

Module 3: Performance Analysis

Upon completion of this module, you should be able to:

- Break a workload down into its basic characteristics
- Use a performance roadmap to focus your analysis on the key issues
- Identify Little's Law and its implications for predicting future performance capabilities



This module focuses on the overall methodology for performance analysis.

Module 3: Performance Analysis

Lesson 1: Workload Profiles and Characterization

During this lesson the following topics are covered:

- Characterizing a workload by time, object, and measure
- Using characterization to become familiar with expected workload performance



This lesson covers concepts of workload profiles and workload characterization.

Knowing Your Workload Helps Answer Analysis Questions

- Large I/Os can overwhelm a Fibre port

- ▶ What size I/O is the workload averaging?
- ▶ What speed are my ports?

- A write heavy workload could be filling the cache with Write Pendings

- ▶ How much writing are my applications doing?

- Response time temporarily grew overnight

- ▶ Does some nightly job change the workload parameters?

- I'd like to start doing something about the heavily used disks

- ▶ Which disks are consistently heavily utilized?

EMC²

Copyright © 2014 EMC Corporation. All Rights Reserved.

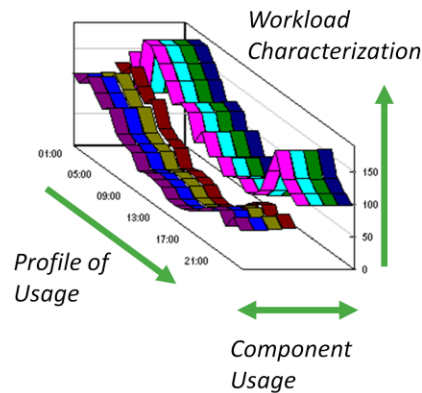
Module 3: Performance Analysis 3

Workload characterization is simply knowing how your workload typically operates. Detailed knowledge of how the I/O traffic affects the data components is essential in answering the questions that come up during analysis.

Characterization might be as simple as looking at a series of daily performance archives to spot common trends. A quick overview will show what components are the most used, what time of day patterns are typical, and the basic composition (I/O size, read/write %, etc.) of the traffic. After this simple overview, you will be able to answer the sample questions shown in the slide, and many others.

Components of Workload Characterization

- Profiles of Usage (When do things happen?)
 - ▶ Peak times of day, week, etc.
 - ▶ Overall traffic levels
- Component Usage (What gets affected?)
 - ▶ How traffic is handled by the storage components
 - ▶ Balances (and imbalances)
- Workload Characteristics (What happens?)
 - ▶ I/O sizes, read vs. write, etc.



EMC²

Copyright © 2014 EMC Corporation. All Rights Reserved.

Module 3: Performance Analysis 4

Workload analysis can be broken into three areas:

1) Overall Usage Profiles:

- When is the heavy use occurring?
- Are there daily, weekly, monthly or quarterly patterns?
- What is the overall level of traffic on the machine?

2) Component Usage:

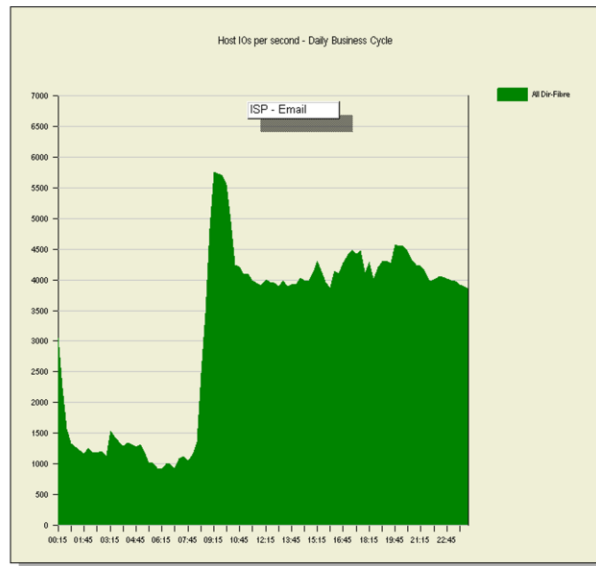
- What is the back-end usage vs. the front-end usage?
- How evenly is the workload spread across the disks, ports, and volumes?
- How is the cache being used?

3) Workload Characteristics:

- What is the mix of I/O sizes?
- How much activity is read vs. write?
- Are the reads sequential or random?
- How many reads are occurring from cache?

By answering these questions, analysts can gain insight into how their system can be improved through tuning or future planning. By understanding their workload, problems can be more easily diagnosed and quickly corrected. Better yet, when combined with a proactive approach, problems can be avoided all together.

Email Application Profile of Usage (Email Traffic)



EMC²

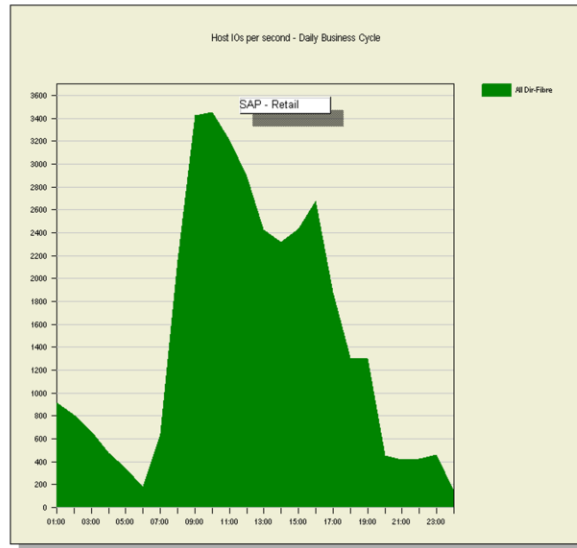
Copyright © 2014 EMC Corporation. All Rights Reserved.

Module 3: Performance Analysis 5

Profiles of Usage defines how the workload changes over days, weeks, or months. Knowing the regular patterns of the workload is important in defining the times of heaviest utilization. Deviations from the known pattern alert an analyst to unexpected changes in the workload.

This graph shows I/Os per second from an email application. You can see a lull in activity in the early morning and then a large ramp-up in activity between 8 – 9 AM (as people arrive and check their messages). The peak is around 9:15, and activity drops to a steady level at 10:30 AM, where it stays the rest of the day, until it falls again at about 10:30 PM.

