machines as well as facilitating payment transactions between people such as online money transfer. Companies operating and managing Financial IT Systems rely on the availability of these services and any downtime might lead to financial losses as well as negative public perception regarding slower execution times for financial transactions. On the other side, IaaS has enabled these companies to focus on lowering the costs of owning and maintaining their own infrastructures as devices, software, and networking for a peak business interval and operating sensitive projects having a very low risk appetite to choose IaaS Solutions with a highly redundant infrastructure for a fee.

The ultimate reason preventing Flexible IT components to displace sensitive custom made systems in the Financial IT domain are safety and reliability. It remains to be seen if the Infrastructure as a Service can break the capital punishment for consistent Financial IT solutions (Samha, 2024; Subramanyam, 2025).

Fig 5.1: Infrastructure-as-a-Service (IaaS) Solutions

5.1.1. Background and Significance

The strategic importance of the Financial IT system for the overall effectiveness of a Financial institution is difficult to overestimate. Security, availability, and scalability are priceless characteristics of such a system. Availability and scalability enable the Financial institution to satisfy the demands of its users during peak periods at a reasonable cost, while IT security has obviously a direct impact on the customer's trust in the Financial institution being able to protect personal and account related information. Any trust crisis usually results in a substantial Financial institution market share reduction, and happens very quickly these days. A multi-million dollar damage usually ensues already from the consequences of substantial service degradation, or main

service interruption after only some hours in case of the Financial institution being not able to cure the problem, or getting hacked.

Over recent times the desire for cost efficiency has become detrimental to performance of important Financial IT characteristics such as security, availability, or scalability. With a share of up to 80% the operational costs are traditionally the highest part of any Financial IT lifecycle costs. In addition to scale effects, those costs can basically be kept low through inexpensive IT components as well as Workforce reduction. The latter can be achieved through an obsessive implementation of the Service Level Management concept, which in addition to IT cost reduction is also supposed to enhance customer satisfaction at given service levels.

5.2. Understanding IaaS

Definition and Overview Infrastructure-as-a-Service (IaaS) refers to the disruptively innovative service model that offers storage and processing resources accessed via the Internet, removing the burden of technology management from organizations. Cloud computing consists of pay-per-use, computing power and on-demand services accessed over a high speed Internet. Considering that on-premises infrastructure requires huge upfront capital investments, which has to be kept up-to-date throughout the entire lifetime, for many organizations IaaS is a natural evolutionary step towards IT Optimization. Ideally, IaaS enables trusted external resources to provide capacity to be dynamically harvested and shared by service requestors, such as enterprises and IT resellers. The requesters can then deliver services to their own clients, offering them immediate access to the latest technologies for a pay-only as-used price. It also allows IT companies to provide tenants with the service levels that they want and are prepared to pay for.

5.2.2. Key Components of IaaS The main components of IaaS solutions consist of the Virtualization Layer, the Network as a Service, the Storage as a Service and the Processing as a Service components. At the top level, the Virtualization Layer manages the hardware. At the managed level, the customer has complete control and total responsibility for maintaining the virtual machines with the guest OS. The customer is ultimately responsible for all aspects of the Virtualization layer, which runs the hypervisor that creates and manages the virtual machines. The only responsibility that the IaaS provider has is to make sure that high-availability machines are available. At the managed level, the IaaS provider creates the virtual machines and the guest OS, configuring them to provide the required services. The IaaS customer accesses and uses the services via the user interface, which is often a Web interface.

5.2.1. Definition and Overview

A lot of different public cloud web stakes are dropped when money is on a table. No wonder then that on the basis of various statistics the financial sector is one of the industries that contributes the most to public cloud hosting revenues. The bulk of the hosted services are traditional services such as storage, databases, etc. in their cloud forms. This made financial and banking IaaS solutions just dynamic reservoirs of computing resources users can use or lease on demand. Financial institutions such as trading and insurance companies saw an opportunity to optimize costs of operating financial applications. They changed the fixed costs of maintaining their own internal data centers with variable costs of leasing external data centers resources on demand. The IaaS financial usage models were focused on cost optimization rather than availability or security. None of the above was however a differentiator of IaaS – all of the functions pioneered in IaaS attempts by Internet companies were implemented in the form of public cloud IaaS solutions. Only after some years of public cloud presence have both positive and negative effects started to attract attention, such as co-mingling of data belonging to different institutions, inconsistent application of security rules, unpredictable performance – the benefits or threats have been acknowledged. Questions about costs were raised and contrary to earlier opinions one could even learn how to discount the initial costs going up and for what stretch of time. All these speed bumps started redefining the needs financial institutions had and the way IaaS would support those needs. These questions previously were ignored or were of secondary importance.

