

What's Old Becomes New Again: A Three-Step Process to Mainframe Optimization in the Era of Social, Mobile and Analytics



If you're reading this article, it's likely for one of two reasons:

1. You are under the steadily increasing cost pressure that has been the rule since the economic crisis of 2008, manifesting itself in regular cries to replace the mainframe with a scale-out x86 architecture, or to just "go to the cloud where everything is cheaper".
2. The demands of your business require mobile applications, you recognize the value of the power and resilience of a mainframe behind the "storefront" – but *still* have demands for cost-containment.

Even in the most mainframe-friendly shops, availability of budget for additional capacity is at a premium. Skilled resources are stretched, such that once- common performance management disciplines have been pushed aside by daily priorities.

“Throwing hardware” at a perceived capacity problem has never been a solution. The best it’s ever been was a short term “band aid,” but today, most shops don’t even have the funding available for that... and if they do, it’s not fun to ask.

We frequently observe shops where performance and capacity management simply has not been a priority. This results in problems that “sneak up” on you. SLAs are missed; helpdesk lines are ringing more often; other performance issues are brought to the forefront which may not even have formal SLAs. We see this both at the infrastructure and application layers.

What’s old is new again. We have found that a number of our clients are looking to reinstate a resource management program consisting of Performance, Workload, and Capacity Management; or *Tune – Manage – Forecast*.

Taken individually or in combination as part of a holistic initiative, these three methodologies form the foundation of a resource optimization program within the enterprise.

Performance Management is the practice of iteratively eliminating or reducing the consumption patterns of a given job, schedule, transaction, or query. “Tuning” is a term commonly used to describe this process; it is MIPS elimination/reduction – plain and simple.

The scope of this analysis includes infrastructure, applications, and scheduling (where elapsed times are of great concern). We take a tactical, task-oriented approach to resource consumption reductions in combination with a more strategic, process-oriented approach for the future. We look to exploit enhancements introduced by the latest generation of mainframes, a current topic since the introduction of the z13.

We are after all consultants, so our approach begins with an assessment, in which we review the current environment, identify and prioritize opportunities based on benefit relative to level of effort, so that you can start to get relief that you need. We know we’ve been successful when you’re performing the same or more work with less computing resource.

Workload Management is the practice of prioritizing workloads and delivering computing resources where they are needed to meet SLAs.

We focus on workload prioritization, SLA achievement, batch schedule optimization, and “doing more with what you have” approaches to distributing computing resources. We leverage the discovery work done in the performance management section described above, assess and reengineer Workload Manager, review critical path batch schedules, and offer insights on obsolete processes & jobs, leveraging new technology where appropriate.

- Workload Management is successful when we deliver enough computing resource to a given workload to allow it to meet its defined SLA (explicit or implicit).

- **Capacity Management** is the practice of modeling future computing resource requirements based upon historical consumption, achievement of SLAs, business forecasts, the effects of the tuning work, and anticipated business events. We take a “business driven” approach, creating a correlation between consumption and business drivers: i.e., what business activity (selling widgets) drives consumption (impact on CPU of every incremental 1000 widgets sold. We consider known future events– acquisition, decommission, business increases/decreases, seasonal impacts, etc. – not just a line that trends by ‘x%.’

We also integrate our Performance and Workload Management efforts into future projections, validating projections as they become actual consumption – and then integrate variances back into the model, as we report to peers and to management.

Capacity Management is successful when it can accurately predict the need for additional computing resource before effective Workload Management becomes impossible.

Obviously, there are close relationships among the three. Workload Management becomes more difficult when the Capacity is improperly planned; Capacity Management predictions are skewed if savings from tuning are not accounted for, and SLAs are missed by inadequate Workload Management. This leads to a false perception – either that the capacity program is flawed or that more hardware is needed. And to go back to where we started... the best outcome in that scenario is a difficult financial discussion.

We’ve been doing this work since 1987. If you’d like to speak with one of our mainframe experts about how we go about this, reach out to us at (877) 467-9885 or mainframe@gtsg.com. Thanks.