Retail Application Profiles of Usage (Sales Transactions)



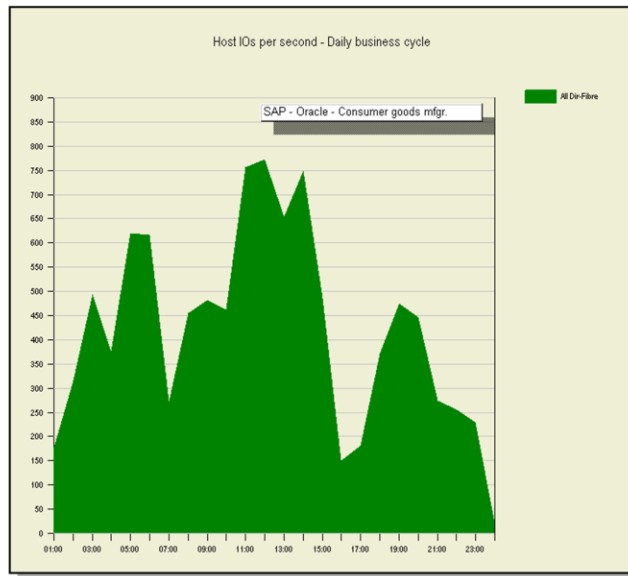
EMC²

Copyright © 2014 EMC Corporation. All Rights Reserved.

Module 3: Performance Analysis 6

The second graph shows I/Os per second for a retail application. In this environment, there is a very large ramp at 7 AM (retailer starts business for the day). There is a reduction in activity after noon, and a sharp ramp-down at 5 PM (close of business). There are a few “residual jobs,” but essentially very little activity at night.

Profiles of Usage (Process Reports)



EMC²

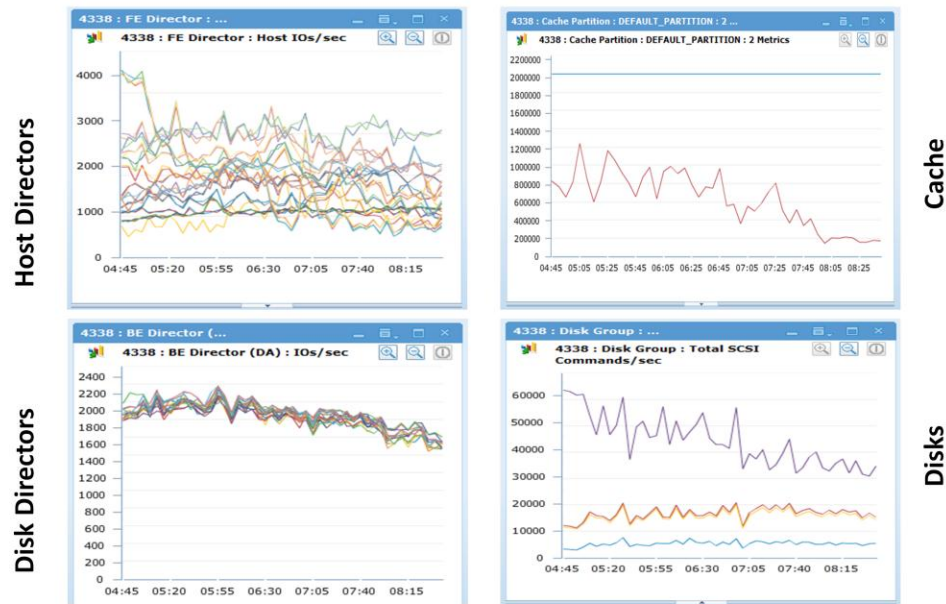
Copyright © 2014 EMC Corporation. All Rights Reserved.

Module 3: Performance Analysis 7

The third graph shows a manufacturing operation. You can clearly see the 3-shift nature of the activity based on the peaks and valleys. The main shift is from approximately 7 AM to 3 PM, the second shift has activity between 5 PM and 11 PM, and there is early morning shift activity from approximately 1 AM to 6:30 AM.

Some of these environments experience dramatic changes in I/O patterns at different times of the day. An analyst can take this into account.

Component Usage



EMC²

Copyright © 2014 EMC Corporation. All Rights Reserved.

Module 3: Performance Analysis 8

The second part of workload characterization is identifying patterns of component usage. This slide shows the four primary component types within a Symmetrix: host directors, system cache, disk directors, and the disks.

Component usage results expose design and configuration flaws by showing imbalances and poor use of directors, cache, and devices. It also shows where insufficient resources could result in poor performance by exposing places where resources are running at or near their technical limits.

Note that all of the host directors are paired. Two components should always be sharing a workload. Some of the host directors are hardly being used. One of the disks is getting about 50% more traffic than any of the others.

Workload Characteristics



EMC²

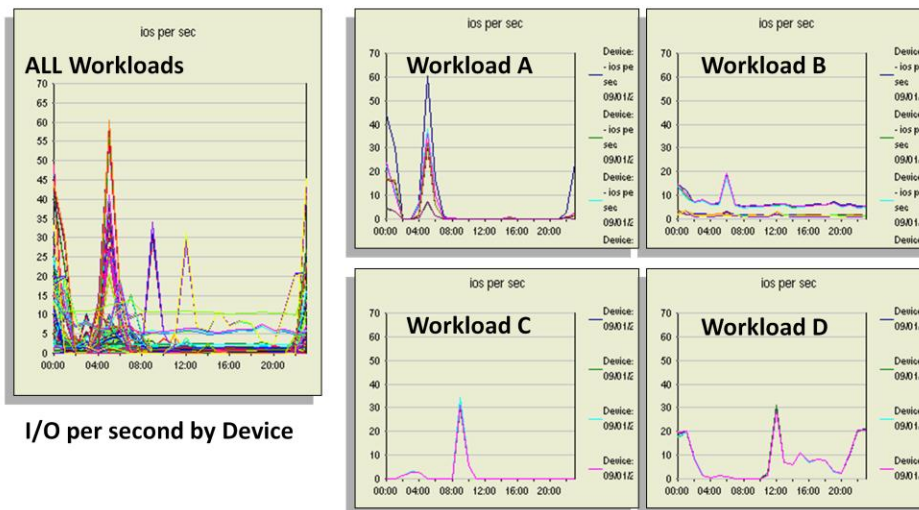
Copyright © 2014 EMC Corporation. All Rights Reserved.

Module 3: Performance Analysis 9

The third piece of workload characterization is the identification of workload characteristics. This is probably the most important area of analysis in regards to performance based capacity planning and component configuration. Discovering application I/O rates and block sizes, as well as read/write and hit/miss ratios, provides you with the data required to determine configuration and layout plans for everything from front and back end directors and number and size of logical and physical devices, to required data channels and HBA queue depth settings.

