

# Cloud and Databases 101: Value, Sizing, and Best Practices

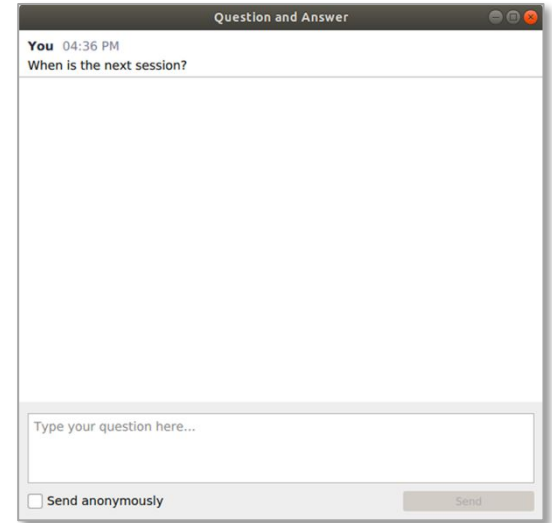
Tom Rieger

Senior Client Engineer



# House Keeping

- We have a lot of materials to cover and plan on keep this within 45 minutes.
- In using Zoom's Webinar technology
  - Everyone is on <MUTE>
  - Questions are entered into the tool
  - Online Polling will be used
  - We are recording the webinar and will share the private link shortly after the session







# Polling question

## Favorite Season?

- Spring
- Summer
- Fall
- Winter
- Don't have one – love them all

# Agenda

What I hope you take away from this session

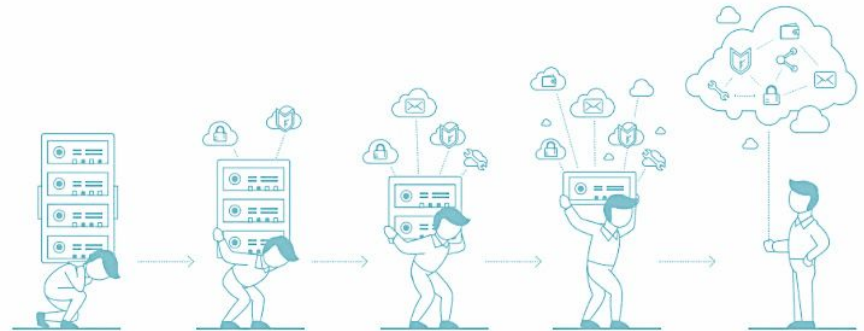
- Cloud
- History of database sizing and measure
- Workloads
- Our 'sizing guide' effort
- Results – thus far
- Lessons learned
- What is next



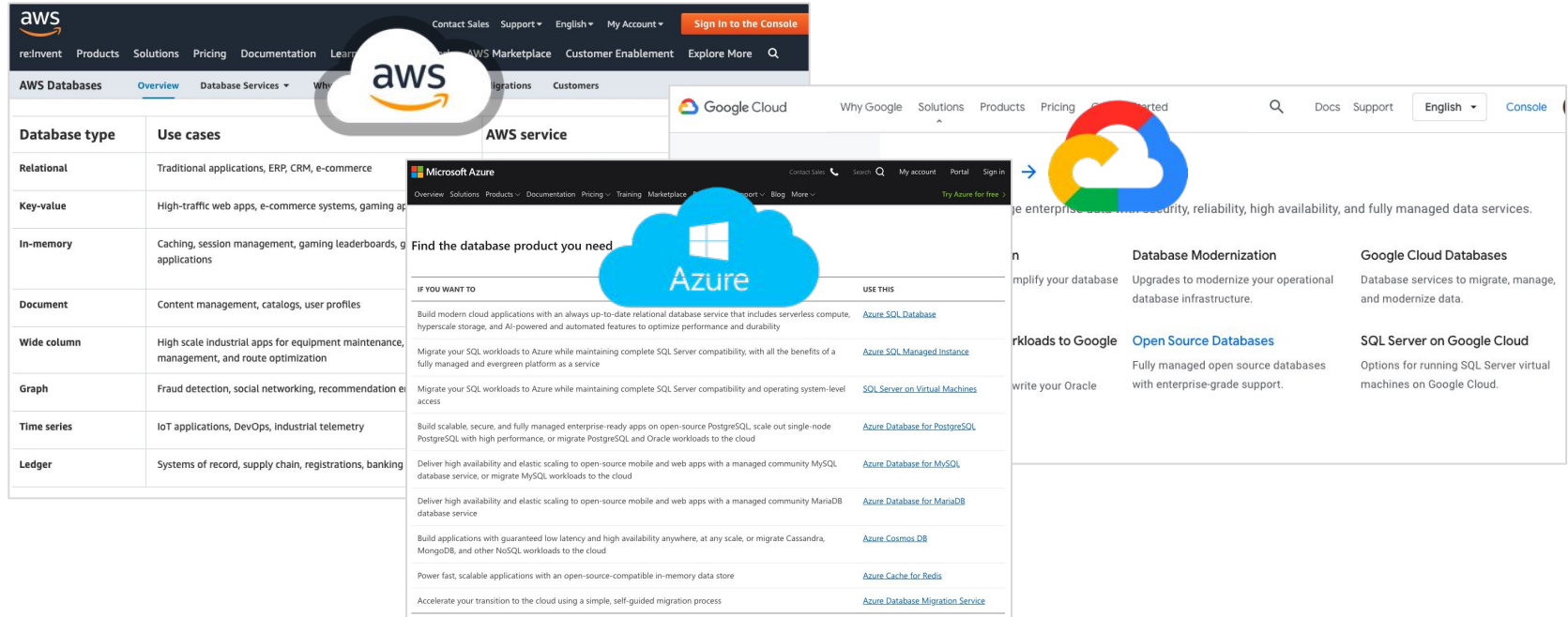
# Cloud

It took evolutionary waves to get us to this point

- Email systems, websites, webcasting/webinars
- HR apps, CRM (Salesforce), file sharing
- Mobile-centric efforts
- Secondary database workloads with smaller requirements
- .....Critical workloads are next



# Cloud platform vendors make ‘databases’ confusing and ‘lock in’



The image shows three overlapping screenshots of cloud provider database service pages. The AWS screenshot on the left features a table of database types and use cases. The Google Cloud screenshot on the right shows a grid of database services. The Microsoft Azure screenshot in the center displays a 'Find the database product you need' section with a list of services and their descriptions.

