

An Introductory Overview and Survey of Business-Driven IT Management

Jacques Sauvé, Antônio Moura, Marcus Sampaio
Universidade Federal de Campina Grande, Brazil
{jacques, antao, sampaio}@dsc.ufcg.edu.br

João Jornada, Eduardo Radziuk
Hewlett-Packard Brazil Research & Development
{joao.jornada,eduardo.radziuk}@hp.com

Abstract - Business-driven IT management (BDIM) is a new, evolutionary and comprehensive IT management approach that aims to improve IT infrastructure, service quality and business results at the same time. To that end, it needs to model and numerically estimate IT-business linkage. BDIM concepts are finding ways into ITIL-based management processes as well as into new IT infrastructure product offerings such as autonomic computing in order to add increased value to the business. In the hope of contributing to define and characterize this new approach, this paper presents an introductory overview of BDIM, discusses its main concepts, illustrates gains over conventional IT management approaches and offers a survey of some recent work on the topic in the literature.

Keywords: Business-Driven IT Management (BDIM), IT governance, Information Technology Infrastructure Library (ITIL), business metrics, modeling, performance evaluation, automation.

I. INTRODUCTION

Over the years, information technology (IT) departments have faced the challenge of providing services efficiently and in a cost-effective way to corporate users and clients. IT operations have now become mission-critical to most businesses. Additionally, demand for IT services has grown due to requirements of corporate investors and regulating agencies. More recently, it is becoming common to expect that IT services also aggregate value to the business, contributing to business goals, results and overall corporate governance. Such expectation is attested to by the introduction of Control Objectives for Information and Related Technologies – Cobit [1] and Sarbanes-Oxley Act (SOX) [2] compliance requirements. To meet this increasing challenge, IT management methods – tools and processes included – have had to evolve in maturity¹.

Early IT Management focused on handling faults of individual IT infrastructure devices: the IT department was essentially firefighting. The introduction of operating system and, later on, network management tools pro-

moted domain-centered IT management: policing systems to protect and ensure proper operations and assisting users through a help desk function. IT management became more user-centric with IT Service Management (ITSM) under which the IT department's role becomes that of a service provider; one then views IT in both a technical and users' perspectives. This has led to a set of popular best practices: the Information Technology Infrastructure Library framework[4]. This led to service catalogs and Service Level Agreements (SLAs) by which certain promises are made to customers about the quality of the service provided. ITSM allows IT personnel to better understand the needs of the business and act accordingly. However, ITSM mostly uses *technical metrics* – usually availability, throughput and response time – to gauge the quality of service (QoS) offered to the user. This is true at least in the very important Service Level Management ITIL process. For other processes such as change management, metrics may be altogether different but they still reflect how smoothly IT operations are running. IT personnel make decisions which may not serve the business best when the only operations smoothness is considered; business needs are best served when *business metrics* – such as cost or revenue – are considered to gauge business impact. BDIM argues that basing IT decisions on IT metrics leads to suboptimal decisions, larger adverse financial impact on the business and, possibly, lower user and customer satisfaction. This paper illustrates some of these arguments.

We now witness the dawn of the next age: Business-Driven IT Management (BDIM). The focus of BDIM is on the business: IT must help the business to achieve its goals, contributing to its results in a measurable way. Business metrics are not only the corporate user's language, they are also the language used by executives. It is not easy to sell IT to senior executives when all one has to report is some percentage of service availability or the attainment of – or failure to achieve – SLA objectives. Reporting to top executives is much easier when business metrics are used. ITSM with business metrics can be a powerful communication tool with the user and with top executives. BDIM is ITSM with business metrics.

BDIM attempts to gauge the impact that IT has on the business and aims at rethinking IT management from this perspective, whether this be in an operational, tactical or strategic context. As such, BDIM also offers support to IT

¹ Maturity evolution in [3] is made equivalent to a four-phase *process* for the IT department to migrate from a reactive to a proactive, business-driven role.

Governance. BDIM involves a new culture, tools and decision-making processes that should aim to *help the business*. This allows IT-related decisions to be taken from a business perspective. A complete Service Management shift to BDIM requires IT personnel or automated tools to use business metrics to gauge the QoS offered to a business user.

BDIM is quite a new area of research, as can be seen by the sparse publications available on the subject and the lack of standard terminology. The term *Business-Driven IT Management* appears to have first been used in [5] where a vision for an adaptive IT infrastructure is described. Alternative terms include: “Business-Oriented Management” [6,7]; “Business Impact Management” [8,9]; “Business Service Management” [10], “Service Impact Management” [11], “Business-Centric Service Level Management” [12], “IT Management driven by Business Objectives” – MBO [13]; recent IBM literature also refers to “Business Objectives” [14]. Although BDIM support tools are being introduced in the market (refer again to [11]) they are still in their infancy and functionally limited. On the other hand, BDIM can cover a lot of ground, in future, adaptive or in current, “conventional” IT infrastructures. One may wish to find the effect of IT outages and degradation on day-to-day business operations; or one may wish to take a bird’s eye view and discover how IT operations may support business strategies. There is growing interest in BDIM because of the gains it can bring to business results, as this paper will show. BDIM is the future of IT Management.

This paper contributes to BDIM by offering a comprehensive introduction to the topic. Its contents reflect both IT and business administration terminology and approaches. The remaining sections of the paper examine BDIM concepts, solutions and gains. Section 2 introduces basic BDIM concepts – including our definition of BDIM; a suggestion of its adoption cycle to address IT problems – particularly those of autonomic or adaptive computing; a discussion on estimating IT-business linkage which articulates any BDIM application; and a comparison to Business Impact Analysis (BIA). Section 3 illustrates BDIM usefulness by applying its concepts to the capacity planning of an e-commerce site. Section 4 reviews the growing BDIM bibliography. Lastly, section 5 offers conclusions.

