SNIA

Education

Six Sigma Performance Analysis for SAN

Dan Iacono, HP

SNIA Legal Notice

- The material contained in this tutorial is copyrighted by the SNIA.
- Member companies and individuals may use this material in presentations and literature under the following conditions:
  - Any slide or slides used must be reproduced without modification
  - The SNIA must be acknowledged as source of any material used in the body of any document containing material from these presentations.
- This presentation is a project of the SNIA Education Committee.
- Neither the Author nor the Presenter is an attorney and nothing in this presentation is intended to be nor should be construed as legal advice or opinion. If you need legal advice or legal opinion please contact an attorney.
- The information presented herein represents the Author's personal opinion and current understanding of the issues involved. The Author, the Presenter, and the SNIA do not assume any responsibility or liability for damages arising out of any reliance on or use of this information.

NO WARRANTIES, EXPRESS OR IMPLIED. USE AT YOUR OWN RISK.

Abstract

Education SNIA

- Six Sigma Performance Analysis for SAN
  - Traditionally, SAN switch performance is measured in average and maximum throughput with a vendor-specific application. There is a better way to measure SAN performance and predict bandwidth. Six Sigma has long been used for measuring and predicting production in manufacturing plants. With SANs, we are in the business of producing data throughput with a finite amount of bandwidth. Six Sigma can then be applied to SAN traffic to understand and predict SAN performance patterns with a 99.73% confidence, which leaves less bandwidth underutilized and increases your SAN ROI.

Agenda

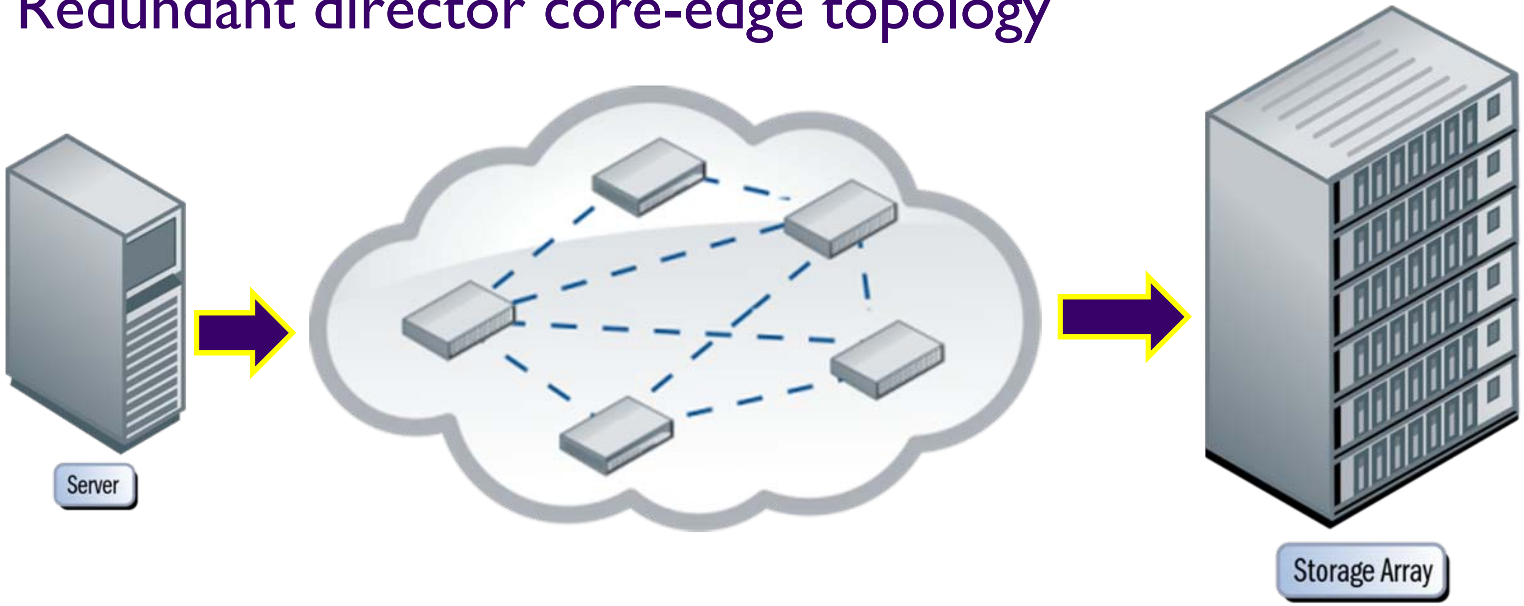
- Need for SAN Performance Analysis
- Requirements for Analysis Solution
- Current Methods For Data Analysis
- Characterize data patterns
- Stat 101
- Six Sigma Performance Analysis
- Case study
- Review

Need for SAN Performance Analysis

- Evolution of SAN from 10's of ports to 1,000+ ports
- SANs have grown in size from 2 to 60+ switches or
- Switch port densities have grown from 8 to 300+ ports per switch
- Topologies have grown from simple core-edge to multiple director cores
- Bandwidth has grown from 1Gb/s to 8 Gb/s
- Not all ports require the same amount of bandwidth

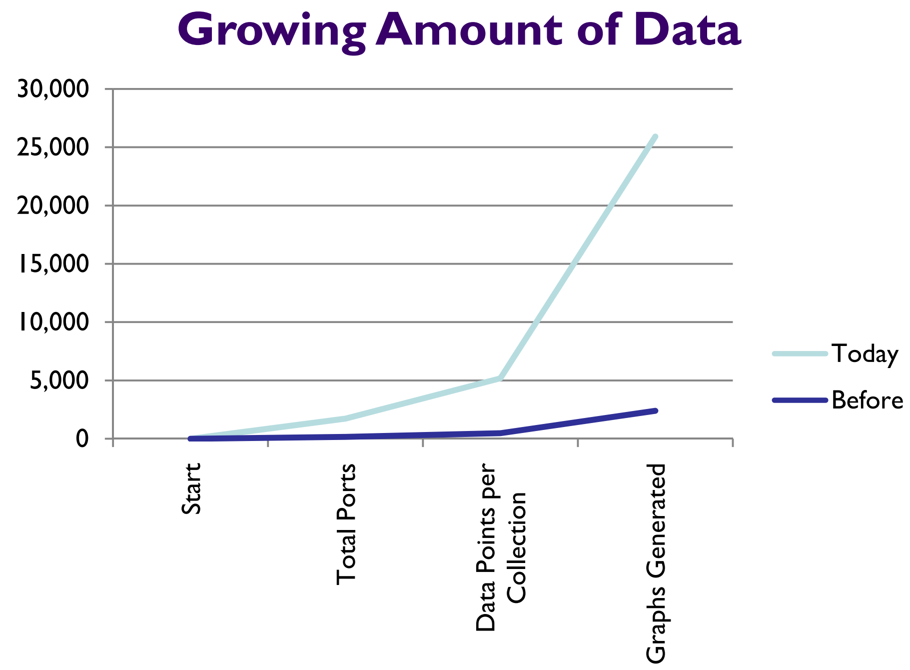
SAN Performance Case Example

- 60 edge switches (16 ports, 960 ports total)
- 6 director switches (128 ports, 768 ports total)
- 1,728 Total SAN ports
- 25,920 individual graphs generated
- Redundant director core-edge topology



