How to use
the hidden RMF treasures
to ensure storage availability

John Ticic

Session 15942

#SHAREorg

SHARE is an independent volunteer-run information technology association that provides **education**, professional **networking** and industry **influence**.

Agenda

- Introduction

- RMF's role in the Enterprise
  - Historical introduction
  - What does RMF provide us with?

- Mine the data
  - Use an example of a single RMF record to explore how we can mine RMF data to extract very useful information

  - The example is disk replication

About IntelliMagic

- A world leader in Storage Performance Management software solutions

- Developing SPM solutions since 1991

- Private, financially independent

- Headquarters in Leiden, NL

- German office in Munich

- US office in Dallas, TX

IntelliMagic Products & Services

IntelliMagic Product Portfolio:

- *IntelliMagic Vision:*
  Disk Performance, Early Warning and Analysis
- *IntelliMagic Vision for z/OS Tape*
  Performance and Analysis for real and virtual tapes
- *IntelliMagic Balance:*
  Improves Storage Hardware Efficiency
- *IntelliMagic Direction:*
  Disk subsystem Predictive Performance Modeling

IntelliMagic (Managed) Services:

Consulting and Educational Services

Storage Performance Managed Services: incl. Consulting

Contact

Email: info@intellimagic.net

RMF's role in the Enterprise

RMF's Original Scope from the '70s

- Performance reporting and capacity planning for MVS processors

- Hardware measurements complemented SMF's accounting-like job and task information

- Runs within MVS, its native scope

- Today, covers the entire sysplex instead of isolated z/OS performance
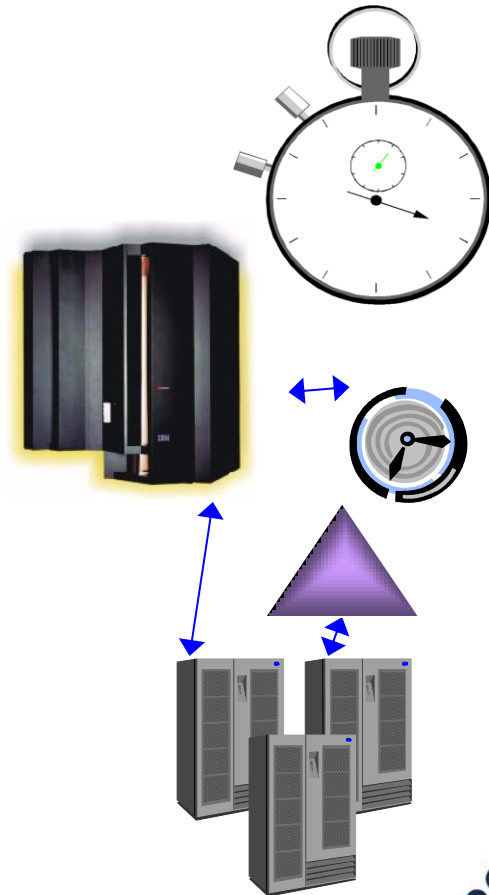
RMF Today

- RMF collects a wealth of performance data
  - Easily over 1 GB per day for large installations

- But data ≠ information
  - The sheer data volume obscures the information

- RMF was Introduced to pro-actively manage resource usage
  - Very hard to do now that information is "buried" in the data

- Many people still use tools built for "1980s RMF"
  - They miss out on a lot of new additions

What is collected?

Consolidated RMF data for a site contains:

- Multiple sysplexes
- 10,000 to 100,000 devices
- Very many CF structures
- Very many XCF groups
- Many WLM service classes
- Many Storage groups

Too much to be able to use RMF printed reports by interval

- We need **smart data mining**

What's happening?

- Lots of money spent on **replication (2/3/4-sites!)** to avoid data loss and service delivery interruptions

- Primary systems get less attention, yet **the primary site is what runs the business** 99+% of the time

- Primary site needs the proper care to avoid **application outages** caused by storage issues
  – Such an outage also interrupts service delivery

- RMF can help assess risks!

Where does RMF fit in this?

Does this mean RMF is no good? Or too good?

- RMF is a very rich source…
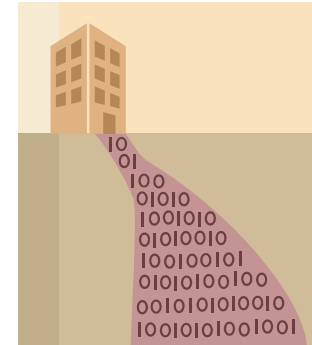
- But you need smart, automated ways to interpret all the information RMF collects today so you can identify risk areas

- And yes, in some areas RMF needs to do more

RMF data sources

RMF data sources

- You may think of RMF as a z/OS product, which collects z/OS data…

- But that is only a portion!

- Mostly it relays measurements from other sources:
  - Some mainframe-centric (e.g. Coupling Facilities)
  - Some mainframe-unaware (e.g. raid group data, zBX data)

Types of RMF data sources

1. "Classic MVS" items from z/OS control blocks
   - Paging, WLM, processor queues, I/O rates, locks

2. IBM architected zSeries hardware items
   - CEC and 'other-LPAR' data such as ICF and ICL
   - Coupling Facility
   - XCF
   - zBX (SMF 104)

3. zSeries-compatible I/O devices through CCWs
   - FICON Directors
   - Storage Systems (IBM DS8000, HDS VSP, EMC V-MAX)
   - Tape Systems (native tape and IBM or Oracle libraries)

RMF Support for IBM I/O devices

Getting counters from I/O devices is not trivial:

- Getting data from the device requires an agreed interface, typically with CCWs

- Need to agree on the structures:
  - z/OS concepts, like logical volumes
  - Storage concepts, like RAID groups
  - Advanced functions, like replication

- Need to define these in RMF such that they can be linked to z/OS concepts and resources

RMF support for EMC and Hitachi

Using the IBM [DS8000] interface to report on VMAX, VSP or P9500 requires **creative thinking:**

- Hitachi/EMC concepts must be mapped to IBM DS8000 concepts

- Mapping must be understood by RMF
  - IBM RMF and EMC VMAX teams need to work together
  - This may not happen…

- Hitachi's architecture is reasonably close to IBM's
  - So HP/HDS are automatically better represented than EMC

Interpreting RMF data

Data from multiple sources requires interpretation:

- **Align** native z/OS and external data
  - E.g. 74.1 device and 74.5 cache counters

- **Deduplicate** redundant data across z/OS images
  - For each system there is unique as well as duplicate data
  - No easy way to 'sum' data from multiple systems

