International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

SUBSTANTIAL BACKING FROM BUSINESSES FOR IAAS CONTROLLING CAPACITY USING MODELS OF SYSTEM DYNAMICS

T. BHARGAVI, Assistant Professor, bhargavitellikapalli@gmail.com, Rishi UBR Women's College, Kukatpally, Hyderabad 500085

Abstract

Providing Infrastructure as a Service (IaaS) is becoming increasingly competitive due to its growing world market. IaaS requires significant physical resources, e.g. network connections, bandwidth, load balancers, and servers. IaaS providers of any size and reach, but especially private and public regional ones, have to manage the capacity of their resources property to reduce costs and meet other business goals. Comprehensive capacity management by JaaS providers involves objective and subjective analyses and it is thus, challenging. Complexity arises, for instance, as one tries to factor management behavior in analyses of business impact and Infrastructure performance to produce indicators in support of interventions related to service capacity. IaaS managers frequently rely on their own expertise, experience, and knowledge to identify whether an apparent capacity deficiency may be due to an infrastructure problem or to an unexpected demand burst. This paper presents a business-driven simulation model based on System Dynamics (SD), Balanced Scorecard (BSC) and Analytic Hierarchy Process (AHP) concepts to support capacity management decisions by IaaS providers in real-life operating scenarios. A case study in a real- world, public regional IaaS provider illustrates the model's support to non-trivial decisions that are aligned with managers' experiences and business expectations.

1. INTRODUCTION

CLOUD computing uses a provider's infrastructure (i.e., hardware and software resources) for service delivery over a network [1]. One may thus talk of Infrastructure as a Service - IaaS. (Henceforth, "IaaS" and "cloud computing" are used interchangeably). With IaaS, customers are relieved from buying and running their own infrastructure, allowing them to concentrate on mission-critical aspects of their business. The cloud services market tends to grow. Indeed, as of 2017, 34.6% of all IT services were already cloud-based, and the worldwide IaaS market is expected to exceed US$ 60 billion by 2024 [2].

For continued business growth, an IaaS provider has to make the infrastructure appear unlimited and be acquired in any quantity and at any time - i.e.. the provider has to plan and manage laaS capacity to satisfy customers and to leverage its own business. For that, the provider may be assisted by the guidelines and recommended practices of IT Service Management (ITSM) - such as those in the IT Infrastructure Library (ITIL) [1].

ITSM usually prescribes what needs to be done; it rarely recommends how- for this depends on local preferences and availability of tools. This is intentional: implementation and tool selection are better taken care of by in loco IT managers who will consider their technical preferences, business policies and market conditions. On the other hand, information on the "how" could ease managers deal with IaaS capacity management challenges. This paper proposes a model to be embedded in a tool to support decision-making in IaaS capacity management. As such, the paper complements related work.

The challenges IaaS managers face to manage capacity are of a technical and business nature. Most related work address the latter; despite caveats that IaaS vendors focus more attention on the former [3], the literature on business-driven IaaS capacity management is scant. Business challenges include: Aligning business strategy with IT-related actions. Adjusting IT resources according to the dynamics of the organization's demands.

Identifying problems related to cloud services capabilities that require changes or improvement actions to support the decision-making process.

The model in this paper addresses IaaS business challenges such as the ones above by considering performance related actions (e.g. those based on Performance Indicators - KPIs and capacity indicators - KCIs) and their effects on capacity design and management of IaaS technical resources - e.g., virtual machine (VM) instances, storage, bandwidth, or staff. Addressing business challenges is likely to be more taxing to corporations and institutions with security-motivated internal (private) IaaS provision or to regional public IaaS vendors (in contrast to world players like Amazon, Microsoft, Alibaba, Google or IBM). Our model could bring more value to them. Such is the case of the public IaaS provider in the case study considered here. (Our case study provider operates regionally from a major city in northeastern Brazil. Henceforth, it will be referenced to as the Alpha provider to avoid exposure of its sensitive operating and strategic details.)

The case study will serve to evaluate the following hypotheses: Preference: Managers prefer the proposed model over the capacity management process they currently use.

•Effectiveness: By applying the model to an laaS scenario, managers can identify elements to support decision-making.

•Utility: Managers consider the model useful.

• Accuracy: Managers consider that the simulation results are sufficiently accurate to support IaaS capacity management.

The model couples system dynamics [4] with the analytic hierarchy process (AHP) method [5] to support multi-criteria decision-making (MCDM) to evaluate the hypotheses. Capacity management is modeled from the IaaS provider's perspective and captures the relationship among process actors in business demands, the role of ITSM service level agreements (SLA), costs, business benefits and return on investment.

The remainder of the paper briefly discusses related work (in Section II); highlights major aspects of the research methodology (III): offers main details of the complete model

(IV): presents the case study (V); discusses verification and validation results (VI); and, finally, brings conclusions with suggestions for future work (VII).

II. RELATED WORK

Technical capacity management is widely discussed in the cloud computing literature. It is desirable to have a system that automatically adjusts the resources to the workload handled by the application. All this with minimum human intervention or, even better, without it-i.e., an auto scaling system. We assume technical capacity management to be autonomic as we model business aspects of the cloud provider's behavior.

