Time Stable Workload Characterization Techniques

Ron Kaminski
Safeway Inc.
How Are IS Decisions Made?

- Internal political pressures
- Maneuvering for future positions
- Status
- Entertaining boondoggles and other ego strokes
- Economic pressure on managers
  - Pay systems based on span of control and budget size
    - Big complex systems need big staffs
    - Big machines cost a lot to buy, a lot to maintain and even more for software
- Extreme management risk aversion
Why Do Technical Staff Have So Little Impact?

- Politically naïve
- Incorrect and short term focus
  - Technical staff think that they are answering this question
  - Managers are thinking years ahead, building empires and relationships
- Answer questions with single numbers or "yes/no", not a framework of information that highlights over-capacity
- Rely on text and huge tables of numbers
  - Management understands charts and graphs
Why Do Technical Staff Have So Little Impact?

• Failure to collect, maintain and have ways to report long term historical context

• Communication skills issues
  – Hiding instead of sharing background information (The guru complex)
  – “Frustrated novelists” bury the answer near the end
    • Nobody reads past the first page, EVER!

• Need to point out past mistakes

• They think that they make decisions
  – Technical staff informs
  – Management decides
How Can We Do Better?

• We need to provide better information in more effective ways, and often more information than they ask for.
• Good business workload characterization is critical to making the right decisions.
• Workload characterized views of consumption are present in almost every well run IS shop.
  – Let’s look at an example of why.
Compelling Need For Upgrades

• The development group is requesting a huge new UNIX box and hundreds of PC upgrades to support their new warehouse system

• They offer these classic CPU charts of their benchmark week as proof
Maybe Not So Compelling!

- The majority of the utilization on the benchmarked UNIX server has nothing to do with the warehouse project!
- Note:
  - Undetected runaway looping process
  - Excessive Tools processing
  - Another project is on the server!
  - Typically, naïve benchmark monitoring swamps real processing
Maybe Not So Compelling!

- These developers are having a good time on that “warehouse” PC!
  - Note all the non-warehouse processing
  - Note how the nightly Diablo tournament takes Tuesday off to watch “Buffy”
- Update, this has recently moved to Wednesday to see “Enterprise”
Is That Realistic?

• Yes!
  – At least a third or more of my consulting engagements and studies of poor performance have found:
    • Excessive monitoring
    • Process pathologies (loop, ramps, etc.)
    • Presence of previously unknown “other resource consumers”

• All are invisible without workload characterized views of consumption!
What Is Workload Characterization?

- A workload is just a grouping of resource consumers
- A workload can have zero, one or many processes and these processes are selected by criteria such as
  - process name
  - directory location
  - owner username
  - any other differentiator that you can reasonably imagine
- Workloads are constructed for a variety of reasons
  - Reporting
  - “What-if…?” modeling growth or hardware choices
Workloads That Don’t Work

• There is considerable “art” in workload characterization
  – Experience
  – Situational awareness
  – In-depth technical understanding of the specific operating system, database, infrastructure components and application design

• When people first discover workload characterization, they generally split into two camps:
  – Zealots
  – Minimalists
Too Many Workloads

• Zealots
  – make too many workloads
    • Dozens
    • Hundreds!
  – Each node is judged “unique”
    • Many complex workloads with intricate criteria
      – Intricate workloads are brittle
    • “Special case mania”
      – The same process can end up in different workloads on different machines
  • These inconsistencies lead to:
    – Increased burdens on the audience
    – Confusion
    – Overburdened tools and capacity planning machines
Too Few Workloads

• Minimalists
  – Are often in a hurry to “get graphs out as soon as possible”
  – Are typically under-trained or unaware of the power that historical views of workload characterized consumption data can provide
    • Wonder why we can’t just use “top” like we used to
  – Tend towards extremely broad workloads
    • Often based on usernames
    • At some point broad workloads aren’t much better than total CPU
  – Get little value for their efforts, and tend to stop
Workloads That Do Work

• The advanced workload characterizer
  – uses consistent criteria to strike a balance between the two camps
  – Intricacy is used only when needed
  – broad, “sweeper workloads” are defined to minimize the clutter

• They also have a stronger weapon - consistency - that makes it easier on them and on the decision makers using their output

• Let’s explore some of their methods
Beginning Hints

We will describe many good reasons to create specific workloads, but keep these general ideas in mind:

• Use “business based” workload names
  – If the database is used for warehouse functions, call the workload warehouse, not database

• Use normal business language and short names
  – Intricate technical names just confuse management
  – Speak business, not “computerese”
  – If you have to explain what the name means, pick another name!
Beginning Hints