Total Data Collection for 5 Days

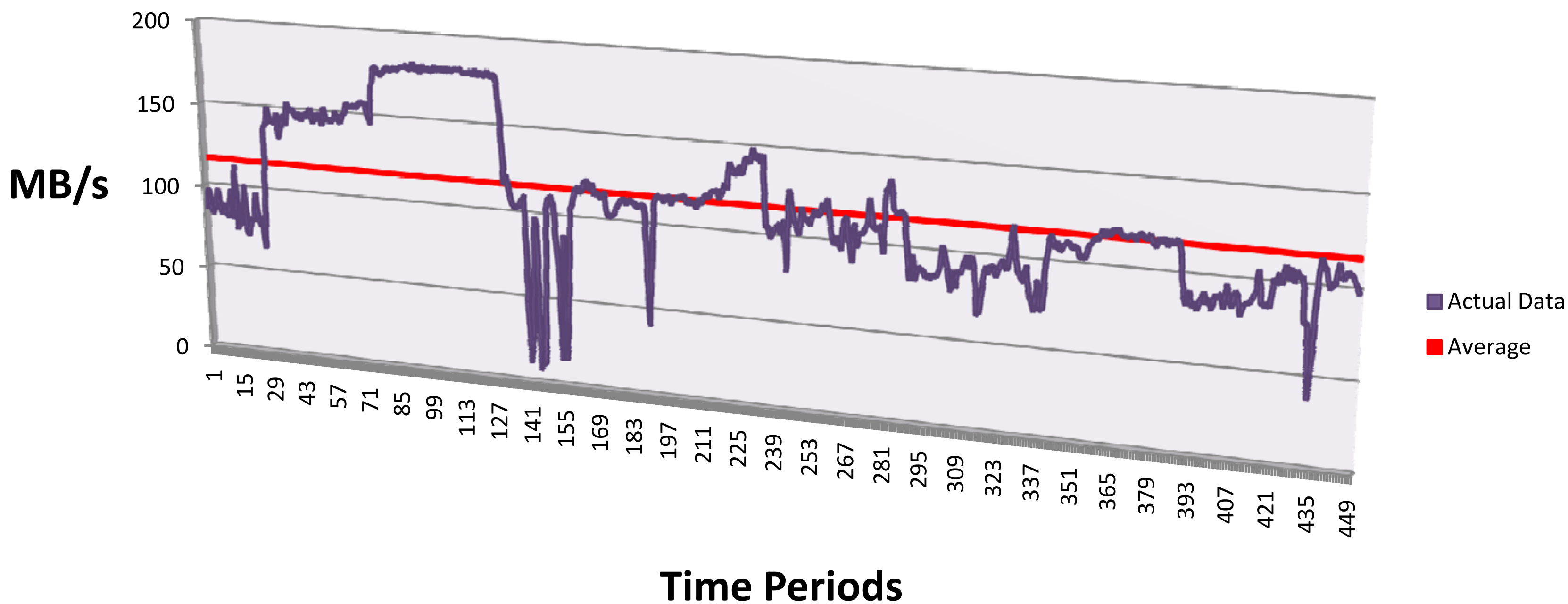
- 5,184 data points at a 2 minute sampling interval (Tx, Rx, Total MB/s)
- 18,662,400 data points per assessment (24 hrs x 5d)



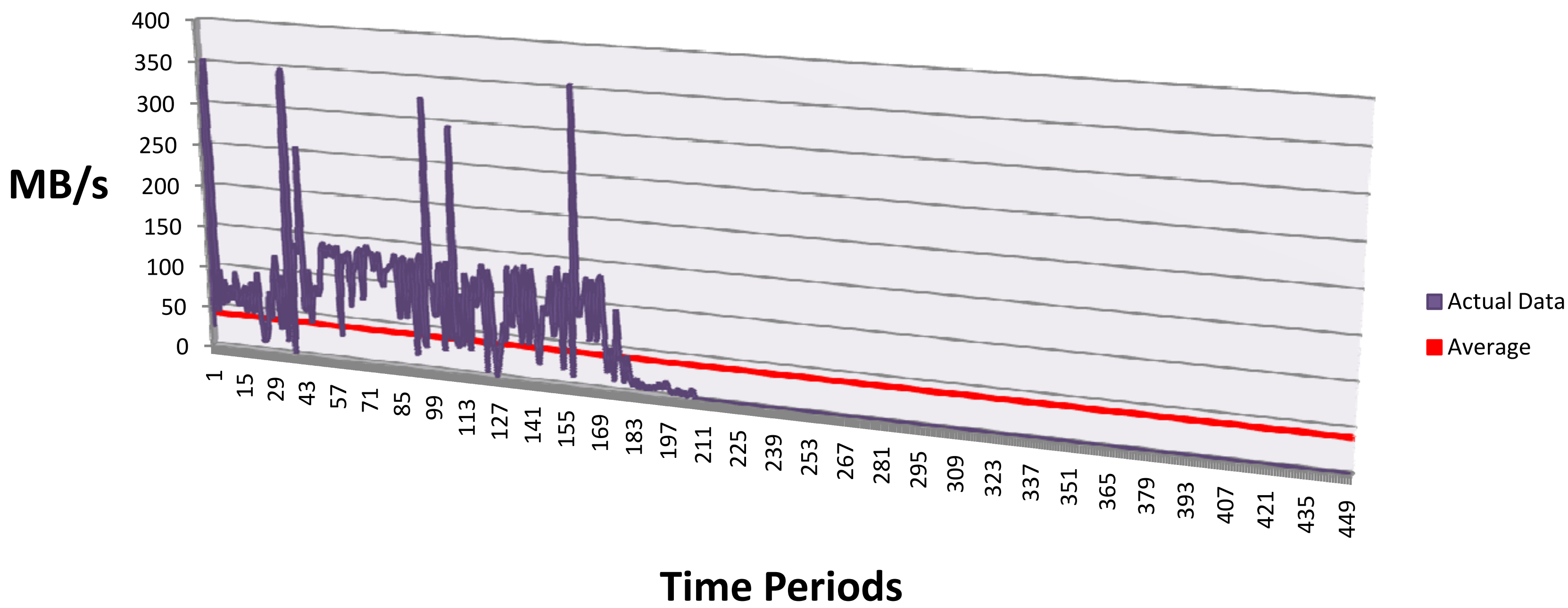
Requirements for Analysis Solution

- The amount of performance data is too large to manually review
- Find two types of data:
  - Sustained – data is consistently at a common level
  - "Spiked" – short periods of time where data throughput is at an extremely elevated level and then returns to normal
- Mine data without looking at graphs or source data
- Data mining needs to be fast and systematic
- Rules for bandwidth decisions

Sustained Data



Spiked Data



MB/s
400
350
300
250
200
150
100
50
0
1
15
29
43
57
71
85
99
113
127
141
155
169
183
197
211
225
239
253
267
281
295
309
323
337
351
365
379
393
407
421
435
449
Time Periods
Actual Data
Average

