Building a benchmark

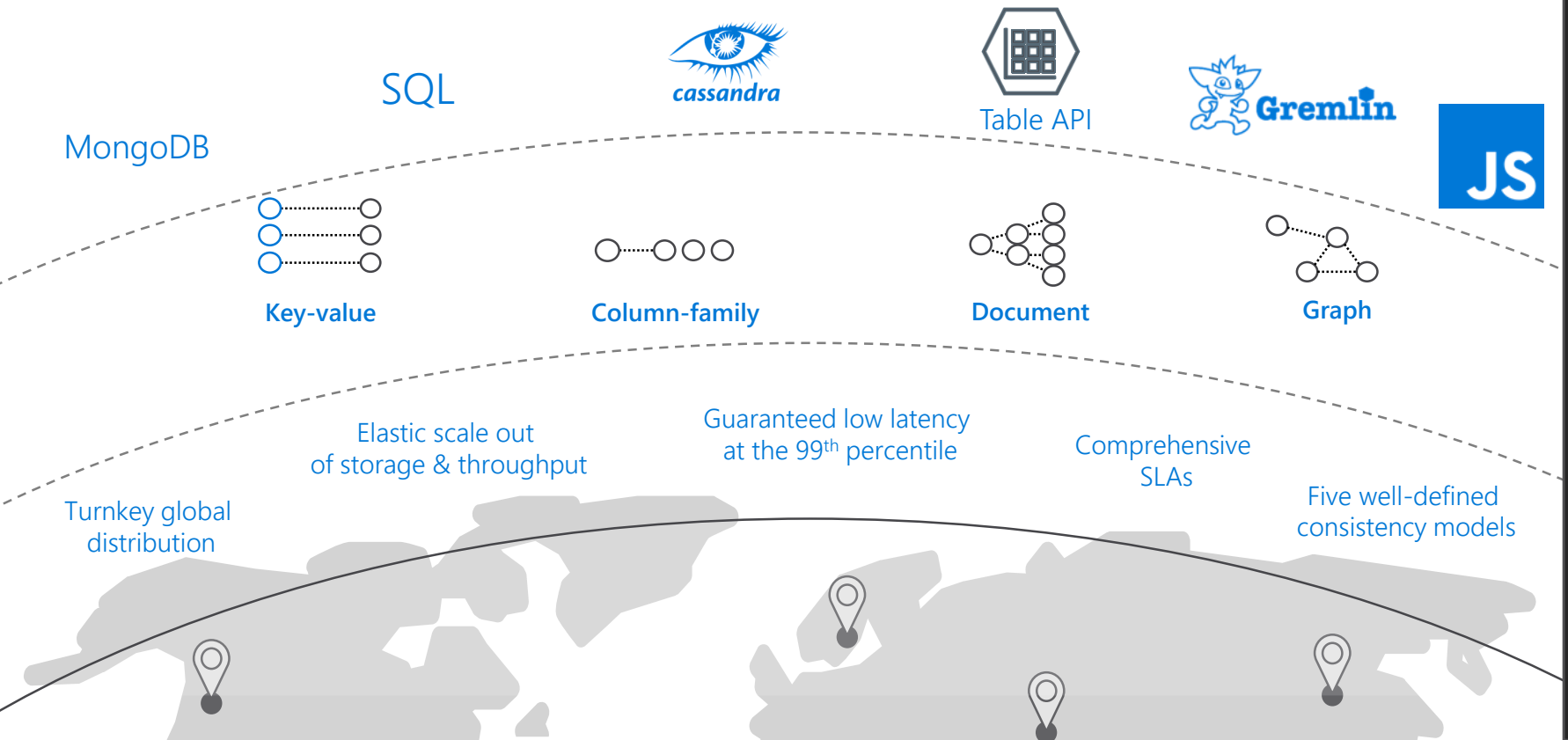
Anthony F. Voellm
Principal Engineer [Lead/Manager]
Microsoft Azure
Twitter: @p3rfguy
tvoellm@microsoft.com


NIGHTMARE
AT BEAVER LAKE

This talk will cover how to build a benchmark; defining it, creating a community, and measuring its success using a real world example.