Offering cloud services to satisfy different user requirements while pursuing business objectives-e.g. keeping costs down -is challenging, due to non-trivial trade-offs associated with service quality and infrastructure costs. IaaS users or providers can choose only two constraints for elasticity, capacity and performance when making decisions on service options, sacrificing the third. For example, providers that prioritize high capacity utilization to reduce costs may have to offer services with low-quality performance [6]. Modelling such trade-offs is indeed complicated. Authors of [7], while discussing optimization of resource scaling in cloud deployments, note they use a proxy for costs (i.e., amount of resources) since per-VM prices depend on factors such as profit margins and market conditions. Our model explores said trade- offs by means of simulation, avoiding the mathematical optimization complexity in [7]. ITSM metrics may be used in validating, justifying.

directing, and intervening actions to align IaaS delivered services with providers' business needs. Business driven approaches to IT services management (BDIM) are presented in [5], [8], [9], [10], [11], [12], [13]. [14], [15], [16], [17], [18]. [19] and [20]. Except for [16], these works do not target (IaaS) capacity management explicitly. The work [16] proposes an approach to e commerce infrastructure capacity design that minimizes the sum of business loss caused by infrastructure malfunctions and infrastructure cost. Conventional, technical in nature capacity design approaches usually try to minimize infrastructure cost only. The mathematical complexity in [16] limits the business perspective to very few variables - e.g... financial loss, and may not easily capture dynamic behavior. The model here, being solved by SD simulation, is flexible enough to consider capacity management and business dynamics as well as multiple business outcome indicators. The authors of the work [21] propose a probabilistic method to measure the business value of IT services. They argue that service capacity is one of the most important components in IT services quality monitoring [21]. Thus, LaaS providers need effective processes for capacity management. This paper addresses IaaS capacity management from encompassing and dynamic business perspective a more

A process asset library (PAL), based on ITIL recommendations and focused on IT Service Capacity Management, is proposed in [22] to facilitate the management of

process assets and to provide a base for business process improvement. Our model differs from PAL as it simulates the capacity management process in IaaS providers to obtain performance indicators and action suggestions to support specific decisions related to capacity management.

Banner and Bellamy [23] present a range of IT capacity management structures that show differences for cloud computing. Differences arise because capacity managers have also to busy themselves with financial and contractual aspects of capacity (because physical aspects are becoming less of a concern in the cloud - for users, that is). Besides including financial aspects in capacity management simulations, our approach also adopts an integrated view of capacity management dynamics that covers demands, benefits, and needs management and alignment with the business.

In addition to the type (physical/logical) of resource being considered, IaaS managers need to work on the allocation, brokering, provisioning, mapping, adaptation, and estimation of required resources in order to assist customers to obtain higher satisfaction levels and for the provider to meet business objectives [3]. Our proposal caters to such need by using the behavior and outputs (Key Capacity Indicators KCIS) generated by the model as feedback for capacity management process improvement.

The authors of [24] propose dynamically provisioning IaaS instances based on the Central Limit Theorem. The minimum number of active instances for the next tasks is set according to demanded quality of service (QoS) - i.e., a low probability of overload. The performed experiments involved simulations based on real overloads. The model proposed here also considers Alpha's real business scenarios but in a wider scope: various actors and resources (technical and managerial) of the IaaS capacity management process are simulated.

In [25], a system based on symbiotic simulation was implemented to support the automated management of distributed virtualized IaaS datacenters. Differently from [25]. our model is non-intrusive.

The research in [26] simulates cloud computing environments to study their survivability and to suggest the reactive-hybrid-to-proactive transition to setting escalation goals for information access control. The model proposed here has a different aim: offer a System Dynamics (SD) tool for managers to obtain performance indicators for (possibly non-autonomic) continuous improvement of services, as well as to subsidize their decision making in the capacity management process to bring about business benefits for the laaS provider.

Among the studies that apply SD to management of IT services in general. Bezerra et al. [18, 27, 28]. present a SD model to support the decision-making process in IT services outsourcing. In addition, in [29], cloud computing performance evaluation predicts and quantifies the cost-benefit of a strategy portfolio and the corresponding quality of service (QoS). The work in [30] proposed a SD model to help manage the web services capacity to ensure fulfillment of the associated SLAs. The focus was on the web services capacity management policies,

service performance and SLA violation penalties. Diverse IaaS scenarios were used in [31] to review research on managing the overhead of virtual machines (VMs). The SD model proposed here adds to these works since it applies SD to yet another scenario: that of IaaS business-driven capacity management.

In general, the research in this paper may be seen to differ from other related work in the state of the art of IT service management and cloud computing literature as it evaluates laaS capacity management behavior from a wider business perspective. The paper reports on research that considers a (dynamic) system behavior where the technical capacity of the provider is increased or decreased autonomously, depending on the demand for VMs, storage, processing, memory, and network traffic. According to the process flow that involves IaaS provider business capacity management, behavioral factors related to human resource (HR) management, process performance, service value addition, and business objectives are analyzed. Although there are works in the literature that address simulations related to some aspects of capacity management of cloud services, we found no models that evaluate all the behavioral aspects of capacity management in an IaaS provider we address in the paper. As such, the paper complements and/or expands the work cited in this section.

III.RESEARCH METHODOLOGY

The research methodology we adopt involves a triangulation of literature review, observation, and the execution of a SD simulator for the case study. For validation of simulation results, we conducted exploratory interviews with IaaS managers, project managers, and infrastructure analysts.

The proposed model was embedded in a software tool [32] that underwent several cycles of refinement until a valid version was reached following these methodological macro steps:

a) Abstraction of essential capacity configuration information through: i) review of the literature on JaaS capacity planning and management; ii) observation of IaaS environments; and, iii) interviews with Alpha managers.

b) Model proposal, instantiation, and implementation; additional meetings with Alpha managers for gathering/ estimating model's variables and metrics.

c) Verification of model implementation and simulation of different scenarios with the model (embedded in the software tool); and,

d) Presentation and discussion of results from simulation runs with Alpha managers for tests and refinement of model (sensitivity analysis) and validation of results.