Database type	Use cases
Relational	Traditional applications, ERP, CRM, e-commerce
Key-value	High-traffic web apps, e-commerce systems, gaming applications
In-memory	Caching, session management, gaming leaderboards, gaming applications
Document	Content management, catalogs, user profiles
Wide column	High scale industrial apps for equipment maintenance, management, and route optimization
Graph	Fraud detection, social networking, recommendation engines
Time series	IoT applications, DevOps, industrial telemetry
Ledger	Systems of record, supply chain, registrations, banking

**Microsoft Azure**

Find the database product you need

IF YOU WANT TO	USE THIS
Build modern cloud applications with an always up-to-date relational database service that includes serverless compute, hyperscale storage, and AI-powered and automated features to optimize performance and durability	<a href="#">Azure SQL Database</a>
Migrate your SQL workloads to Azure while maintaining complete SQL Server compatibility, with all the benefits of a fully managed and evergreen platform as a service	<a href="#">Azure SQL Managed Instance</a>
Migrate your SQL workloads to Azure while maintaining complete SQL Server compatibility and operating system-level access	<a href="#">SQL Server on Virtual Machines</a>
Build scalable, secure, and fully managed enterprise-ready apps on open-source PostgreSQL, scale out single-node PostgreSQL with high performance, or migrate PostgreSQL and Oracle workloads to the cloud	<a href="#">Azure Database for PostgreSQL</a>
Deliver high availability and elastic scaling to open-source mobile and web apps with a managed community MySQL database service, or migrate MySQL workloads to the cloud	<a href="#">Azure Database for MySQL</a>
Deliver high availability and elastic scaling to open-source mobile and web apps with a managed community MariaDB database service	<a href="#">Azure Database for MariaDB</a>
Build applications with guaranteed low latency and high availability anywhere, at any scale, or migrate Cassandra, MongoDB, and other NoSQL workloads to the cloud	<a href="#">Azure Cosmos DB</a>
Power fast, scalable applications with an open-source-compatible in-memory data store	<a href="#">Azure Cache for Redis</a>
Accelerate your transition to the cloud using a simple, self-guided migration process	<a href="#">Azure Database Migration Service</a>

**Google Cloud**

Database Modernization

Upgrade to modernize your operational database infrastructure.

Database services to migrate, manage, and modernize data.

Open Source Databases

Fully managed open source databases with enterprise-grade support.

SQL Server on Google Cloud

Options for running SQL Server virtual machines on Google Cloud.



# Polling question

Q. Do you feel there is a degree of confusion in the database offerings from cloud vendors?

- YES – very confusing
- Somewhat
- Not really – very clear
- Not sure – to be honest



# History of database benchmarking and sizing

- General workloads
  - Transactional
  - Analytical
- Vendor-specific workloads
  - Database-specific
  - Vendor-specific – SAP SD
- More fit-for-purpose
  - Load testing
  - Performance testing
  - IO performance
- Historically benchmarks were focused on **performance NOT economics**

TPC®

-VS-

\$\$



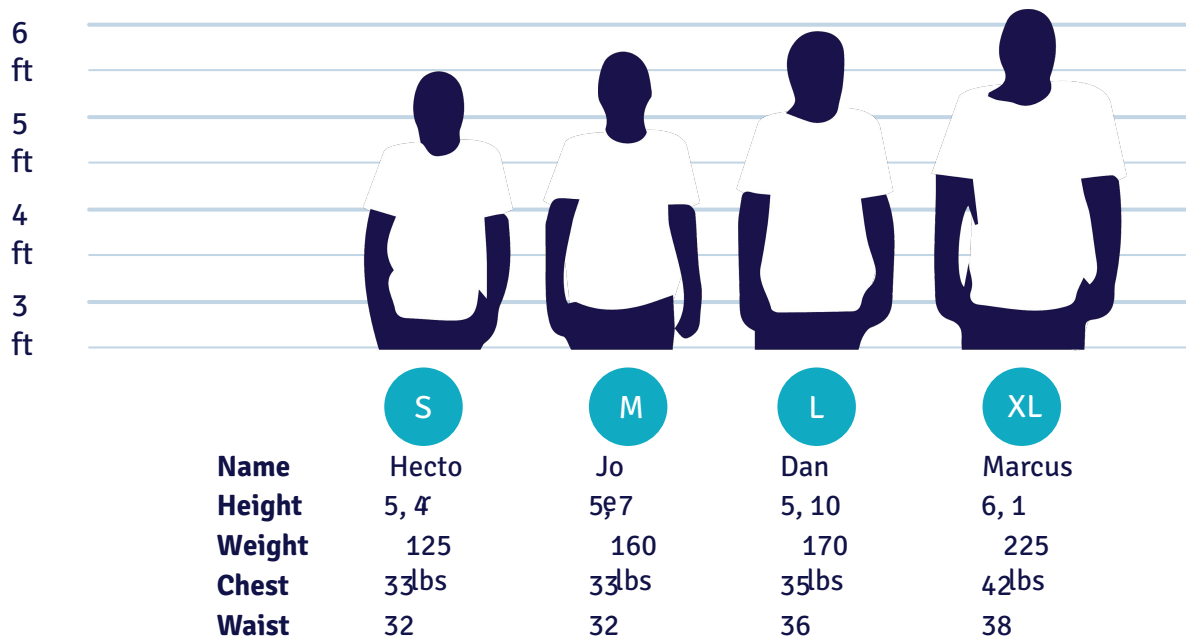
# Traditionally, IT sizing was like a ‘tailored suit’...