• Avoid defining “nit” workloads
  – Set a threshold (we use 2.0% - 0.5%), and determine that anything smaller is a nit that doesn’t deserve a unique workload

• Consider your audience and vary your workload characterizations accordingly
  – Basic workloads for daily reporting
  – More intricate workloads for accurate modeling and problem resolution
Beginning Hints

• Anticipate the “next questions” and always answer them before being asked
  – The unanswered “next question” can be a huge time waster
    • Often a stall technique used by the politically astute
      – It raises temporary doubt in your findings, and builds their case for swift purchase, before you answer their question
      – Often a way for the old guard to show that they still are the “top dogs” to management
    • Impatient or frightened management might run off and buy something!
Beginning Hints

– If you are going to shoot down someone’s hypothesis that lack of CPU was the cause of a problem, you’d better find out what really caused the problem before the meeting.

– Your goal:
  • One meeting or phone call per issue!

– They may say “We just want a quick and dirty answer” but they never really do! Always cover:
  • CPU
  • Memory
  • Disk IO
  • Workload response time changes
Beginning Hints

• Cultural differences are real and might affect your workload choices
  – Some cultures avoid direct blame or information that would cause someone to “lose face”
  – Any workloads are better than none

• Be consistent!
  – Always use the same groupings on all similar nodes
    • Use the same colors if you can!
  – Reduce the burden on your audience
  – Multiply the value of your workload creation efforts
  – Use consistent precedence order to decide where to put a process that meets the criteria to be in several different workloads
Workload Types

• The Heavy Hitters
• The Usual Suspects
  – The “Must Have” workloads
• Sweepers
  – No Nits Allowed!
The Heavy Hitters

- In every firm, there are usually a few well-known monster applications that receive the lion’s share of the attention.
- As a new capacity planner, you will often see these on the top of your “to-do” list:
  - Major databases
  - Payroll applications
  - Integrated accounting automation
  - Customer analysis packages
- If the vice president’s phone rings when there is a problem, it probably belongs on this list.
The Heavy Hitters

• Investigate…
  – Does consumption match up well with work done?
  – Are backups running at inopportune times?
  – Are there any “well intentioned” but now bloated “home-grown” monitors present
    • Is their output worth the cost?

• Lean towards zealotry with these
  – Often the effort to subdivide huge workloads into smaller parts will yield the answers to long-standing questions
  – There are often millions in savings available
The Heavy Hitters

- Often have large “black box” components
  - Look for vendor supplied information to further subdivide consumption
  - Function counts
    - Find or create periods where only a single function is running, then divide to get function impact
  - Factor analysis
    - Keep a change history
    - Watch for noise!
The Heavy Hitters

• Use recent data!
  – Small coding, database or file layout changes can have huge impacts
  – Solving the previous problem is no fun!

• Reinvestigate after all changes

• Question any processing that does not consume resources like the underlying business
  – Ramping consumption monitors
The Usual Suspects

• Most firms deploy similar tools across all nodes of a given operating system
  – Monitoring tools
  – Consumption collectors
  – Disk Defragmenters
  – Security sweepers
  – Anti-virus products
  – Remote administration/access
  – Backup and Restore

• Most firms have well known rogues
  – Music downloads and streaming audio
  – Excessive web surfing on that speedy company LAN
The Usual Suspects

• Some of them can bloat-up in resource intense ways
  – A memory leak or CPU intense bug in a common tool can ruin your whole summer
  – Get to know their peccadilloes
    • Automate detection and notification of their common issues
• Set a maximum budget for non-application code and manage to it
• Bonus answers:
  – Are you really backing up all the machines?
  – Are administrators still using insecure or banned applications?
The “Must Have” Workloads

• Backup/Restore
• Tools
• Administration accounts
• Political workloads
  – Is the __________ department hogging the machine again?
• Heavy Hitters
The Joys and Perils of Sweeper Workloads

• After you’ve characterized your major applications, usual suspects and infrastructure, what do you do about that pile of “nits”?
  – You make sweeper workloads!

