

Load Testing Performance for SQL Server

Anthony E. Nocentino

He/Him

Senior Principal Field Solution Architect

Everpure – Formerly Pure Storage



Anthony E. Nocentino

Senior Principal Field Solution Architect @ Pure Storage

- Specialize in system architecture, performance, SQL Server, Kubernetes, Containers, Microsoft Azure and VMware

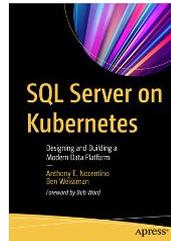
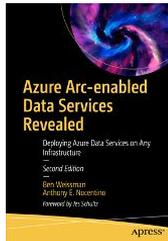
email: anocentino@purestorage.com

Blog: www.nocentino.com

GitHub: <https://github.com/nocentino/>

Pluralsight Author: www.pluralsight.com

BlueSky: <https://bsky.app/profile/nocentino.com>



Agenda

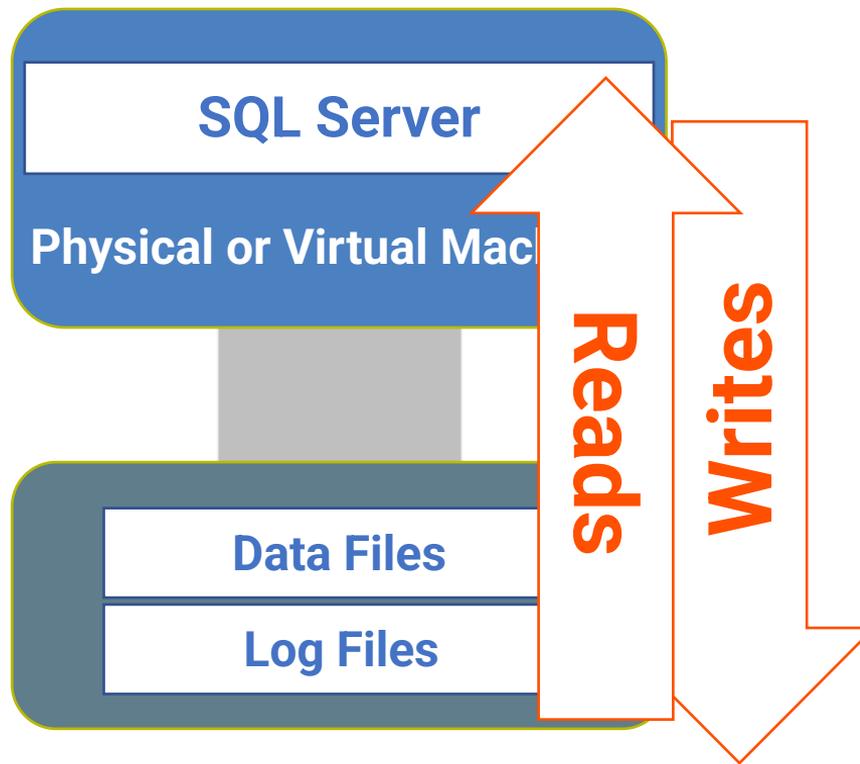
- Understanding Your Hardware's Capabilities
- Load Testing Strategies
- Interpreting Performance Data

Understanding Your Hardware's Capabilities

- You need to know what's possible with the gear you have
 - Physical Systems
 - Virtual Machines
 - Cloud
- What about higher up in the stack?
 - CPU / Transactions / Application
 - ABSOLUTELY!