Current Methods For Data Analysis

- Min, Max, and Average do not characterize SAN performance well
- Average
  - Finds sustained throughput
  - Misses spiked data
- Maximum
  - Might catch spiked data,
  - Finds many false positives
  - Misses sustained throughput
- Minimum – Doesn't really help

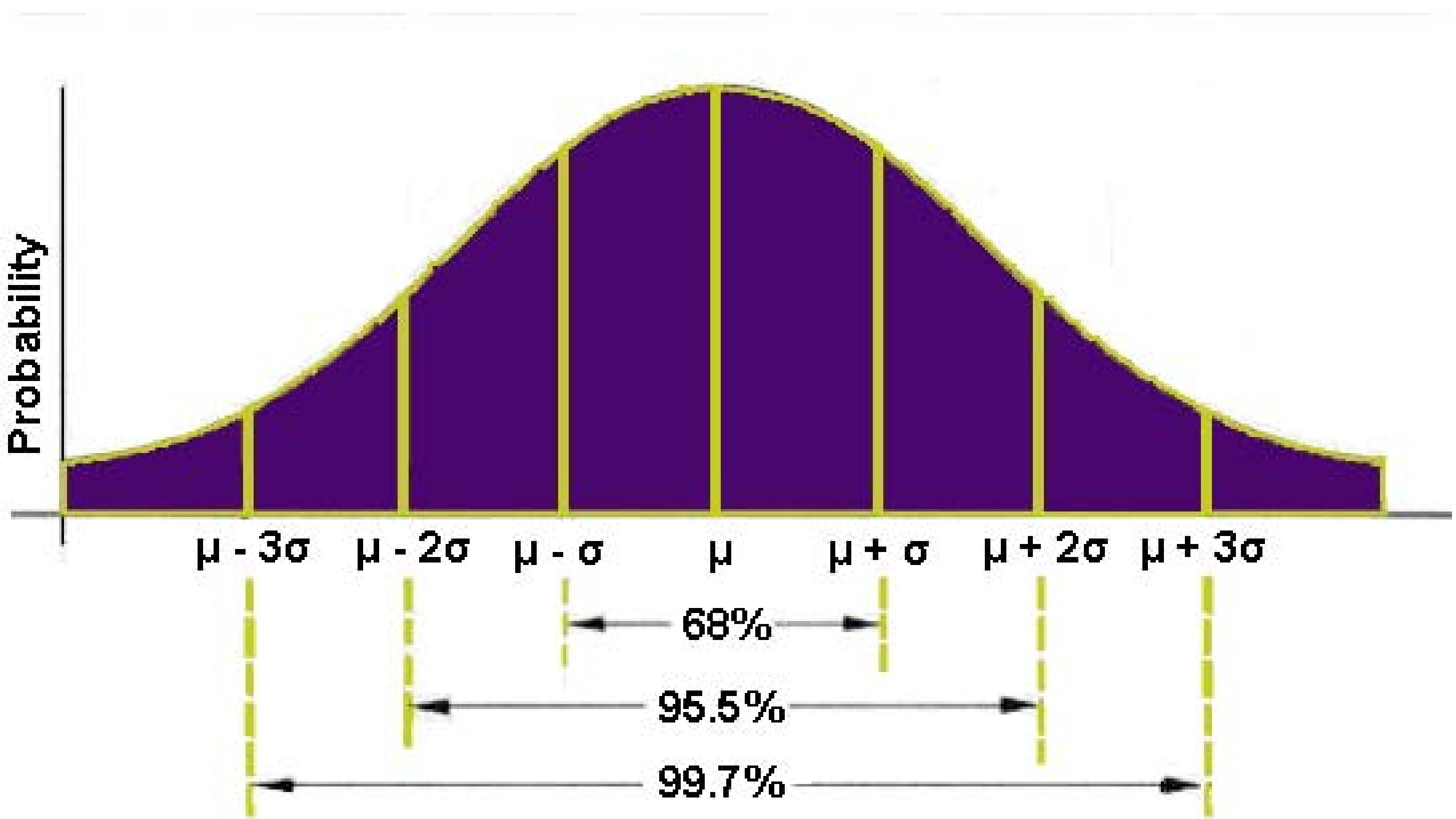
Using a Statistical Method

- Classifying and predicting sets of data is what statistics is all about
- Disk array analysis solutions commonly use 95th percentile calculation today for performance analysis
- With the SAN we are taking it one step further with the 99.73th percentile
- Six Sigma gives us a proven statistical method to apply guidelines to our data

Stat 101

- **Set** – is a collection of data
- **Sample** – is a set of data that is a representation of a larger set of data
- **Variance** – how much fluctuation in values between data points
- **Sum (∑)** – the addition of a set of numbers together
- **Average (μ)** – the sum of a set of numbers divided by amount of numbers in the data set
- **Standard deviation (σ)** – a set unit to measure variance of data in a set (also referred as sigma)

The Bell Curve



Disk Performance

Education SNIA

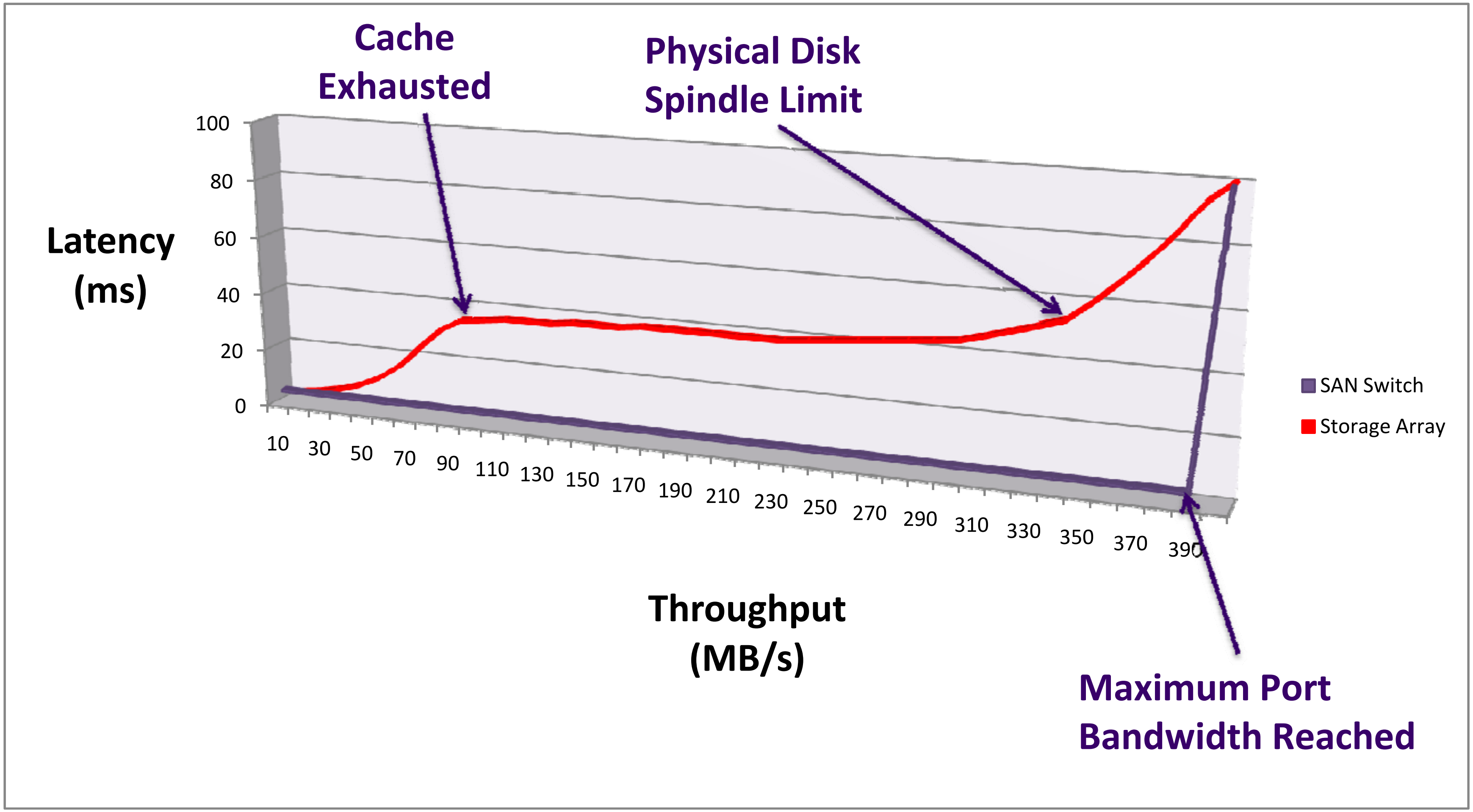
- Performance varies due to many issues:
  - Cache
  - Disk speed
  - RAID type
  - Number of disks
- Disk arrays have diminishing returns when approaching maximum bandwidth or performance capacity
- Disk performance analysis use 95^{th} percentile (2σ) because the last 5% is where problems usually occur