II. BDIM – CHARACTERIZATION AND CONCEPTS

A. A definition and adoption of BDIM

Business-Driven IT Management is the application of a set of models, practices, techniques and tools to map and to quantitatively evaluate dependencies between IT solutions and business performance and using the quanti-

fied evaluation to improve the IT solutions’ quality of service and related business results.

Our definition above implies that BDIM is actually an IT quality improvement and control approach with metrics for the IT-business interdependencies as objective functions; we here refer to these metrics as *IT-business linkage metrics*.

One can recommend applying BDIM practices and techniques according to the steps of the *Plan, Do, Check and Act* – PDCA cycle, initially proposed by W. Shewhart from Bell Labs in the 1930’s and later adopted and promoted by W.E Deming for quality improvement programs [15]. The PDCA cycle is used dynamically: the completion of one cycle triggers the start of the next.

The initial *Plan* step is where changes or tests to improve a solution or system are envisaged. The goals of the changes are specified here, after the problem is formalized, studied and cause-effect relationships are analyzed. In the BDIM context, goals are set to align IT to the business; the problem can be any IT-related decision, e.g., optimal capacity planning, optimal IT resource allocation, prioritizing incidents or changes, etc. Planned (corrective) changes are implemented in the *Do* step. Data gathering tools or modules may have to be integrated to collect information for subsequent control and adjustments of the system. In the *Check* (or *Study*) stage, system performance is monitored, gathered data is analyzed and compared against set goals. The *Act* (or *Control*) stage triggers a new spin in the cycle in case deviations from the set goals are unacceptable.

Figure 1 illustrates the application of the PDCA cycle to BDIM. In control theory terminology, the PDCA cycle is equivalent to *closed-loop control* [16]. BDIM proposes to introduce feedback about business results into IT products and services to drive IT-related decision making and thus attain QoS that maximizes business value.

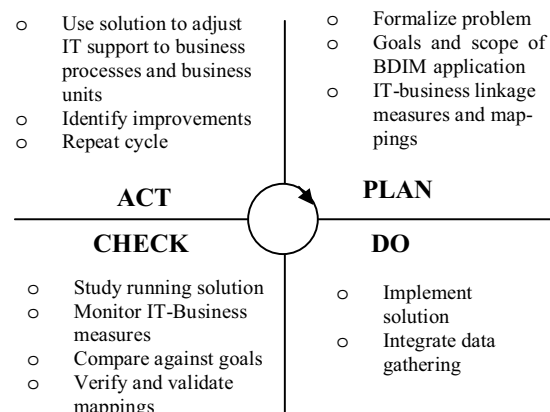


Figure 1. BDIM PDCA cycle

A closed-loop control view of BDIM is given in Figure 2. In this figure, IT services support business processes (BPs) which are used to produce business results (solid lines in the figure); the dashed lines represent information used by the BDIM control mechanism to adjust IT decisions and QoS to improve business results. IT practitioners and researchers may prefer the control view of BDIM. Business administration experts may opt for the BDIM PDCA cycle. For continuous quality improvement using BDIM – whose aim is to improve business results by making changes to the IT infrastructure and services – the process can be re-examined and a new change or test can begin.

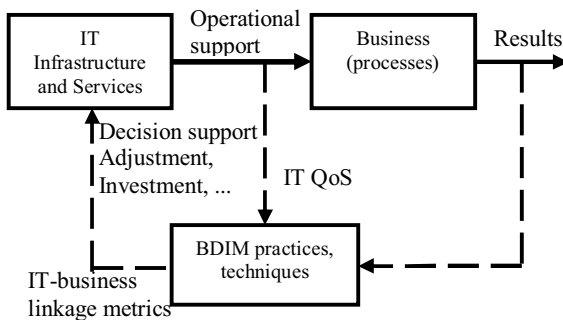


Figure 2. BDIM as a closed-loop control mechanism

The application of BDIM concepts extends the planning, design, engineering, deployment, operations, usage and management of IT solutions – products and services – to explicitly include business measures as optimization functions. BDIM techniques may be applied either off-line, to design optimal capacity IT solutions (see section 3) or to tune a system during down periods, for instance; or, on-line, to automatically tune an IT infrastructure to meet varying workload conditions [6] or adaptive/autonomic computing requirements [13,14].

B. BDIM and autonomic computing

Automated use of BDIM techniques may allow for real-time monitoring, notification, reporting and on-the-fly control of IT solutions in support of new, self-managed computing platforms as envisioned by IBM with its Autonomic Computing Initiative [17], by HP in its Adaptive Enterprise Program [18] and by Microsoft’s Dynamic Systems Initiative [19]. An encompassing and detailed discussion of adaptive IT infrastructures driven by business goals is offered in [5] – whose authors, however, are quick to ascertain that such “IT infrastructures ... will not be ready for a few years”. Conventional IT systems have shown gains when subject to business-driven adjustments (see for example [14,33,35,36,]). Since automated adjustments can be made faster, it is reasonable to expect that autonomic systems that embed real-time control mechanisms with business measures as objective functions will add higher value to corporate users. Embedded BDIM techniques will thus cloak auto-

nomic systems in an attractive selling proposition making these systems quickly embraced by manufacturers.

BDIM tools should automate decision support in settings of interest to manufacturers, providers, users and managers of autonomic IT products and services. BDIM support – in the form of utility functions for automatically ranking incidents and changes to be handled by self-reconfiguration and self-healing features – has already been proposed for autonomic products [20,21]². An IT service or resource provider may do BDIM-based dynamic allocation of server resources in order to maximize customer revenue and hence, its own bonus commission – in fact, such an “autonomic self-optimization according to business objectives” web site scenario is the focus in [14].