Up to now:

‘Educated guesses’ and  
‘release note analysis’  
to decide  
sizing and cost.



# The cloud allows a more elastic 't-shirt' sizing



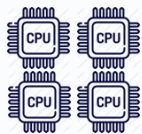
# Think of cloud provisioning as 't-shirt' sizes

SMALL

MEDIUM

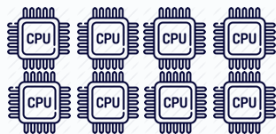
LARGE

X-LARGE



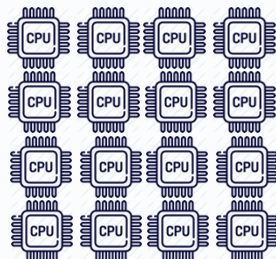
4 CPU

32 GB RAM



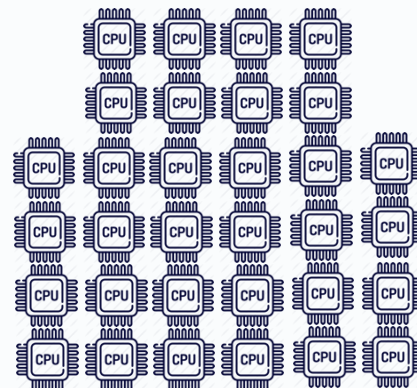
8 CPU

64 GB RAM



16 CPU

128 GB RAM



32 CPU

256 GB RAM

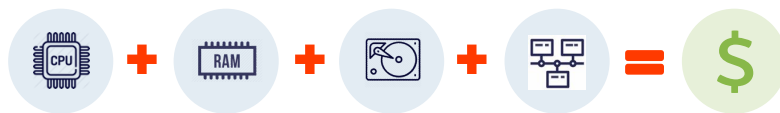
VARIABLE

CONSTANT

Same IO/storage, network, database size, concurrency, workload

# The effort at hand - OPTIMIZATION

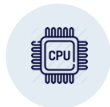
- Evaluate the optimal mix of CPU, RAM, storage and network available
- Use a defensible, standards-based workload that clients can also utilize in their own environment
- Measure technical results, time studies and financial business case
- Publish application works for EDB clients – scripts, videos, calculators, best practices, how-tos, business case assessment templates





# Example: Amazon brings a lot of choices

44 different  
choices with '8  
cores'



Purpose	Instance Name	CPU	RAM	EBS bandwidth	Network bandwidth	Storage	PER MONTH
General Purpose	a1.2xlarge	8	16		Up to 10	EBS Only	\$93.81
General Purpose	t4g.2xlarge	8	32	Up to 2,780	Up to 5 Gbps	EBS Only	\$123.08
General Purpose	c6g.2xlarge	8	16	Up to 4,750	Up to 10	EBS Only	\$125.12
General Purpose	t3a.2xlarge	8	32		Up to 5 Gbps	EBS Only	\$137.68
General Purpose	m6g.2xlarge	8	32	Up to 4,750	Up to 10	EBS Only	\$141.04
General Purpose	c6gd.2xlarge	8	16	Up to 4,750	Up to 10	1 x 474 NVMe SSD	\$141.25
General Purpose	c5a.2xlarge	8	16	Up to 3,170	Up to 10	EBS Only	\$141.62
General Purpose	t3.2xlarge	8	32		Up to 5 Gbps	EBS Only	\$152.28
General Purpose	c5.2xlarge	8	16	Up to 4,750	Up to 10	EBS Only	\$156.22
General Purpose	c6gn.2xlarge	8	16	Up to 9.5	Up to 25 Gbps	EBS Only	\$159.21
General Purpose	t2.2xlarge	8	32		Moderate	EBS Only	\$167.90
General Purpose	c5d.2xlarge	8	16	Up to 4,750	Up to 10	1 x 200 NVMe SSD	\$176.66
General Purpose	m5.2xlarge	8	32	Up to 4,750	up to 10Gbps	EBS Only	\$176.66
General Purpose	m4.2xlarge	8	32	1,000	High	EBS Only	\$180.89
General Purpose	c4.2xlarge	8	15	1,000	High	EBS Only	\$183.96
Memory Optimized	r6g.2xlarge	8	64	Up to 4,750	Up to 10	EBS Only	\$185.42
General Purpose	c5n.2xlarge	8	16	Up to 4,750	Up to 25 Gbps	EBS Only	\$198.56
Memory Optimized	r5a.2xlarge	8	64	Up to 2,880	Up to 10	EBS Only	\$208.05
General Purpose	m5d.2xlarge	8	32	Up to 4,750	Up to 10Gbps	1 x 300 NVMe	\$208.05
Memory Optimized	r6gd.2xlarge	8	64	Up to 4,750	Up to 10	1 x 474 NVMe SSD	\$211.92
General Purpose	m5n.2xlarge	8	32	Up to 4,750	Up to 25 Gbps	EBS Only	\$219.00
Storage Optimized	h1.2xlarge	8	32		Up to 10 Gigabit	1 x 2,000GB HDD	\$232.14
Memory Optimized	r5.2xlarge	8	64	Up to 4,750	Up to 10 Gbps	EBS Only	\$232.14
Memory Optimized	r5ad.2xlarge	8	64	Up to 2,880	Up to 10	1 x 300 NVMe SSD	\$240.90
Memory Optimized	r4.2xlarge	8	61		Up to 10	EBS Only	\$245.28
General Purpose	m5dn.2xlarge	8	32	Up to 4,750	Up to 25 Gbps	1 x 300 NVMe	\$250.39
Memory Optimized	r5d.2xlarge	8	64	Up to 4,750	Up to 10 Gbps	1 x 300 NVMe	\$264.99
Accelerated Computing	inf1.2xlarge	8	16	Up to 4.75 Gbps	Up to 25 Gbps	EBS Only	\$268.64
Accelerated Computing	f1.2xlarge	8	122		Up to 10	470 GB	\$268.64
Memory Optimized	r5n.2xlarge	8	64	Up to 4,750	Up to 25 Gbps	EBS Only	\$273.75
Memory Optimized	r5b.2xlarge	8	64	Up to 10,000	Up to 10	EBS Only	\$274.12
Memory Optimized	r5b.2xlarge	8	64	Up to 10,000	Up to 10 Gbps	EBS Only	\$274.12
General Purpose	m5zn.2xlarge	8	32	3,170	Up to 25 Gbps	EBS Only	\$303.83
Memory Optimized	r5dn.2xlarge	8	64	Up to 4,750	Up to 25 Gbps	1 x 300 NVMe	\$307.33
Storage Optimized	i3.2xlarge	8	61		Up to 10 Gbps	1 x 1,900 NVMe	\$312.44
Memory Optimized	z1d.2xlarge	8	64		Up to 10 Gigabit	1 x 300 NVMe SSD	\$342.27
Accelerated Computing	g4dn.2xlarge	8	32		Up to 25 Gbps	225 GB	\$346.02
Storage Optimized	i3en.2xlarge	8	64		Up to 25 Gbps	2 x 2,500 NVMe SSD	\$449.68
Storage Optimized	d3.2xlarge	8	64	1,700	Up to 15	6 x 2TB HDD	\$459.17
Storage Optimized	d2.2xlarge	8	61		High	6 x 2000 HDD	\$586.92
Memory Optimized	x1e.2xlarge	8	122	Dedicated 500 Mbps	Up to 10 Gbps	EBS + 120GB SSD	\$750.44
Accelerated Computing	p3.2xlarge	8	61	14 Gbps	Up to 25 Gbps	EBS Only	\$1,524.24
General Purpose	c5ad.2xlarge	8	16	Up to 3,170	Up to 10	1 x 300 NVMe SSD	??
Storage Optimized	d3en.2xlarge	8	32	1,700	Up to 25 Gbps	4 x 14TB HDD	??