The model outputs (KPIs, KCIS and suggested decisions concerning IaaS Capacity management) were analyzed by managers. During model implementation and testing, sensitivity analyses were conducted to determine the laaS provider's variables that most influence results and which changes should be made to the configuration to improve laaS capacity management performance.

The simulator uses tolerance levels (in %) for design and capacity metrics, as defined

by Alpha managers. When a capacity metric nears its tolerance value, managers are signaled about the corresponding business impact.

Verification of adherence of the SD simulation solution to the model's aims involved testing the model's internal correctness, i.e., checking whether the model was constructed to correctly mimic the functioning of the modelled system. Tests of structure, behavior, and learning [33, 34, 35] were also carried out. Validation, on the other hand, aimed to test the external correctness of the model, i.e. whether it is appropriate to address the target problem. According to [33], the validation of a model can be defined as "establishing confidence in the usefulness of the model with respect to its purpose". We investigated the model's structure and relationships before proceeding with conceptual model validation over several scenarios. Simulation scenarios were configured through the model input variables, which were collected from estimates by Alpha managers who also defined business tolerance levels. Simulation results were then assessed by Alpha managers who were interviewed on their confidence in accepting them.

IV. THE PROPOSED MODEL

A. Scope, variables and architectural components

The proposed model allows investigating JaaS capacity management complex dynamic behavior in terms of stocks (the accumulation of things), flows (the motion of things) and feedback links at any level of aggregation. Model simulation scenarios relate to the dynamics of actors' behaviors. The IaaS provider's behavior is related to the demand it receives, and subsequently, to valued services it delivers to its clients. The modelled processes and related activities are made to align with the provider's business needs. A Balanced Scorecard (BSC) map [36] is used as a base for modeling the IaaS provider scenarios, covering financial, customer, learning and growth and internal processes perspectives. By using the model, managers can find the conditions under which the capacity management process should evolve to contribute to business goals. The model also recommends improvement actions to better support decision-making in capacity management.

The model's scope encompasses:

1. Analyzing IaaS provider dynamics, with the purpose of supporting capacity management, service delivery and decision-making.
2. Enabling understanding of IaaS services as they are provided towards achieving specific IT objectives.
3. Accounting for cause and effect relations, amongst IT goals and business objectives.

Four interactive activities, inherent to capacity management [38] and which can be performed in a PDCA (Plan-Do-Check- Act) cycle [9, 10], were considered: monitoring, analysis, tuning and capacity implementation. The proposed model maps capacity management activities into the PDCA cycle as follows: plan includes analysis and modeling; execution (do) consists of monitoring, while check consists of tuning and act Includes implementing changes that may be required for improvement. In a continuous cycle, new

analyses are performed, and the process repeats itself.

Fig. 1 shows a general view of the IaaS provider activities that are supported by the model and its main architectural components with their respective interrelationships.

Variables are divided into four types: input, calibration, mediator, and output. Inputs (1) represent resources, expected performance, and the characteristics of the demand. Input values are provided by managers or collected from cloud monitoring tools. Calibration variables (C) serve to customize the (generic) simulation model so that it matches the IaaS organization's operating characteristics, management policies, and guidelines. Mediating variables (M) represent information endogenous to the system, obtained from the inputs, from calibration or, in situations involving feedback links, from outputs. An example of a mediating variable in the model is the resource performance, a ratio of consumption of IaaS resources in relation to demand for capacity (in percentage). The output variables (O) are values resulting from the cause and effect relationships between the input, calibration and mediator variables over time. (It should be noted that, depending on the scenario at hand, the required capacity percentages of the IaaS environment - modelled as a set of KPIs, KCIs and objectives - serve as inputs and sometimes, as mediators). Table I exemplifies the types of variables for demand management.

The model simulates the behaviors and interactions amongst the four processes of demand management, capacity management, business management and benefits management. Fig. 2 offers an overall view of the model's I, C and O variables and its SD simulation module for the interplay of these processes.

B. Model diagrams and implementation The model was implemented using the Stella visual programming language for SD [32], according to a hierarchical design (Business objectives - IT objectives Key Capacity Indicators - Key Performance Indicators). Highlights of the model's design are shown in Fig. 3.