In order to build and operate solutions for BDIM, either for autonomic or off-line usage, one needs to choose and evaluate appropriate *IT-business linkage metrics*, discussed next.

C. IT-business linkage metrics

Business metrics measure results of a company as it operates in the marketplace. In order to capture *IT-business linkage* we look into variations (positive or negative) in business metrics caused by IT performance. We refer to these variations as *BDIM metrics*. Good BDIM metrics must satisfy the following requirements:

- **Clear business semantics** – the metric must be immediately understood by business people (i.e., in areas other than IT) and related to their day-to-day activities – the use of numerical attributes *already* employed by the business entity, instead of inventing new metrics, is likely to satisfy this requirement.
- **Business impact** – of certain IT events (say component faults) on business results must be captured. In ITIL literature [4], *impact* is a measure of business criticality, often equal to the extent to which an incident leads to distortion of agreed upon or expected QoS. With BDIM, impact corresponds to *IT-business linkage* which must be numerically evaluated with respect to business results.
- **Comparison property** – allows two situations to be compared. For example, the impact of two contingent events can be compared using the numerical value of the BDIM metric under scrutiny. Section 2.5 discusses metric accuracy.
- **Additive property** – allows the sum of metrics. As an example, the total impact of all contingent events occurring over a one-month period can be calculated by adding the impact metric of each event.
- **Instantaneous impact** – some situations may call for a BDIM metric to provide an *instantaneous* rate of

² By allowing the time scale to stretch - say with human intervention - a similar BDIM tool may be used to prioritize incidents and schedule IT infrastructure changes that disrupt business the least.

impact. This means that, at any point in time, things are getting worse (or better) at the given rate. The total cumulative impact can be simply calculated by integrating the rate metric over the desired time period. One example: an e-commerce site presents a potential revenue loss rate (say in \$/hour) whenever it becomes unavailable. The longer it takes to bring the site back up, the greater the potential loss. The total estimated financial loss will be given by the integration of the instantaneous loss rate over the total predicted unavailability resolution time.

Revenue, profit, costs, financial loss, earnings-per-share, production level, inventory turnaround, time from order to fulfillment, time-to-market, market share, frequency of new product launches, frequency of delayed deliveries, average time for collection, age of accounts receivables [22] are but a few examples of business metrics. Some of these metrics measure beneficial results (e.g., revenue); others measure adverse results (loss). Loss is a business metric that accounts for missed revenue opportunities due to contingent events in the IT infrastructure or services. It is actually a *rate* at which loss is instantaneously accumulated at any given instant in time. Accumulated loss is another important BDIM metric since it captures the adverse impact on business when IT faults or degraded performance are not resolved over a period of time. While rate of loss may remain constant over this period, accumulated loss will steadily worsen.

Note that a *financial* BDIM metric satisfies all requirements listed above and is clearly desirable. However since financial quantities may be difficult to determine, other metrics can be used. One can envisage non-financial metrics such as number of affected people or business processes or missed deadlines – even though these measures lead to a dollar value indirectly. In some instances, certain non-financial metrics can be converted to a financial value: this is called *monetizing the impact*. Financial metrics and monetized impact estimates need to be used with caution. In most cases they are just estimates of *potential* financial gain or loss and should not be accounted for as hard dollar figures. Nevertheless, in general and as noted in [23 – p.142], monetizing performance measures seems to sharpen one's responses remarkably and are thus appropriate for decision making.

BDIM evaluates business metrics by mapping their dependencies on IT performance. Such mapping is done through the use of what we call an *IT-business linkage model*, or simply a *BDIM model*.

D. BDIM models

Quantifying BDIM metrics is a complex undertaking and cannot be done intuitively by IT personnel without the help of proper tools. Paraphrasing the remark in [24] that "... theoretical work in system administration is

about building models which link cause and effect", one can state that much theoretical work in BDIM consists of building models that link IT causes and business effects. The business results of IT performance depend on complex relationships between the IT infrastructure, IT services, business processes (BPs) using these services and the business units that employ business processes to generate aggregate value to the business. What is needed is a formal model that captures these relationships and can estimate business results numerically, to steer manual or automatic IT management by business terms. We identify the following model requirements:

- **Estimate business impact** – map IT metrics or events into business metrics (i.e., the model shows the impact of IT behavior or condition couched as business measures).
- **Drill-down** – provide a means whereby the dependencies of estimated business results on IT performance levels can be traced down – first to related business processes (BPs) and from there to supporting IT services / components or infrastructure.
- **IT measurements**. – accept as input measurements on the IT infrastructure and relate those values to business metrics. Thus the model must include IT entities on which measurements may be performed or attributes gathered by other means.
- **Low intrusion** – A BDIM model is said to be low-intrusion if it requires:
 - little modeling effort (the model is simple);
 - little instrumentation on the IT infrastructure;
 - little or no access to other business data (such as financial data, BP execution data, etc.);
 - little or no modification to legacy software;
 - little input to define final-user requirements;
 - little or no configuration (parameterization, etc.);
 - little effort to calibrate or to establish baselines;
 and if it is robust to changes in the business and IT entities and their relationships in the sense that changes can rapidly be effected over time to keep the model synchronized with reality.
- **Flexibility** – allows adding or removing entities to make model simpler (by removing entities) or more accurate (by adding entities).
- **What-if analysis** – allows the analysis, over any time frame, of the impact of simulated IT QoS levels (or equivalently, corresponding BP performance levels), in business metrics. IT performance levels may be made to improve or to degrade, as a result of changes in the characteristics of the IT infrastructure or services. BDIM what-if analysis tools will allow for the observation of changes in business results (*impact*) as IT QoS changes.
- **Business changes** – corporate executives may change business priorities or strategies without incurring costly modeling overhauls.