5.2.2. Key Components of IaaS

As the layer of the cloud stack that is closest to the hardware, IaaS systems expose a number of core hardware components through a cloud service model, supporting either arbitrary access to the full physical resources of the data center pool, or – more frequently – a virtualization layer that allows users to deploy software services that transmit their provisioning and configuration to the hypervisor. Providing access at this layer requires significant management localization and overhead, but allows users the maximum flexibility for customizing the core services their applications and workloads expose: concierge-like customization of raw, physical resources is expensive, but it is still the option that attracts the most demanding, highest-paying cloud customers. Accordingly, some IaaS systems expose access to some capabilities that lower-tier cloud services don't or can't, such as on-demand assembly of hardware machines, or special actuators like FPGA or GPU hardware acceleration. We will return to these issues in later sections, after taking a step back and explaining how IaaS services provision their core services: hardware resource provisioning. In this latter, essential layer, IaaS vendors provide cloud resources in the shape of standardized units that are optimized for commonly

encountered workloads. These units come equipped with resource modules of specific types, quantity, and capacity, and are designed to be added in bulk to the data center to match capacity demand, and scaled up or down quickly and cost-effectively during the lifecycle of a service.

5.3. Importance of Agility in Financial IT

The agility of created software applications and IT systems is an important property of modern financial IT systems. The core financial IT system is involved in almost every product, service, and process executed within a financial organization, often creating a bottleneck in the organization's flow of products and services. Agile infrastructure is an important part of agile systems. There are three levels of agility and optimization for the infrastructure of a financial organization: PC or workstation optimization, server optimization, and a center infrastructure optimization for many applications with many users. Business processes and new products are often changing in financial organizations according to business strategy and mission. Because money is the core focus of every financial organization, products and services are created and changed quickly according to the economic situation. New risks, such as credit risk, liquidity risk, and operational risk, can come to the banking industry from many factors, causing banks and other financial organizations to change product portfolios. Financial companies are focusing more on risk management and customer service. The banking system is also changed according to rules of the central banking system. Therefore, the core banking system has to be changed according to the needs of the company.

Agility is defined as the ability to develop and support systems and applications that address constantly changing business conditions and recreate and eliminate business processes, operational applications, and third-party trading and messaging connections as a reaction to the changing economy. Modern financial organizations may have hundreds of applications used in back office, front office, risk management, and treasury department automation. The products of a company are executed through those applications in IT systems, helping automate the complete product life cycle, monitor transactional processes, and manage cash flow, operational processes, and risks.

5.3.1. Defining Agility in IT Systems

The term agility is widely used in contemporary terminology to capture the scene of rapidly changing business and technology landscapes. However, there are many definitions of agility. For example, an item-based and similar definition states that being "agile" means that a company can quickly respond to changes. This is often explained as changes in the market, in competitors' reaction, or change in the customer's needs, and

is one aspect of the term. Another, more pithy, definition of agility is that it is "the ability to respond effectively to unpredictable changes". There is also a third definition of agility which states that agility is the quality of being adaptable or resilient or able to respond to change in a nimble or well-judged manner. This definition is more closely aligned to IT systems. It covers both quick responses to predictable events as well as unpredictable ones. However, the term responds does not imply the same set of actions in both cases. In the case of predictable changes, responding means anticipating the changes and planning for them. In the case of unpredictable events, agility implies that the organization has the ability to quickly and effectively assess and react to the changes made necessary by the event.

Fig 5.2: Agility in Financial IT of Optimizing Infrastructure-as-a-Service (IaaS)

Although IS agility is often defined in terms of the ability to quickly change IS functionality, research primarily measures IS agility by the time taken to change IS functionality. There are three categories of IT changes that are more closely aligned to IS agility: changes to make system performance improvements, changes to make system enhancements, and changes to make system corrective actions.

5.3.2. Benefits of Agile Infrastructure

Organizations can realize specific benefits by providing IaaS solutions that are agile. First, agile infrastructure allows for dynamic adjustment of owned capacities based on business requirements. Old technology may utilize inflexible capabilities that need to be carefully planned and obtained long in advance of predictable workloads. Older

solutions accommodate workloads over a long period without infrastructure adjustments. If workloads exceed planned capacity or slump below anticipated needs, low utilization of some resources or workload underperformance still costs business units money as work gets queued waiting for resource availability. Older systems often require very costly infrastructure purchases.