93.81/  
month



\$1,524.24/  
month

# The real abstract topic - STORAGE



Google Cloud



- No performance control
- Quoted for **600GB**

Type	MAX 'sustained' IOPS	MAX 'sustained' Throughput
Standard	Read: 450 Write: 900	Read: 72 MB/sec Write: 72 MB/sec
Balanced	Read: 3,600 Write: 3,600	Read: 72 MB/sec Write: 72 MB/sec
SSD	Read: 15,000 Write: 15,000	Read: 240 MB/sec Write: 240 MB/sec

[https://cloud.google.com/compute/docs/disks/?&\\_ga=2.62767211.-658045011.1598996595#pdperformance](https://cloud.google.com/compute/docs/disks/?&_ga=2.62767211.-658045011.1598996595#pdperformance)

- All about which level and what you are willing to pay – **600GB sizing**

Type	MAX 'sustained' IOPS	MAX Throughput
Optimized HDD (st1)		MAX 147 MB/s max
General Purpose SSD (gp2)	3000	
General Purpose SSD (gp3)	3000	500 MB/sec
Provisioned IOPS SSD (io1)	MAX 30,000	
Provisioned IOPS SSD (io2)	MAX 64,000	

[https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ebs-volume-types.html?icmpid=docs\\_ec2\\_console](https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ebs-volume-types.html?icmpid=docs_ec2_console)

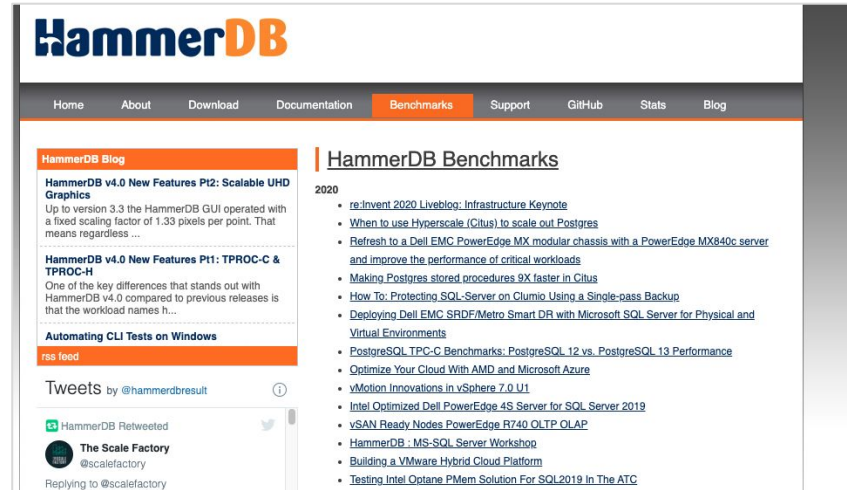
- Azure has 'binary' pricing = 128GB, 256, 512GB, etc.
- **Assume 1024 GB**

Type	MAX 'sustained' IOPS	MAX Throughput
Standard HDD	500	60 MB/sec
Standard SSD	500	60 MB/sec
Premium SSD	5000	200 MB/sec
Ultra Disk	51,200*	768 MB/sec*

<https://docs.microsoft.com/en-us/azure/virtual-machines/disks-types>

# Workloads for this effort – Why?

- HammerDB is open-source – evolving, community and free
- Standards based
  - TPC-C for transactional
  - TPC-H for analytical
- Works with all major databases – Oracle, SQLServer, DB2, MySQL, Postgres