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

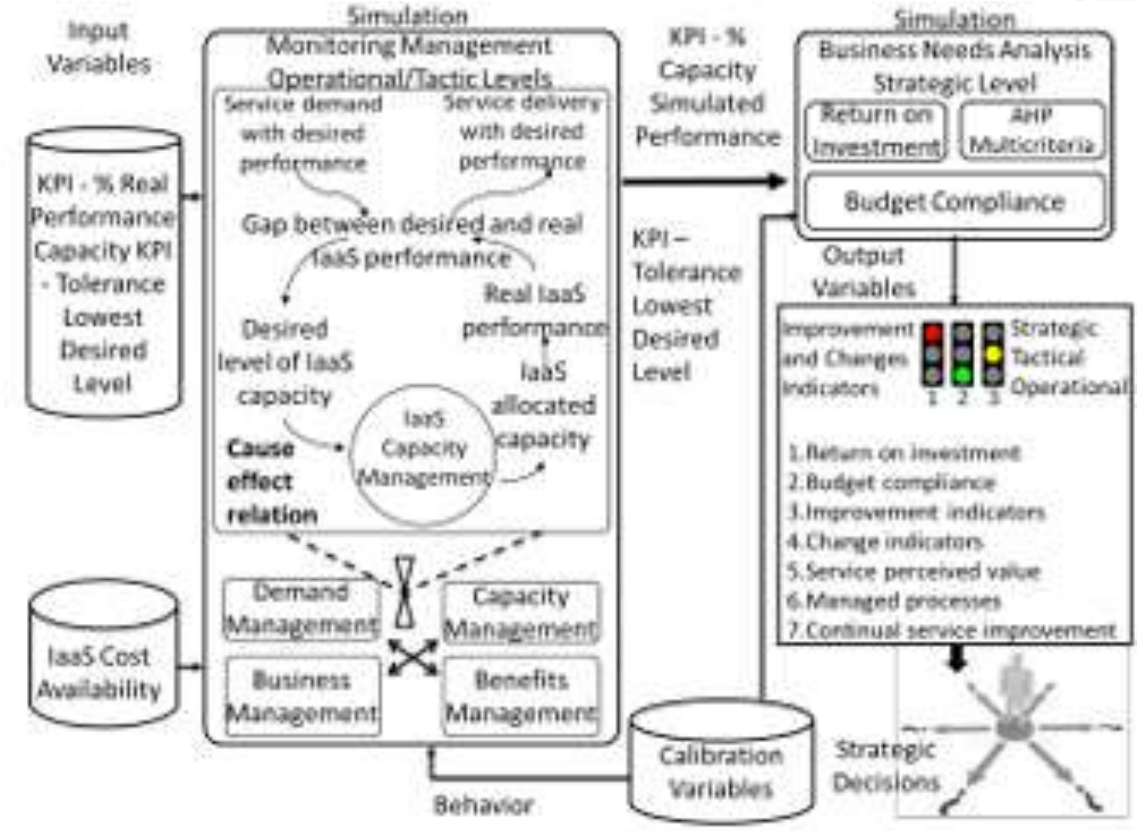

Fig.1. Integrated view and main architectural components of the model

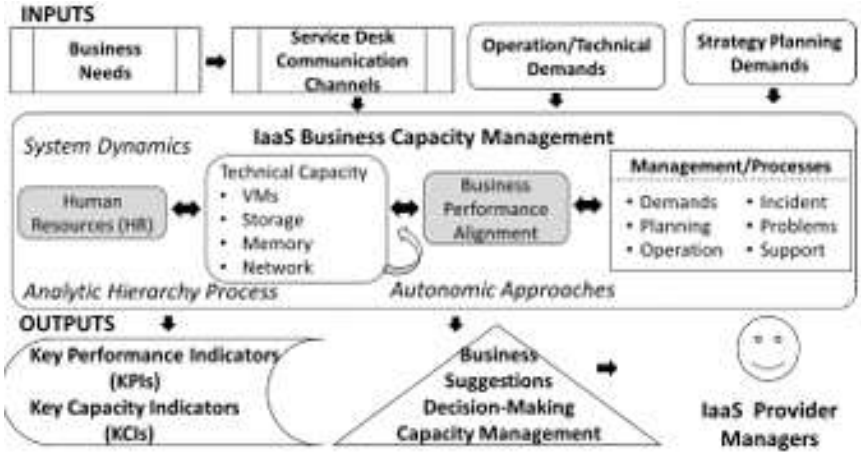

Fig. 3. Management behavior dynamics in the proposed model

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

TABLE I
VARIABLES IN DEMAND MANAGEMENT

Name (Type)	Description (Unit)
Bandwidth / link consumption (I)	Capacity that the link offers for data traffic (Mbp
Memory Consumption (C)	Amount of memory consumption (MB)
Storage Consumption (C)	Amount of storage consumption (MB)
Processing Consumption (C)	Amount of processing consumption (Mflops)
Resource performance (M)	Consumption ratio of *IaaS* resources in relation t demand for capacity (%)
Availability (O)	Percentage of availability of services (%)
API KPI (I)	Availability Performance Indicator (%)
CCP *KPI* (I)	Contingency / Continuity Plan Indicator (%)
DS *KPI* (I)	Storage Performance Indicator (%)
DM *KPI* (I)	Memory Performance Indicator (%)
DP *KPI* (I)	Processing Performance Indicator (%)
DV *KPI* (I)	Virtualization Performance Indicator (%)
LPI *KPI* (I)	Link Performance Indicator (%)
MPCI *KCI* (O)	Machine Physical Capacity Indicator (%)
MTTR (I)	Mean Time To Repairs (%)
MTBF (I)	Mean Time Between Failures (%)
No Interruptions (O)	Relationship between Mean Time Between Failu and Mean Time to Repair (%)
Capacity Feature (I)	Amount of capacity for the *IaaS* resource (MB)

Demands trigger simulation. Demand management is represented by parallel flows, where the resource flow is fed by its capacity stock and the demand for services flow is fed by the demand management stock. The output variables of the demand management.

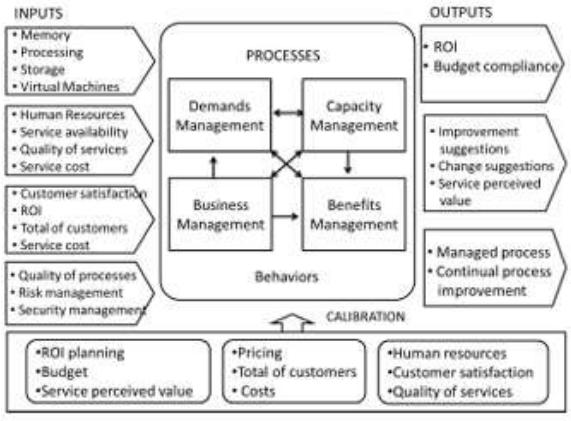

Fig. 2. Overview of model variables and simulated processes (based on [27])

t process are used to analyze and verify the achievement of strategic objectives, and to calculate and to monitor the evolution of monitored, color coded indicators (green when the output value complies with agreed SLAs and/or tolerance % levels; yellow signals the need for attention because the output is nearing a critical level; and, red informs that the indicator is compromised and an intervention is necessary).