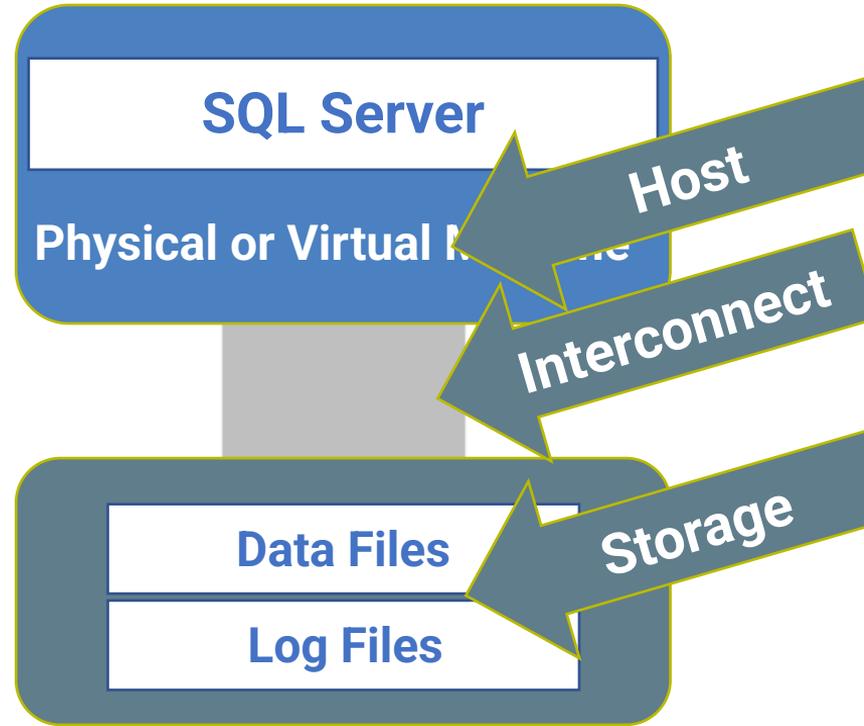
Understanding the IO Stack

- **SQL Server stores data...**
- **On disks**
 - In files
 - Local
 - SCSI
 - PCIe
 - NVMe
- **Remote**
 - Fibre Channel
 - iSCSI
 - NVMe-oF