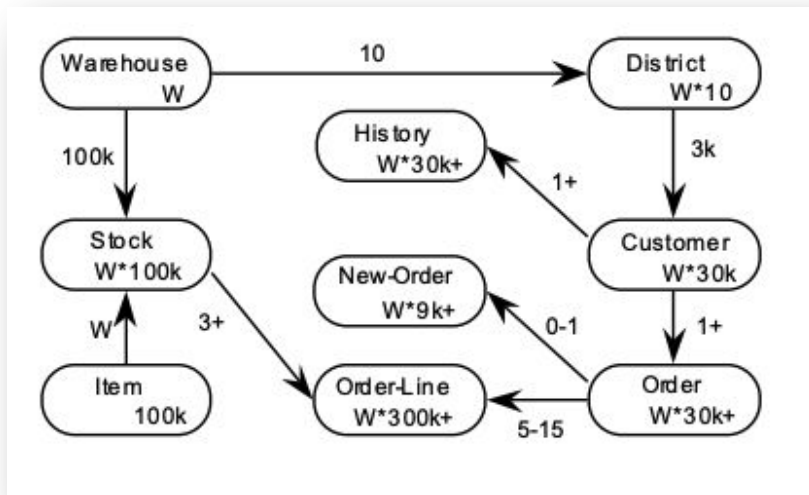


The screenshot shows the HammerDB website with a navigation bar including Home, About, Download, Documentation, Benchmarks, Support, GitHub, Stats, and Blog. The main content area is divided into two columns. The left column features a 'HammerDB Blog' section with articles such as 'HammerDB v4.0 New Features Pt2: Scalable UHD Graphics' and 'Automating CLI Tests on Windows'. Below the blog is a 'Tweets by @hammerdbresult' section showing a tweet from 'The Scale Factory'. The right column is titled 'HammerDB Benchmarks' and lists various benchmark articles from 2020, including 're:invent 2020 Liveblog: Infrastructure Keynote', 'When to use Hyperscale (Citrus) to scale out Postgres', and 'Testing Intel Optane PMem Solution For SQL2019 In The ATC'.



# Quick Anatomy of TPC-C

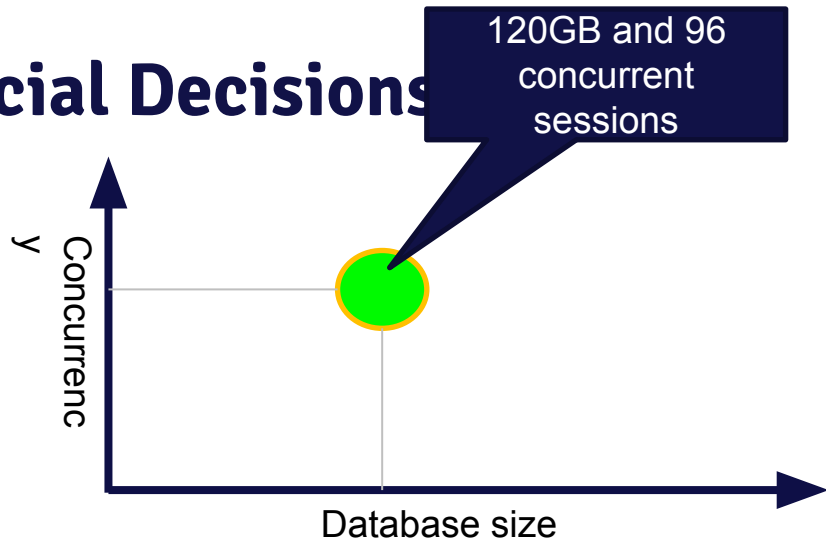
- Order Processing Use-case
- Typical Transactional use case
  - Simple SQL
  - Large in volume
- Two primary ‘knobs’ to size the effort
  - Warehouses – database size
  - User Concurrency
- **Resulting measure:**  
Transactions-per-minute (TPM)



Action Type	Mix
SELECT	75%
INSERT	8%
UPDATES	16%
DELETE	1%

# NEXT STEP: Technical/Financial Decisions

- Two Dimensions – database size and concurrent sessions
- What virtual machines to use
- What storage to use and how to tune it (where possible)
- **NOTE:** Not all clouds offer the same CPU