SAN vs. Disk Performance

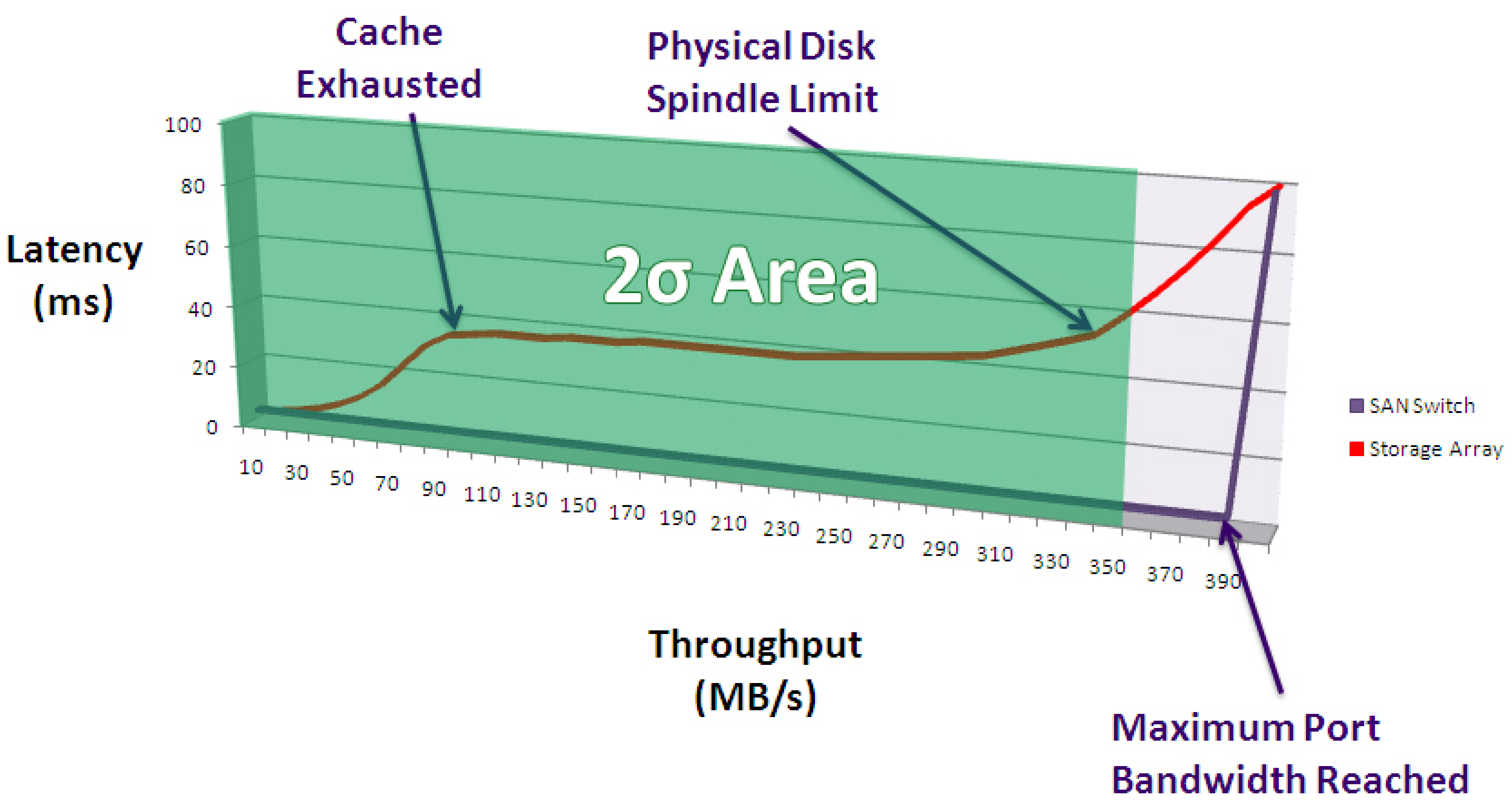
Education SNIA

- SAN performance does not vary because data is switched the same way
- Only when full bandwidth is met does a SAN switch restrict performance on an application
- The last percentile is when data may be effected which is between 3σ and 6σ
- 99.73rd Percentile (3σ) is used for SAN
- 95th (2σ) in disk arrays