An example may clarify some details of the dynamics in Fig. 4. Suppose that a large customer of an IaaS provider promotes a large sales campaign, which will generate an increase in demand at certain times, requiring elasticity of bandwidth, storage, processing capacity and memory. In addition to automatically managing technical resources, the provider's managers still need to consider aspects related to staff, processes and service management, so that the capacity of the service is in accordance with the needs of the business. This scenario can change dynamically in a short period, due to competition, with an abrupt decrease in the use of resources by the large customer. Sometimes, the capacity management process may be affected by variables identified by the service desk, service monitoring or in staff management activities. In the model, a monitored metric will affect the business if its capacity tolerance percentage value is greater or lower than the estimated tolerance level (depending on business guidelines).

IaaS provider managers may consider varying levels of

As for employee performance, the model considers the annual turnover, the level of

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

knowledge and motivation of the team, as well as the degree of professional experience of each team member.

Planning of capabilities consists of correctly dimensioning the extent that services meet demands. The proposed model seeks to support decisions on strategy to provide the necessary capabilities which are to be correctly aligned with the planned benefits Le.. intended business results. For that, the model mimics a collection of behaviors" in capacity management together with their interactions and implications (results). The main underlying mechanisms that serve as base for the causality diagrams are the actions needed to reach an objective, as Fig. 4 illustrates, but also:

"Target-driven archetype, implemented as balancing links: Dynamics of allocation of IT resources (considering capacity and elasticity consumption) to the execution of services, as in the diagram of Fig. 5: and, Qualitative criteria, as shown in Fig. 6, for there may be influence of the external environment on the behavior of the model (qualitative factors).

The stock flow diagrams in Figs. 5, 6 and 9 were drawn using Stella facilities [32] and reflect the SD model's actual coding.

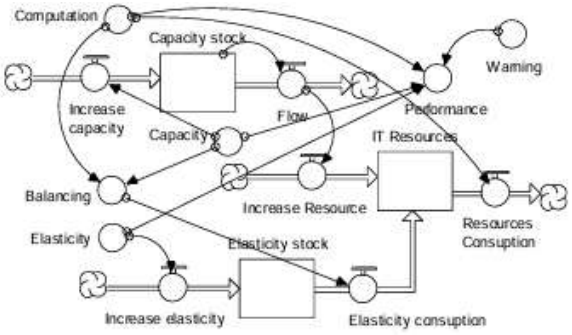

Fig. 5. Resources allocation implementation in the proposed model

Benefits management behavior in the model addresses and implements the benefits to the

The demand management behavior in the model implements the demand for services over time, and the scheduling of the necessary capabilities. These demands are treated as services that demand resources. The model represents services, capabilities of the IaaS provider, and the flows that relate these two entities. One can simulate different demand arrival patterns (continuously or in bursts, for instance) and associate them to different IT functions. The demands generation is done by providing consumption attributes (as a percentage between 0 and 100%) for storage, memory, processing, and virtual machines.

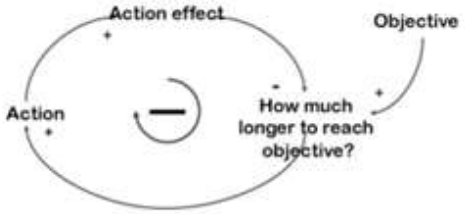

Fig. 4. Balancing link "objective reaching"

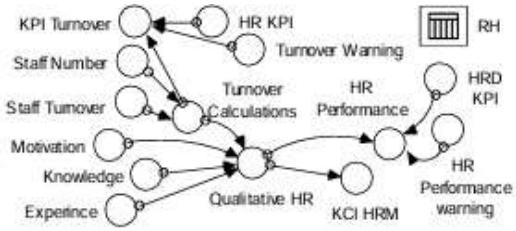

Fig. 6. Some qualitative criteria of the proposed model

IaaS provider, such as consolidating the integration of information from demand

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

management, capacity management, strategic business management and infrastructure resources. The benefits tracked in the model are related to reducing costs, achieving customer satisfaction through the quality and agility of services, return on investment (RO), and profitability. The model considers budget conformity and costs such as those of utilities, administration, staff and marketing. The model also considers the IaaS provider pricing and revenue strategies. Customer satisfaction is treated as a relation between the number of service interruptions and the satisfaction indicator. To validate obtained benefits, the model estimates a relationship (service-aggregated value) between ROI and customer satisfaction with delivered services. The higher the obtained benefits percentage, the greater its impact on customer satisfaction and business objectives achievement. The higher the aggregated value percentage, the greater the impact on alignment with business strategy and the better will the delivered services be.

Fig. 7 indicates that meeting the level of capabilities needed for provisioning laaS services requires both knowledge of and experience in demand management and its interactions with capacity management. If the required capacity to satisfy demand for IaaS services is not met by allocating resources, capacity management will have to bridge the gap.

In turn, capacity management will be affected by the cost of investment in IaaS resources. The demands in the model are dimensioned based on information for the consumption of IaaS resources (storage, memory, processing and virtual machines). The consumption of IaaS resources interacts with the demand management process. An increase in the capacity of a resource feeds the stock of capacity of that resource, through the resource flow process, updating the demands for the management stock.