Azure Cosmos DB

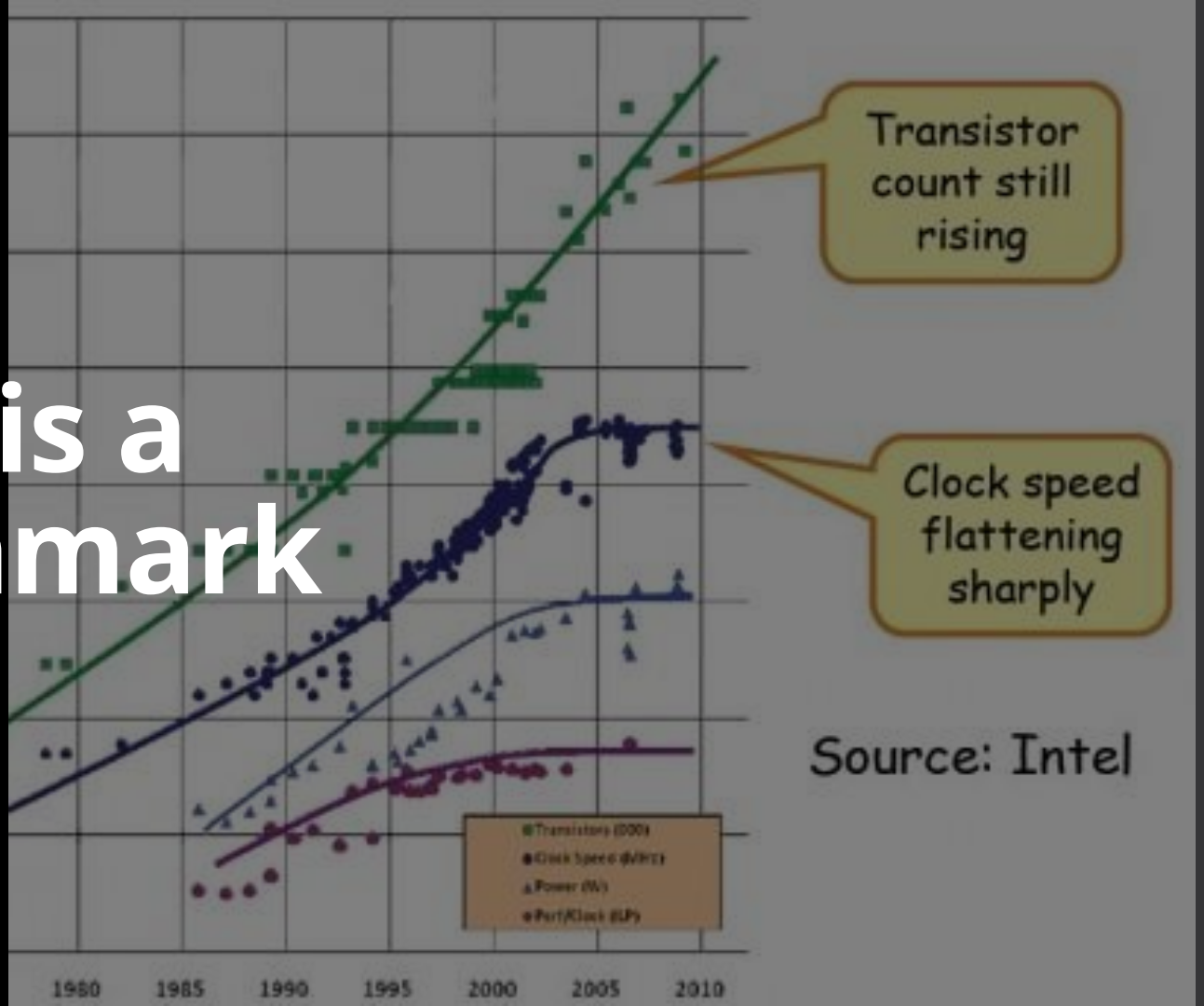
A globally distributed, massively scalable, multi-model database service



Overview

- What is a benchmark
- Defining a benchmark
- Building a community
- Measuring Success

What is a benchmark



In computing, a benchmark is the act of running a computer program, a set of programs, or other operations, in order to assess the relative performance of an object, normally by running a number of standard tests and trials against it. The term ...