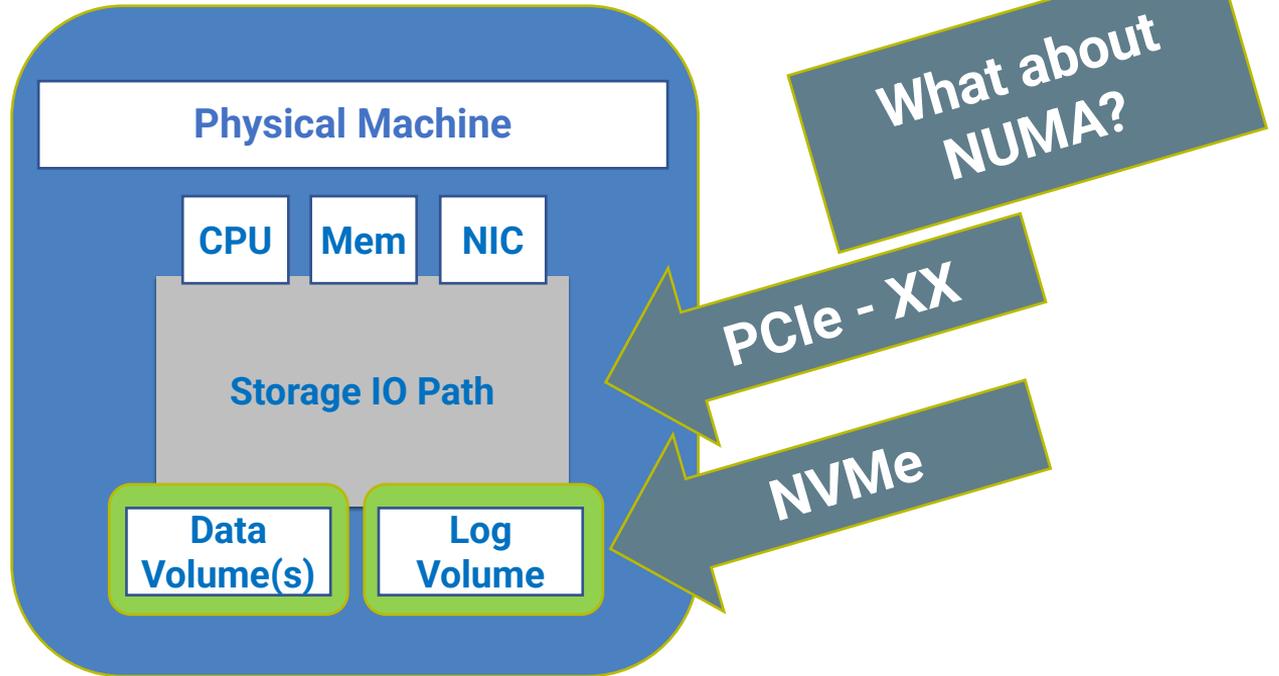


Understanding Your Hardware's Capabilities

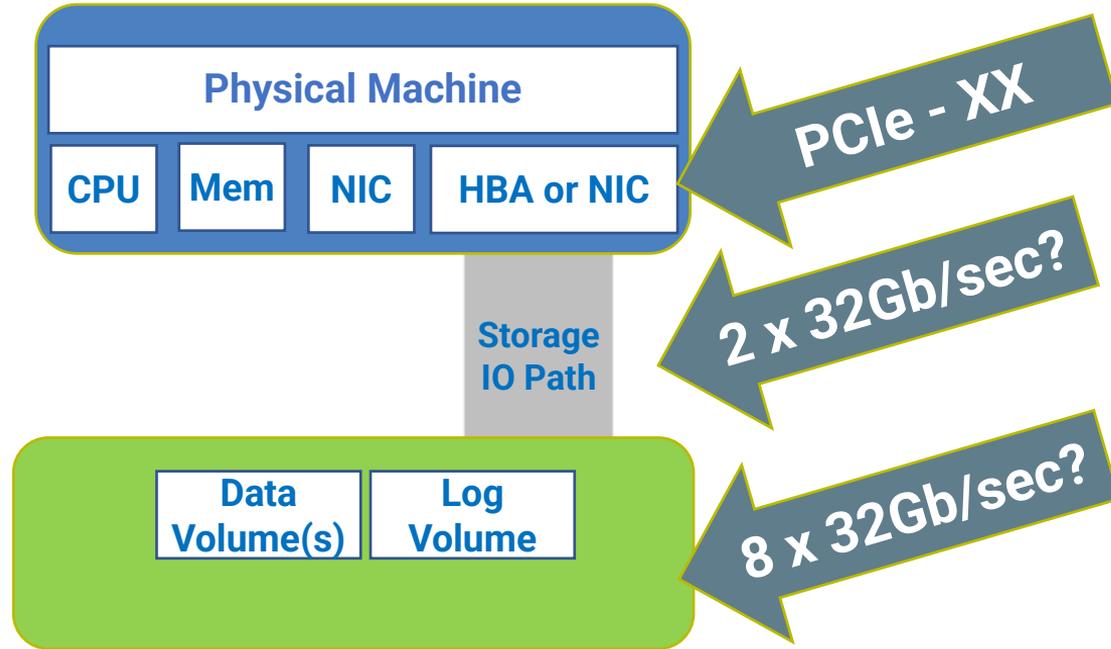
- Understand how data flows through the machine
- CPU as is important as the storage performance
- Dive deep, read the manuals



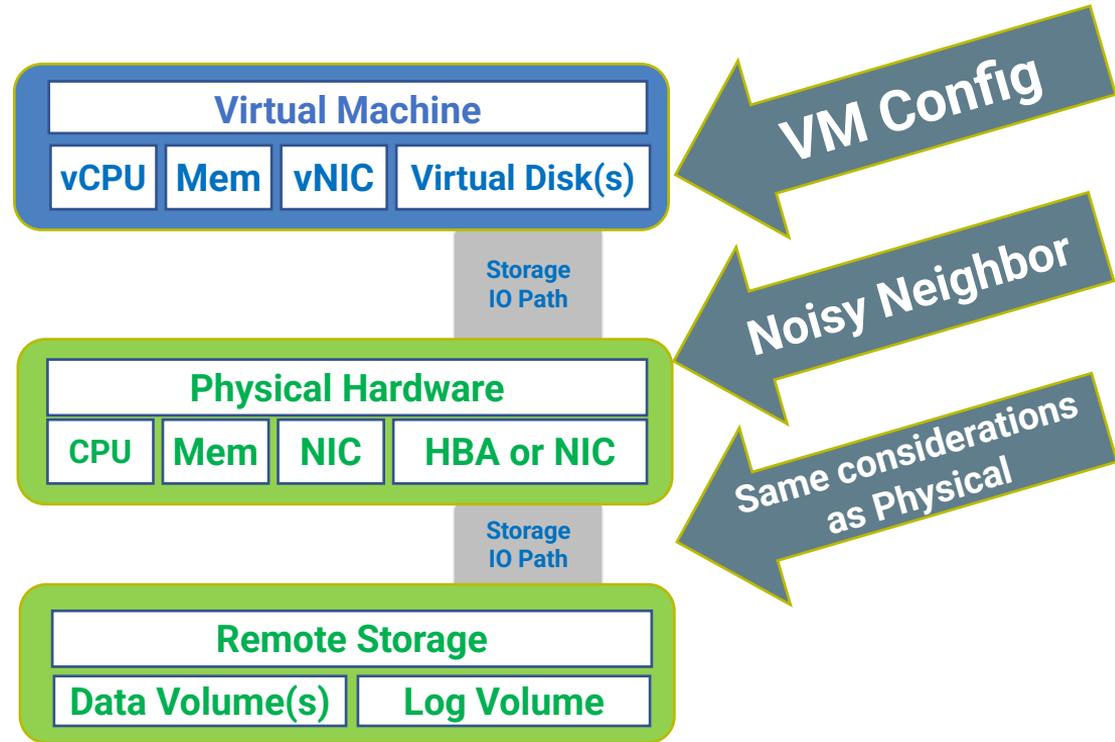
Physical Machine Considerations – Local Storage