Next, the connection with the capacity management process is established, where actions related to elasticity are taken. All capabilities of the IaaS provider resources are represented in the same capacity management flow, as shown in Fig. 8.

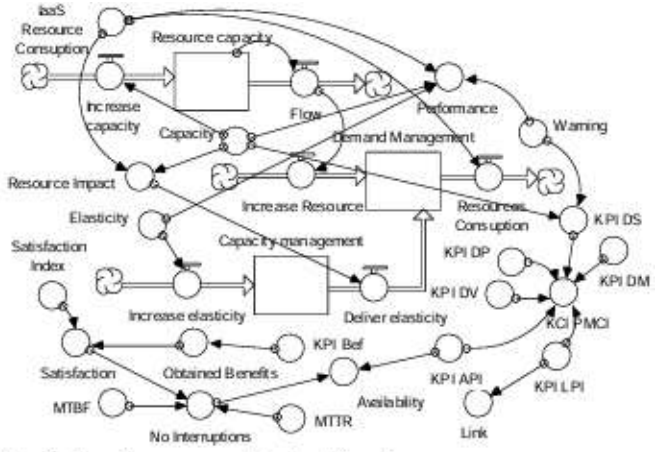

Fig. 8. Capacity management stock and flow diagram

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

Strategic business decisions are influenced by established benefits, which, in turn, are influenced by external factors. The benefits to be tracked in the model are related to cost reduction, obtaining customer satisfaction, through the quality and agility of the services provided, and the retention of organizational capacities. The business management process considers information from demand management and capacity management processes. The benefits management process addresses the benefits for the provider, in an integrated manner with the other processes.

C. Simulation runs

Usage (running) of the simulation model is triggered by a capacity demand to the service provider in a simulated scenario and happens in 5 major steps:

1) After examining logged data for input variables from the monitored IaaS environment and estimating quantitative and qualitative data for the calibration variables, managers must define values for the model's other variables. Managers can use as support to input estimates, results from service monitoring (please refer to [9]), properly projected to reflect the IaaS scenario, or any tool that can estimate percentages relative to IaaS indicators (inputs) for the proposed model. A business tolerance percentage level is to have been previously set for each indicator (Objective, KCI and KPI) by the managers.

2) Calibration variables' values may be adjusted to make the simulated behavior more closely reflect the target laaS provider's reality. Once satisfactorily calibrated, this step may be omitted. Different, additional calibrations may happen as part of running or designing experiments.

3) The simulator is then run for a given scenario to produce outputs to support decisions on IaaS capacity changes that will

Managers may feed weights (real numbers between 0 and 1) according to their perception of the importance that each of the ROI, Profitability and Customer Satisfaction variables should have in the BL: HR motivation, professional knowledge and experience; business processes and service improvement contribution to customers' satisfaction; and relative importance of memory. processor, storage and virtualization activity to infrastructure.

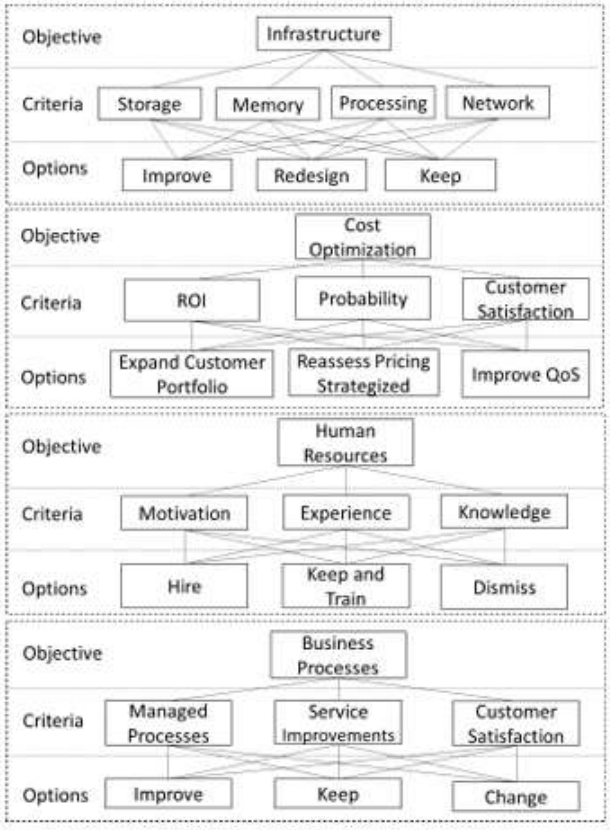

Fig. 9. Multi-criteria decisions using AHP

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

V. CASE STUDY

The model was evaluated in a case study at the Alpha IaaS provider. Alpha organizes its clients in two categories: 1) Corporate with companies, including those that classify as sole proprietorship; and, ii) Personal - comprised of individuals. When a client wants to change a service, s/he can choose options defined by the provider. Clients do not have the option to assemble required services, but s/he chooses the option that best suits its/her/his needs. Available options present different values for core, RAM capacity, disk space, traffic speed and bandwidth.

A. Acquisition of input values and calibration

Simulations of three business scenarios (numbered 1, 2 and 3 in Section "B. Results" to follow) were performed to study whether the model, with real-world input data and in a basic configuration of calibration variables, would be able to produce credible performance indicators. The indicators considered in the simulations were: costs and process optimization; investments in infrastructure and in Human Resources HRS: and perceived service value for the provider. As a preliminary step to the simulation runs, the model was calibrated with variables that, in the view of interviewed Alpha managers. reflected the then current operating environment of the provider. The interviewees estimated all calibration values. Alpha uses formally established control procedures, with the support of tools, for monitoring the consumption of internet, memory, processing, and storage resources. There were no formal records related to the knowledge and motivation of its IT staff, however. The customer satisfaction indicator in the model depends on the desired return on investment. Alpha keeps detailed accounting of revenue from personal and corporate plans.