Benchmarks provide a method of comparing the performance of various subsystems across different chip/system architectures.

http://en.wikipedia.org/wiki/Computer_benchmark

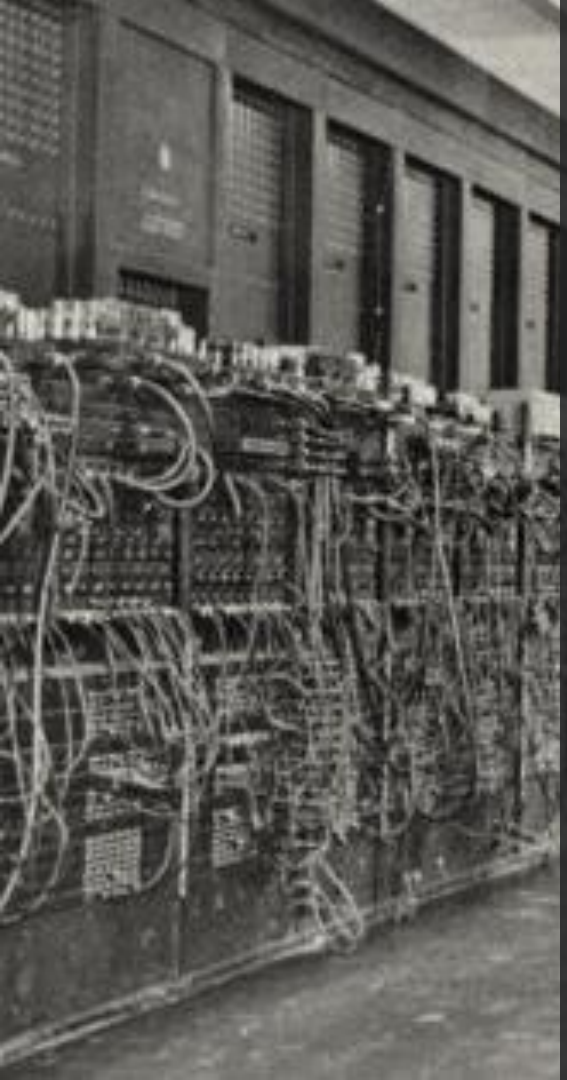
Lies, Damn Lies, and Benchmarks

What Does One Have To Do To Find Performance Truth?

By Alexander Carlton
Hewlett-Packard
Cupertino, Calif.

spec

http://www.spec.org/osg/news/articles/news9412/lies.html



A very brief history of benchmarks...

Instruction level Profiling Gibson Mix, ADP Mix, Process Control Mix

1957–1971

1985

Computer Graphics benchmarks are mosty BitBlt, Line, Ellipse, Box, and Text

TPC-C

1992

2004

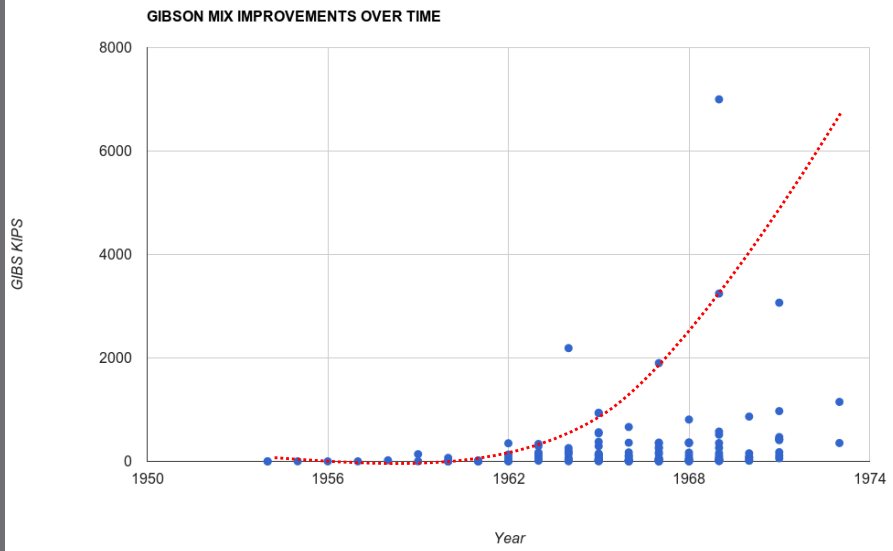
HPL v1.0 - High-Performance Linpack Benchmark for Distributed-Memory Computers

vConsolidate - The first virtualization benchmark

2007

2016

PerfKit Benchmarker is born!