• Take advantage of
  – user groups
  – naming conventions
  – regular expressions
    • Country-specific language patterns
    • .*[Aa]dmin.* works great in the USA
  – operating system specific common names
The Joys and Perils of Sweeper Workloads

• Perils
  – Creeping innovation bloats your little sweeper
  – The blame game
    • “What did it” is usually more useful than “who did it”
    • Focus on the technical problem that you can solve

• Solutions
  – Keep sweepers small
    • Set a threshold, we use 5%
  – Use workload naming conventions to help you find your sweepers
    • .*_logins
Time Stability

• Time Stability is a measure of the resilience of your workloads in the face of changes and entropy

• Workloads that remain consistent over long periods reduce the burden on your audience
  – They like it when “Tools” always means the same thing
  – Inconsistency limits your effectiveness

• You time is a scarce resource
  – Lots of fiddling with workloads means less nodes serviced
  – Consistency eases automation efforts
  – Consistency helps you reduce the cycle time between “question asked” and “answer delivered”
Time Stability

• Time stability will become even more important
  – Machine proliferation
  – Operating system proliferation
  – Increased overseas outsourcing
  – Increased use of purchased packages
    • No “in-house” developer expertise
  – Shrinking IS staff spread over more machines
  – You will be responsible for thousands of nodes someday very soon, if not already
    • There is simply no time to fiddle with twitchy workloads
Using Workloads for Analysis

• Graphical methods
  – Workload consumption over time
    • Weeks or months, not just a day
  – Workload consumption versus Business Metric of Interest (BMI)
    • Vital to forecast an modeling accuracy
    • Need to validate consumption versus your Candidate BMIs (CBMIs)

• Automated methods
  – Pathology signatures
    • Friday morning, 8:00 AM “Automating Process and Workload Pathology Detection”
Workload Consumption Over Time

- Great for seeing consumption patterns
  - Look for:
    - When are employees using the system?
    - When do Backups run?
    - Are there any troughs?
    - Are we monitoring ourselves to death?
Workload Consumption Over Time

- Look for:
  - Are there weekday versus weekend differences?
  - Are there “batch-like” functions present?
  - Do some workloads rise when others appear?
    - Database
    - Nightly batch
    - Backups
    - Users
Workload Consumption Over Business Metric

- Great for seeing what consumption follows your CBMI
  - And what doesn’t!
- Look for:
  - Static workloads whose consumption appears as a horizontal line
    - Backups
    - Tools
    - Parts of nightly batch?
Workload Consumption
Over Business Metric

• Look for:
  – CBMI related workloads whose consumption appears as a rising line
    • Database
    • Middleware
    • Interesting if you are planning to grow this system, aren’t they?
Workload Consumption Over Business Metric

- Look for:
  - Workloads that aren’t either static or CBMI
  - Undiscovered CBMIs?
    - Workloads involving human activity
      » Employee logins
    - Are you adding people?
Workload Consumption Over Business Metric

- The three basic patterns:
  - Linear
  - Static
  - None
  - Combinations
    - Most workloads are combinations
Workload Consumption Over Business Metric

• Why care about the three patterns?
  – Validate your CBMIs
  – Improve your workloads
    • Watch out for that y-intercept!
  • The zero CBMI test
  • Increase accuracy, split workloads into:
    – Static
    – Linear
Extra For Accurate Modelers

- The Naïve Growth Trap
  - Let's add 20% more accountants and warehouse transactions
- Naïve growth:
  - Highest point times 120%
  - $79.79\% \times 1.2 = 95.78\%$
- Workload growth %
  - $(95.78\% / 79.79\%) = 20\%$
- Wrong!

![Diagram showing CPU utilization and warehouse transactions with static source of error]
Extra For Accurate Modelers

- Accurate modelers remember that all workloads can have:
  - Linear components
  - Static components
- Accurate growth:
  - Linear component times 120% plus unaltered static component
  - $29.79 \times 1.2 + 50 = 84.58\%$
- Workload growth %
  - $(84.58\% / 79.79\%) = 6\%$
- Correct!
Extra For Accurate Modelers

• The Naïve Growth Trap is very common
  – When you don’t graph versus BMIs
  – When your workloads contain static components
• Precise modeling tools will give the wrong answers if you use incorrect growth estimates
Extra For Accurate Modelers

• Other issues
  – Dealing with inevitable fuzziness
  – Deciding between CBMIs

• Want more detail?
  – Friday morning, 10:30 AM “Business Metrics and Capacity Planning” coauthored with Yiping Ding

![Graph showing warehouse transactions vs. CPU % utilization with markers for samples, naive 20% growth estimate, and proper growth estimate.]
Using Workloads to Solve Problems

- Accountants are unhappy with performance on the warehouse system Tuesday through Friday mornings
  - Management reports late
  - Response times double
- Monday mornings are fine
  - Maybe because of lower weekend warehouse volume?
- Hardware vendor is proposing a $350,000 server as the solution
Using Workloads to Solve Problems