aws	Google Cloud	Azure
3.2Ghz Intel	2.8 Ghz Intel	2.6Ghz Intel

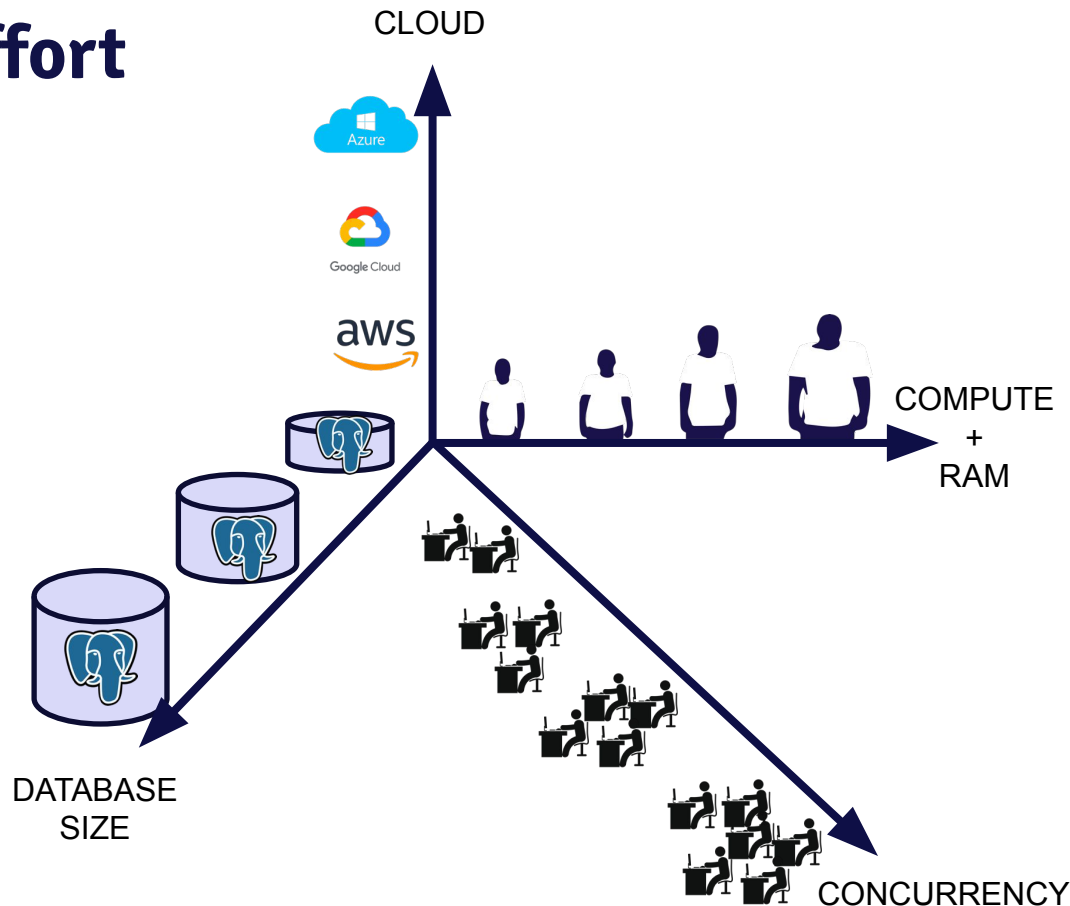
# How to best tune the database

- Postgres Configuration Settings
- Disk volume utilization
- Monitoring and measuring
- Attempts and mistakes trying the different tunings, cloud pieces-and-parts and the definition of 'good'
- Internal 'mock trials' to test the thesis
- Engaged customers to validate process and results



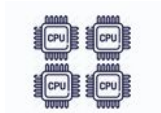
# The dimensions of the effort

- HammerDB 3.3 running on its own 4-way server
- Ran over 360 separate benchmark runs across AWS, Azure and Google
- Generated over 33.3TBs of data across different sized instances
- Evaluated different storage, Postgres configurations and more

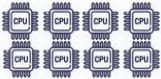
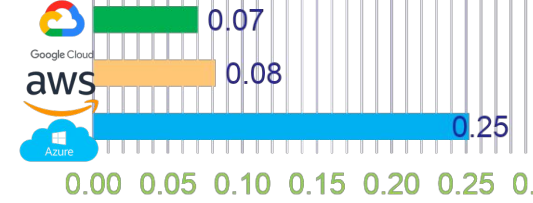
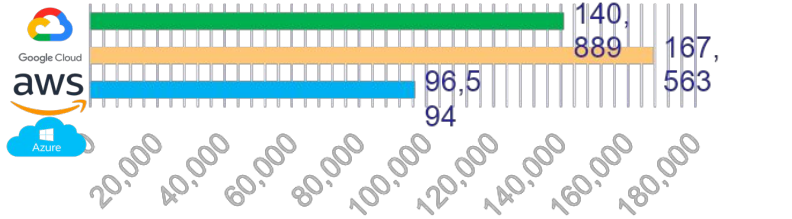




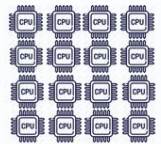
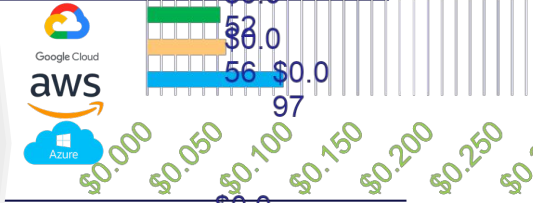
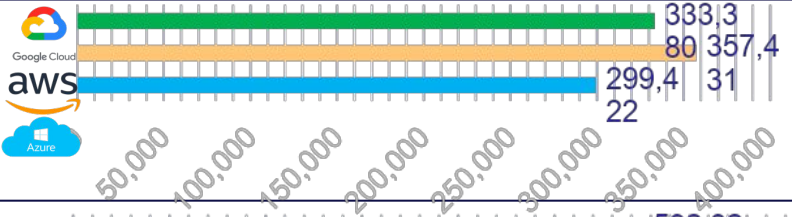
# The Results



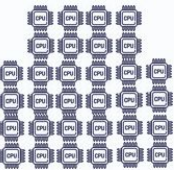
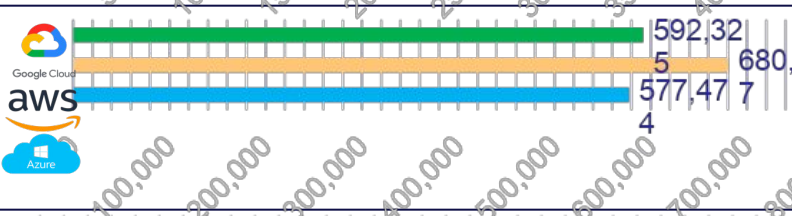
4 CPU  
32GB RAM



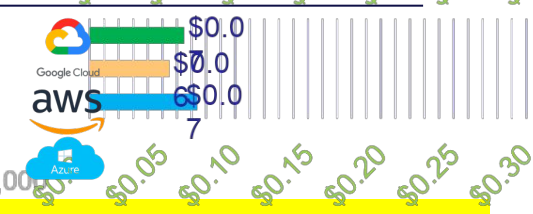
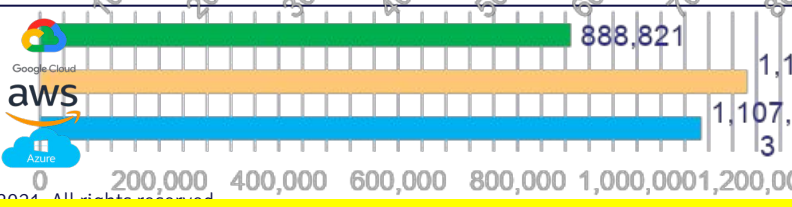
8 CPU  
64GB RAM