Fixed Point Add/Subtract	0.330
Fixed Point Multiply	0.006
Fixed Point Divide	0.002
Branch	0.065
Compare	0.040
Transfer 8 characters	0.175
Shift	0.046
Logical	0.017
Modification	0.190
Floating Point Add	0.073
Floating Point Multiply	0.040
Floating Point Divide	0.016

The Gibbs Mix

Data From: http://roylongbottom.org.uk/cpumix.htm

Defining a benchmark

Why are you building a new benchmark?

- Are there no other benchmarks to use?
- What will be unique in your benchmark?
- What are you trying to demonstrate?
- What new value are you adding?

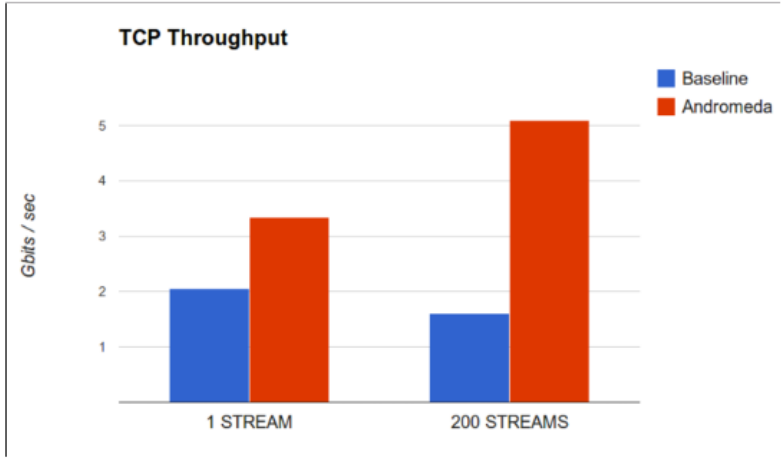
For whom are you designing the benchmark?

- Marketing to get attention for the product?
- Competitive analysis to define which customers to address?
- For the feature developer to drive an improvement?
- For management to bake-off different architectures?

Enter the Andromeda zone - Google Cloud Platform's latest networking stack

Wednesday, April 2, 2014

We have recently made the latest networking technology that powers our internal services available to Cloud Platform users across the world. Andromeda - the codename for Google's network virtualization stack - now powers two Google Compute



What are you building?

- Simple benchmark that runs in minutes?
- Benchmark that requires a lot of time and hardware?
- Representative customer workload?
- Actual customer code?