When simulating the business scenarios, input values reported by Alpha's managers and by the support and monitoring systems used by its IT staff were applied. Because of the lack of historical records, some KPIs were valued based on the experience and expertise of the managers. The Cost Optimization KCI results from ROI assessment and customer satisfaction. Cost optimization is achieved when there is good KPI performance related to ROI and customer satisfaction. Customer satisfaction is achieved through customer satisfaction surveys when Alpha's customers inform how satisfied they are with IaaS services.

The HR KCI considers the professional experience of the staff, the level of employee knowledge, the level of employee motivation, the number of employees and staff turnover. The Service Perceived Value KCI considers non disruption of services and data quality. The service quality KCI indicates the quality of the service at a given moment. Budget compliance is estimated using the ratio between the amount that was planned for the expenses of the laaS provider and the expenses that were actually incurred.

Periodic monitoring dynamics for Alpha's services were considered in accordance with

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

ITIL's capacity management cycle. During the collection phase of the records and historical information (execution of the Dynamic Scorecard), the Alpha provider was generous to offer information from systematic and detailed monitoring records of its strategic objectives and management processes, which provided information that allowed following the steps of the Dynamic Scorecard plan. The simulations were based on real scenarios proposed by Alpha managers. The system dynamics tool that implemented the proposed model generated the dashboard information in this experimental research.

The results of the simulations were compared with Alpha's actual data and decisions its managers made (or would make). The proposed model, based on the results of the information generated post-simulation, presents guidelines to support decision-making. Said guidelines are influenced by the outputs of the KPIs, by the generated average of the KCIS and by the choices in the AHP method.

B. Results

1) Baseline scenario simulation

The first evaluated scenario was based on Alpha provider's baseline. The simulation objective was to analyze the IaaS capacity in face of its then current business demand. Business processes' situation and human resources' needs were also evaluated. Alpha's costs to maintain its IaaS infrastructure in operation were mainly due to electricity bills, marketing campaigns, staff payroll, and taxes. ROI was used to evaluate profitability of the business.

For this first scenario, managers expected the model to signal problems and recommend decision(s) they could make on related solutions. If a given indicator (KCI) did not adhere to the established standard, managers expected results to allow for elicitation of cause and effect relationships.

The input data set for the baseline simulation included the following provider variables: estimated capacity: estimated elasticity; desired KPIs (default) and defined KPIs, infrastructure costs; and, size, knowledge level and motivation of staff. Simulation results allow for visual evaluation of the graphical indicators for KPIs, KCIs and decision making support (recommendations). Graphical indicators are useful in validating behavior in SD modelling and some we use are displayed in Figs. 10 to 23 below to display our model's output.

Table II shows the KCIS generated by the model in the baseline scenario simulation in relation to tolerance thresholds, defined by Alpha managers based on their experience.

TABLE II
KCIs - FIRST (BASELINE) SCENARIO

KCI	Value	Tolerance threshold
Cost Optimization	93.80%	≥ 90%
Process Optimization	85.70%	≥ 80%
IaaS Infrastructure	67.50%	≥ 70%
Staff Investment	88.50%	≥ 90%
Services Perceived Value	80.80%	≥ 85%

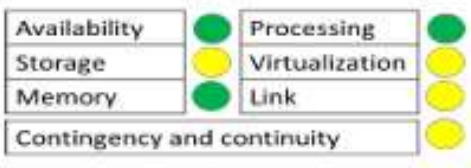

Fig. 10. Baseline scenario - Demand and capacity management indicators

International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

VI. VERIFICATION AND VALIDATION

Table VII shows results for verification tests of the model and its implementation. In this table, the entries for the first 4 lines followed a strategy of performing tests with the model and associated software; the remaining four lines used a strategy of literature review, observation and checking for face validity (39). Continued usage of the model may reveal unsuspected areas in which the model can be improved.

TABLE VII
MODEL VALIDATION TESTS

Computerized Model Verification	Assess whether model implementation is error-free
Dimensional consistency	Dimensions of variables are consistent and the units are correct
Syntax validation	Behavior-governing model estimates are free of syntax errors
Semantic validation	Behavior-governing model estimates are free of semantic errors
Conceptual model validation	Evaluate the model's structure
Structure confirmation	Model's estimates correspond to the actual process relationships
Parameter confirmation	Evaluating model's variables against process knowledge
Extreme conditions	Assessing model's behavior under extreme conditions
Sensitive behavior	Determining the process variables to which the model is highly sensitive

Evaluating and validating a complex simulation model, such as the one here, is a multiyear effort. Twelve Alpha managers have made actual usage of the model and then answered at questionnaire on their preferences and impressions about the model's effectiveness, utility and accuracy. Additionally, other nine Alpha IT managers were presented (but did not use) the proposed model and simulation results and answered the same questionnaire. Their impressions are summarized in Table VII From this Table, we may claim face validity [39] for the proposed model. Statistical inference [40] was used to test the preference, effectiveness, utility and accuracy hypotheses. A binomial statistical test with a 5% significance level was used to produce the results, as shown in Table VIII.