If too many logical volumes are to be considered, it might be better to perform this study at the Symmetrix system level instead.

Isolating Workloads



EMC²

Copyright © 2014 EMC Corporation. All Rights Reserved.

Module 3: Performance Analysis 10

Identifying devices that have workloads in common, such as databases and their corresponding logs, makes it easier to understand the sometimes confusing amount of information that Performance Viewer can produce. Identifying common devices also allows a storage administrator to fine tune the configuration in order to benefit specific applications without negatively affecting other co-existing workloads. Isolating these common resources, of course, requires a bit of investigation in order to identify volumes/devices that work together. However, once these common resources are identified, you can use Performance Viewer to build views of important metrics that assist you in characterizing the workload generated by the source application.

By examining the graphs on this slide, you can see that it is easiest to isolate common resources by identifying groups of lines that follow a common pattern. These groups of resources represent workload from what is most likely the same host, if not the same application.

Examples of Characterization

	Line of Business	Hosts/OS	Compute Style	Environment or Application(s)	Database(s)
1	Web		web	video streaming	
2	Car Rental	Sequent	OLTP		Oracle
3	Retail pharmacy	Open Systems			
4	Web		web	home sites	
5	Electronic				
6	Auto Mfg				
7	Software Vendor	Unix	OLTP	Oracle Financials	Oracle
8	Banking	SUN E10K/Solaris	DW		Oracle
9	Banking - bill paying	HP/UX/SUN/AIX	OLTP + DW		Oracle/UDB
10	Bank				
11	Bank	Windows	e-mail	Lotus Notes	
12	Discrete Mfgr.	Unix	OLTP	ERP/MRP	Oracle
13	Banking - bill paying	MF Sysplex	OLTP		DB2
14	Systems Vendor	Open Systems	e-mail, file server	MS Exchange	
15	Software Vendor	Unix	OLTP	multiple applications	Oracle
16	Retail/Supermarket	Open Systems	DW	SAP	Oracle
17	Chemical		DW		
18	Electricity				
19	Auto Mfg				
20	Electronic mfgr.	Unix	DW	SAP	Oracle
21	Telecom				

EMC²

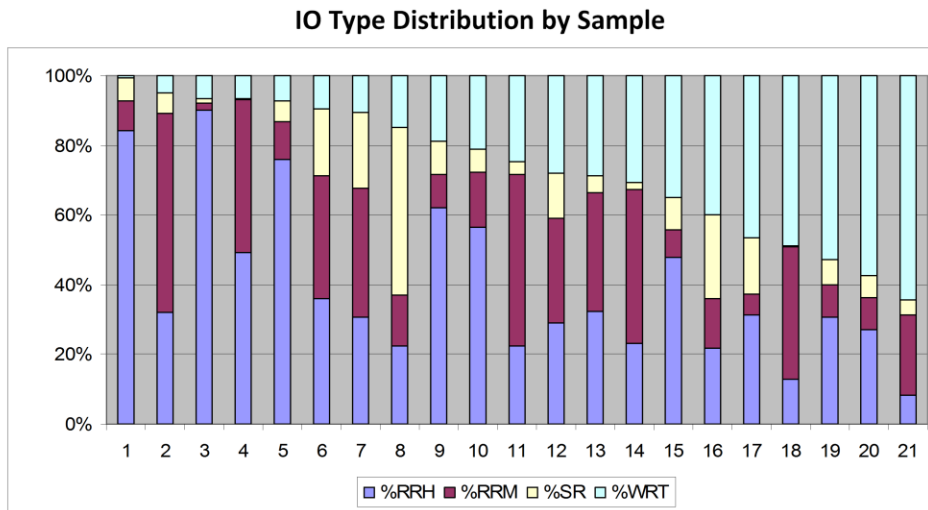
Copyright © 2014 EMC Corporation. All Rights Reserved.

Module 3: Performance Analysis 11

In situations where “live” data is not available, you can use “best guess” estimates in order to get a ballpark idea of what required resources will be necessary to support a planned environment. EMC engineering looked at a set of systems from 21 different businesses to see how their overall system profiles vary. The chart above provides business use, platform, and application information that can be used as a template for other similar configurations. The following slides will provide workload statistics based on the list above.

Note: Any use of these numbers is intended to help estimate the expected load on a configuration. As stated above, they allow you to make a “best guess” evaluation only. The actual patterns created by the projected application may vary. Use of these numbers should only be considered when actual data is not available.

Example Workload Characteristics



EMC²

Copyright © 2014 EMC Corporation. All Rights Reserved.

Module 3: Performance Analysis 12

By matching up the numbers shown here from those on the previous chart, we can begin to characterize the workload of these applications based on percentages of I/O types. This chart shows the percentage of overall application I/O based on percentages of random read hits, random read misses, sequential reads, and writes.

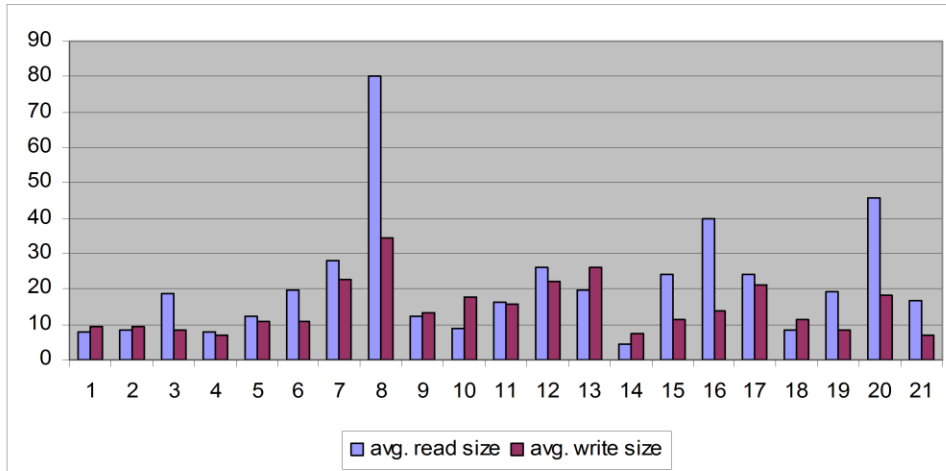
So how could I use this?

Let's say you are planning on installing an Exchange server into your environment and you do not know what the I/O patterns of that system will be. By looking at the table on the previous slide, you can see that the example Exchange server is #14. Looking at the chart on this page, we can examine the workload pattern of the sample Exchange system. Here, we see that the workload breaks down as follows:

Random Read Hit	22%
Random Read Miss	46%
Sequential read	2%
Writes	30%

Example Workload Characteristics

Average IO Sizes Distribution by Sample



EMC²

Copyright © 2014 EMC Corporation. All Rights Reserved.