The business impact, drill-down and IT measurement requirements suggest a layered model with at least two basic layers: the IT services layer and the business layer.

The IT services layer is below the business layer and feeds information from the IT infrastructure to calculate business metrics³. Intermediate layers can make it easier to generate final business metrics since they offer intermediate metrics that make mapping IT-business linkage more intuitive; also, additional layers can provide a better drill-down model, as long as the entities encountered in the intermediate layers are familiar to the tool users.

Figure 3 presents a basic, general three-layer hierarchical architecture from which some reported work in the literature build BDIM models – see for instance [14,25]. The advantage of the hierarchical modeling technique is twofold: first, it provides a structured approach to BDIM model development; and, second, it allows the combination of multiple techniques to develop the model solution. The topmost, business layer supports the calculation of business metrics. The intermediate, BP layer captures how IT-related metrics (such as response time, throughput, and availability) or events (service stoppages, SLA violations, and infrastructure changes) affect changes in BP-oriented measures (such as BP throughput, turnaround time, etc.). The IT Services layer may typically take into account:

- existing hardware – e.g., number of servers, of CPUs on each server, network configuration;
- applications and services⁴ – details include required resources and operating behavior) provided by the IT infrastructure; and,
- user behavior synthesizing the way the user utilizes the IT services supported by the infrastructure.

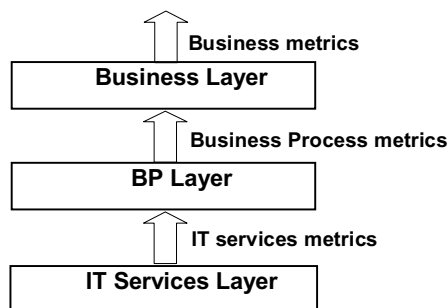


Figure 3. A basic BDIM model

For a cleaner structure, the contents of the lowest layer can be organized as a collection of sub-layers: one with the user behavior sub-model; another for the sub-model of

³ A layer obtains information from the layer below, adds value, and provides information to the layer above it. “Above” and “below” merely refer to the way the layers are drawn in figures.

⁴ We follow [5] and differentiate applications and services as follows: applications are accessed by human users through an appropriate interface such as a GUI or a browser, whereas services are accessed by programs.

application and service operating details and yet another sub-model for the hardware organization.

Input to the BDIM model is provided at the IT Services layer in the form of information on the state of the corporate IT infrastructure. Entities included in this layer are routers, switches, host computers, and other basic IT components (e.g., database services). As said, this layer comprises final IT services – including software applications – as accessed by users or other systems.

The Business Process (BP) Layer consists of those processes that make use of IT services – or through which a user interacts with IT services. The BP layer is the stepping stone to link IT behavior to business results.

Finally, the Business Layer is where business entities are modeled and final business metrics calculated. The model here could be built from a business performance evaluation methodology such as standard accounting; Return On Investment (ROI) Analysis, Activity-Based Costing and Economic Value Added (EVA) [22]; or the Balanced Scorecard [26].

The final structure of the BDIM model should mimic the stratification of a company’s vision of the IT infrastructure and policies [27] it uses to run its BPs and its business rules [29] to deliver services (for a more thorough discussion on this, please refer to section 2 in [5]). The model serves to analyze how BDIM may affect business operations or results statically or dynamically – as in the case of adaptive / autonomic computing.

So far, the discussion is quite abstract and generic in that no particular metric or formula used in calculating metrics is imposed by the architecture in Figure 3. The solution of an actual model for a given BDIM scenario provides mapping formulas for IT to business metrics according to real corporate characteristics. Note also that a generic architecture is required in order to account for different instrumentation realities found in different enterprises; for example, some enterprises have Business Process Management Systems (BPMS) in place while others do not. It is very important that the BDIM model enable a given enterprise with very little instrumentation to be able to perform BDIM, with as little intrusion as possible. In general, with more instrumentation, metrics are more likely to be measured; with little instrumentation, they are more likely to be calculated.

At this point, one may consider *when* to perform BDIM. Two alternatives are discussed in [5]:

1. *A priori* – starting from a model of the BPs, IT services and resources, derive necessary requirements (changes needed to meet business objectives), either by mathematical analysis or through simulation. Then carry out the required changes. In this case, we have model-based approaches, which are less intrusive.
2. *A posteriori* – when one treats the system as a black box and tries to infer the system behavior by mining data or doing statistical analysis from operating logs.

Note that *a posteriori* requires measurements which are more intrusive. Note also that 1 and 2 may go hand-in-hand for validation purposes, an activity in the Check step of the BDIM PDCA cycle. Logged data may also serve to feed models in 1.

Closer examination of the model in Figure 3 also serves to indicate *where* BDIM may be applied: to adjust the IT infrastructure, to an individual BP or sets of BPs, to Business Units (BUs) or even to the entire company. This breadth of application is categorized in [5] as *local* (to a given layer) and *global* (over all layers). Most efforts towards BDIM so far concentrate on the IT layer. The model suggests however, feasibility of application to the BP layer – but this is yet to be explored. As remarked in [5] in an adaptive computing context, few business managers will let anyone change a BP on the fly. In general however, BDIM may assist Business Process Management – BPM [28,30]. Interest in pairing BDIM to BPM will increase as BDIM tools mature and their embedded models are solved and validated using data from real cases.