- **Supplement** with vendor-specific external data
  - Necessary to get the complete picture
  - EMC SQ Mirror to compute back-end data
  - Hitachi TGZ to understand concatenated parity groups
  - DCOLLECT data to get VTOC statistics
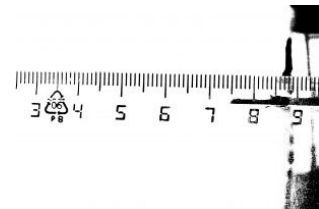
Mine the Data

How to go from big data to useful information

- Let's look at an example:

   What can link information tell us about our replication performance and risk level

   - RMF provides disk link statistics (SMF 74.8)
     – ECKD, PPRC, SCSI
     – Read/write MB/s, Ops, response time, Intensity

   - Data is source from disk system, <u>not</u> a z/OS control block
     – Data from one LPAR that has access to the disk system is all that is needed

   - RMF post processor command is ESS

   - Supported by all current enterprise disk vendors

RMF Report

Let's start at the beginning:

```
E S S   L I N K   S T A T I S T I C S

           z/OS V1R12            SYSTEM ID SYS1           DATE 10/31/2013            INTERVAL 05.00.051
                               CONVERTED TO z/OS V1R13 RMF    TIME 11.00.00              CYCLE 1.000 SECONDS
SERIAL NUMBER  0000012345       TYPE-MODEL  002107-961    CDATE 10/31/2013   CTIME  11.00.00   CINT   05.00
------ADAPTER------    --LINK TYPE--     BYTES       BYTES         OPERATIONS      RESP TIME         I/O
SAID   TYPE                             /SEC        /OPERATION      /SEC           /OPERATION      INTENSITY
0200   Fibre 4Gb      PPRC SEND         6.8M        10.1K          667.6          0.5             337.8
0200   Fibre 4Gb      PPRC RECEIVE      0.0         0.0            0.0            0.0             0.0
                                                                                                 ------
                                                                                                 337.8

0201   Undefined      NO DATA TO REPORT OR ZERO
0202   Fibre 4Gb      PPRC SEND         6.7M        10.1K          666.1          0.4             241.5
0202   Fibre 4Gb      PPRC RECEIVE      0.0         0.0            0.0            0.0             0.0
                                                                                                 ------
                                                                                                 241.5

0203   Undefined      NO DATA TO REPORT OR ZERO
0230   Fibre 4Gb      ECKD READ         6.5M        10.4K          625.6          0.1             33.5
0230   Fibre 4Gb      ECKD WRITE        3.7M        8.4K           441.8          0.5             213.8
                                                                                                 ------
                                                                                                 247.4
```

$HASP375
IMUSJT1R
ESTIMATE
EXCEEDED BY
306,000 LINES 2
%
SPOOL ….

RMF post processor ESS (link) report for a single day for 6 disk systems approximately 400,000 lines!

ESS report is a <u>small</u> report (single LPAR only). No consolidation is needed.

Cache, Device, Channel, IOQ, … make the printed reports almost unusable!
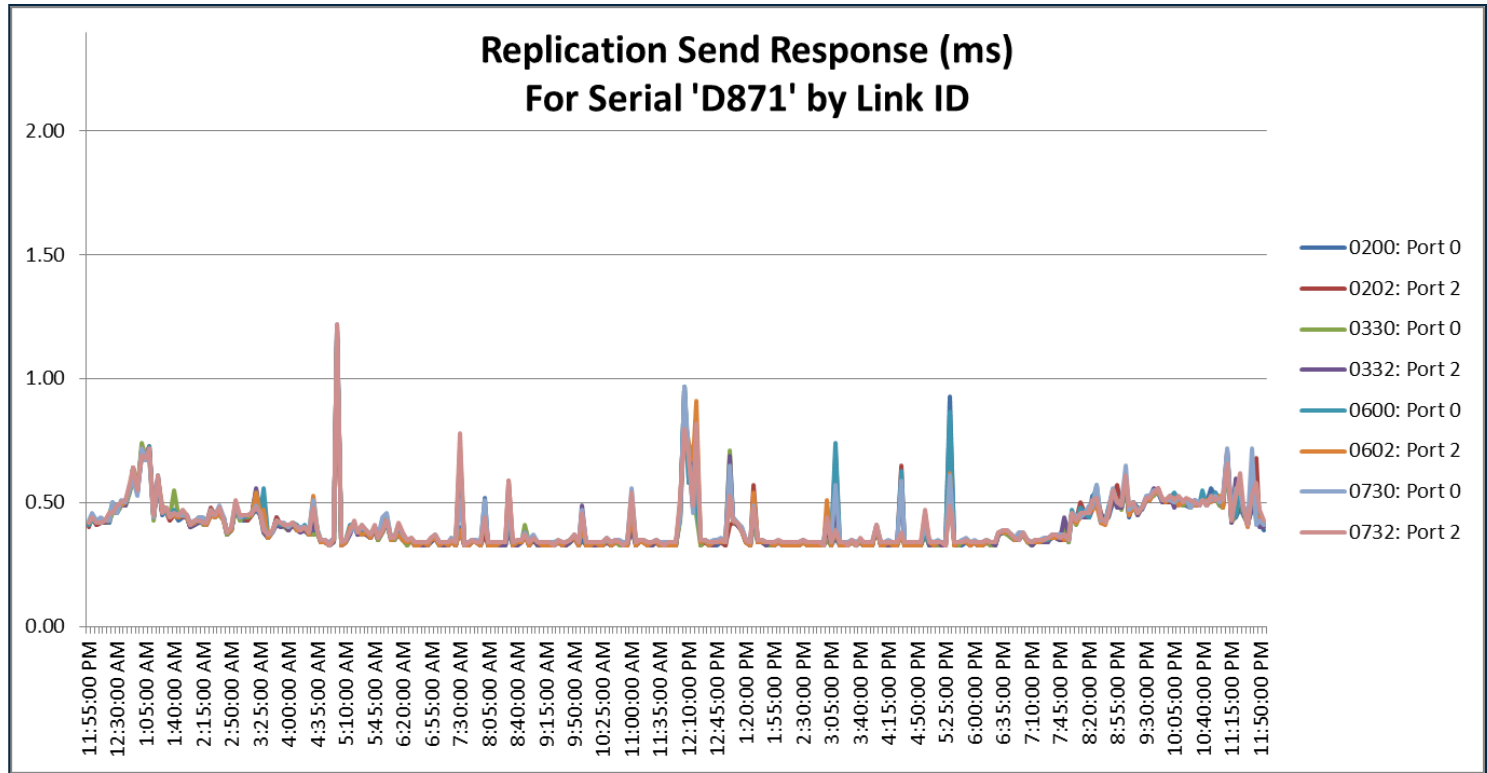
Printed Report

- The printed report has all the required information, but the format doesn't allow us to interpret the data

  - We have to look at each line to see something important

- So, let's go and look at the data visually

  - For the following charts, let's only look at the send response time (ms) for a single disk system

  - Let's now use Microsoft Excel to view the data

Visualize the Data

Replication Send Response (ms)
For Serial 'D871' by Link ID

Legend:
- 0200: Port 0
- 0202: Port 2
- 0330: Port 0
- 0332: Port 2
- 0600: Port 0
- 0602: Port 2
- 0730: Port 0
- 0732: Port 2

Simply plot the data points

24 hours

We see 8 links (ports), and can see a few peaks.