Module 3: Performance Analysis 13

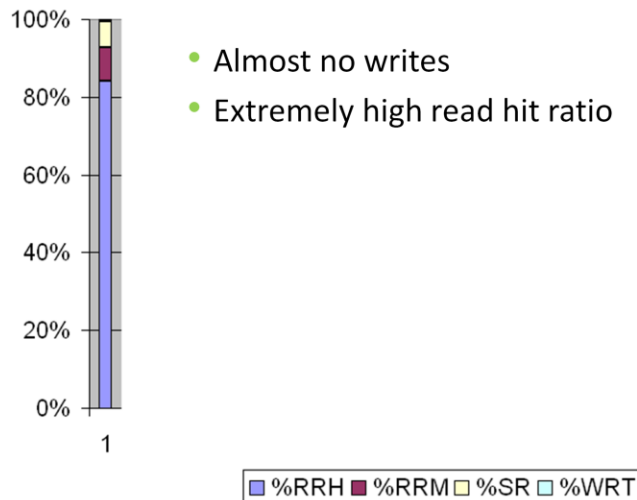
This chart allows you to take the next step of characterization, which is predicting workload based on expected I/O size. As we will see later, I/O size has a large impact on overall channel throughput and system efficiency, therefore, this type of characterization is crucial. When we look at this chart we find, as anticipated, that the larger I/O sizes come from data warehouse applications, whereas smaller I/O sizes are found in applications, like corporate e-mail systems. Again, using the Exchange example, we can see that the sample I/O size for a similar system would be as follows:

Average read size 4k

Average write size 7k

Note: These numbers are averages based on a given time window. A average write size of 7k does not mean that all write I/O will be 7k. Some I/O could be 4K, 8k, 12k, etc., however, the average over time is 7k.

Video Streaming Profile



- Almost no writes
- Extremely high read hit ratio

EMC²

Copyright © 2014 EMC Corporation. All Rights Reserved.

Module 3: Performance Analysis 14

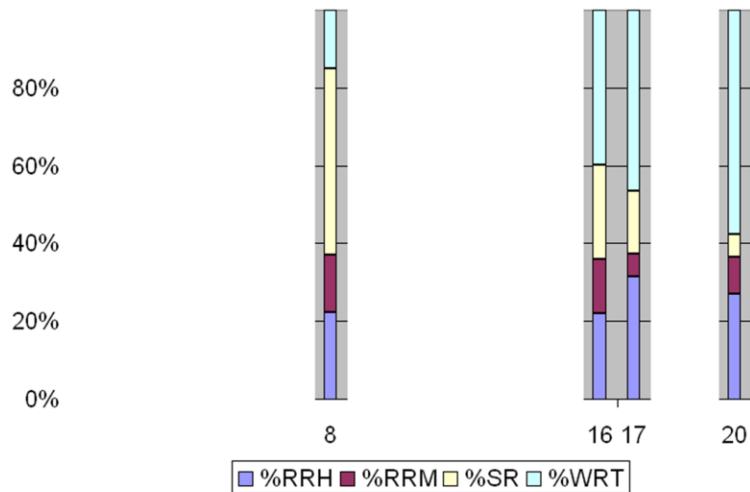
One of the most extreme profiles is the “web video streaming” profile. There are almost no writes, meaning the data on the server does not change often. Perhaps no new videos were written to the disks during the study period.

Note the extremely high random read hit ratio: around 85%. This is very good, but if the prefetch operations are working properly, this is perfectly expected for this type of data.

You would normally expect a very high sequential read measure for this type of media server, but the graph shows a value of only about 6%. If Performance Viewer was used to collect the data, then the analyst probably based the “%SR” value on the “seq reads per sec” metric. Unfortunately, this metric is not defined using the common definition of “sequential,” it is actually measuring the events that trigger a new prefetch task. For very sequential workloads, like video streaming, you would expect this measure to be quite low, since a single extended task might be all that is needed to prefetch many megabytes of data. This topic will be discussed in detail later in the course.

Data Warehouse Profiles

- Low hit rate
- Varying write rate



EMC²

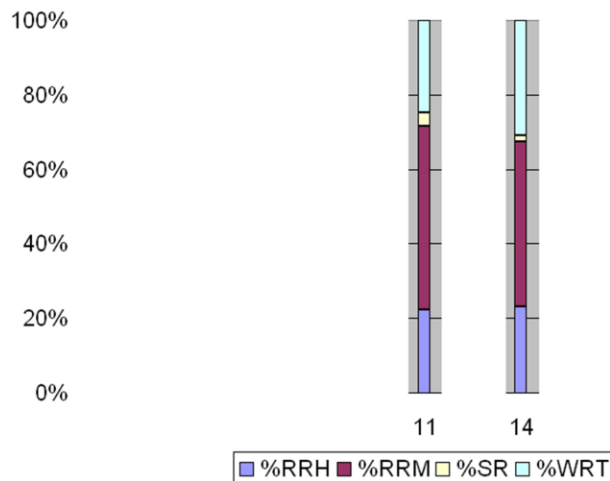
Copyright © 2014 EMC Corporation. All Rights Reserved.

Module 3: Performance Analysis 15

The four data warehouse profiles are shown above. They have a low hit rate in common, and a widely varying write rate.

Email Profiles

- Low hit ratio
- Little sequence



EMC²

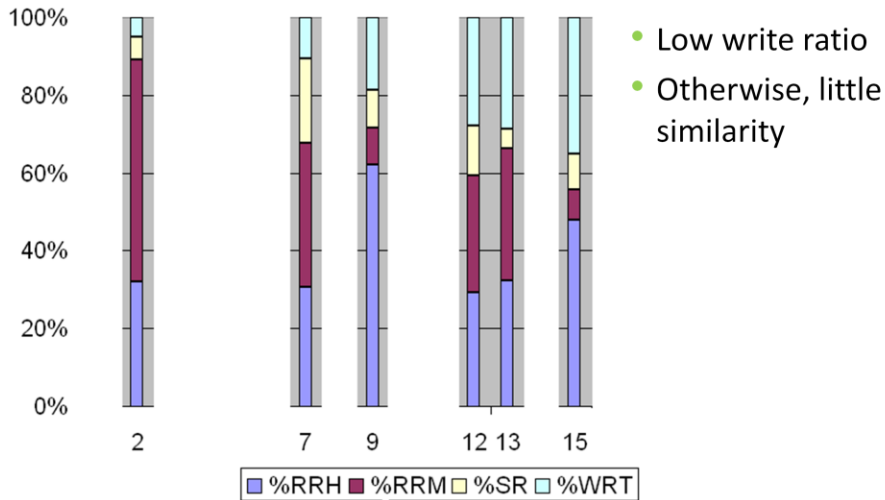
Copyright © 2014 EMC Corporation. All Rights Reserved.