SAN vs. Disk Performance

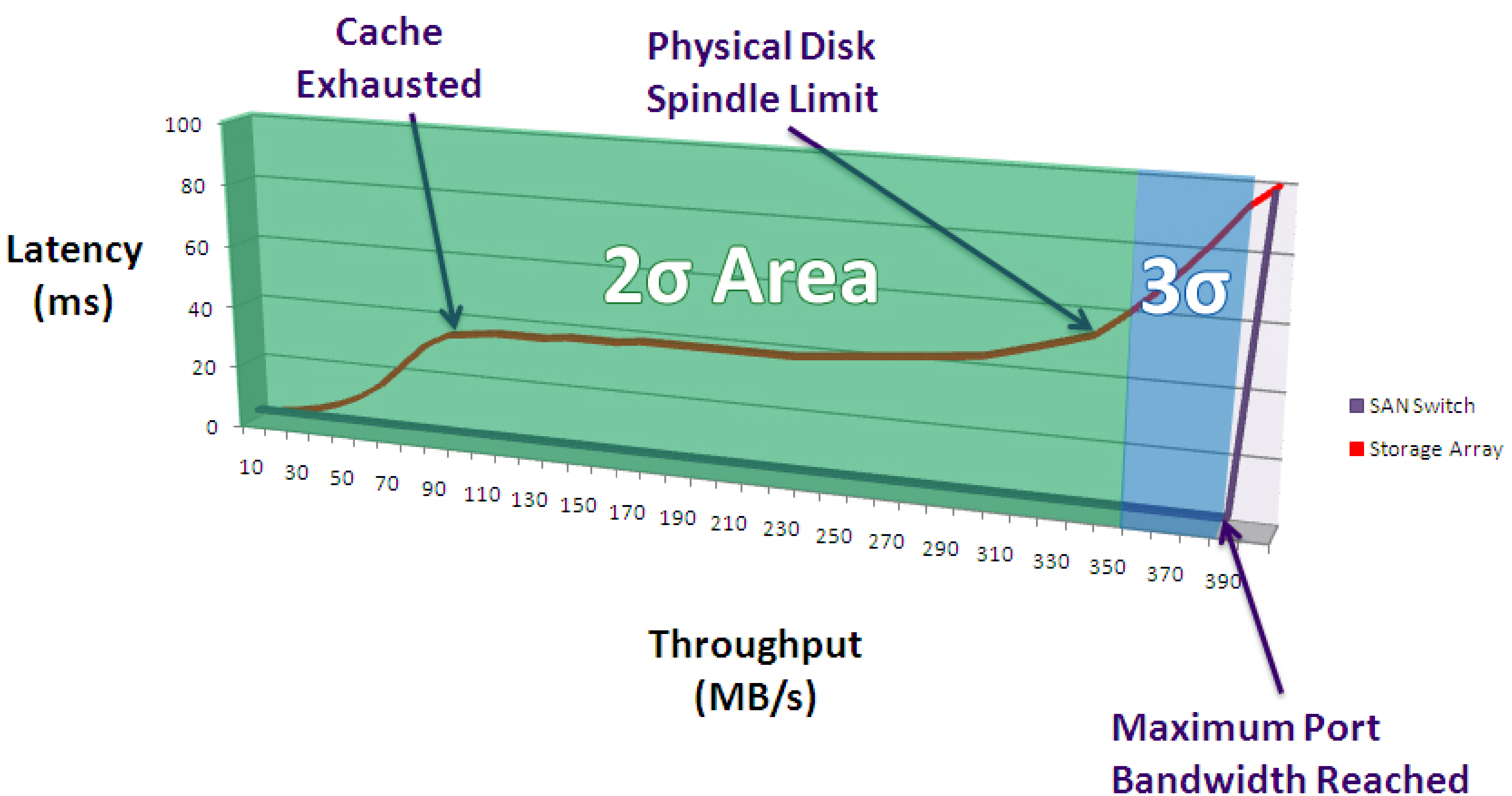


Disk Array Area of Concern



SAN Area of Concern

Education SNIA



What is Six Sigma?

- Six Sigma (6σ) literally means six standard deviations from the mean
- It's a method to improve quality through effective measuring and monitoring to attain zero defects
- 6σ is equivalent to the 99.9999998027th percentile or only 1 defect per 3.4 million
- Mostly used in manufacturing industries to obtain cost savings from operations

Six Sigma and SAN Performance

- Six Sigma is usually applied in environments where something is manufactured
- SAN's are in the business of processing and moving data
- Goal of six sigma is no defects above 6σ
- Goal of SAN no performance data is above 6σ
- 6σ can be thought of as the "Theoretical Limit" or maximum capable throughput of a port
- Cost savings can be realized by understanding and predicting throughput requirements

Six Sigma Bandwidth Prediction

- **Example:**
  - 2Gbit port (200 MB/s TX & RX)
  - Average = 150 MB/s
  - Standard deviation (σ) = 45 MB/s
- Maximum port bandwidth is 400MB/s
- 6σ is 420 MB/s (150 + (6 * 45)) which translates into the port being theoretically over-throttled
- Recommendation would be to upgrade port to 4Gbit technology to guarantee bandwidth

HA Watermark

Education SNIA

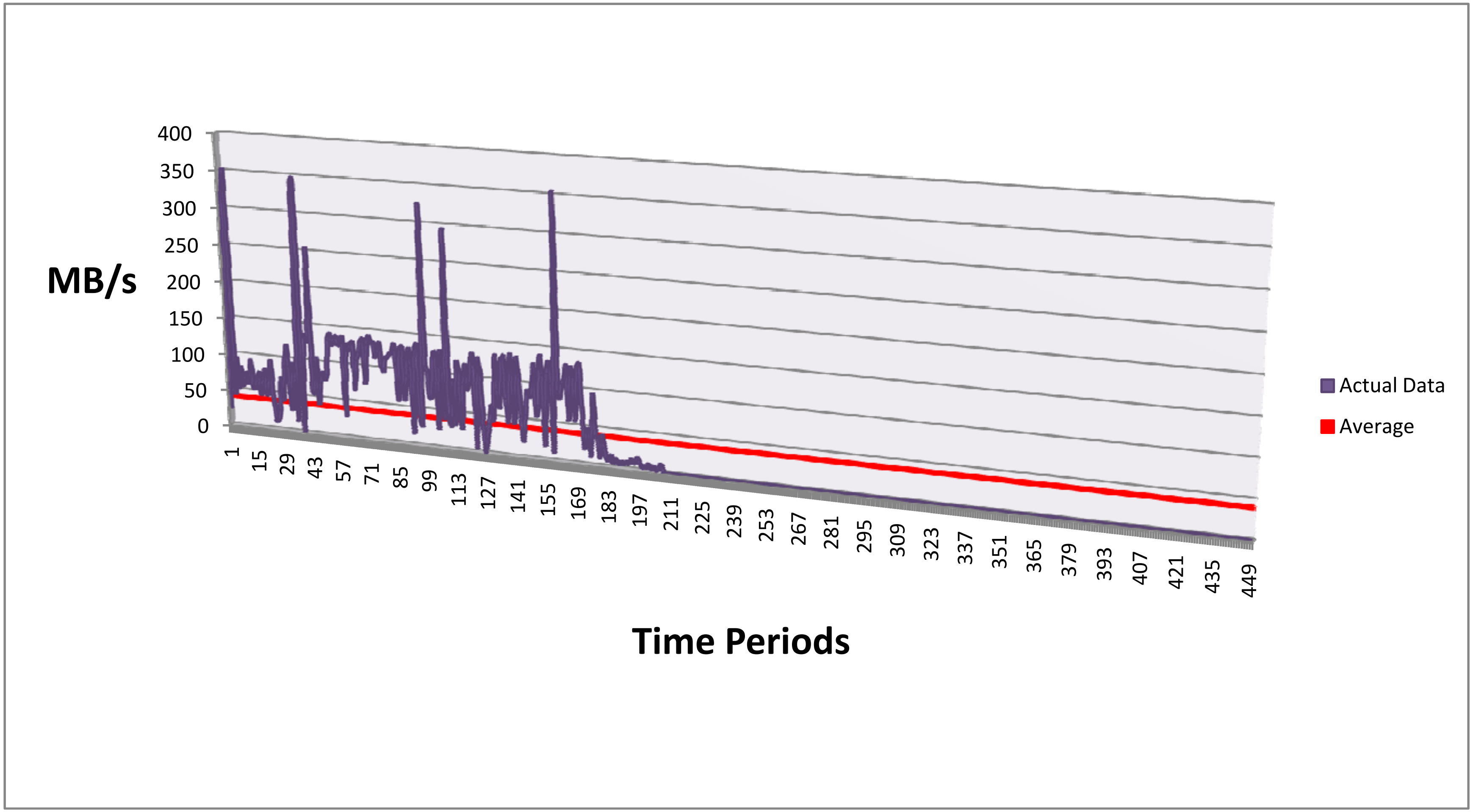
- In Highly Available (HA) environments bandwidth is reserved for failure scenarios
- Example:
  - Server has 2 x HBA's
  - Redundant SAN
  - SAN bandwidth is 100/00 or 50/50
- HA Watermark is a SAN performance level that will guarantee enough bandwidth during failures
- HA Watermark should ½ the total bandwidth and less than the 3σ mark