Physical Machine Considerations – Remote Storage



Virtual Machine Considerations



Cloud Considerations

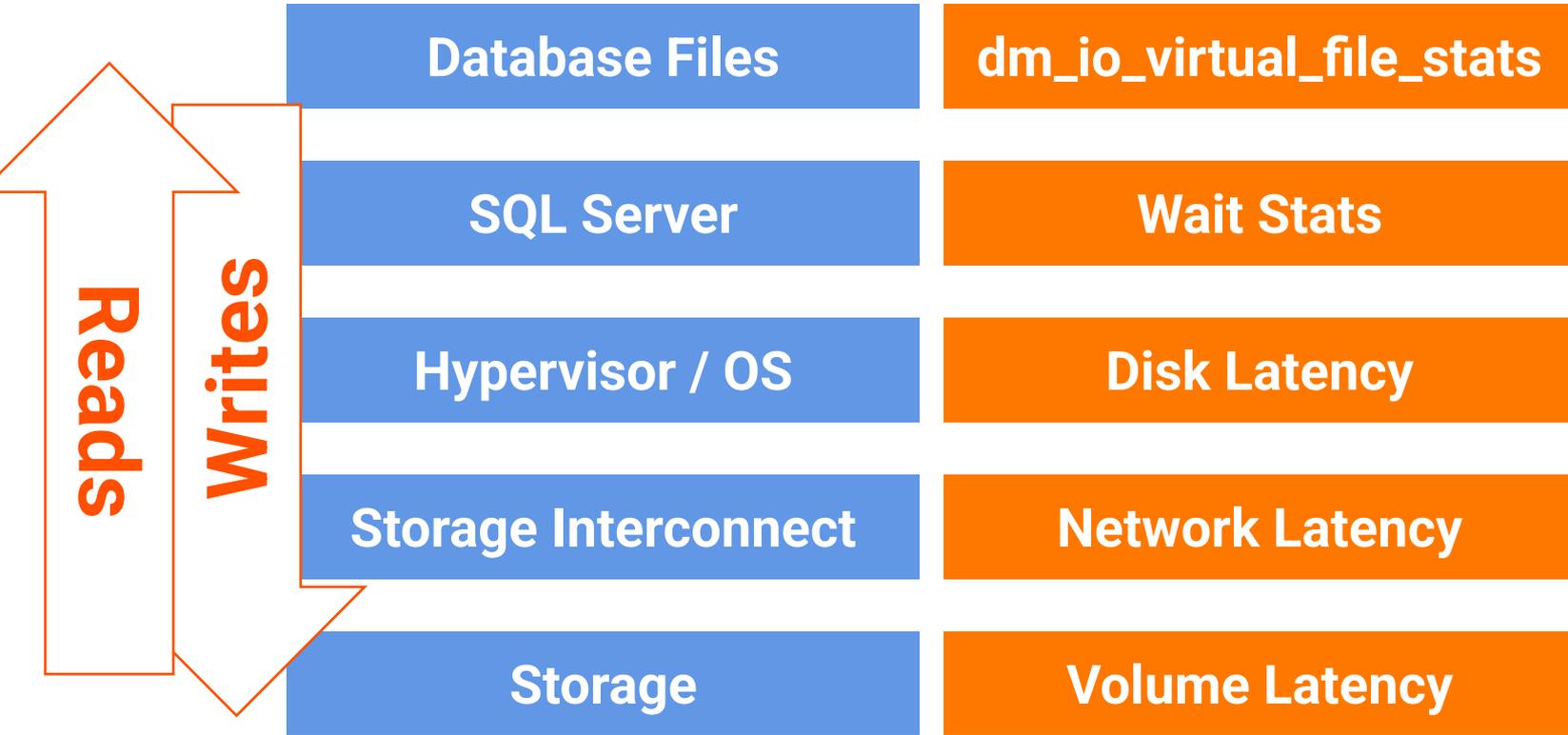
- **Understand your SKU's capabilities**
- **CPU – Number of Cores**
- **Virtual Machine IOP/Throughput Limits**
 - Local Storage
 - Remote Storage
- **Disk IOP Limits/Throughput**
 - Local Storage
 - Remote Storage
- **Network Throughput**