Module 3: Performance Analysis 16

The two email profiles are shown here, with the Lotus Notes application (11) on the left and the Microsoft Exchange application (14) on the right.

The low read hit ratio is common for email servers. Their I/O patterns are often very random and are not typically long enough to take advantage of prefetching. A read hit percentage of slightly better than 20% is perfectly acceptable under these circumstances.

OLTP Profiles



EMC²

Copyright © 2014 EMC Corporation. All Rights Reserved.

Module 3: Performance Analysis 17

The OLTP profiles have a low write ratio in common, and little else. The type of transactions processed will have a large bearing on the workload characteristics.

Are the client applications primarily querying data, or are they making changes as well? This would have an impact on the write ratio.

Do the queries tend to be repetitive? Is the same record accessed frequently over a short time period? This would have an impact on the hit ratio.

Benchmark Workload Distribution

OLTP1 - Medium Workload Database (Exchange)			
	Block Size	Alignment	% of Total
Random Read Hit	4k	4k	40
Random Read Miss	4k	4k	24
Random Write	4k	4k	16
Sequential Read	4k	4k	10
Sequential Write	4k	4k	10

OLTP2 - Heavier Workload Database with larger block size (Oracle)			
	Block Size	Alignment	% of Total
Random Read Hit	8k	8k	20
Random Read Miss	8k	8k	45
Random Write	8k	8k	15
Sequential Read	16k	32k	10
Sequential Write	16k	32k	10

EMC²

Copyright © 2014 EMC Corporation. All Rights Reserved.

Module 3: Performance Analysis 18

When EMC engineering and other groups conduct performance testing on arrays, they use tools that simulate real-world applications. These simulation tools are used in controlled environments and in a consistent manner to ensure the most accurate representation possible of a system's performance capabilities, regardless of array type. The simulation tools create a workload based on industry standard workload predictions based on varying types of On-Line Transaction Processing (OLTP) databases and Decision Support Systems (DSS). Decision support includes backups, migrations, and other bulk transfer operations.

This slide shows the values that we use to generate workload based on the type of OLTP database. OLTP1 represents medium workload databases that tend to produce smaller block 4k I/O. An example of this type of database would be Microsoft Exchange. OLTP2 represents heavier workload databases that tend to use larger block sizes. An example of this type of system would be Oracle.

Comparing the benchmark workload distribution to the workload characteristics of your environment helps you identify which benchmark most closely matches your application's activities. Once you know that, you can monitor the benchmark reports that EMC (or some other manufacturer) makes about a new product. For instance, a company might announce a new product that can perform up to 5000 I/O per second under the OLTP2 benchmark. If your environment matches the OLTP2 characteristics and can currently do only 3000 I/O per second, you will know that the new product can process requests considerably faster.

Benchmark Workload Distribution (Continued)

OLTP3 - Database characterized by 2k block sizes (Sybase)			
	Block Size	Alignment	% of Total
Random Read Hit	2k	2k	20
Random Read Miss	2k	2k	40
Random Write	2k	2k	20
Sequential Read	16k	16k	10
Sequential Write	4k	4k	10



OLTP3 represents medium workload databases that tend to produce very small block 2k I/O. An example of this type of database would be Sybase.

Benchmark Workload Distribution (Continued)

DSS1 - Medium load Decision Support Application			
	Block Size	Alignment	% of Total
Random Read Hit	32k	32k	18
Random Read Miss	32k	32k	18
Random Write	32k	32k	4
Sequential Read	32k	32k	48
Sequential Write	32k	32k	12

DSS2 - Heavy load Decision Support Application			
	Block Size	Alignment	% of Total
Random Read Hit	4k	4k	0
Random Read Miss	4k	4k	15
Random Write	4k	4k	5
Sequential Read	64k	64k	70
Sequential Write	64k	64k	10

EMC²

Copyright © 2014 EMC Corporation. All Rights Reserved.

Module 3: Performance Analysis 20

Decision support systems (backup, migration) tend to be far more efficient with their use of storage resources due to the fact that their I/O sizes tend to be significantly larger. This is much more efficient in environments that perform a higher percentage of sequential tasks.

DSS1 represents medium sized decision support environments.

DSS2 represents heavy load decision support environments.

Module 3: Performance Analysis

Lesson 1: Summary

During this lesson the following topics were covered:

- Characterizing a workload by time, object, and measure
- Using characterization to become familiar with expected workload performance



This lesson covered concepts of workload profiles and workload characterization.

Module 3: Performance Analysis

Lesson 2: Performance Analysis Roadmap

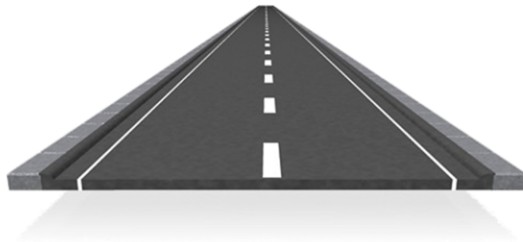
During this lesson the following topics are covered:

- Developing a roadmap of key metrics



This lesson covers developing a roadmap for performance analysis.

Roadmap



- **Roadmap:** Sequence of metrics to analyze
- Most important metrics to the system are represented
- Check the roadmap on a regular basis to get a feel for the system “normal” state
- Check all of the roadmap before drilling down into a problem to avoid getting sidetracked on one symptom of a larger problem

EMC²

Copyright © 2014 EMC Corporation. All Rights Reserved.

Module 3: Performance Analysis 23

Analysts use roadmaps of one kind or another for nearly all performance analysis. A roadmap is simply a list of the key metrics used to diagnose performance problems. Each type of Enterprise system has a different roadmap, and different workload environments might have slightly different roadmaps. Some analysts will examine the metric data in a top down fashion, while others will prioritize the analysis so that the most critical issues are seen first.

Check the roadmap on a regular basis to get a feel for the “normal” state of the system. This will make it easy to identify abnormal situations or rapid growth.

A good roadmap presents a broad view of the system. Check the entire roadmap before drilling down into a specific problem. Viewing the whole system first helps identify the most severe problems, instead of just the ones that are spotted first.