Example HA Watermark Calculation

- Example:
  - 4Gbit environment
  - 400 MB/s Tx & Rx = 800 MB/s total bandwidth
  - HA Watermark = 800 / 2 = 400 MB/s
- To guarantee true HA for 50/50 environments HA Watermark must less than 6σ level
- For 99.73% level of certainty HA Watermark should be less than 3σ level (Recommended for most environments)

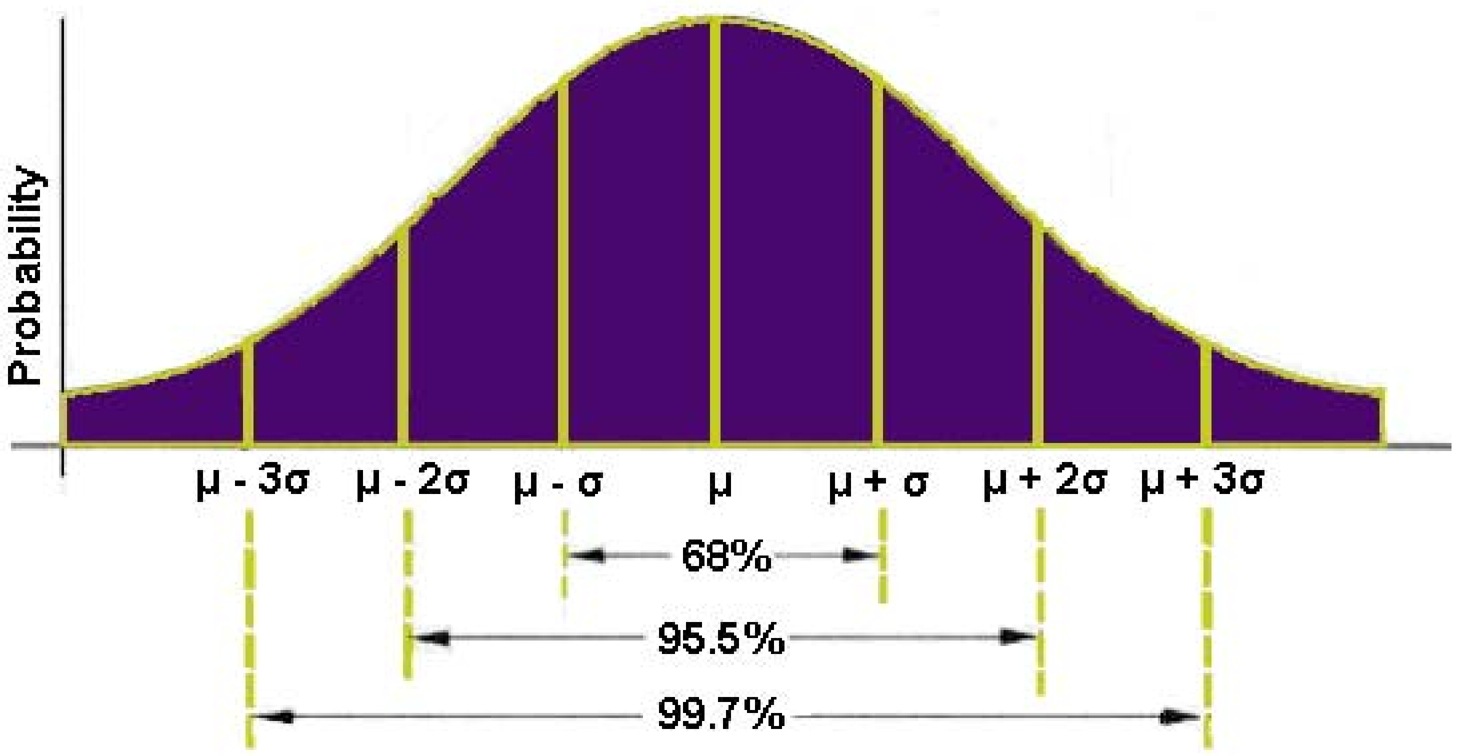
Case Study

Sample Data

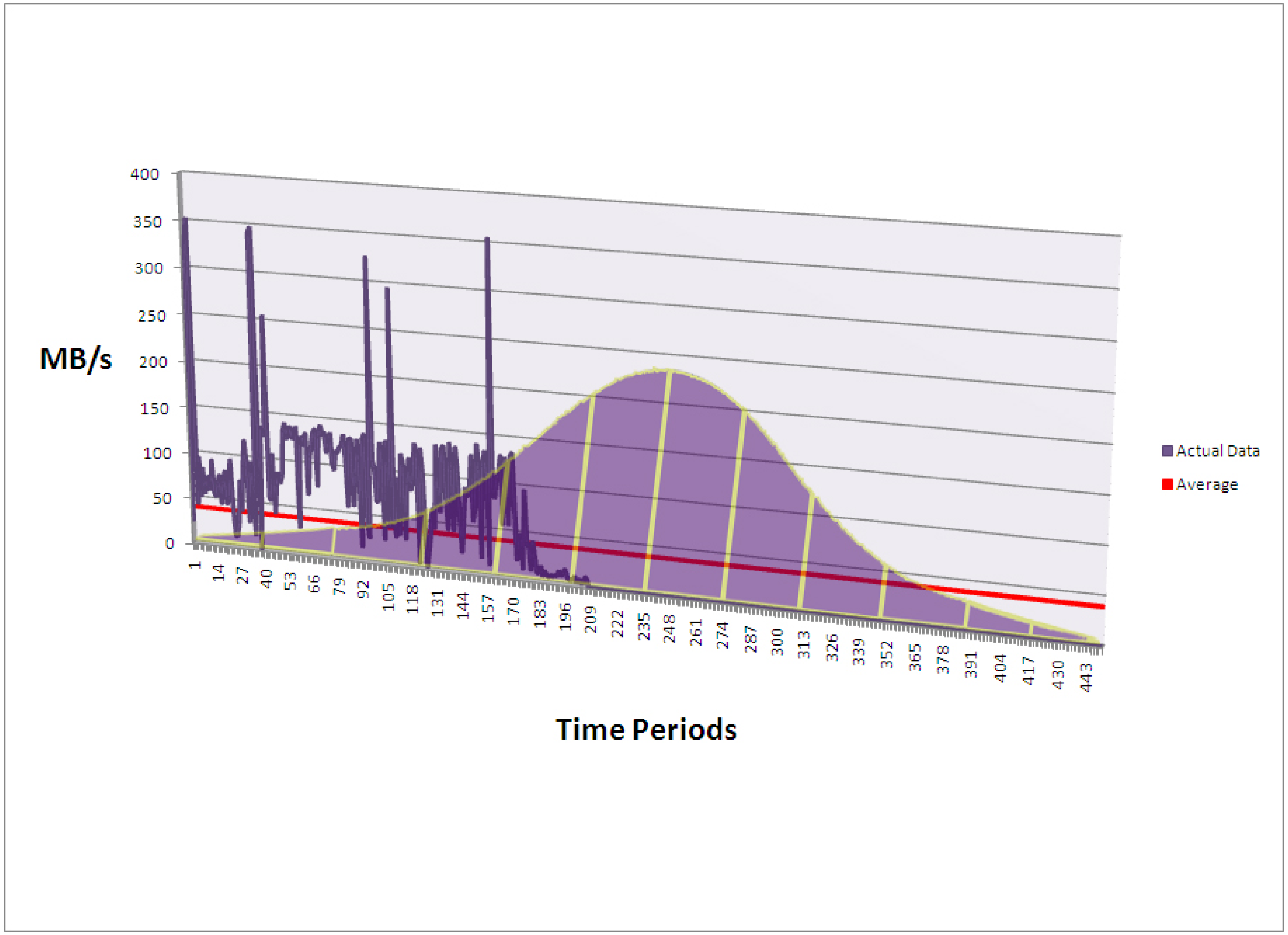


MB/s
400
350
300
250
200
150
100
50
0
1
15
29
43
57
71
85
99
113
127
141
155
169
183
197
211
225
239
253
267
281
295
309
323
337
351
365
379
393
407
421
435
449
Time Periods
Actual Data
Average