16 CPU  
128GB RAM



32 CPU  
256GB RAM



# Be prepared for...



Google Cloud

- No way to control disk performance
- Cheapest disk for the performance – but no top tier storage
- Smallest number of choices in CPU/RAM
- Most proactive is showing how to save money



- STORAGE is EXPENSIVE: The pre-reserved IOPS storage is confusing. How much do you need?
- VM: Chose the same sized VMs as used by RDS – on purpose!
- NVMe storage does not persist after server bounce
- DBaaS config settings are obfuscated



- STORAGE is EXPENSIVE: Their 'Ultra disk' and the IOPS and throughput math = confusing
- Problems with provisioning
- Make sure you understand 'resource quotas'
- DBaaS config least control of Postgres config file

They all offer fundamental 'monitoring' of mostly hardware resources – and little in the way of database *disk and oldest Postgres*

# Storage in the cloud – and price/performance



Example:  
Microsoft Azure

- Top performing storage: **Ultra Disk**
- You pay for it in discrete increments: 256GB, 512GB, 1024GB, etc
  - i.e. Even if you want a volume of 300GB you pay for 512GB
- You configure:
  - IOPS = operations/sec
  - Throughput = MBps



- Each VM has published ‘limits’

- Pricing
  - \$0.15/mon per GB
  - \$0.06/mon per IOPS
  - \$1.23/mon per MBps
- For our sizing efforts we used the following VMs:

Name	Size	“Max uncached disk throughput IOPS/MBps”
E4ds_v4	4 CPU/32 GB	6400/96
E8ds_v4	8 CPU/64 GB	12800/192
E16ds_v4	16 CPU/128 GB	25600/384
E32ds_v4	32 CPU/256 GB	51200/768

# Storage in the cloud – and price/performance



Example:  
Microsoft Azure

- MEDIUM T-Shirt
- 120GB database
- 96 concurrent sessions
- 5 minute 'warm-up' and 30 minute run
- Started with a fresh database each time
- ONLY VARIABLE CHANGE WAS DISK PERFORMANCE!

IOPS	MBps	Result: TPM	Monthly Storage Cost	\$/TPM
35,000	2,000	303,163	\$9,378	\$0.42



# Storage in the cloud – and price/performance



AWS EC2 and  
lo2 disk

- MEDIUM T-Shirt
- 120GB database
- 96 concurrent sessions
- 5 minute ‘warm-up’ and 30 minute run
- Started with a fresh database each time
- **ONLY VARIABLE CHANGE WAS DISK PERFORMANCE!**

IOPS	Result: TPM	Monthly Storage Cost	\$/TPM
35,000	352,863	\$2,291	\$0.12
30,000	351,260	\$2,025	\$0.11
20,000	357,431	\$1,375	\$0.09
15,000	354,238	\$1,050	\$0.08
10,000	353,212	\$725	\$0.07
7,000	355,717	\$530	\$0.06
5,000	353,743	\$400	\$0.05
4,000	349,428	\$335	\$0.05
3,500	340,057	\$302	\$0.05
3,000	327,774	\$270	\$0.06

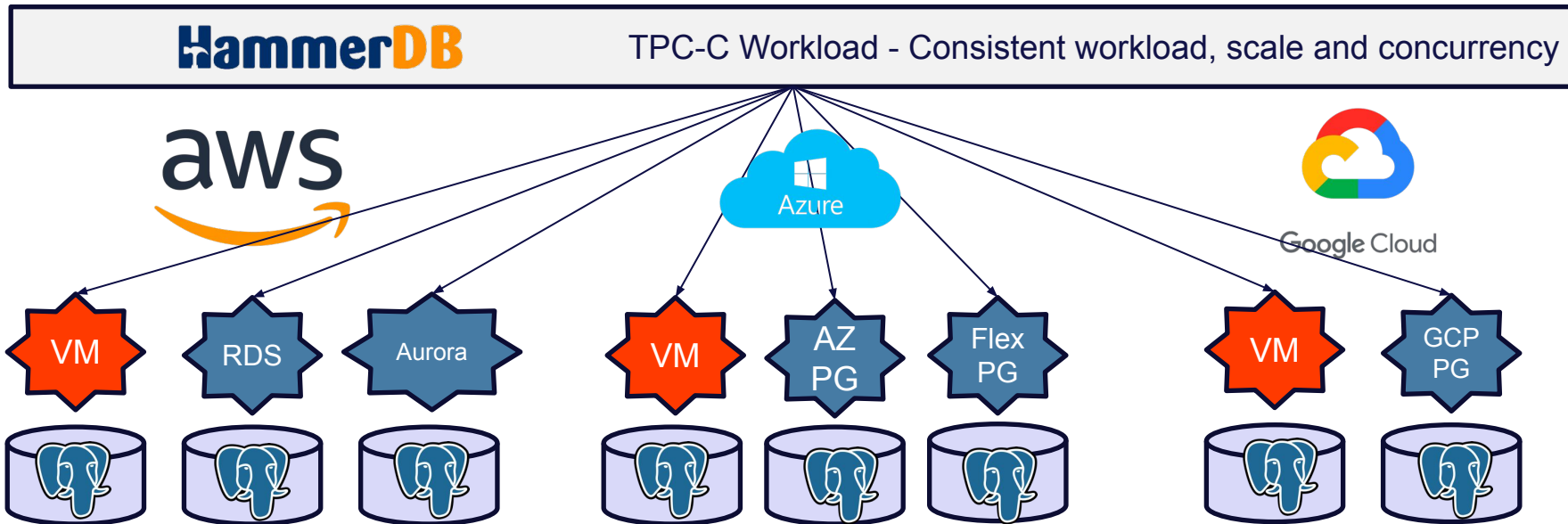


# Polling question

Q: You you be interested in this level of findings on these cloud offerings:

- AWS RDS Postgres?
- AWS Aurora Postgres?
- Azure Single Server Postgres?
- Azure Flexible Server Postgres?
- Google Cloud Postgres?
- Oracle?
- SQLServer?