Roadmap Benefits

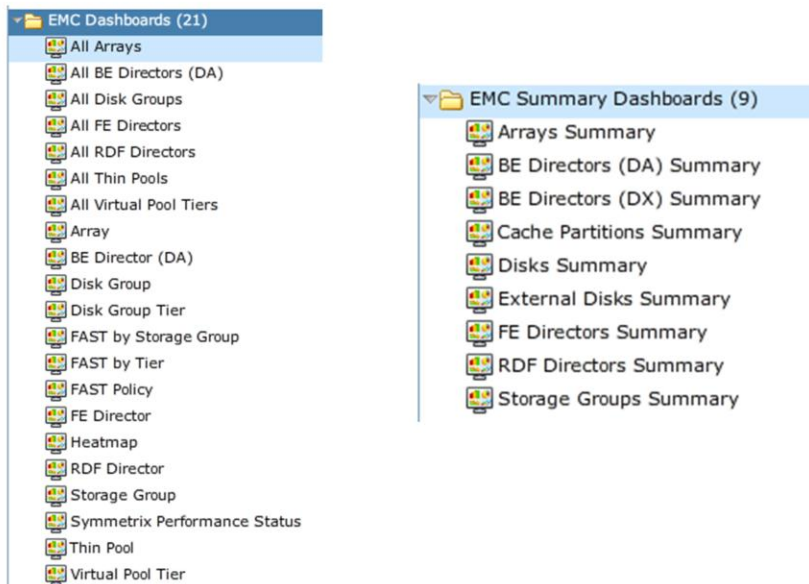
- Condenses hundreds of metrics into just a few to make analysis manageable
- Capacity planning: compare past Roadmap metrics with current to determine amount of increase per week/month
- Combine what you learn from the charts with an understanding of the applications, hosts, and the environment to resolve any problems or to understand system requirements
- This analysis does not give you the answers directly, rather it provides clues, insight, and a good place to start



Roadmaps are useful because they condense a large number of metrics into just the most important few. This is an important consideration in the large environments that analysts are typically asked to manage. A single Symmetrix might consist of thousands of objects in Performance Viewer, each of which might have 10-20 performance measures. A roadmap helps focus on just the important measures.

Roadmaps are also useful in providing a baseline of data for capacity planning. Since the roadmap measures are the most representative of the environment, they will show the effects of increased traffic load. Comparing the same few measures over the course of weeks or months will show the rate of growth.

Performance Viewer Dashboards



EMC²

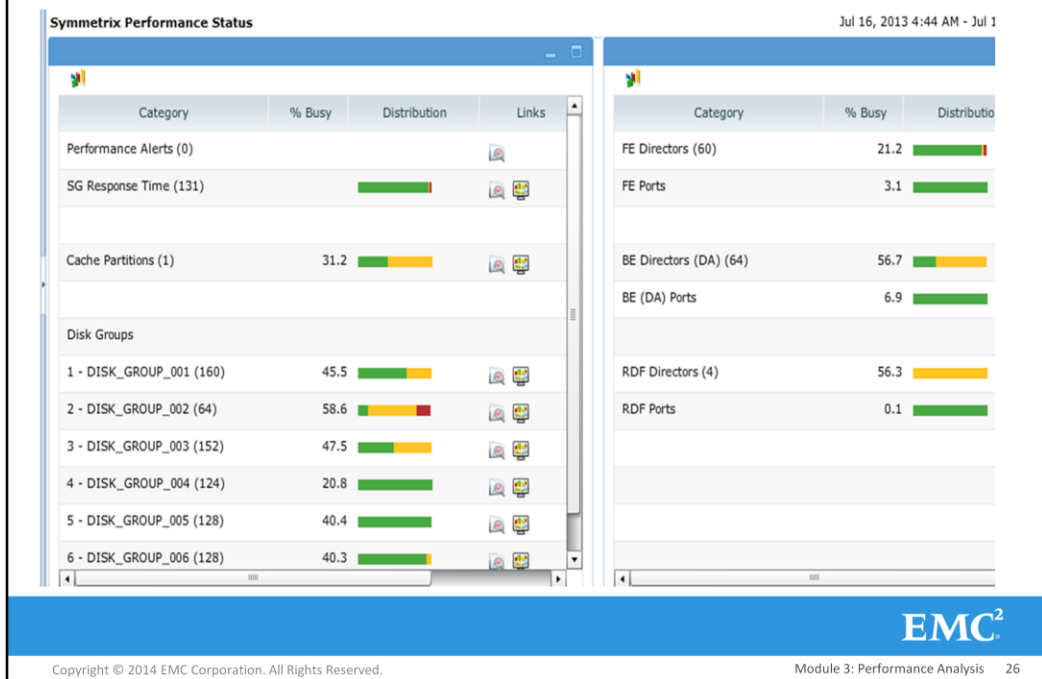
Copyright © 2014 EMC Corporation. All Rights Reserved.

Module 3: Performance Analysis 25

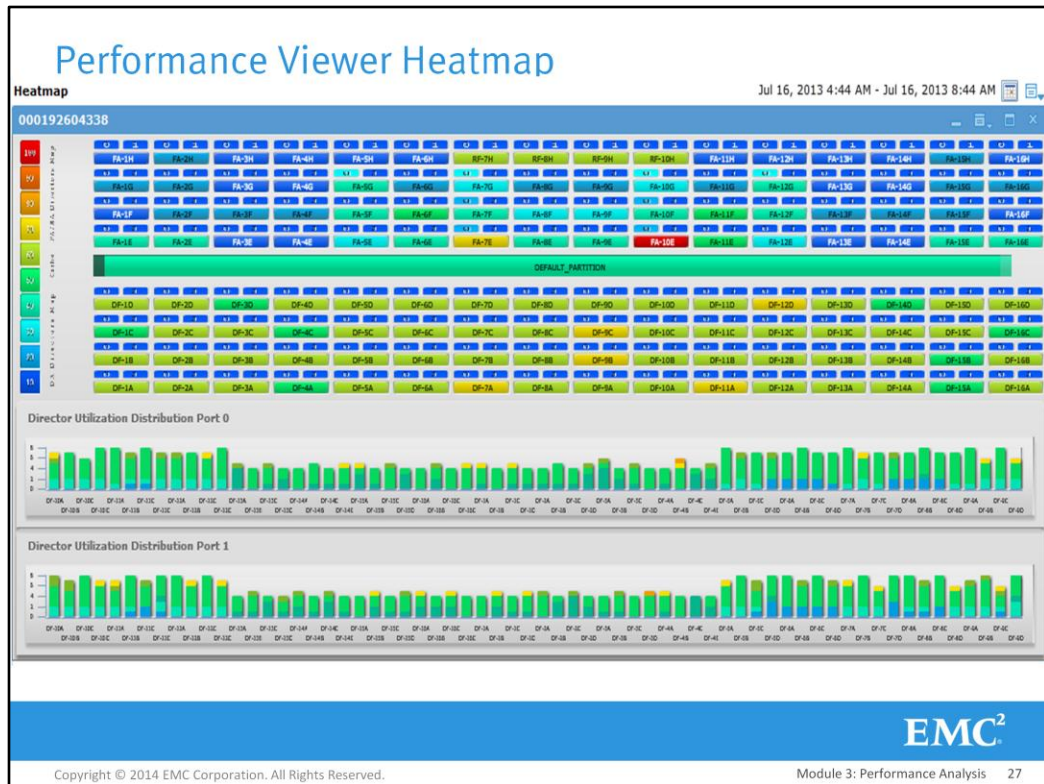
This slide displays the Performance Viewer Dashboard lists. The Dashboards are a great resource for getting the high level picture of what is going on in the systems. You can create your own Dashboard that reflects your own personal roadmap.