E. BDIM model solution and metric accuracy

A business metric is a numerical value. The solution of a BDIM model is sought in order to value business metrics of interest (output of the top layer). In some instances, it will be possible to measure the metric, but it is expected that the more frequent case will involve estimating (through mathematical calculations or simulation) the metric value based on IT-business linkage considerations through the hierarchy of layers. The hierarchical layering characteristics of a BDIM model allow for the application of a hybrid solution technique when solving the model: one may use measurements in some layers and mathematical calculations and simulation in others. In fact, if the interfaces between model layers are maintained, one may actually change the solution technique employed within a given layer without any consequences to the solution techniques used in other layers. Measurements and simulation techniques may require the most coding effort when compared to mathematical calculations but, those techniques seem to be more credible to managers and business executives. Measurement results are the most credible (or accurate) but unfortunately, they also demand the most intrusive tools to be obtained. Mathematical calculations – once formulas have been derived and properly validated – usually offer the fastest and cheapest alternative (in terms of software deployment and execution costs). On the down side, a mathematical technique (such as queuing or reliability theory) usually requires simplifying (and restrictive) assumptions for analytical tractability. As a result, its estimates may not be as accurate. A hybrid approach thus allows for the most appropriate combination of techniques in order to achieve

an acceptable balance of intrusion, complexity, cost and accuracy for the solution of a BDIM model.

It is important to realize that not all scenarios where BDIM is used will require the same accuracy when estimating metric values. For example: “How much is the company losing per minute when e-mail service is down?” is probably not answerable with a one-dollar accuracy. However, estimating an order of magnitude for the metric (thousands of dollars? tens of thousands?), may be sufficient for certain decision making. In other instances yet, a simple *impact level (IL)* may be sufficient:

- $IL = 1$ for the case of No Immediate Impact;
- $IL = 2$ for Moderate Impact; and,
- $IL = 3$ for Critical Impact.

When solving BDIM models, one should be aware of the fact that “quantifying a link between a leading indicator and a dependent variable is partly trial and error” [23 – p. 107]. Trying to link IT faults (such as those that lead to an SLA violation) to adverse business impact (such as a potential increase in cost or loss) is no different. A given BDIM solution presupposes the combination of the rigor of mathematical analysis with the strength of managerial experience. Even if results are estimates at best, the quantitative technique developed to produce the required mapping functions will provide a framework for thinking about the problem the BDIM solution is being applied to. As commented in [23 – p.140]: “Decision making becomes much tighter as soon as the uncertainties can be quantified, no matter how vaguely”. Also, note that for comparison purposes, metric accuracy need not be high as long as the partial ordering of alternatives is correct.

F. BDIM and BIA

It is important to note that evaluating *business impact* as an IT management task was *not* originally proposed by the BDIM community. Information security practices recommend that one carry out Business Impact Analysis – BIA [31] to elicit security requirements of corporate assets. Although, both BDIM and BIA evaluate business impact, they differ in several aspects.

While BIA is a *qualitative risk* assessment process, BDIM is quantitative in general, although in some situations, BDIM qualitative estimates may do. BIA uses mainly interviews for analysis / validation. As such, BIA does not lend itself to automation easily and definitely not for real-time, autonomic applications, as BDIM does. From the point of view of the business, BDIM is a holistic IT management approach. BIA is a task within IT security management, a single process within the realm of IT management, as proposed in ITIL [4]. From this point of view, once can say that BDIM encompasses BIA. BIA identifies time-critical BPs to determine the impact on these BPs of a significant interruption or disaster; it considers no partial IT capability, such as high response time. On the other hand, BDIM solutions cover a continuum of

IT performance possibilities. BDIM is thus more adequate for optimization techniques under budget restrictions. By considering worst case scenarios, BIA yields the maximum tolerable downtimes (MTDs) or equivalently, the recovery time objectives (RTOs), for each critical BP. Hence, BIA is applicable mainly for prioritization of incidents and continuity of services (a subset of IT management processes) while BDIM may be applied to all ITIL management processes. BIA it is not used by manufacturers as an embedded feature of IT products; it is typically performed by the IT department for disaster recovery and business continuity planning. BDIM techniques and procedures are being embedded in IT manufacturers' offerings, including security solutions [32].

III. A MOTIVATING BDIM APPLICATION EXAMPLE

Gains by using BDIM are illustrated here by considering the design of the data center infrastructure for an e-commerce site from a business perspective as carried out in [33]. It differs from non-BDIM methodologies in that it evaluates and compares alternative designs by calculating the financial loss imposed by a conventionally designed infrastructure. The methodology provides the optimal infrastructure that minimizes monthly financial outlays as calculated by the infrastructure cost plus the business loss incurred during failures and performance degradations. The infrastructure itself consists of three tiers: a web tier, an application tier and a data tier. Each tier is served by a load-balanced cluster with a certain number of machines, sufficient to handle the applied load. Varying this number of machines affects response time and thus the business loss due to customer defections occurring because of high response time. Furthermore, additional machines are available in standby mode to improve site availability and hence reduce business losses due to service unavailability.

The problem in [33] is to choose the number and type of machines in each tier's load-balanced cluster and the number of standby machines, that is, the configuration that minimizes cost of the infrastructure ($C(T)$) plus business financial losses due to imperfections in the infrastructure ($L(T)$) over a time period T , i.e.,

Find:	For each tier j ($j=1,2$, and 3), the total number of machines n_j and the number of load-balanced machines m_j
By minimizing:	$C(T)+L(T)$, the total financial impact on the business over a time period T
Subject to:	$n_j \geq m_j$ and $m_j \geq 1$

To solve the problem, one needs to derive expressions for $C(T)$ and $L(T)$ first. $C(T)$ can be calculated as the sum of individual costs for all components used in the design. Estimating $L(T)$ is done through a BDIM model (Fig. 4).