We can also see that all links are basically receiving the same response time.
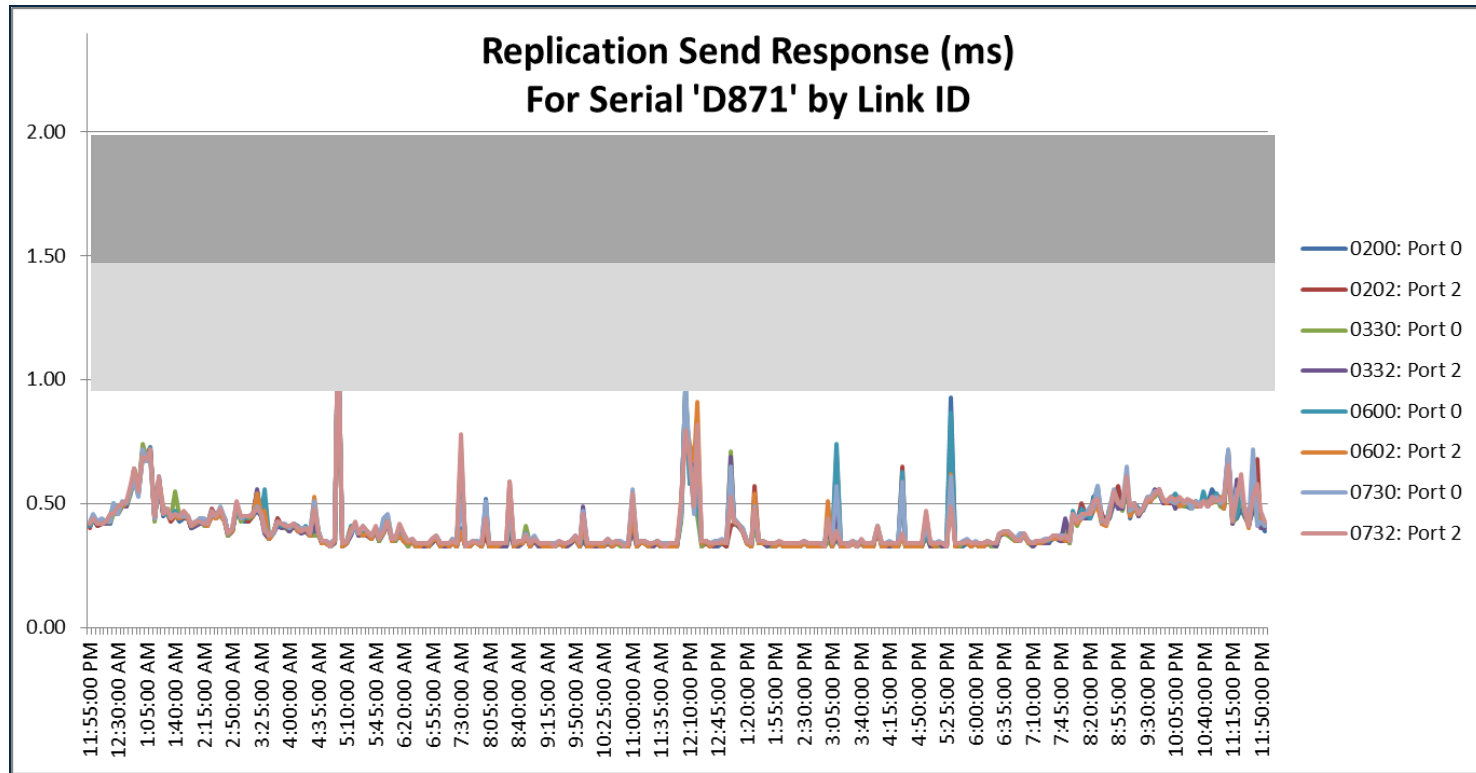
Visualize the Data (II)

- Looking at the data visually certainly makes it easier to understand the information

  - We see the response time for 24 hours
    - Could be ok
    - Has a few peaks
    - Seems to be the same for all links (ports)

- Can we get more out of this data?
  - Is this a good response time for these links?
    - Different link speeds will give us a different response time
    - An average of 0.4 msec – but how fast should we be?
    - Do the peaks indicate problems? Are they hurting us?
  - What is the risk assessment of our replication infrastructure?

> Looking at the chart creates more questions than answers!

Thresholds?

- We know what kind of replication links we have so we can define <u>hardware</u> based thresholds for this chart

  - For example: Link speed (2/4/8 Gb) and distance (km) impact the possible response time

  - Now we can tell if the send response time is good or bad for this link type

- What does the chart look like with threshold indicators?

- **Note**: I've used two thresholds to indicate if things are "bad" or "really bad"

Define Thresholds

Replication Send Response (ms) For Serial 'D871' by Link ID

Legend:
- 0200: Port 0
- 0202: Port 2
- 0330: Port 0
- 0332: Port 2
- 0600: Port 0
- 0602: Port 2
- 0730: Port 0
- 0732: Port 2

We have set thresholds that now give us valuable information as to how to interpret our response times.

Why Thresholds?