The ordering of the charts and graphs presented in this class follow a logical “front-to-back” sequence. When doing actual analysis, you might still view the measures in that order, but over time, your sequence might change. For instance, if you repeatedly find that Disk utilization is your most common problem, it would be natural for you to look at those related measures first, and then check the rest of the array’s indicators.

Symmetrix Performance Status



This slide displays the Symmetrix Performance Status Dashboard. This gives a quick overview of the whole Symmetrix with color coded bar graphs based upon the metrics you have defined. View these as individual Heatmaps for components. From here, you can go to more specific summary pages or navigate to analyze the particular components by clicking on the icons in the Links columns.



Displayed above is the Performance Viewer Heatmap Dashboard. In addition to the Performance Status Dashboard, the Heatmap might be a good place to start analysis of systems. The Heatmap will show, by color coding, any hotspots. The top to bottom display correlates to the front-end, the cache, and the backend of the Symmetrix.

All Arrays Dashboard



Similar in concept is the All Arrays Dashboard, showing key metrics for all of the registered Symmetrix on this Unisphere server.

Performance Viewer Roadmap

- Front-end
 - ▶ Monitor > EMC Dashboards > All FE Directors > Hosts IOs/sec and % Busy
 - ▶ Monitor > EMC Summary Dashboards > FE Directors Summary
 - ▶ Analyze>Diagnostic>Director > Host I/Os / Sec> Trend Line
- System/Cache
 - ▶ Monitor>EMC Summary Dashboards>Cache Partitions Summary
 - ▶ Analyze>Diagnostic>Cache Partition>Single Dynamic Cache Partition>% Hit
 - ▶ Analyze>Diagnostic>Cache >% Hit>Trend Line
 - ▶ Analyze > Diagnostic> Cache Partition
 - ▶ Analyze>Cache Partition>WP Count>Trend Line
 - ▶ Analyze Diagnostic> Cache Partition> Meta(s)> WP Count
- Backend
 - ▶ EMC Dashboards>ALL BE Directors
 - ▶ EMC Dashboards>BE Directors
 - ▶ EMC Summary Dashboards>BE Directors Summary
 - ▶ Analyze>Diagnostic>DA> Ios/Sec >Trend Line
 - ▶ EMC Summary Dashboards>Disks Summary
 - ▶ EMC Dashboards>All Disk Groups
 - ▶ EMC Dashboards>Disk Groups
 - ▶ Analyze>Diagnostic>Disk Group
 - ▶ EMC Dashboards>All Disk Groups
 - ▶ Analyze>Diagnostic>Disk(s)>Reads/sec >Writes/sec
 - ▶ Analyze>Diagnostics>BE Director(DA)>DA(s)>% Busy

EMC²

Copyright © 2014 EMC Corporation. All Rights Reserved.

Module 3: Performance Analysis 29

Displayed above is the Performance Viewer Roadmap that will be discussed in the next few modules of this course.

The details about why each of these measures is included, and exactly what to look for in each measure, are given in those modules. It is presented here, just to give a feel for the checklist nature of a roadmap.

The ordering of the graphs presented in this class follows a logical “front-to-back” sequence. When doing actual analysis, you might still view the measures in that order, but over time, your sequence might change. For instance, if you repeatedly find that disk utilization is your most common problem, it would be natural for you to look at the related measures first, and then check the rest of the array’s indicators.

Module 3: Performance Analysis

Lesson 2: Summary

During this lesson the following topics were covered:

- Developing a roadmap of key metrics



This lesson covered developing a roadmap for performance analysis.

Module 3: Performance Analysis

Lesson 3: Little's Law and its Impact on Response Time

During this lesson the following topics are covered:

- Using Little's Law to describe the non-linear relationship of utilization and response time



This lesson covers the concept of Little's Law and its impact on response time.

Non-Linearity Misconceptions

Note: It is a misconception to assume that a linear increase in traffic load will produce a linear change in performance measures.

- **Current workload:** 3000 I/O per second, 10ms avg. response time.
 - ▶ If we get a temporary spike to 6000 I/O per second, will response time go to 20ms?

- I/O traffic is growing at a rate of 10% per month
- We can tolerate no more than a 50% increase in response time
 - ▶ Will the system stay within bounds for another 5 months?

- Our 1Gb FA ports are about 50% utilized with a throughput of about 0.5 Gb per second
 - ▶ Can we increase to 80% without jeopardizing our applications?

EMC²

Copyright © 2014 EMC Corporation. All Rights Reserved.

Module 3: Performance Analysis 32

One of the difficulties in performance analysis and capacity management is the non-linear nature of data systems. A linear increase in traffic load will typically not produce a linear change in performance measures, such as response time. This makes it difficult to accurately answer questions like the ones asked in this slide. The obvious answer (yes) would be true only if every component scaled linearly. However, in most cases, “no” is the correct answer to these questions (or more appropriately, “it depends”).

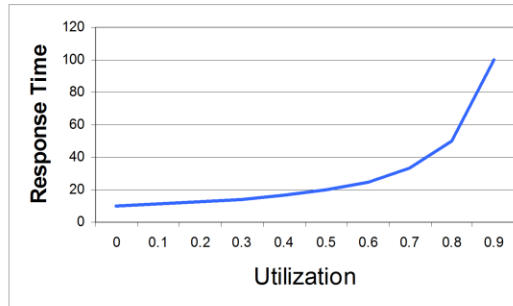
The examination of Little’s Law and the benchmark examples that appear over the next pages will help explain the non-linear characteristic of performance.