Cloud Considerations – Understanding VM SKUs

Size	vCPU	Memory: GiB	Temp storage (SSD) GiB	Max data disks	Max temp storage throughput: IOPS / MBps	Max uncached Premium SSD disk throughput: IOPS/MBps	Max burst uncached Premium SSD disk throughput: IOPS/MBps	Max uncached Ultra Disk and Premium SSD V2 disk throughput: IOPS/MBps	Max burst uncached Ultra Disk and Premium SSD V2 disk throughput: IOPS/MBps	Max NICs	Network bandwidth
Standard_E2bds_v5	2	16	75	4	9000/125	8000/230	10000/1200	12000/300	15000/1200	2	12500
Standard_E4bds_v5	4	32	150	8	19000/250	16000/460	20000/1200	21400/600	30000/1200	2	12500
Standard_E8bds_v5	8	64	300	16	38000/500	33000/930	40000/1200	44200/1200	60000/1200	4	12500
Standard_E16bds_v5	16	128	600	32	75000/1000	66000/1800	70000/2000	88400/2300	96000/2600	8	12500
Standard_E32bds_v5	32	256	1200	32	150000/2000	130000/3700	140000/4000	174200/4800	180000/5200	8	16000
Standard_E48bds_v5	48	384	1800	32	225000/3000	190000/5600	200000/6000	253300/7300	260000/7850	8	16000
Standard_E64bds_v5	64	512	2400	32	300000/4000	220000/6000	230000/6500	294800/7800	310000/8500	8	20000
Standard_E96bds_v5	96	672	3600	32	450000/4000	260000/7500	260000/8000	390000/8500	390000/9000	8	25000
Standard_E112ibds_v5	112	672	3800	64	450000/4000	260000/8000	260000/8000	400000/10000	400000/10000	8	40000

Elastic Series (NVM)

Understanding Queuing



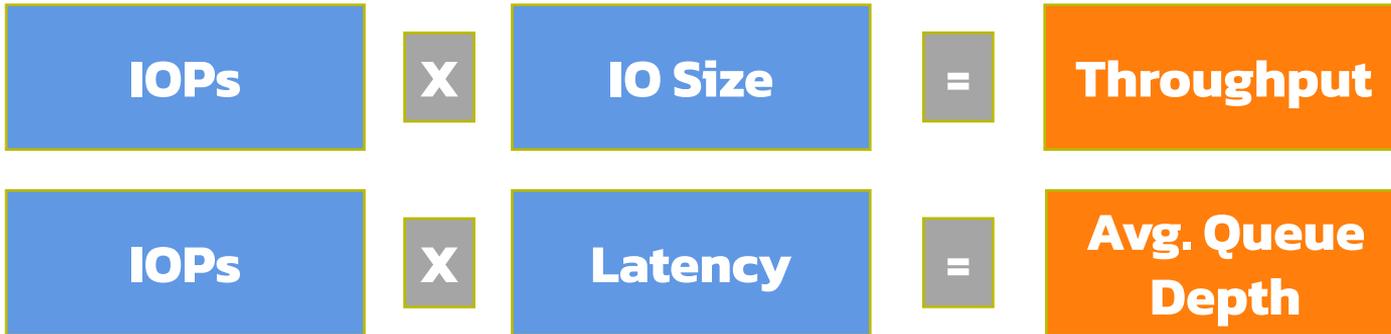
Each layer presents a potential to inject latency