The BDIM model built in [33] uses response time (RT) and availability (A) as IT service metrics (layer 1 outputs). Layer 2 models the fact that only a fraction of the customers who visit the e-commerce site buy; some defect; others just browse. Customer behavior is modeled through the use of a Customer Behavior Model Graph (CBMG) [34]. Revenue-generating sessions are initiated at a rate of $f \cdot \gamma$ sessions per second, where γ is the average number of sessions initiated per second and f is the fraction of sessions that generate revenue. If availability were perfect and response time were always low, this would also be the revenue-generating throughput (sessions ending without customer defection and producing revenue) of the e-commerce BP. However, due to IT imperfections, the actual throughput is X transactions per second, with $X < f \cdot \gamma$. The lost throughput in transactions per second is ΔX (the layer 2 e-commerce BP metric). In the business layer of the model, BP throughput is monetized. When the average revenue per completed revenue-generating session is ϕ , one may express the business (revenue) loss over a time period T as: $L(T) = \Delta X \cdot \phi \cdot T$.

$L(T)$, the layer 3 output or BDIM metric, has two components: loss due to unavailability and loss due to high response time. Thus, we have:

$$L(T) = (\Delta X^A + \Delta X^{RT}) \cdot \phi \cdot T \quad \text{Equation (1)}$$

where ΔX^A is the throughput lost due to service unavailability and ΔX^{RT} is the throughput lost due to high response time (customer defections). In [33], ΔX^A is solved using reliability theory and ΔX^{RT} is solved using queuing theory.

For a given set of values (such as average response time of 8 seconds or higher; \$1 revenue per transaction with a load of 14 transactions per second; and $T = 1$ month), the BDIM approach yields a design for the site infrastructure which produces a total monthly Loss + Cost of just over \$ 50 thousand; an average response time of 0.26 s; and, availability = 99.98%. For the same set of values, an ad hoc, non-BDIM design for the site, produces a Loss + Cost which is two orders of magnitude higher (\$ 5 million), average response time of 1.76 s. and identical service availability (99.98%). In this case, a BDIM design results in millions of dollars being saved every month. Other studies show similar BDIM results.

IV. SOME BDIM RESULTS IN THE BIBLIOGRAPHY

Some publications are beginning to address BDIM – albeit without always referring to the term *BDIM* explicitly. This section offers a non-exhaustive review of BDIM bibliography, to illustrate the increasing attention being devoted to BDIM research in recent years, in several areas of IT management.

The application of BDIM to ITIL processes has been gaining ground. As already discussed, incident prioritization has been addressed in [20] and change management in [21]. The capacity planning and management of a Web site is carried out in a series of related papers [25,33,35] – one of which [33] was discussed in the previous section. The papers are based on the model that estimates IT-business linkage metrics from the loss of throughput due to high response time and service unavailability. Service Level Objective (SLO) definition is considered in [25] by pondering the cost of the IT infrastructure needed to support the IT services and the losses incurred from service degradations. The capacity planning of an e-commerce site when it is subjected to load surges – which happen around special dates, such as the end-of-year buying season or during planned sales promotions – is done in [35]. Again, this work compares the BDIM approach to the conventional cost-oriented one. Gains over \$100 thousand/year in favour of BDIM were estimated in a typical scenario.

Papers on cost-benefit analysis in network security settings can be considered early BDIM studies. In fact, Security Management is an ITIL-defined process and comparing the replacement cost of a corporate asset to the acquisition price plus the cost incurred to bring an IT protection solution back to working order may not be full BDIM, but it definitely takes business-oriented measures into account. As such, paper [32] addresses intrusion detection using a BDIM perspective. The paper argues that trying to block all intrusion attempts can be a very expensive security defense operation. Alternatively, one may deploy intrusion detection devices that estimate the business value of the asset under attack and then decide what to do about it – note that one such device is an example of a BDIM tool. As security threats are detected, the device will compare response cost against damage cost. Only if the damage cost is higher will it stop an attack. In other words, the device itself decides if it is cost-effective to launch an emergency response using a financial objective measure. One such device, called *hummer*, has been built by the University of Idaho [32].

By using a CBMG obtained from http logs, paper [36] describes a customer session at an e-commerce site and presents a family of priority-based resource management policies for the servers. Priorities are made to change dynamically as a function of the state a customer is in (browse, search, select, add to cart, pay) and as a function of the amount of money the customer has accumulated in the shopping cart. The policies attempt to keep customers satisfied with the site's QoS (response time) – particularly those customers that would cause high revenue losses (considering the total items already in the cart) if they departed due to poor QoS. The policies aim at optimizing two business-oriented metrics at peak times without changing server capacity: revenue throughput (re-

nue/sec, which is to be maximized); and potential lost revenue/sec (to be minimized). The adaptive policies pursue their optimization goals by simply attributing higher service priority for customers buying potentially higher volumes. By simulating an electronic bookstore, the study shows gains for the adaptive policies when compared to resource management done by tracking “conventional performance metrics such as response time, throughput and availability”.

E-commerce policy modeling is also addressed in [14] by considering the dynamic and autonomous optimization of IT infrastructure parameters (such as policies that attribute traffic handling priorities) according to high-level business objectives (such as revenue increase). A three-layer BDIM model (although not identified as such) coupled with an optimizing algorithm which searches over the space of possible actions/policies is solved by simulation and was used to tune the operations of a stock trading web site. The user behavior sub-model uses a CBMG and a set of user classes, each having its own session initiation frequency, to capture the traffic pattern of the requests that enter the site. Attribute values for the CBMG and classes are derived from web logs. User attributes, such as purchase amount in a purchase request, are modeled using a set of probability distributions generated from logs and database values on the customer's site using statistical methods. Tier-level service demand sub-models are built by applying business process analysis and distribution-fitting algorithms to a log produced by tracing and monitoring tools. IT resource (CPU, disk) requirements service times are modeled by a load-dependent distribution function also generated from data gathered by performance tools at the web site. The site was built using IBM technology and incorporating the above BDIM techniques. The site services two types of transactions: gold and platinum, each carrying its own response time objective and penalty – to be paid by the site provider whenever the objective is not met. The provider collects a commission of either 4% of the total price or \$25 (whichever is greater) for each buying/selling of a stock. Through BDIM, the site was able to turn a loss of \$400 thousand into a positive income of \$180 thousand over an unspecified time frame.