- So our thresholds now show us if the peaks are really problems or not

  – This additional information is very valuable and can be used to understand the response times we are looking at. Now we can relate the measured response time to a hardware defined maximum

- Ok, we have used the specific hardware to create thresholds. But, we have other information available to us

  – Let's see if we can come up with thresholds that are hardware <u>and</u> workload dependent
  – Both are specific to our environment

Hardware + Workload Thresholds

Replication Send Response (ms)
For Serial ' D887 ' by Link ID
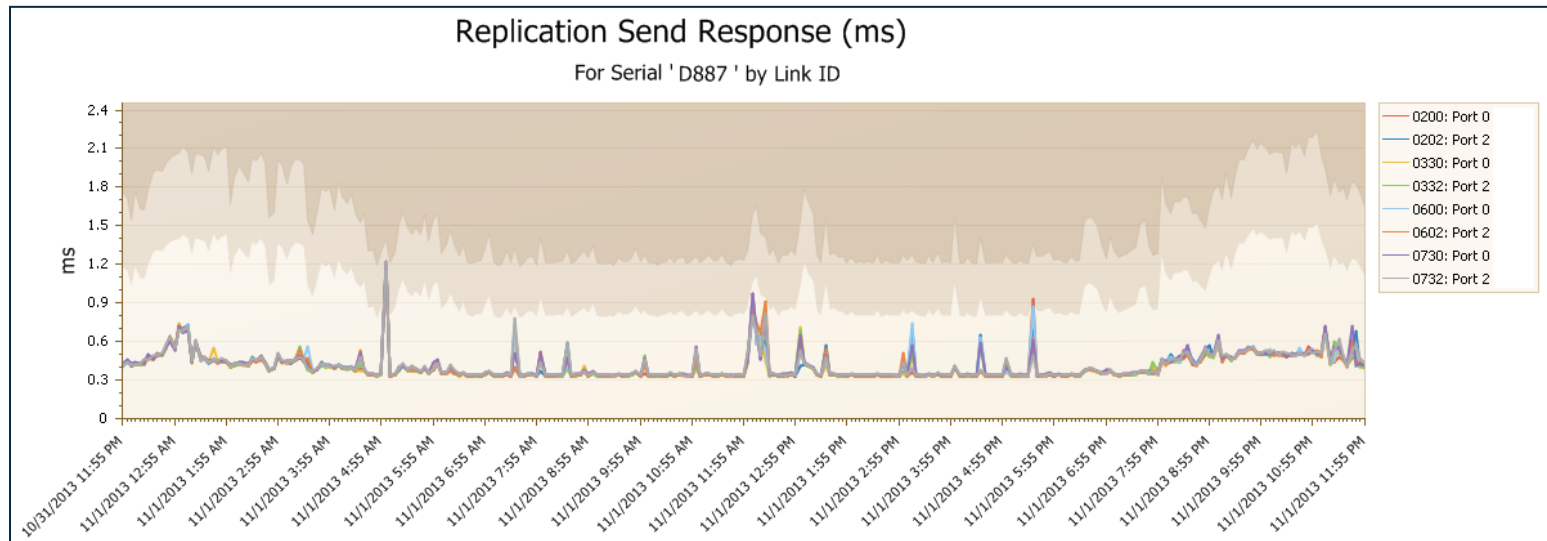
Note how our threshold allows the response time to be higher during the batch window.

We have added a workload component to our purely hardware thresholds.

The workload component causes our thresholds to no longer be straight lines, but to match the data being looked at.

Workload Thresholds

- The workload component added to the hardware threshold value is the effective <u>connected data rate</u>

- Connected data rate:
  - We expect a higher send response time for a larger block of data. A lower response time is expected for smaller blocks

  - So, we need to know how large the data is that corresponds to this send response time

  - Well, RMF has this information ☺

- Great, so we now know how good (or bad) our response time is, <u>for our workload and hardware</u>

Rating

- Now that we know what is good/bad, let's make it easier to see
  - We shouldn't be required to look closely at the chart. Let our reporting system tell us

- Look at each data point and determine if it is above the threshold
  - In this example we have two thresholds, so assign data points in the higher threshold (worse response time) a higher rating

- In this example, each link gets a rating for the 24 hour period and the worse rating is shown in the chart title
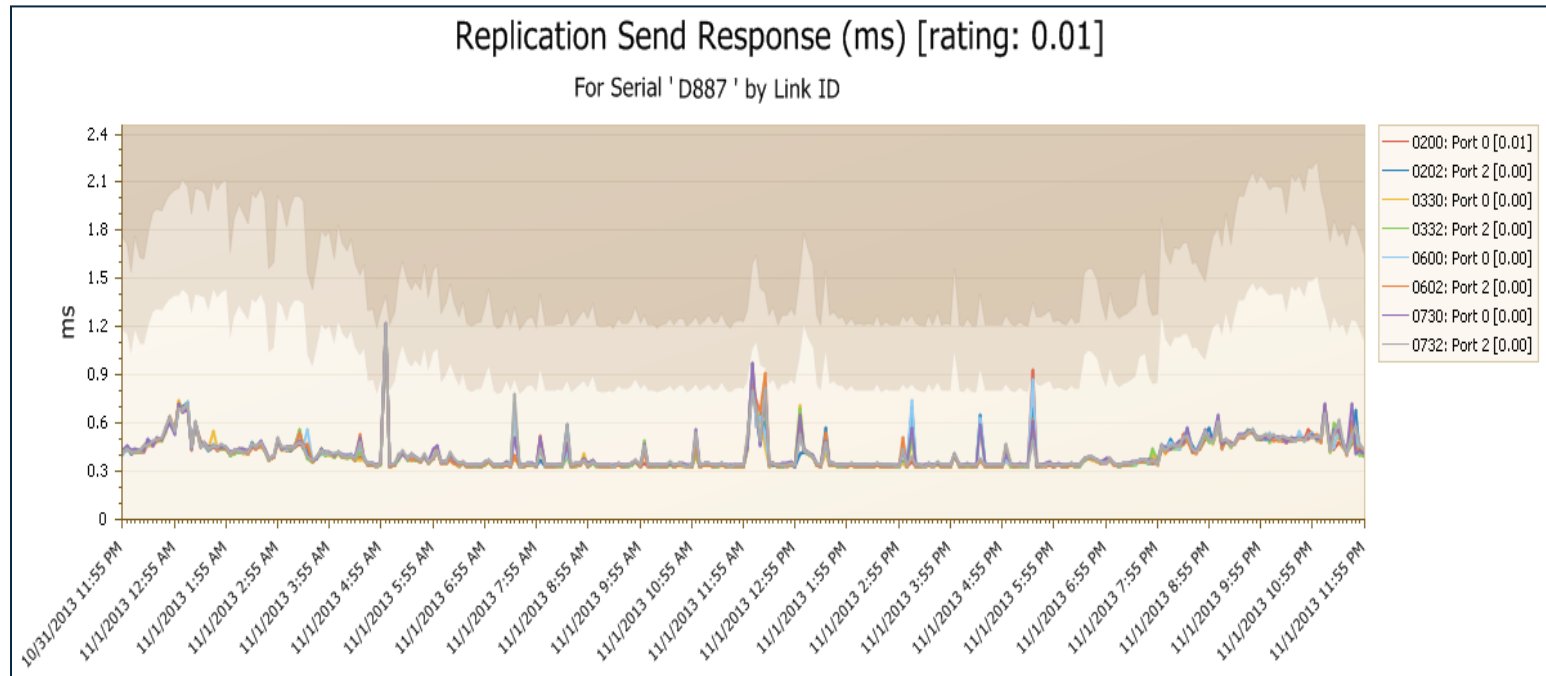
Rated Chart

The rating is in the legend for each link.