The Bell Curve

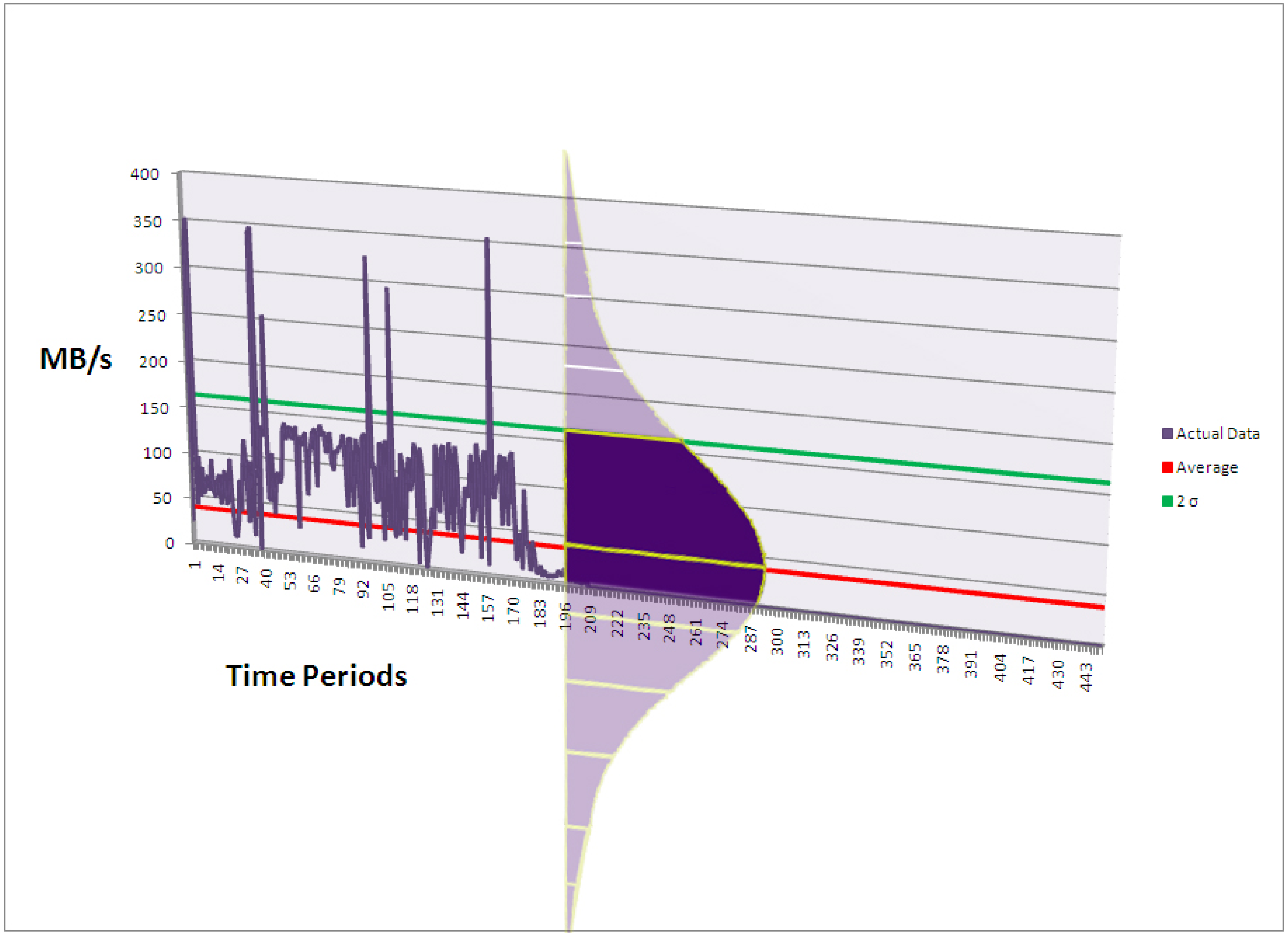


Wrong Way for SAN and Bell Curve

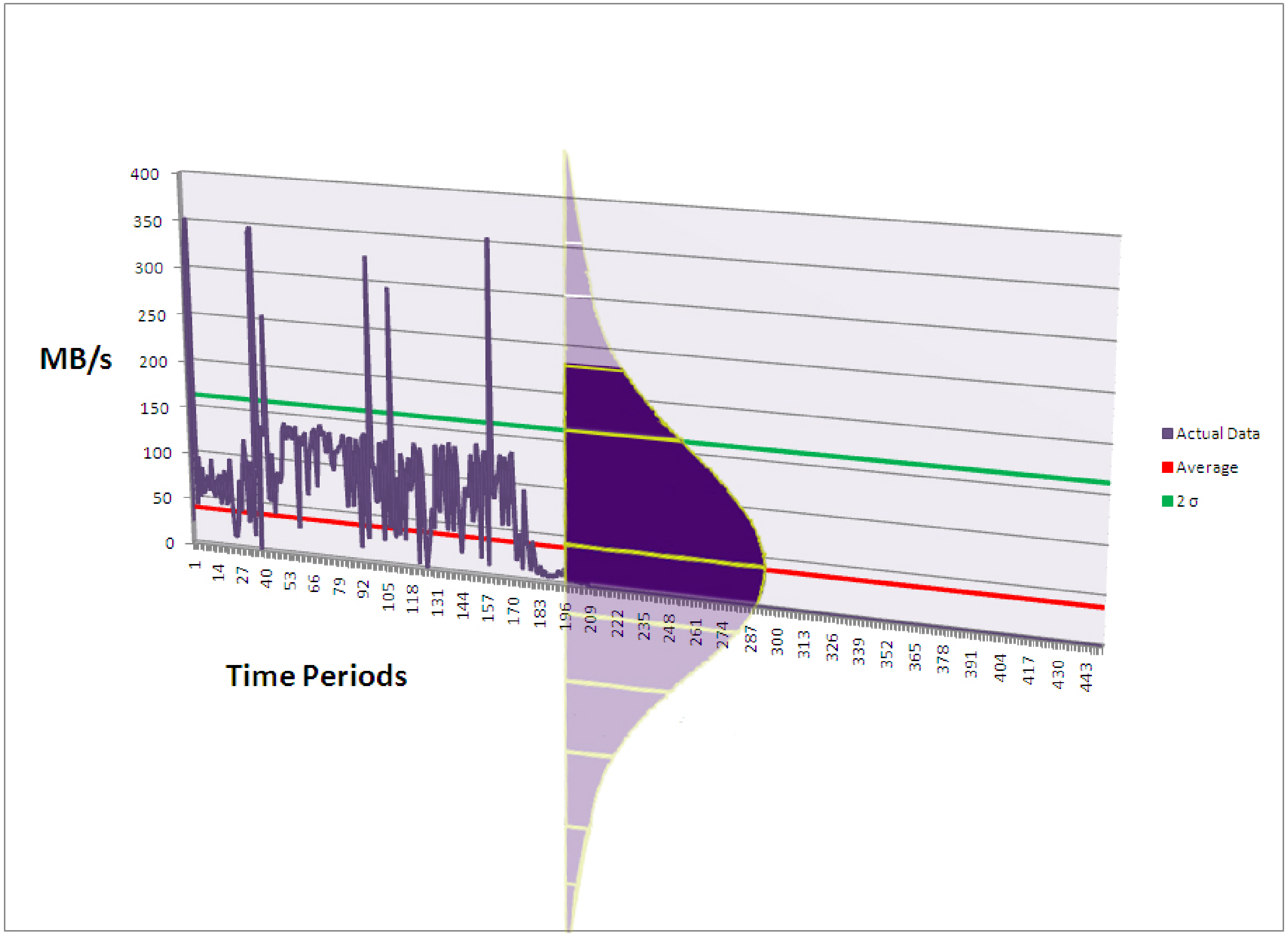


2σ (95th Percentile)

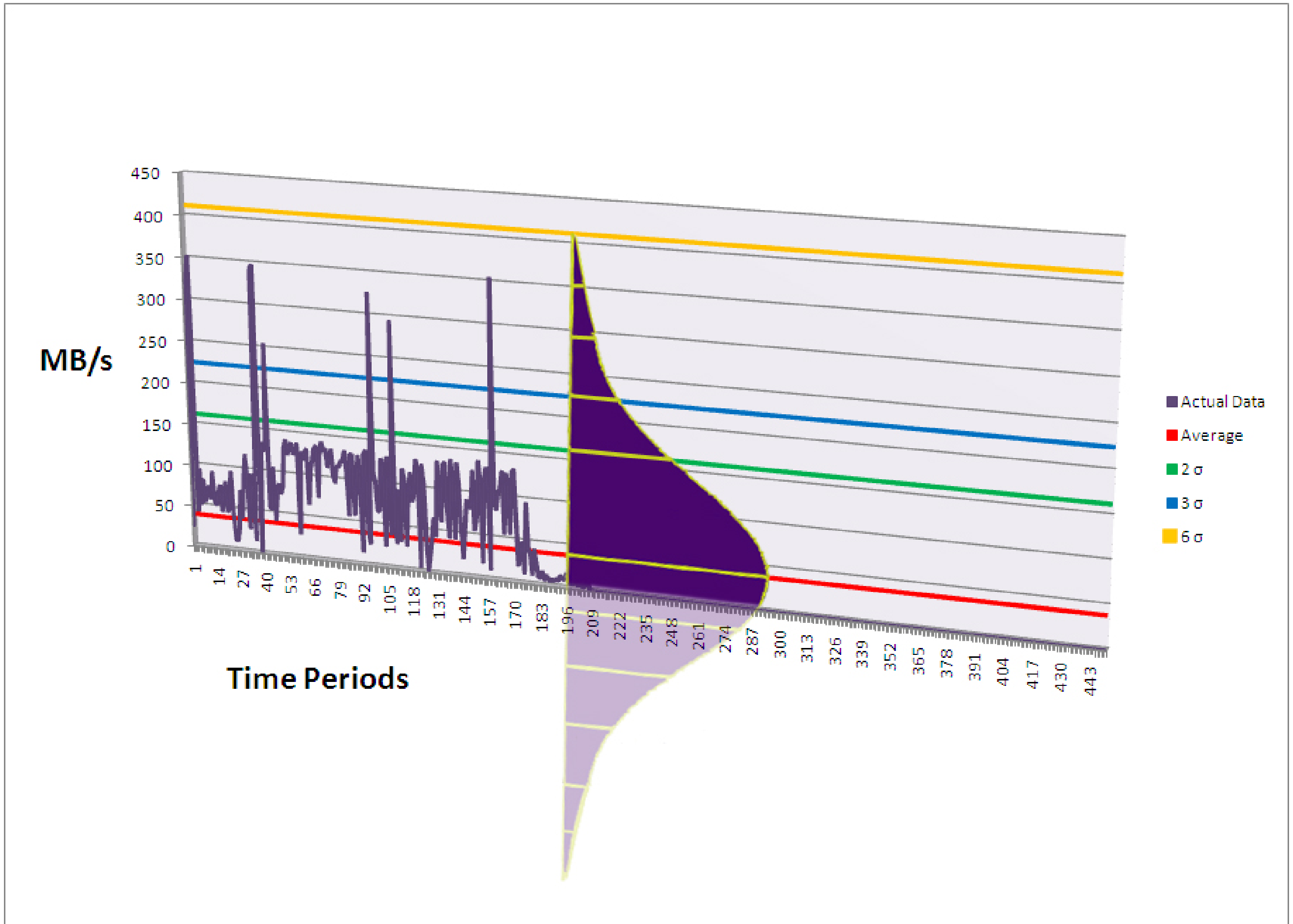
Education SNIA



3σ (99.73th Percentile)



6σ (99.999th Percentile)



Conclusion

- We are using statistics mine the SAN data
- Six Sigma is used to understand data patterns and a level of industry accepted tolerances
- σ = standard deviation or called sigma
- HA watermark = Port maximum capabilities / 2
- 6σ (99.999%) $<$ Port maximum capabilities
- 3σ (99.73%) $<$ HA watermark

Q&A / Feedback

- Please send any questions or comments on this presentation to SNIA: *tracknetworking@snia.org*

Many thanks to the following individuals for their contributions to this tutorial.

- SNIA Education Committee

Brad Wayland
Carl Nadrowski
Heon Jo
Jarred Jasper
Martin Shinners