Optimization policies for a shared utility computing environment which supports multiple third-party-applications subject to SLA performance targets – in terms of maximum throughput and minimum response times are considered in [37]. The resulting model is solved using mathematical and simulation techniques. (Liu et al. [38] solve the problem of optimizing SLA profits using a queuing theoretic model but does not address dynamic workload characteristics). Overall business loss is shown by simulation to be lower when the proposed adaptive (versus the static) resource management policy is used – implying that capacity assignments made

under SLA performance targets *and* business-oriented policies yield better business results.

In addition to references [5,20,21] on BDIM and autonomic computing already mentioned in section 2.2, reference [39] by HP reviews the different open and proprietary IT Management/Governance frameworks that support IT organizations in their transition to utility computing. It also proposes that strategic alignment be put at the core of utility computing by introducing the discipline of Management by Objectives (MBO). Again, BDIM is not addressed explicitly, but the BDIM context is certainly the backdrop of the discussion. In fact, when one ties in IT Utility and IT Management/Governance to strategic goals, a need according to many CIOs [40], one is, in effect, proposing the use of business-oriented performance measures as guidelines. Such is the case of paper [41] which proposes reviewing SLAs and the enterprise IT balanced scorecard in order to keep the cost of eliciting knowledge about the business value of the service low (information on SLAs may be obtained from the Configuration Management Data Base – CMDB as defined in ITIL's Configuration Management process). Diao, Hellerstein and Parekh [42] propose an elegant approach for automated enforcement of SLAs (see also [43]) by using IT-level feedback loops to maximize profits – revenue as percentage of completed transactions minus rebate to customers who experience low-quality service.

CONCLUSIONS

This paper presented an introductory overview of a new, business-driven IT management (BDIM) approach. We hope the paper has contributed to this new approach by: defining and characterizing BDIM as well as identifying and illustrating major BDIM concepts, requirements, needs and applications. The paper also offered a brief survey of some recent work in the growing BDIM literature. The survey showed BDIM gains for some ITIL processes and futuristic autonomic computing infrastructures.

Central to BDIM is the IT-business linkage model. As in any area where modeling and solution tools are applied to evaluate cause-effect relationships, there is no single or standard way for estimating business measures through specific models. The literature reviewed indicates that BDIM models and their solutions depend on the particular scenario being considered and on the modeling and solution techniques preferred by the researchers involved. The papers reviewed tend to define business metrics from a service provider perspective and to address aspects of IT operational processes (incident and change management processes are recent highlights). It is only natural that it be so since major IT service providers function frequently in firefighting mode and utility computing requires on-the-fly self-management features. As BDIM R&D efforts

evolve to encompass a fuller set of ITIL processes and migrate to a deeper, longer-term business perspective, with validated business gains and providing support to autonomic computing, BDIM may become the ubiquitous approach for managing IT.

Acknowledgements – This work was developed in collaboration with HP Brazil R&D. We also thank the reviewers whose comments much improved the contents.

REFERENCES

1. IT Governance Institute, “*Cobit 3rd Edition*”, 2000, www.isaca.org/cobit.htm
2. Public Company Accounting Reform and Investor Protection Act of 2002 – “The Sarbanes-Oxley Act 2002”, www.legalarchiver.org/soa.htm
3. “Managing the Wan for Application Delivery with Visual UpTime Select”, Enterprise Management Associates Product view (www.enterprisemanagement.com), Oct. 2003.
4. IT Infrastructure Library, “ITIL Service Delivery” and “ITIL Service Support”, OGCommerce, UK.
5. Machiraju, V., Bartolini, C. and Casati, F., “Technologies for Business-Driven IT Management”, in “Extending Web Services Technologies: the Use of Multi-Agent Approaches”, edited by Cavedon, L., Maamar, Z., Martin, D. and Benatallah, B., Kluwer Academic, 2005, pp. 1-28.
6. Menascé, D., Almeida, V.A.F., Fonseca, R. and Mendes, M.A., “Business-Oriented Resource Management Policies for e-Commerce Servers”, *Performance Evaluation* 42, Elsevier Science, 2000, pp. 223-239
7. Casati, F., Shan, E., Dayal, U. and Shan, M-C., “Business-Oriented Management of Web Services”, *ACM Communications*, October 2003
8. Dubie, D., “Never-fail business services”, *Network World Fusion*, <http://www.networkworld.com/buzz/2002/bim.html>
9. Mason, P., “A new Culture for Service-level Management: Business Impact Management”, IDC White Paper, <http://www-900.ibm.com/cn/software/tivoli/products/download/analystreports/ar-service-level-mgt.pdf>, 2002.
10. Darmawan, B., Cox, K. and Ragab, B., “Business Service Management Best Practices”, IBM Redbook, 2004.
11. BMC Software, “Align your IT to your Business”, White Paper, <http://whitepapers.silicon.com/>
12. The Mercury Interactive Corporation, “Topaz for SLM – The Solution for Business-Centric Service Level Management (SLM)”, http://www.mercury.com/us/website/pdf-viewer/?url=/us/pdf/products/whitepapers/MI_SLM_Brief.pdf
13. Bartolini, C., Boulmakoul, A., Christodoulou, A., Farrell, A., Sallé, M. and Trastour, D., “Management by Contract: IT Management driven by Business Objectives”, *In Proc. HP Openview University Association – HP OVUA*, 2004.