Second, agile infrastructure allows workloads to be migrated to different machines that possess the guidelines to handle workload needs, when needed. Providing simple guidelines for workloads to follow to migrate between different machines utilizing new technologies can eliminate major job performance issues and avoid additional cost impacts on business units. Such workload migrations may occur based on periodic schedules showing predictable changes in demand for different types of workloads, or in automatic fashion when needed monitored by specialized management systems. Companies that provide IaaS systems with little or no associated scheduling may still be utilizing older technologies that would be seen as too inflexible by today's advanced IT professionals.

Third, agile infrastructure allows loads on machines, and infrastructure capabilities themselves, smartly to adjust to the demands posed on the infrastructure by the workloads that flow across the infrastructure. Dynamic models of workload loading, for each machine or component of capability, can constantly oversee the differing resource needs of workloads, and their impacts on resource or capability performance. These dynamic forecasting models are able to predict what enhancements will be necessary within what timeframe windows, and communicate such needs to intelligent hosts or switches that control the resources needed, and the dynamically evolving workloads.

5.4. Scalability in Financial IT Systems

The creation of service-oriented architectures has enabled the financial services industry to work towards overcoming the criticisms that it has faced from regulators and customers alike regarding lingering technology constraints upon core services. These constraints, where they exist, are a product of decades of investment in distributed data center architectures and the traditional deployment of transactional systems that consist of networked servers running proprietary hardware and software with limited deployment flexibility. Such limitations mean that many financial IT systems cannot accommodate the required levels of System Non-Functional Properties, such as scalability.

In the Massively Parallel Processing Hybrid Transactional Analytical Processing layer of IaaS solutions, the most obvious scaling mechanism is an increase in the number of commodity blades to form a single large Massively Parallel Processing cluster. For on-

line transaction processing, replication to provide an additional level of redundancy will provide clustering at little added cost. A permanent solution in a growing market will demand greater levels of both transactional and operational processing per node. The main driver for this is the capability of the commodity blades which, despite the ever-increasing levels of vertical scaling apparent in the microprocessor families, has shown little sign of reduction in the on-going transition to chip multiprocessors which, in turn, are showing little sign of technological maturity. Hybrid transaction analytical processing architecture layering with a loosely coupled structure in the presentation tier will seek to lower costs as well as increase fault tolerance. Scalability in HTAP systems is not without its challenges as it may require more complex partitioning layers and expose data and users more readily to the implicit sharing of an Elastic Pool.

5.4.1. Types of Scalability

Scalability is a property of a system that defines its ability to increase system performance, including throughput, by providing funds to expand the system's capacity such as the addition of hardware resources in the network, servers, and storage. In the different branches of systems research, different definitions of scalability are used. The computer performance community uses a "strong scalability" and "weak scalability" definition to engage much of parallel computing systems that are inspired by limits on parallelism. The network community primarily concerns the limits of central shared components that engage definitions: "sublinear (or superlinear) and superlinear scalability". In essence, scalability is either the property of specific software or a class of computer utilization or a property of a system the economy of a quantity of employing computer software into its utilization and in the performance of applications. There are several aspects by which the definition of scalability may differ: What kind of system is there being a definition, what kind of performance, what kind of software or system, the use at hand, and doing what?

Scalability becomes a central property when the growth of demand increases faster than the reduction of cost of provisioning resources and can become economically and practically undesirable, requiring considerations to determine whether engaging in scaling relations are met when developing business systems and neural networks and operating them to establish solutions. Cloud computing arises as one in a class of approaches for resolving the issues behind scaling relations through exploiting the scalability of resources, with little regard for the efficiencies of speed or other specific conditions that would determine. The types of scalable, hierarchical infrastructure resources particularly apply for cloud services with live provisioning capabilities, specifically computing power in the form of servers and storage.

5.4.2. Challenges in Scaling Financial Applications

There are a plethora of applications which characterize the financial systems landscape. The main concentration of issues for financial applications resides with the more systems who specialize in the settlement, clearing, or exchange of assets between entities, as these are usually critical due to the centralized role played. The different characteristics of scalability stem from the specialization of the services offered. Hence, market makers implementing quoting systems at the critical speed to support transactions for high frequency traders on low latency markets have to deal with different issues than those opened by an asset management application. Each type of application also has its own characteristics of input, output, and expected performance. They include the various type of bids for equity or commodity achieved over the various time intervals, the periodic space and time clusters achieved, as well as the loops trading over the order-depth book.