One link has a 0.01 rating (one data point above the threshold).

This is shown as the maximum in the chart title

Rated Chart (II)

- Let's make it easier to understand the rating value – use a color

    - **Green**
        - No problems. No risk
    - **Yellow**
        - Some issues. Performance is probably ok, but it should be looked at. Some/low risk
    - **Red**
        - Oops! Need to look at this now because there are performance issues. Significant/high risk

- Let's use this color as a border to make it stand out in our reporting system. We want to know when things are not running properly

Colored Rated Chart
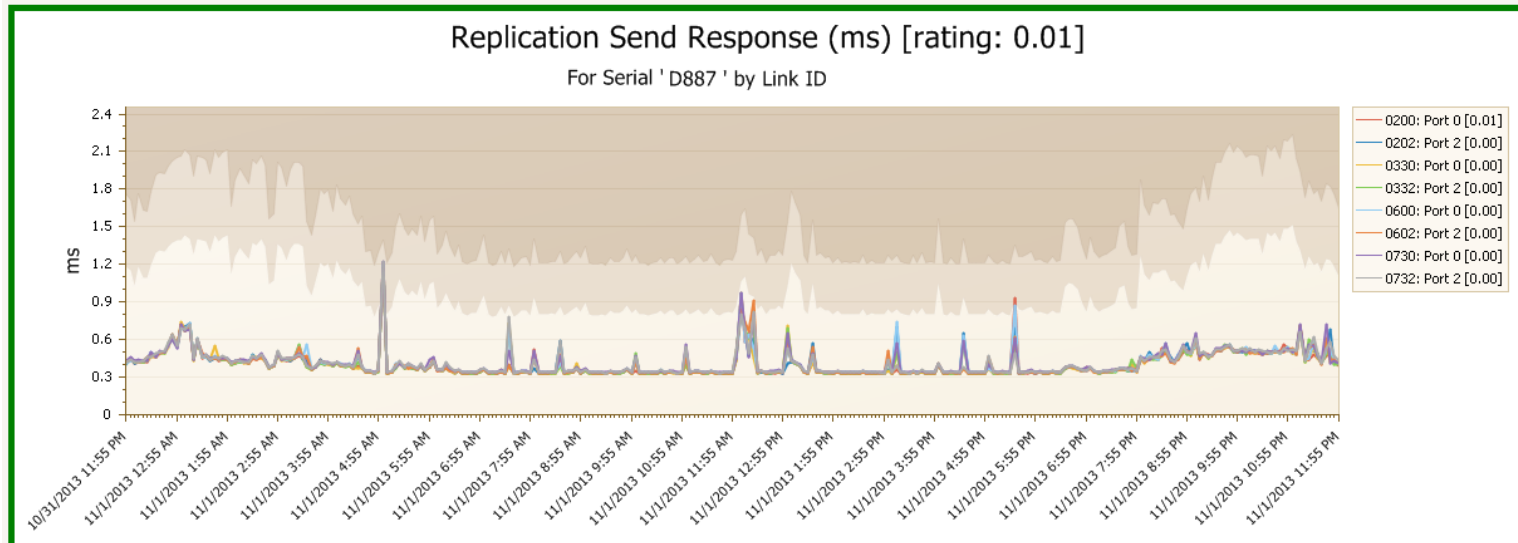
When we look at this chart we can immediately see that there are no problems with these links and our replication. The green border is the quick indicator that everything is ok.

There is no risk of replication issues affecting our service levels.

Even though there are some peaks in the data.

What about Service?

- We have rated our chart and now know if the link response times are good or not

- For some metrics, we can make the information even easier to understand

- Let's define service levels (basically rate the data) and see if our agreed service level objective is being met

- **Note**: I've changed the service ratings to make the chart visually more interesting

Service Level
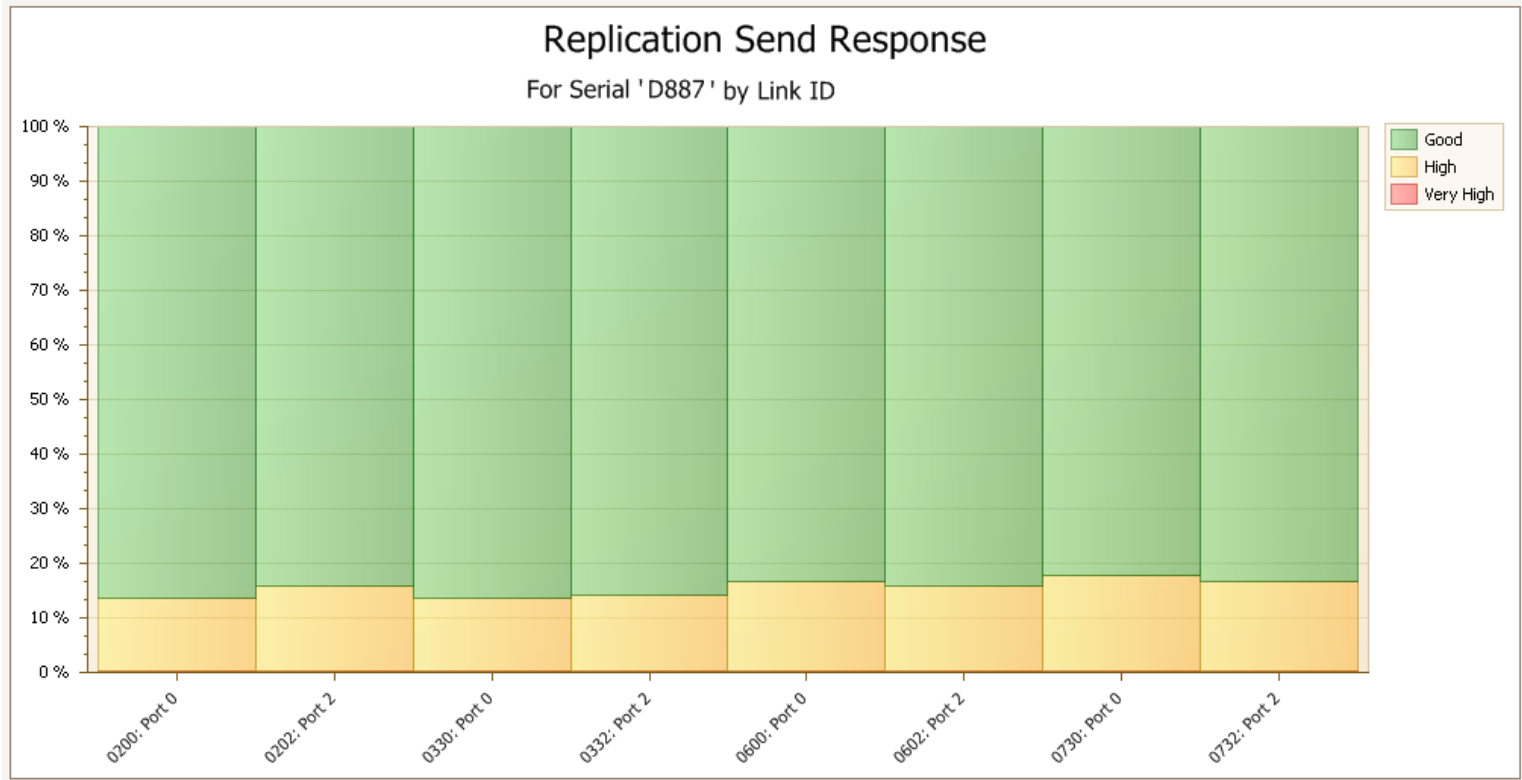
We are still looking at link response time.