Step 1: Setup

Step 2: Warm up

Step 3: Pre-execute

Step 4: Execute

Step 5: Post-Execute

Step 6: Cleanup

Step 7: Publish results

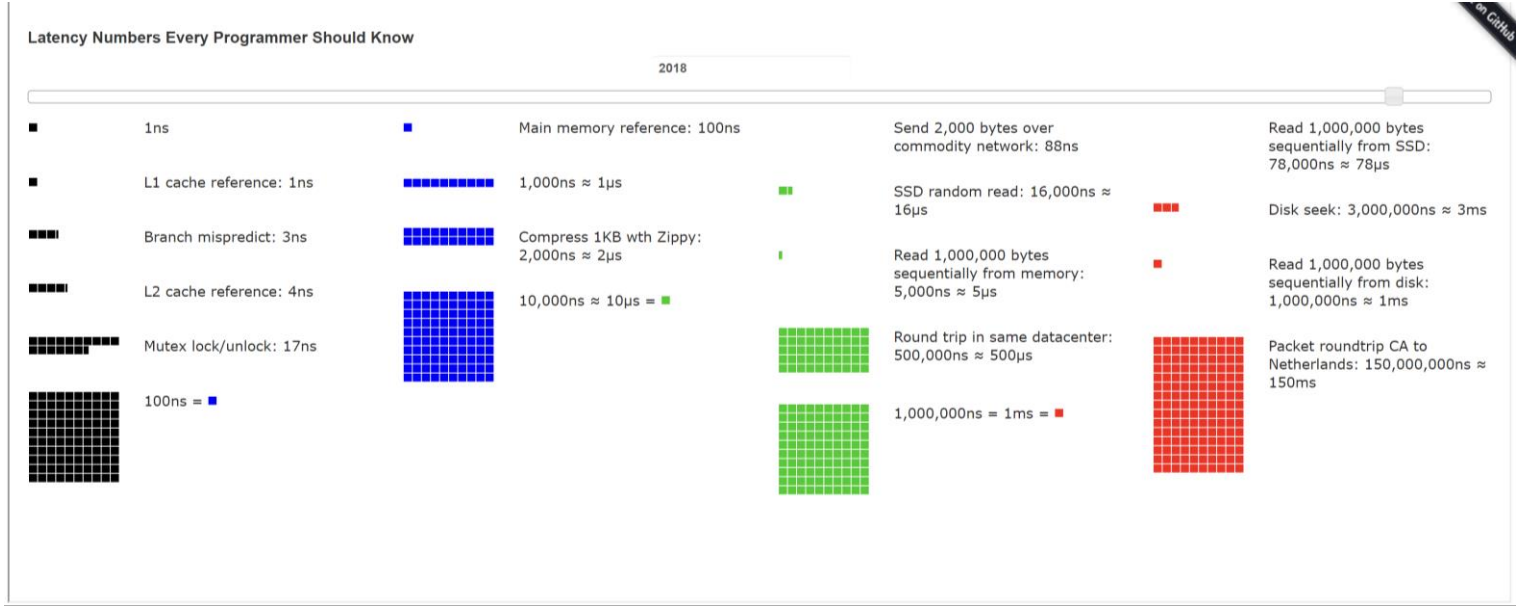
Where will it run?

How will it be built?

- Will it run everywhere?
- Focusing on ease of use?
- Looking for extensibility?

How will it be built?