# Multi-Cloud Sizing and Benchmarking – to help you decide



Is this interesting? Tell me in the Q/A

# Deliverables to you

Make it all more predictable and prescriptive



“How to” –  
Legacy workload  
evaluation  
documents and  
reproducible scripts



Best practices around  
cloud setup and  
database configuration



Create time study  
and formulation  
documents and tools

- How long to accomplish each task
- Based on workload size, what size and configuration recommendations



Financial business  
case formulas and  
documents to help  
speak in ‘debit/credit’  
lexicon



# Polling question

Would you find this useful?

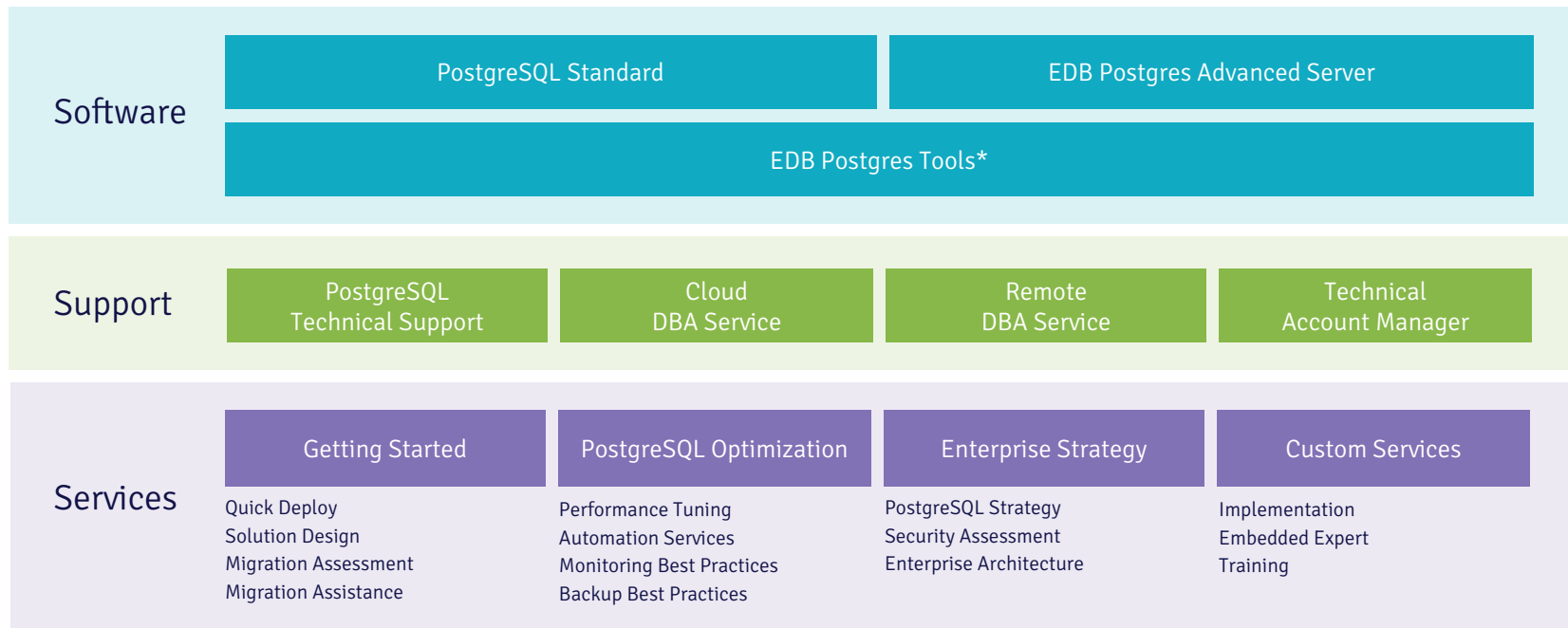
- Yes - very much
- I think so
- Not really
- Not sure

# Key Takeaways

- It is difficult to technically compare cloud database offerings – but cost-per-transaction is the most important measure.
- Most enterprises inadvertently over-provision and over-pay for Postgres in the cloud. Consider the storage examples.
- Some cloud offerings restrict your flexibility to tune, configure and optimize based on your unique workloads.
- Some DBaaS offerings are built on obsolete hardware and older releases of Postgres
- Let us know if you want to have a specific conversation by cloud and need



# EDB offerings



\* Postgres Enterprise Manager, Backup and Recovery Tool, Failover Manager, Replication Server, Containers, Kubernetes Operator, PostGIS, Pgpool, PgBouncer, Connectors, Foreign Data Wrappers, Migration Toolkit

POSTGRES  
VISION 2021

Register Now

POSTGRES VISION 2021

JUNE 22 - 23

[postgresvision.com](https://postgresvision.com)

## The Future is Postgres



Global event



2 full days



Multiple tracks



PostgreSQL experts and customers



Interaction with colleagues



PostgreSQL community connection



# In closing



## Questions

Please make sure if you have questions to type them into the tool



## Exit survey

There is an exit survey that will pop up – please tell us your thoughts on this webinar



## Contact us

**Tom Rieger** –  
[tom.rieger@enterprisedb.com](mailto:tom.rieger@enterprisedb.com)  
CALL ME – 952-221-6514  
<LINKEDIN>  
[www.enterprisedb.com](http://www.enterprisedb.com)