On average, our links are providing the agreed service level for 85% of the time (green).

What else?

- What else can we get from this data?

- We have looked at the response time, added smart workload/hardware thresholds and rated our chart

  – This shows us the response time relative to some upper limit (a threshold). We can also look at the service level delivered

- But, how do the links look compared to one another?

  – We could see in the line chart that the 8 links almost have the same response times
  – Can we see this better in a different chart format?

Balance
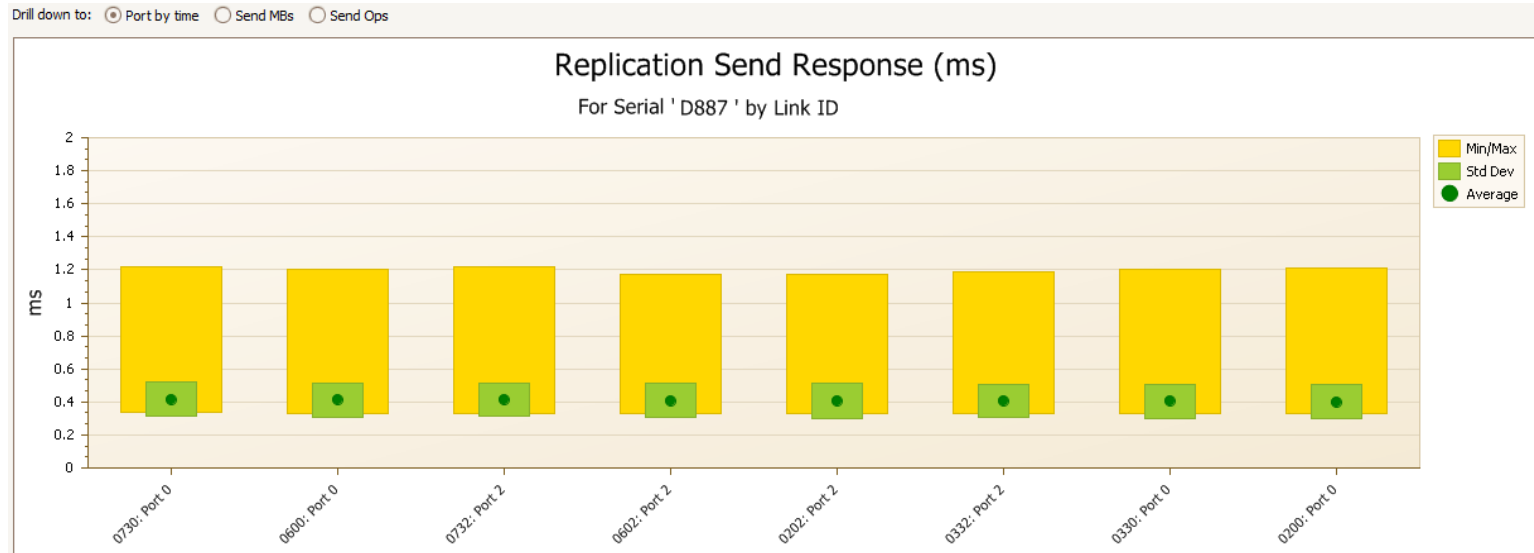
For each link, we can see the average for 24 hours (green dot), the maximum und minimum values (yellow block) and the standard deviation (green block).

Balanced Links?

- Using the <u>same</u> send response time data, we can now clearly see that the links are very well balanced

- We can also see that there are some peaks, but the standard deviation is close to the average response time

- Why do we care?

Balanced Links?

- You get the most out of your hardware when you balance the load and use all like resources evenly

    - Our example is for replication links, but it is also valid for RAID ranks, HDD, adapter cards, volumes, …

- Replication imbalance can indicate problems elsewhere:

    - One of the main trunk lines could be having hardware issues
    - A local disk component may be having hardware problems
    - Configuration errors

- These errors may be local or at the remote site

Another Example
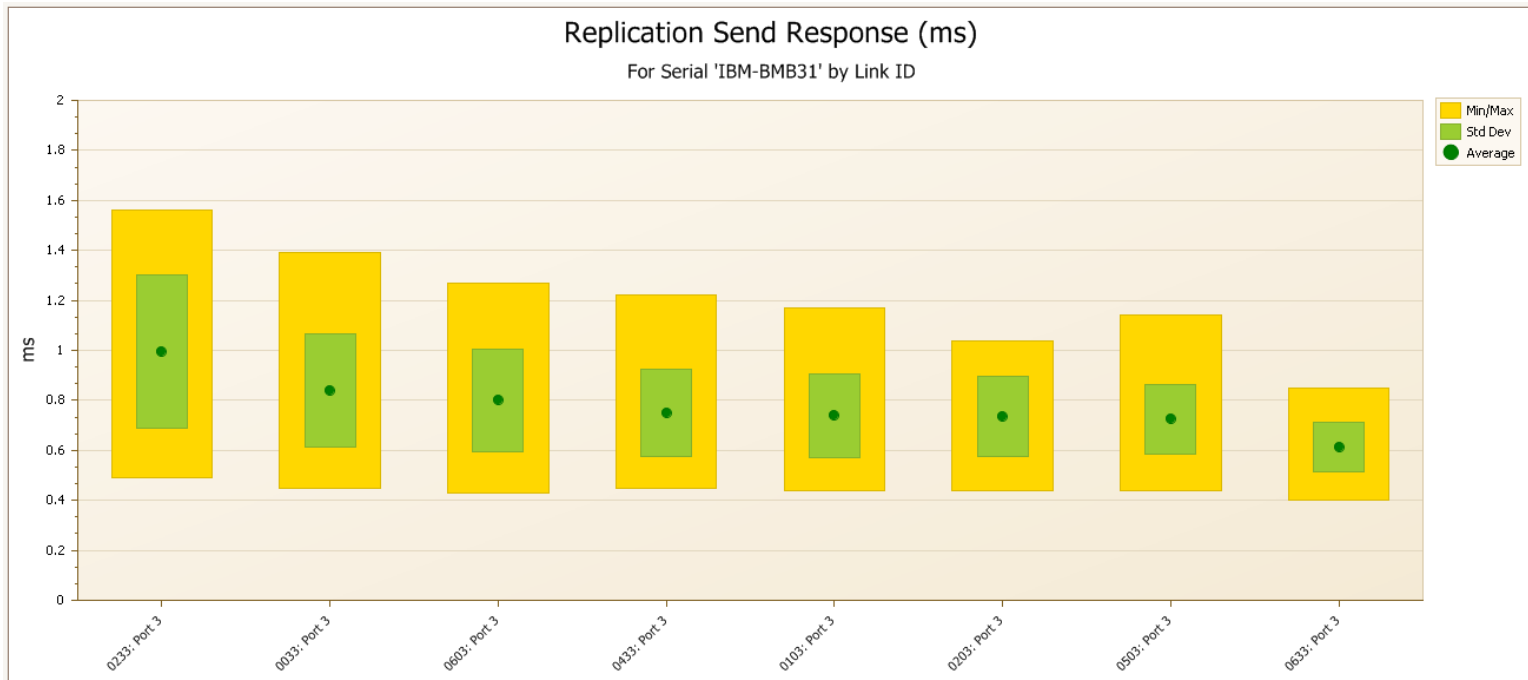
In this example, the customer was using different trunk lines for the PPRC links.