If a layer can't support the IO demand, then queuing can occur

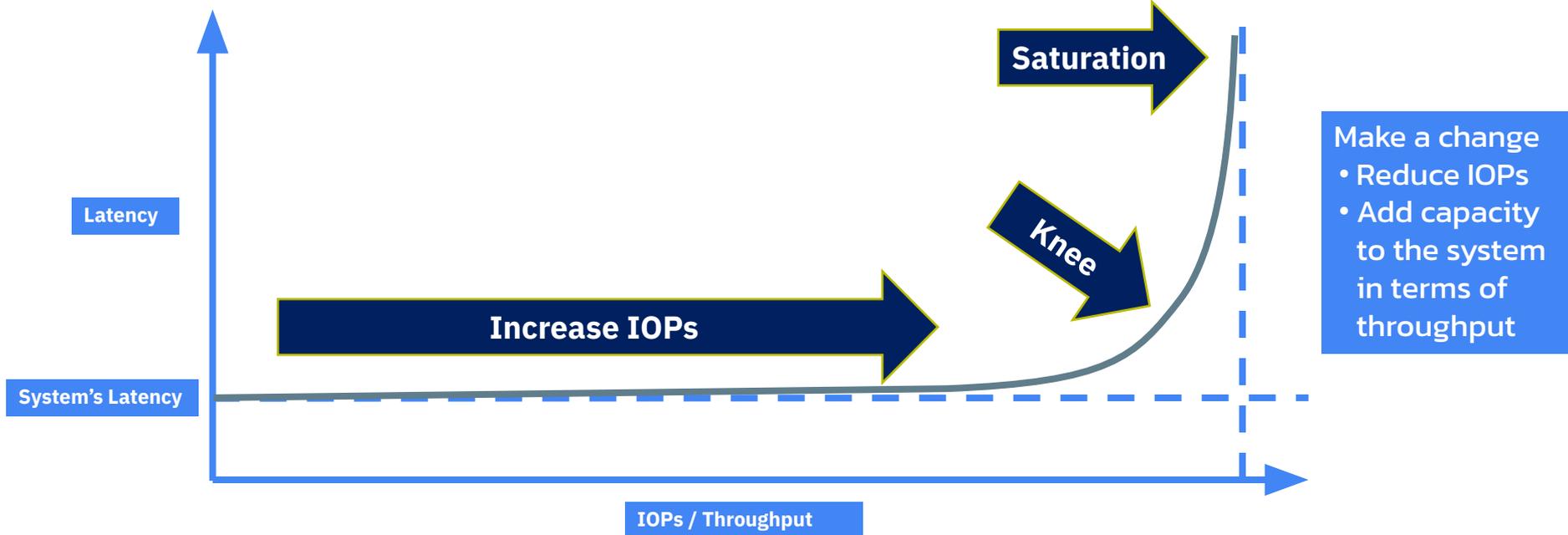
Little's Law

- **Three Key Characteristics of Process Behavior:**
 - **Throughput (IOPS):** Input/output operations per second
 - **Queue Time (Latency):** Time data spends waiting in the queue
 - **Inventory (Queue Depth):** Number of requests in the queue
- **Influence on System Performance:**
 - You can change one characteristic to impact the other two
 - Use Little's Law to analyze and optimize processes

Finding the Maximum in a System



You Goal When You're Designing a Load Test – Increase IOPs until latency is unacceptable



Load Testing Strategies

- **Synthetic**
- **Real Workloads**

What Are You Testing For?

- **Synthetic**
 - Throughput
 - Latency
- **Real Workloads**
 - Transactions per second
 - Critical stored procedure
 - Key Performance Indicator
 - Add to cart?
 - Checkout?

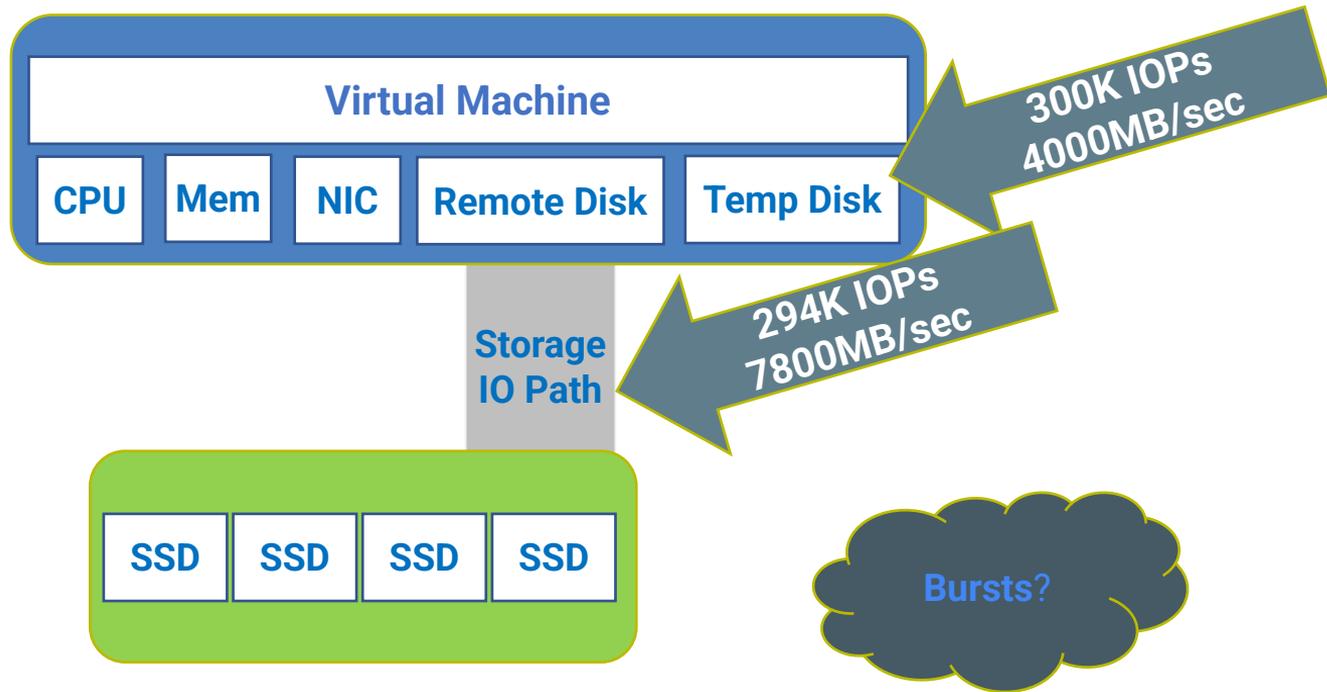
What Are You Testing For?