14. Aiber, S., Gilat, D., Landau, A., Razinkov, N., Sela, A. and Wasserkrug, S. "Autonomic Self-Optimization According to Business Objectives"; *In Proceedings of the International Conference on Autonomic Computing*, 2004.
15. Deming, W.E., "Out of the Crisis", MIT Press, 1989.
16. Ogata, K., "Modern Control Engineering", 4th Edition, Prentice Hall, 2001.
17. IBM, "Autonomic computing: IBM's perspective on the state of information technology", <http://www.research.ibm.com/autonomic/>
18. HP, "Adaptive Enterprise Program", <http://www.hp.com/go/adaptive>
19. Microsoft, "Microsoft Dynamic Systems Initiative Overview", <http://www.only4gurus.com/v3/download.asp?ID=5276>
20. Bartolini, C., and Sallé, M., "Business Driven Prioritization of Service Incidents", In Proceedings of the 15th International Workshop on Distributed Systems: Operations and Management (DSOM 2004).
21. Keller, A., Hellerstein, J., Wolf, J.L., Wu, K. and Krishnan, V., "The CHAMPS System: Change Management with Planning and Scheduling", In *Proceedings of the IEEE/IFIP Network Operations and Management Symposium (NOMS 2004)*, IEEE Press, April 2004
22. Morgan, J. N., "A Roadmap of Financial Measures for IT Project ROI", *IT Pro Magazine*, IEEE Computer Society, January/February 2005, pp.52-57.
23. Numbers Guide – "The Essentials of Business Numeracy", The Economists Books Ltd., Third Edition, 1997.
24. Burgess, M., "Analytical System Administration 3", www.iu.hio.no/SystemAdmin/res3.html, 2000
25. Sauv , J., Marques, F., Moura, A., Sampaio, M., Jornada, J. and Radziuk, E., "SLA Design from a Business Perspective", In *Proceedings of the 16th International Workshop on Distributed Systems: Operations and Management DSOM 2005*
26. Kaplan, R., and Norton, D., "The Balanced Scorecard: Measures that Drive Performance", *Harvard Business Review*, 70(1), 1992, pp.71-79.
27. Flegkas, P., Trimintzios, P., Pavlou, G. and Liotta, A., "Design and Implementation of a Policy-based Resource Management Architecture", In *Proceedings of Integrated Network Management 2003*, pp. 215-229.
28. Smith, H. and Fingar, P., "Business Process Management (BPM): The Third Wave", Meghan-Kiffer Press; 2003
29. Ungureanu, V. and Minsky, N., "Establishing Business Rules for Inter-Enterprise Electronic Commerce", In *Proceedings of Proceedings of the 14th International Conference on Distributed Computing, 2000, Lecture Notes In Computer Science; Vol. 1914*, Springer-Verlag
30. Lazovik, A., Aiello, M. and Papazoglou, M., "Associating Assertions with Business Processes and Monitoring their Execution", In *Proceedings of the 2nd International Conference On Service Oriented Computing*, 2004.
31. Jackson, C. B., "The Business Impact Assessment Process", Chapter 140, pp. 1709-1723 in *Information Security Management Handbook*, 5th Edition, H. F. Tipton and M. Krause, Editors, Auerbach Publications, 2004
32. Wei, H., Frinke, D., Carter, O., "Cost-Benefit Analysis for Network Intrusion Detection Systems", In *Proceedings of the 28th Annual Computer Security Conference*, Oct. 2001
33. Sauv , J., Marques, F.T., Moura, A., Sampaio, M., Jornada, J. and Radziuk, E., "Optimal Design of E-Commerce Site Infrastructure from a Business Perspective", In *Proceedings of the 39th Hawaii International Conference on System Sciences (HICSS)*, 2006.
34. Menasc , D. and Almeida, V., "Capacity Planning for Web Services: Metrics, Models and Methods", *Prentice Hall*, Upper Saddle River, New Jersey, USA, 2002.
35. Sauv , J., Marques, F.T., Moura, A., Sampaio, M., Jornada, J. and Radziuk, E., "Business-Oriented Capacity Planning of IT Infrastructure to Handle Load Surges", In *Proceedings of IEEE/IFIP Network Operations & Management Symposium – NOMS 2006*.
36. Menasc , D., Almeida, V.A.F., Fonseca, R. and Mendes, M. A., "Business-Oriented Resource Management Policies for e-Commerce Servers", *Performance Evaluation* 42, *Elsevier Science*, 2000, pp. 223-239.
37. Abrah o, B., Almeida, V., and Almeida, J., "A Self-Adaptive SLA-Driven Capacity Management Model for Utility Computing", Elsevier Service, to appear, 17 pp.
38. Liu, Z., Squillante, M.S., and Wolf, J.L., "On maximizing service-level agreement profits", In *Proceedings of the ACM Conference on Electronic Commerce*, 2001.
39. Sall , M., "IT Service Management and IT Governance: Review, Comparative Analysis and their Impact on Utility Computing", HPL-2004-98, Palo Alto, CA, USA, 2004.
40. Reich, B. and Nelson, K., "In their own words: CIO visions about the future of in-house IT organizations", In *ACM SIGMIS Volume 34, Issue 4 Fall 2003*, pp.28-44.
41. Abi, I., Sall , M., Bartolini, C., Boulmakoul, A., "A Business Driven Management Framework for Utility Computing Environments", *HPL-2004-171*, HP Labs Bristol, 2004.
42. Diao, Y., Hellerstein, J.L., and Parekh, S., "A Business-Oriented Approach to the Design of Feedback Loops for Performance Management," In *Proceedings of the 12th International Workshop on Distributed Systems: Operations and Management*, Nancy, France, October 2001, IEEE, New York (2001), pp. 11-22.
43. Governatori, G. and Milosevic, Z., "Dealing with contract violations: formalism and domain specific language", In *Proceedings of EDOC 2005 IEEE Press*, 2005, pp. 46-57.

ULRs were last checked in March, 2006.