Little's Law Definition

- Little's Law predicts non-linear increases in response time with increasing utilization of shared access queuing systems
- Useful to illustrate relationship and to make rough estimates

$$\text{Response} = \frac{\text{Service}}{1 - \text{Utilization}}$$

- **Service:** Time to service a single work unit
- **Utilization:** Percentage of time spent servicing work units (0.0 to 1.0)
- **Response:** Actual time to initiate, service, and return a single work unit



EMC²

Copyright © 2014 EMC Corporation. All Rights Reserved.

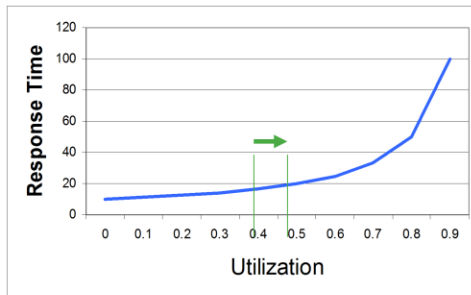
Module 3: Performance Analysis 33

Little's Law provides a theoretical formula for a well known performance relationship seen in queuing systems (take inputs and produce outputs). As shown in the chart above, the law predicts that response time for a queuing system will increase non-linearly as utilization increases. Furthermore, the law states that 100% utilization cannot be achieved without experiencing infinite response time.

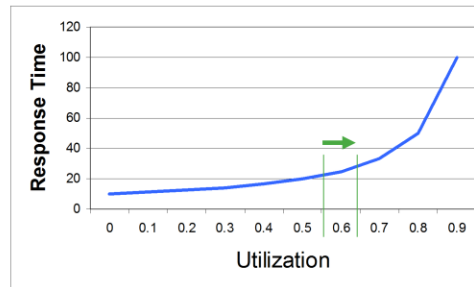
This formula cannot be used for exact predictions in the complex environments studied in this class. However, Little's Law is very useful for rough estimates and as a general illustration of the relationship of response time and utilization.

Implications of Little's Law

- A small increase in traffic may cause a small or big increase in response time, depending on current utilization
- Each element in the pathway might be experiencing a different utilization, making predictions even more difficult



Increase in utilization from 0.3 to 0.4 (+10%)
Increase in response time from 14 to 16 (+2)



Increase in utilization from 0.7 to 0.8 (+10%)
Increase in response time from 33 to 50 (+17)

EMC²

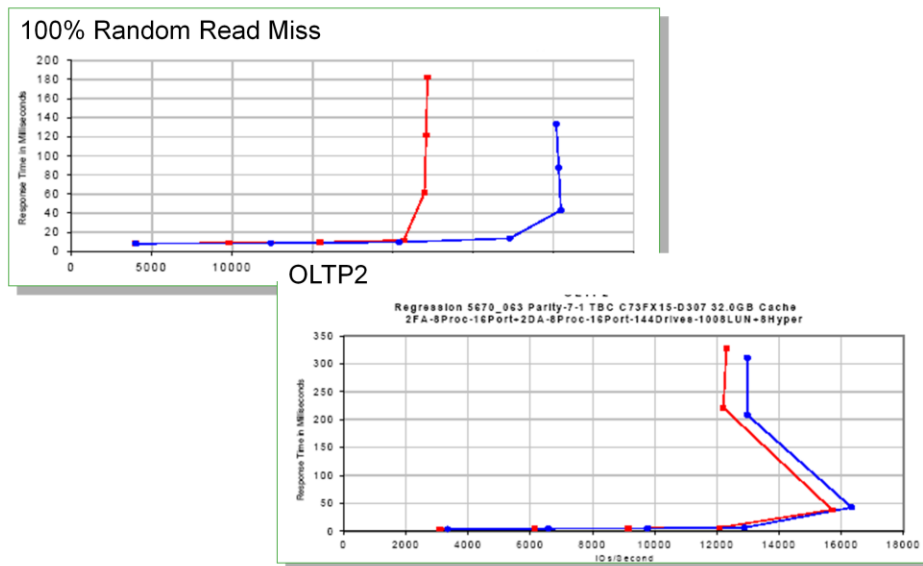
Copyright © 2014 EMC Corporation. All Rights Reserved.

Module 3: Performance Analysis 34

An important implication of Little's Law is that a fixed increase in traffic will not always cause a fixed increase in response time. Both of the charts shown on this slide illustrate an increase in utilization of 10% (to keep the math simple, let's assume this was a result of a 10% increase in traffic). The two charts clearly show that the utilization increase has a much greater effect on the response time when utilization is already high. Keeping utilization low while still performing needed data operations will be a repeated theme in this course.

Little's Law applies to all elements in a data path: host software, connectivity channels and end points, caching systems, and disk drives all respond to increases in utilization in this way. Predicting response time changes, as a result of traffic increases, is an extremely difficult task in light of this complexity.

Benchmark Examples of Little's Law Curve



EMC²

Copyright © 2014 EMC Corporation. All Rights Reserved.

Module 3: Performance Analysis 35

The two benchmark examples on this slide show benchmark results on a pair of Symmetrix arrays (one a slightly older model). The first shows the overall array performance in a Random Read Miss Test. All (or nearly all) reads are randomly addressed and calculated to produce misses. Under this unrealistic circumstance, the Symmetrix cache is not really a factor. All the reads must be retrieved from physical disks and then delivered to the host. The benchmark curves show the characteristic Little's Law shape: response time rises only gradually until the "knee" of the curve, and then shoots dramatically up. Little's Law and this benchmark agree: more I/O can be easily added only until a certain point.

The second graph shows a slightly different type of curve that is common with Symmetrix benchmarks. This OLTP2 (Oracle-like) test is more realistic and makes better use of the Symmetrix cache. This is evident from the "spike" at the knee of the curve. As I/O traffic increases, the cache absorbs some of the overflow temporarily. But, all caching systems have their limit. As the I/O increases further, that limit is reached and the response time, again, rises dramatically.

Module 3: Performance Analysis

Lesson 3: Summary

During this lesson the following topics were covered:

- Using Little's Law to describe the non-linear relationship of utilization and response time



This lesson covered the concept of Little's Law and its impact on response time.

Module 3: Summary

Key points covered in this module:

- Breaking a workload down into its basic characteristics
- Using a performance roadmap to focus your analysis on the key issues
- Identifying Little's Law and its implications for predicting future performance capabilities



This module covered the overall methodology for performance analysis.

This slide is intentionally left blank.



This slide is intentionally left blank.