Cross site replicas of an application and data are also used to counterbalance the reason for their periodic behavior. They allow re-routing to other sites in case of local failure or excess demand. However, transparency is usually partial via rerouting maps, leading to local caches of the configuration map. This solution which is considered a good practice finds its limits during local loss of service.

5.5. Resilience in Financial IT Systems

Understanding Resilience Resilience ensures continuous business operations according to business requirements despite changing environments and extreme events, e.g., system failures, software bugs, infrastructure issues, communication channel errors, and operational mistakes. In the context of IT system design and operations, resilience can be understood as the survival under anticipated normal and abnormal conditions. The design, implementation, and operation of individual financial IT systems contribute to these attributes of sector resilience. Financial IT systems should neither create the potential for error leading to problems in other systems nor amplify propagating shockwaves. When the internal level of resilience is exceeded, errors should be detected as quickly as possible, bridging of gaps executed, and recovery actions performed. For IT systems critical to business operations and providing external resilience, it is essential to avoid unplanned downtimes. In that case, errors should be detected, limited, and contained independently of the error cause, and recovery should happen online without human intervention.

5.5.2. Strategies for Enhancing Resilience Business and IT system design should aim for a high degree of natural resilience. A fundamental principle to enhance resilience is to not depend on any IT system component. Service clustering, load balancing, and automated failover processes allow diverting service requests from a failing component

to a fallback copy. However, it should not be forgotten that the quality of the external service provided by the system as a whole is only as good as that of its components supporting the respective function.

Fig : Achieve Agility, Scalability, and Resilience in Financial IT Systems

5.5.1. Understanding Resilience

Resilience generally refers to the ability of a system to absorb shocks without major degradation or collapse, and it is an old idea. The last twenty or so years have witnessed increased interest in the idea of resilient systems. The terrorist attacks, the hurricanes, and the cyber attacks stimulated interest in research on different aspects of resiliency, leading to a gradual expansion of its use and focus in the literature. The upper bound on a system's ability to continue "normal operations" is often known as the "intended behavior," while loss of control or dramatic degradation is often referred to as the "loss of system" ("system" being the collection of components and their relationships). In the literature, there is usually an explicit or an implicit time horizon. Moreover, loss of system adverse consequences can be serious, quarterly financial reports can be affected, national security can be compromised, and physical suffering can take place.

Resilience is related to reliability and is most often confused with robustness, but the three concepts should be treated separately. Resilience is a temporal measure, related to survivability – how long the system can keep loss of intended operations effect under control. Generally, if one looks at performance measures, reliability is usually the probability of "no failure," robustness is usually the worst case (minimum) performance during loss of intended operations (especially over all possible uncertainties), and resilience is the expected performance during loss of intended operations (the average over all possible uncertainties). For financial IT systems, loss of intended performance can have a serious negative effect.

5.5.2. Strategies for Enhancing Resilience

Resilience is seen as a core property of financial IT systems. Therefore, IT organizations seek to increase the resilience of their infrastructures that host the infrastructure components. An example for a resilience-enhancing mechanism is the use of several redundant nodes in the cloud, represented by different physical machines in different geographic dispatching centers in different parts of the world. Such redundancies are costly to the service provider to install and maintain. In addition to that, adding redundancy is not the only risk management strategy cloud consumers can apply, nor do they even have to resort to that. One issue is that increasing resilience primarily combats the effects of normal accidents, more robust infrastructures are very costly, and the market pricing of cloud services is based on the assumption that cloud infrastructures undergo normal accidents. It is also debatable whether the mere risk-smoothing effect of redundancy is sufficient.

There are also alternatives to investing in additional infrastructure capacity. Cloud customers are able to compose systems in such a way that damage from rare, but consequential accidents be limited. For this, there are several strategies that depend on the main dimensions of IT system structure. First, modern Cloud services are usable as infrastructure horizontal layers, providing services from the infrastructure layer of the OSI model up to the PaaS level plus security. Hence, the network of virtual machines interacting for a specific business process can be established on demand by the cloud customer. This gives room for avoiding single points of failure which might not be avoidable if the underlying physical infrastructure is operated in-house. The customer has the choice of the number and location of nodes to take the risk of being blocked by rare accidents.