TABLE VIII
HYPOTHESES TO TEST FACE VALIDITY

Hypothesis	% who agree there is enough evidence to accept hypothesis
Preference	95
Effectiveness	90
Utility	100
Accuracy	90

Further comments on the model were obtained from additional meetings with Alpha managers and include:

"The model allows conscious decision-making and having a visual representation is what most helps In that regard."

• "The model shows the flow of information, where it comes in, and how it is processed. It allows us to analyze a capacity indicator's repercussion within the system, its influences and its impacts on the process."

"It is a novelty to be able to see the integration of the company's strategic planning with IT actions. The model shows this integration in practice and thus, we can see the impact of our decisions on the performance of services."

"The model allows us to make more sensible decisions towards desired goals. It draws attention to investment needs, customer base expansion."

These results indicated the usefulness of systems dynamics theory for modeling complex processes and management behaviors, such as in IaaS capacity management, to support decision-making, as recommended in [38] and [41]. On the other hand, the single (Alpha) case study and the

short periods of simulation imply deficiency in statistical significance of our conclusions and are thus, threats to validity. As for construct validity, there is always doubt whether the variables are well understood by the managers who must attribute values to them. This subjectivity leads to the threat that one may not be obtaining simulation outputs that match reality.

VII. CONCLUSION AND FUTURE WORK

Decision-making within IaaS capacity management is complex for it needs to consider the interplay among different levels of management, business, benefits, and demand.

This paper presented a novel system dynamics model to support multiple criteria decision-making (MCDM) in capacity management by JaaS providers, pushing forward the state of the art. Other contributions our research brought include:

A way to align business capacity management practices and technical capacity management in IaaS provisioning contexts.

Integration of business and technology needs.

Recommended decisions based on multiple simulated criteria Aligunt of performance monitoring, capacity management

behaviors and decisions suggested by the model with the faaS provider's business objectives

The proposed model is valuable from the cloud providers' perspective because it captures capacity behavior based on business impact; it is not intrusive because it simulates the IaaS provider's real scenarios offline; and, it can be used as a complement to IaaS provider capacity management and performance tools. The model is applicable to IaaS provider's scenarios that use ITIL processes and it has been tested in real-world business scenarios at a case IaaS provider ("Alpha provider") in Brazil.

The model requires a good amount of input and calibration variables. Alpha managers agree that such amount is justified by the complexity of a capacity management scenário in any IaaS provider. During refinement tests, the modeling process was continuously re-evaluated, and the input and calibration variables deemed less influential were removed. The proposed model provided support for decision-making by identifying which capabilities could be improved, changed, or maintained and how such capabilities behave in a complex network of interactions, among many other factors involved in an laaS capacity management process. The model is flexible and allows the inclusion of new management behaviors, In addition to other processes that may be proposed by managers in order to improve it.

Future work could investigate replication of this study to other JaaS service providers with different cloud environments [42] or service offerings; integrating system dynamics models into capacity frameworks in order to consider their maturity level; and, exploring simulation models for training capacity management professionals. To make the effectiveness of MCDM support in the proposed model evolve, one could try to

extend the classic AHP method to treat other types of decisions. or to evaluate the possibility to switching to other decision support methods such as the Analytic Network Process (ANP). We plan to embed the experimental tool that implemented the proposed model in a new software tool with a library of cloud IaaS providers' business models not just Alpha's service options offering. That will make the new tool more generic and as such, more readily applicable to other cloud scenarios.

ACKNOWLEDGMENTS

The authors thank the Alpha provider for granting access to its service provisioning documentation, logs, and its managers for their time. Constructive criticism from anonymous referees helped us improve this paper. Their contributions are much appreciated.

REFERENCES

[1] S. S. Manvi; G. K. Shyamb, "Resource management for Infrastructure as a Service (IaaS) in cloud computing: A survey," Journal of Network and Computer Applications vol. 41, pp. 424–440, 2014.

[2] Infrastructure as a Service Market Analysis, Market size, application analysis, regional outlook, competitive strategies, and segment forecasts, 2016 to 2024. Available from www.grandviewresearch.com/industry analysis/infrastructure-as-a-service-iaas-market. Accessed July 8th 2020.

[3] A. L. Mesquida; A. Mas,"Integrating IT service management requirements into the organizational management system," Computer Standards & Interfaces, vol. 7, pp. 80-91, 2015.

[4] J. D. Sterman, "Business Dynamics. Systems Thinking and Modelling for a Complex World," McGraw-Hill, 2000.

[5] A. S. Lima; J. N. de Souza; J. A. B. Moura; I. P. da Silva, "A Consensus Based Multicriteria Group Decision Model for Information Technology Management Committees," IEEE Transactions on Engineering Management, vol. 65, no. 2, pp. 276-292, 2018, doi: 10.1109/TEM.2017.2787564.

[6] M. Carvalho; D. A. Mesnace. F. V. Brasileiro, "Capacity planning for IaaS cloud providers offering multiple service classes," Future Generation Computer Systems, vol. 77, pp. 97–111, 2017. Softw

[7] A. Gandhi; P. Dube, A. Karve, "Model-driven optimal resource scaling in cloud," Syst Model, vol. 17 pp. 509, 2018, https://doi.org/10.1007/s10270-017-0584-y.

[8] C. Bartolini; C. Stefanelli, "Business-driven IT Management," IFIP/IEEE International Symposium on Integrated Network Management (IM 2011), pp. 963–969, 2011.

[9] A. S. Lima; J. N. de Souza; J. A. Oliveira; J. Sauvé and J. A. B. Moura, "Towards business-driven continual service improvement," 2010 Network Operations and Management Symposium Workshops (NOMS Wksps), pp. 95-98, 2010.