- **OLTP vs OLAP**
 - Latency
 - Throughput
- **SQL Server does 8KB IOs, right? What about 64KB IOs?**
- **Average IO sizes**
 - Read and writes
 - You can get these from your current SQL Server
- **Log flush – 512B-60KB**
- **Read-ahead – up to 512KB**
- **Backups – 1MB-4MB**

Load Testing Tools

DiskSpd	HammerDB	Workload Tools	DEA	SQL Server
Synthetic	Synthetic	Real Workload	Real Workload	Hybrid
Only Storage	Standard Workloads	Playback Your Workload	Playback Your Workload	Actual Workload
				Also “Synthetic”

My Test Rig – Azure VM – Edbsv5 (NVMe)



HammerDB

- Open-source database benchmarking and load testing tool
- Supports major databases: SQL Server, Oracle, PostgreSQL, MySQL, MariaDB, Db2
- Implements industry-standard workloads (TPC-C, TPC-H)
- Simulates OLTP and OLAP environments
- Scales from single-user tests to thousands of virtual users
- CLI and GUI interfaces for automation and CI/CD integration
- Widely used for performance validation, capacity planning, and regression testing

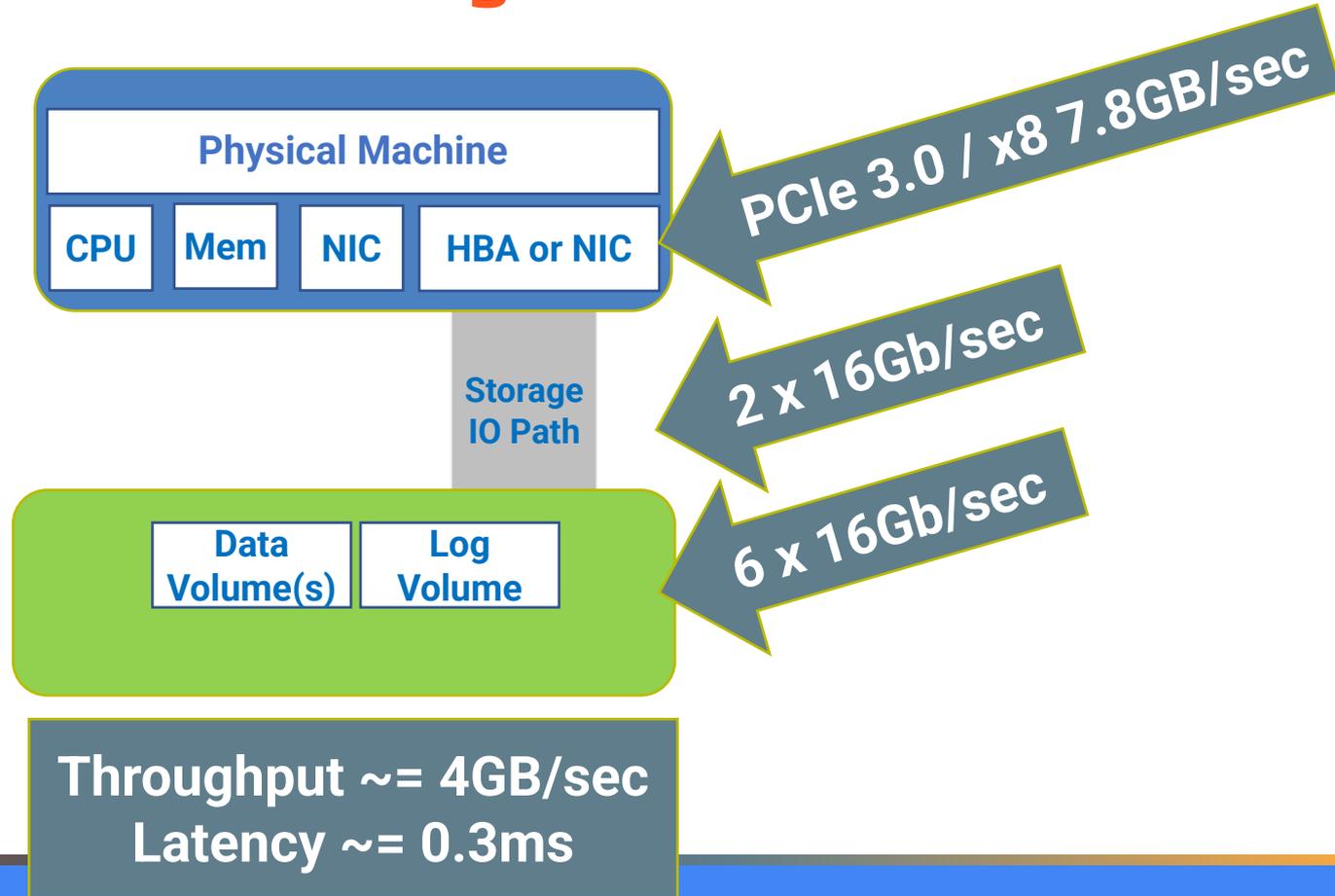
Interpreting HammerDB Metrics – Core Metrics