The same amount of data and the same number of I/O's were being sent on each link, but the links were delivering different response times.

How am I Configured?

- How many links do you need to ensure stable data replication?

  - Actually - all of them!
  - You (or your vendor) sized your links for response time AND availability

- The balance charts show us if all links have the same workload and also if all links are running!

- But, some RMF data gives us configuration information, if we know the hardware type/model
  - This is the case with link data ☺

Configuration

DSS Links: For Serial 'D887' by Link ID

Drag a column header here to group by that column

Link ID	Serial	Type	Link type	FICON Read (op...	FICON Write (...	Fibre Re...	Fibre Wr...	Replication Se...	Replicatio...
▶ 0030: Port 0	IBM-12345	4Gbs	FICON	1328.40	664.23	0.00	0.00	0.00	0.00
0031: Port 1	IBM-12345	Unknown	Inactive	0.00	0.00	0.00	0.00	0.00	0.00
0032: Port 2	IBM-12345	4Gbs	FICON	1064.53	630.16	0.00	0.00	0.00	0.00
0033: Port 3	IBM-12345	Unknown	Inactive	0.00	0.00	0.00	0.00	0.00	0.00
0100: Port 0	IBM-12345	4Gbs	FICON	1328.38	664.93	0.00	0.00	0.00	0.00
0101: Port 1	IBM-12345	Unknown	Inactive	0.00	0.00	0.00	0.00	0.00	0.00
0102: Port 2	IBM-12345	4Gbs	FICON	1063.32	629.91	0.00	0.00	0.00	0.00
0103: Port 3	IBM-12345	Unknown	Inactive	0.00	0.00	0.00	0.00	0.00	0.00
0200: Port 0	IBM-12345	4Gbs	PPRC	0.00	0.00	0.00	0.00	908.99	0.00
0201: Port 1	IBM-12345	Unknown	Inactive	0.00	0.00	0.00	0.00	0.00	0.00
0202: Port 2	IBM-12345	4Gbs	PPRC	0.00	0.00	0.00	0.00	908.49	0.00
0203: Port 3	IBM-12345	Unknown	Inactive	0.00	0.00	0.00	0.00	0.00	0.00

Single adapter

This is only an excerpt of the complete configuration table

We can see 3 IBM host adapters, each with 2 active ports.

The first two host adapters are only for FICON (4Gbs).

The third is for PPRC (also 4Gbs). These are our replication links.

Why is Configuration Important?

- It is very important to spread the PPRC and FICON load across all host adapters (front-end directors)

- Sometimes, replication links (PPRC) are placed on the same adapter card as FICON links
  - This is not a problem, but they do represent different workloads

- Why do we care?
  - A heavily used FICON port can cause poor PPRC response time if both ports are on the same adapter card

- We looked at link (port) balance. What about adapter card balance?