- Examine a day when they are happy (Monday) and a day when they are not (Tuesday)
  - What is different?
- Monday:
  - Nightly batch finishes in an hour
  - Backups run at 3 AM
  - No Batch and Backup contention
Using Workloads to Solve Problems

• Tuesday:
  – Nightly batch runs much longer
  – Backups run much longer
  – Batch and Backups are fighting for the same IO bandwidth
  – The accountants poor morning response is due to contention with nightly batch running late
• There is a big empty trough just before the busy period…
Using Workloads to Solve Problems

- The Tuesday pattern repeats over the next three problem days!
- Why can’t we move the backups to the trough?
  - No more IO contention
  - Key backup data (day’s warehouse transactions) are present after 7 PM
  - Nightly batch will end before the accountants arrive!
- *We don’t need a $350,000 upgrade! We need to move the backup start time!*
What to Look for in Vendor Products

• Ask your vendor how you can subdivide consumption. Favor vendors that can characterize consumption:
  – by username
  – by process name
  – by command line parameters
  – by directory
  – Using regular expressions
  – in AND combinations of the above
  – in OR combinations of the above
What to Look for in Vendor Products

• Ask to see examples of workload characterized CPU consumption over large spans of time
  – Days
  – Weeks
  – Months

• Make sure that workload characterization can be done after the data is collected
  – Otherwise, you are stuck with your first guess
What to Look for in Vendor Products

• Ask to see examples of workload characterized CPU consumption
  – Don’t settle for total CPU
  – On *NIX, don’t settle for “sar” CPU components
    • We want to make business workloads, %SYS and %USR hardly ever helps solve a business issue

• Ask to meet their other customers
  – They are probably sitting next to you!
  – CMG is a great place to do it!
  – Have dinner with someone here that you don’t know…
What to Look for in Vendor Products

• No vendor is perfect, but there are some great ones out there!
  – Compare talent, commitment and technical excellence
  – Great support and continuing technical improvements are expensive
    • Be willing to help pay for it!
What We Wish Vendors Would Provide

• The Über Workload
  – Multi-level groupings of workloads
  – Less detail for general reporting
  – More detail underneath when you drill in
  – Great for post-consolidation nodes

• This OR That
  – Boolean OR relationships could reduce the number of workloads needed to describe certain business functions
What We Wish Vendors Would Provide

- “Generate On Demand” Web Graphics
  - Traditionally we’ve had “do all ahead” graphics
    - Nightly race to make morning reporting
    - Few nodes actually ever examined
    - Huge graph files, large consumption to produce them
  - Modern web technologies like CSS and XHTML make this much more practical
  - Multiple output formats supported
    - Screen
    - Print
    - Handheld
What We Wish Vendors Would Provide

• Workloads That Apply To All Nodes
  – Most products create/apply/store workload information in the context of a single machine or small set of machines
  – Maintaining consistency as workloads evolve is hard for humans
  – Consider workloads that apply to all nodes of a given operating system
    • We use operating system based default workloads on new machines until we get a chance to look at them
What We Wish Vendors Would Provide

- **Workload Color Consistency**
  - All the graphs in this presentation have consistent colors
  - All the workloads your viewers see should too
    - Once they know that brown means backups, the next time they see brown, they already know what it is
  - Colors should be assigned at workload creation
  - If an application spans many nodes
    - If the same workload on each node is the same color, meetings go well
    - If not, try to explain why the pink workload here is the green workload there and the puce workload there…
What We Wish Vendors Would Provide

• Use more colors!
  – The “web safe” pallet of 256 colors does not allow uniqueness in environments of thousands of nodes
  – Modern browsers and machines support millions of distinct colors, use them!
  – Calculate color pallets to account for common forms of color blindness
    • Many men are at least partially color blind
  – The limitations of the 1990s should no longer restrict us today
What We Wish Vendors Would Provide

• Make it easier!
  – Automate
  – Make libraries of common application workloads (and hopefully BMIs)
    • 85% of WINDOWS machine workloads are always the same
  – If BMIs are available, how about pre-calculating accurate growth percentages for us?
  – Add more graphics to the modeling tools
  – Automate precedence order choices via rule sets
Summary

Workload characterizations are incredibly powerful ways to increase the information quality available to decision makers. Whether you use a commercial product, or you program a solution yourself, you owe it to your audience to provide the power of significant, consistent, and business relevant characterized consumption views.

We look forward to large-scale innovation and improvements in workload characterization and reporting technologies in this decade.
Questions?

Have a Great CMG!