- TPM (Transactions Per Minute)
 - Total committed TPC-C transactions across all types
 - Includes Payment, Delivery, Order Status, Stock Level, and New Order
- NOPM (New Orders Per Minute)
 - TPC-C compliant primary metric
 - Represents committed *New Order* transactions only
 - Used for cross-platform performance comparison
- QPHH (Query Per Hour @ Scale Factor)
 - TPC-H composite metric
 - Derived from Power Test + Throughput Test

Interpreting HammerDB Metrics - Scaling Analysis

- Throughput vs VU curve (linear region -> saturation -> contention)
- Knee-point identification
- CPU-bound vs I/O-bound determination
- Locking and latch contention analysis

My Test Rig – Remote Storage



Key Performance Metrics

Latency

How Long a Request Takes

Transaction Log I/Os

Saturation Leads to Queuing

Throughput

Amount of data moved

Function of the storage interconnect

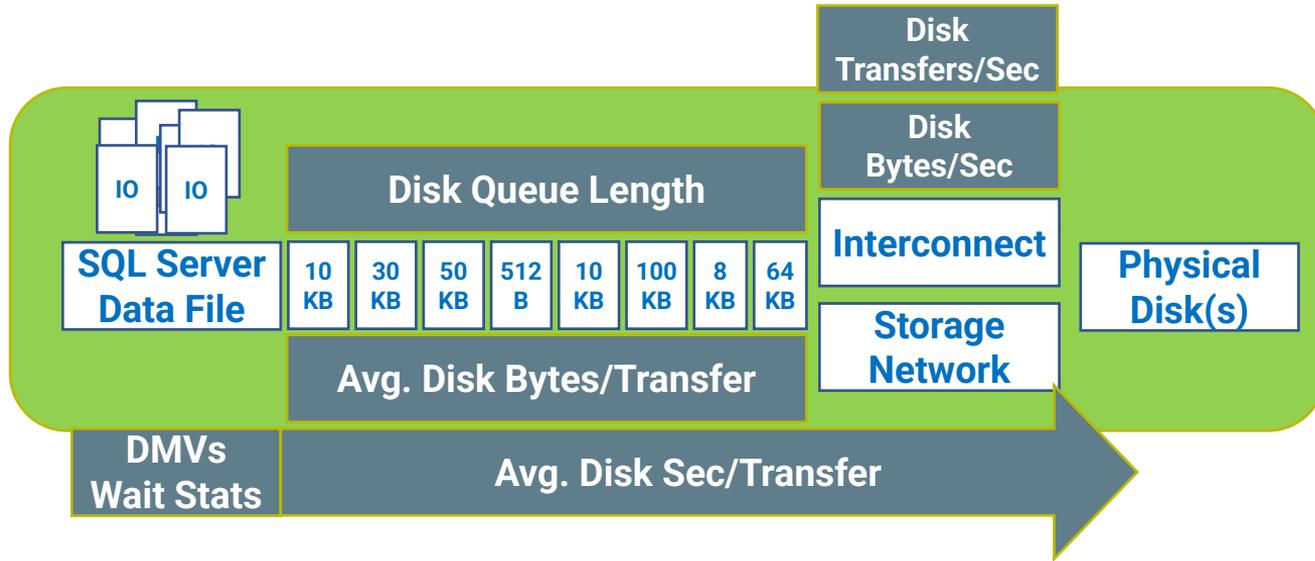
Type of storage used

IOPs

Number of Requests

Size Depends on the Application

Understanding the Timeline of an IO



Collecting Data

- **Perfmon**
- **Resource Monitor**
- **PowerShell**
- **SQL Server DMVs**
- **Hypervisor / Cloud**

Interpreting Data

- **Understanding What's Good and What's Bad?**
- **Hardware Capabilities**
- **You're going to have to read the manuals!**

So...How Do YOU Design a Load Test

- It depends...

Review

- Understanding Your Hardware's Capabilities
- Load Testing Strategies and Tools
- Interpreting Performance Data