Adapter Utilization Chart

Front-end Adapter Utilization (%) [rating: 0.00]
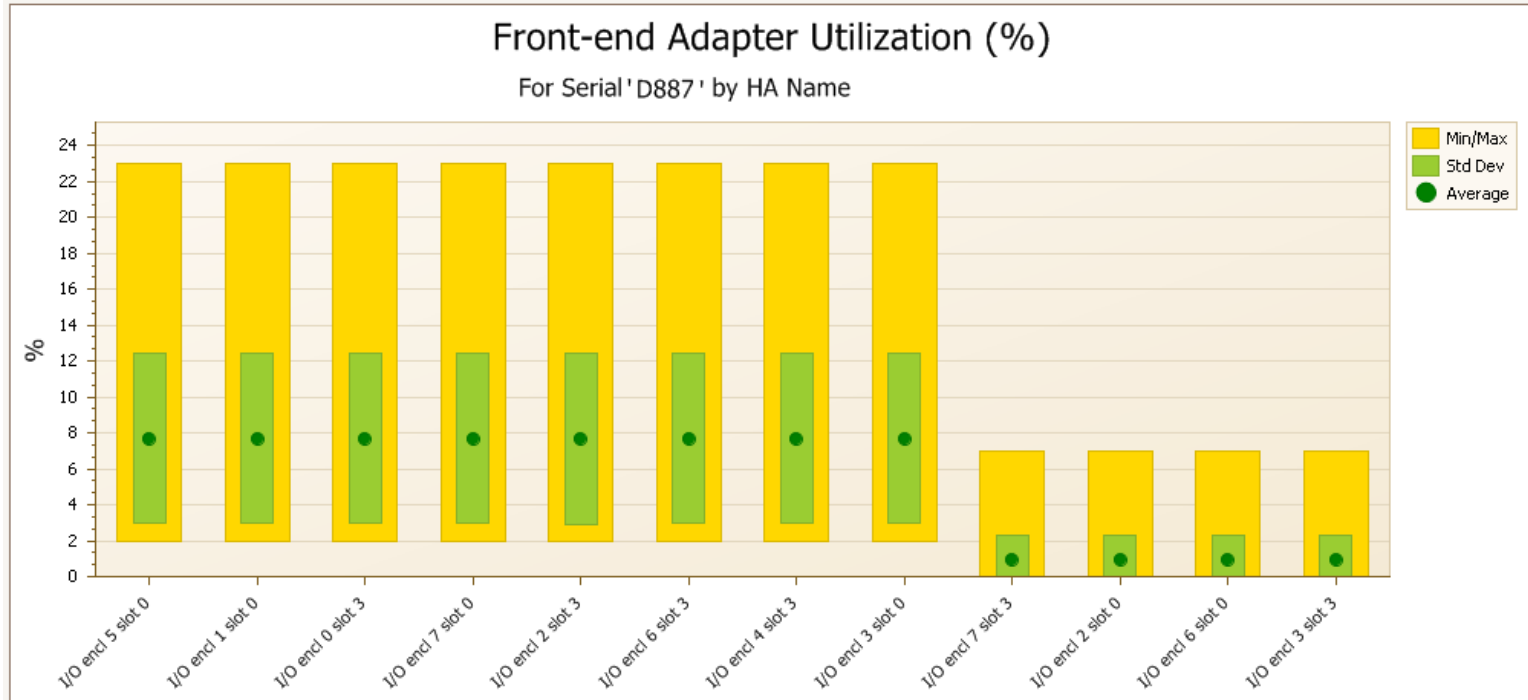For Serial 'D887' by HA Name
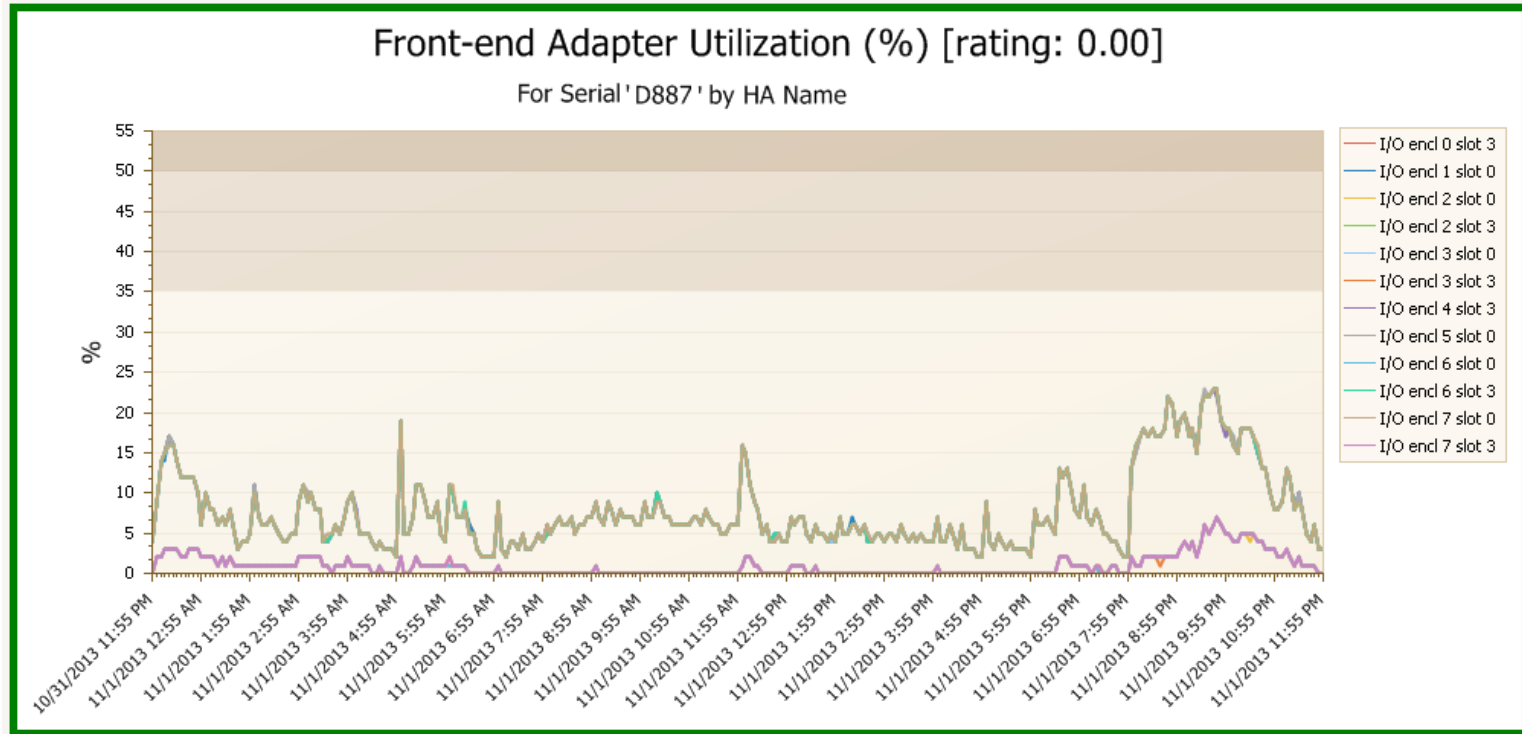
We see that the 12 adapter cards are not at their limits (good). But, one or more cards are hardly being used (bad).

Note: This is where a "balance" type of chart can be used.

Adapter Balance Chart

Front-end Adapter Utilization (%)
For Serial 'D887' by HA Name

Yes, very clearly we have an imbalanced configuration. 4 adapter cards are doing very little!

Luckily it is not affecting our replication (all PPRC links were balanced).

Icing on the Cake

- We have various types of rated charts. What else do we need?

- Automation is critical because we need to look at these chart regularly

  – Automate the creation of charts

  – Automate the sending of reports to specific people/groups
    - Only send charts if there is a problem!

What is needed

- Process a lot of data quickly

- Get the right reports to the right people
  - Not everyone needs to know about link error rates
    - But, the <u>right</u> person certainly does!

- Once the data is rated, use alerts to quickly and accurately get the right data visible

- Keep the implementation simple or it will not be done properly!

- It must be fully automated

Summary

- We only looked at a single RMF record (ESS link)!

  - We can see how well our replication is running (response time) and if the agreed service is being adhered to

  - Using thresholds, ratings and various different chart types, we can evaluate the <u>RISK</u> level of our replication

  - **Green** – no problems – no risk
  - **Yellow** – potential problems – low/medium risk
  - **Red** – definite problems – high risk

- **Protect your investment in replication by being aware of where you are today! High risk can quickly become an expensive outage!**

Amazing what kind of replication related information we can get from RMF.

One record type of many!

Questions?

RMF has many more treasures.

Start hunting now!