5.6. Conclusion

This chapter explored how the Infrastructure-as-a-Service (IaaS) paradigm can be optimized and leveraged within the scope of the Financial IT systems. Followingly, the specific characteristics of infrastructure-related management costs define the way how the concept of Financial IT governance must be specifically configured and ultimately how it impacts the optimization of the IaaS service utilization and requisitioning strategies. The key finding is that for any organization, monetary costs are paramount regarding the optimal utilization of IaaS solutions defined in this chapter. The way how costs evolve, affecting optimal decisions, ultimately determine the direct behavior of Financial IT departments and any client organization. In this regard, a cost focus is necessary, but not sufficient on its own. Specific dynamics of cost development define the behavior patterns of the Financial IT department and client departments. The demand for investment in self-developed Financial IT solutions would normally be zero if the services offered were free.

Thus, we find numerous characteristics in our analysis: (a) Cost-Free-Service Collusion: Due to the growing availability, the IaaS requirement costs must pose a minimal space for the service collusion; (b) Demand for Internal Developments: A pronounced share is demanding to develop solutions internally; (c) IT Governance Focus on Investment: In our population, Investment in own Development is in Focus; Be user of Services offered: A necessary requirement to optimize such a service offering is that all departments must offer services to use. It de facto leads to de facto increasing costs of accounting departments.

As Financial IT-related solutions increase, and the complexity of cooperation increases, the services offered can evolve to a strategic competitive advantage. Given the strategic relevance, a much clearer pattern of imputing costs must be defined. The emerging probability of mainly external vendors creates growing criticality. A growing criticality ultimately means that these services must be continually monitored for developing implied market solutions by vendor companies. Do they exist already, and at which point in time could they be preferable compared to the internal developed controls?

5.6.1. Emerging Trends

The growing offering of IaaS capabilities also has positive effects. Potentially, Federal Reserve Banks may be able to deliver better services to the customers and redistribute total cost savings to its smaller customers in the form of temporarily lowered IaaS prices. Meanwhile, competition drives providers to offer an IaaS cloud that is more focused on financial transactions, increasing the service options. Security becomes a growing commodity, mostly because financial institutions are tightly regulated, a factor that leads

to a result where active cloud management platforms exist to mitigate the risk of mitigating security services. Banking technology, in its function as the electronic layer and meaning the communication, the security, and the caching that transfers the values between banks and citizens, becomes commodity. In a hosted snowball effect, IT, as the electronic layer, becomes a commodity basis for any bigger financial institution to deliver bank products transactions that sit on top of them. Wages for IT services providers collapse and the amounts of these services go on to price the services as a percentage of the value transactions flow. The central bank policy response may be to offer the large banks an option: create or take over a larger infrastructure layer or set up a "governmental" institution to do this.

In a virtual cloud banking world, it may become possible for a new bank to decide to build its product and service offering in such a way as to not have an internal IT layer, either in terms of some of the major services or the entire basis of the transaction price set for bank companies. In such a case, prices collapse. Banks either try to keep selling higher margin products and services that do not focus on the transaction layers or banks digitize the entire range of product offerings and focus on the transaction layers, with a price of delivery that is more of a straight value transfer price model. Prices for higher margin products offered by banks then have to collapse some, because potential customers feel that it is unfair for the banks to keep profits so high, while the repetitive costs of bank transaction processing now are at much lower levels.

References

Chowdhary, M. A. M. (2025). Financial Network Infrastructure: Scalability, Security and Optimization.

Subramanyam, S. V. (2025). Cloud-based enterprise systems: Bridging scalability and security in healthcare and finance. *IJSAT-International Journal on Science and Technology*, *16*(1).

Oye, E., & Matthew, A. (2024). AI-Driven Cloud Evolution: Transforming Infrastructure for Future-Ready Solutions.

Arogundade, O. R., & Palla, K. (2023). Virtualization revolution: Transforming cloud computing with scalability and agility.

Samha, A. K. (2024). Strategies for efficient resource management in federated cloud environments supporting Infrastructure as a Service (IaaS). *Journal of Engineering Research*, *12*(2), 101-114.

Samha, A. K. (2024). Strategies for efficient resource management in federated cloud environments supporting Infrastructure as a Service (IaaS). *Journal of Engineering Research*, *12*(2), 101-114.