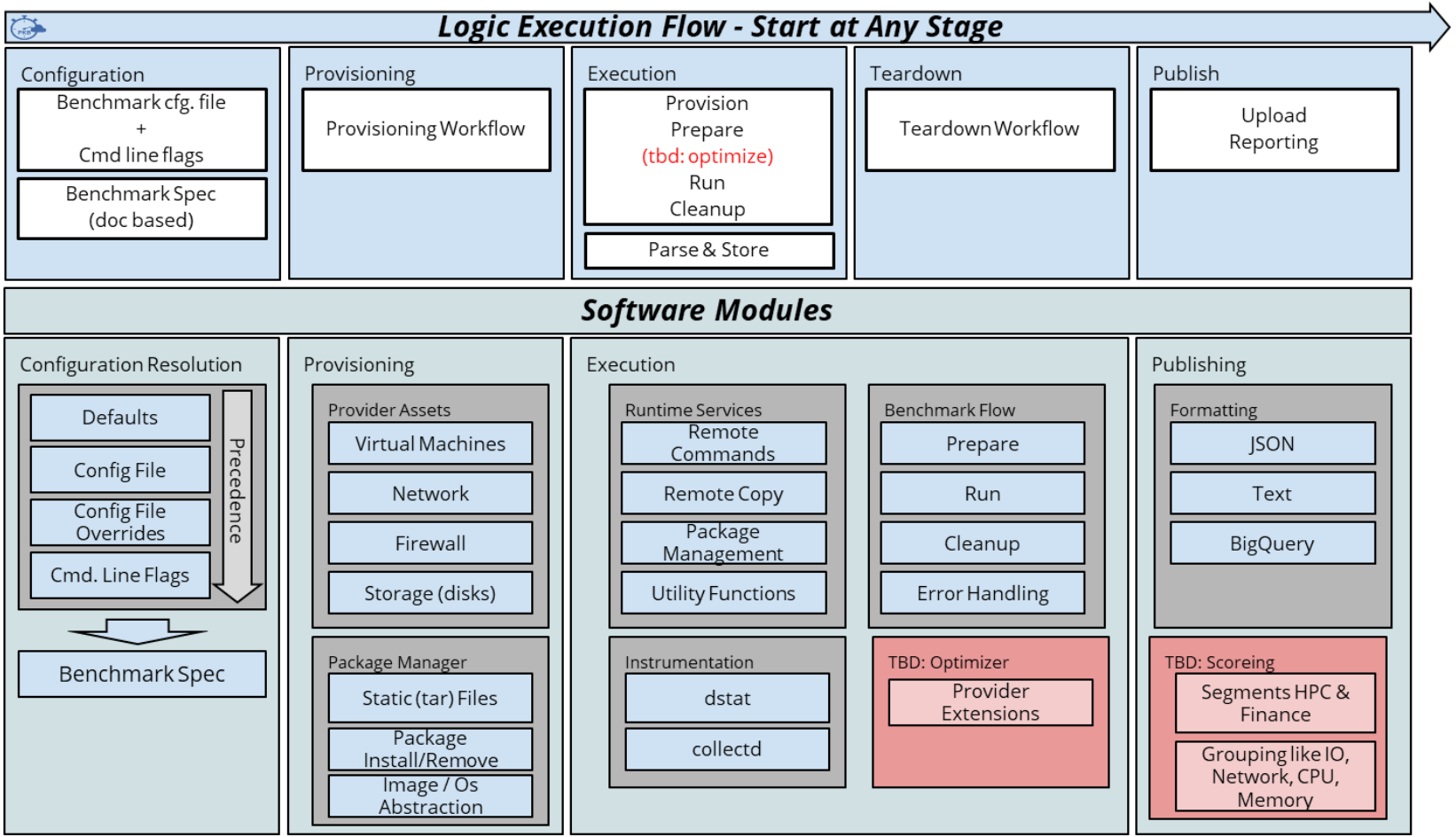
The Basics	New primatives	Meta operations	The Metrics
Network	No SQL	Start / Stop VM + Cluster	Throughput
Storage	Memcache	Snapshot a workload	Latency
Memory	Hadoop [Sort]	Provision a disk	Overhead (CPI + Power Consumption)
CPU	Boundary+Cluster network ops	Query status	Cost!
[Graphics]			



How will it be built?

http://www.eecs.berkeley.edu/~rcs/research/interactive_latency.html

How will it be built?



When?

- Is your product complete?
- Is the market ready?
- Do you have the people to support it?
- Do you have partners lined-up?
- What is the launch vehicle?

Building a community

Community Firsts Since Launch

(Googlers can't claim these :))

- First change to a benchmark - Rackspace - meteorfox
- First Bug - Tata - ksasi
- First pull request - <self> - yshalabi
- First change to a benchmark set - Intel - hungvelo
- First new benchmark - MIT - hkasture
- First new provider - Intel - kivio
- First new package (image) - CloudHarmony - Jason Reed
- First academic paper to use PKB - UNCLAIMED?
- First conference talk mentions PKB - EPFL EcoCloud - Babak Falsafi

How do you build a community?

Start small – Get one partner.

Find the right platform for engagement.

Get a prototype in place.

Do tech talks to get buy in on the ideas –before- its complete.

Tell the community what problems are still be solved.

Layout some challenges.

PerfKit Benchmark Partners in 2015

Academia

- Stanford
- MIT
- EPFL
- Berkeley

Cloud Providers

- RedHat
- Rackspace
- Qualcomm
- **Microsoft**
- Mellanox
- Google
- Intel
- Canonical
- Cisco
- CenturyLink
- Broadcom
- ARM

Industry

- Tradeworx
- CloudSpectator
- CloudHarmony

Community Firsts Since Launch

(Googlers can't claim these :))

First change to a benchmark	-	Rackspace	- meteorfox
First Bug	-	Tata	- ksasi
First pull request	-	<self>	- yshalabi
First change to a benchmark set	-	Intel	- hungvelo
First new benchmark	-	MIT	- hkasture
First new provider	-	Intel	- kivio
First new package (image)	-	CloudHarmony	- Jason Reed
First academic paper to use PKB	-	UNCLAIMED?	
First conference talk mentions PKB	-	EPFL EcoCloud	- Babak Falsafi

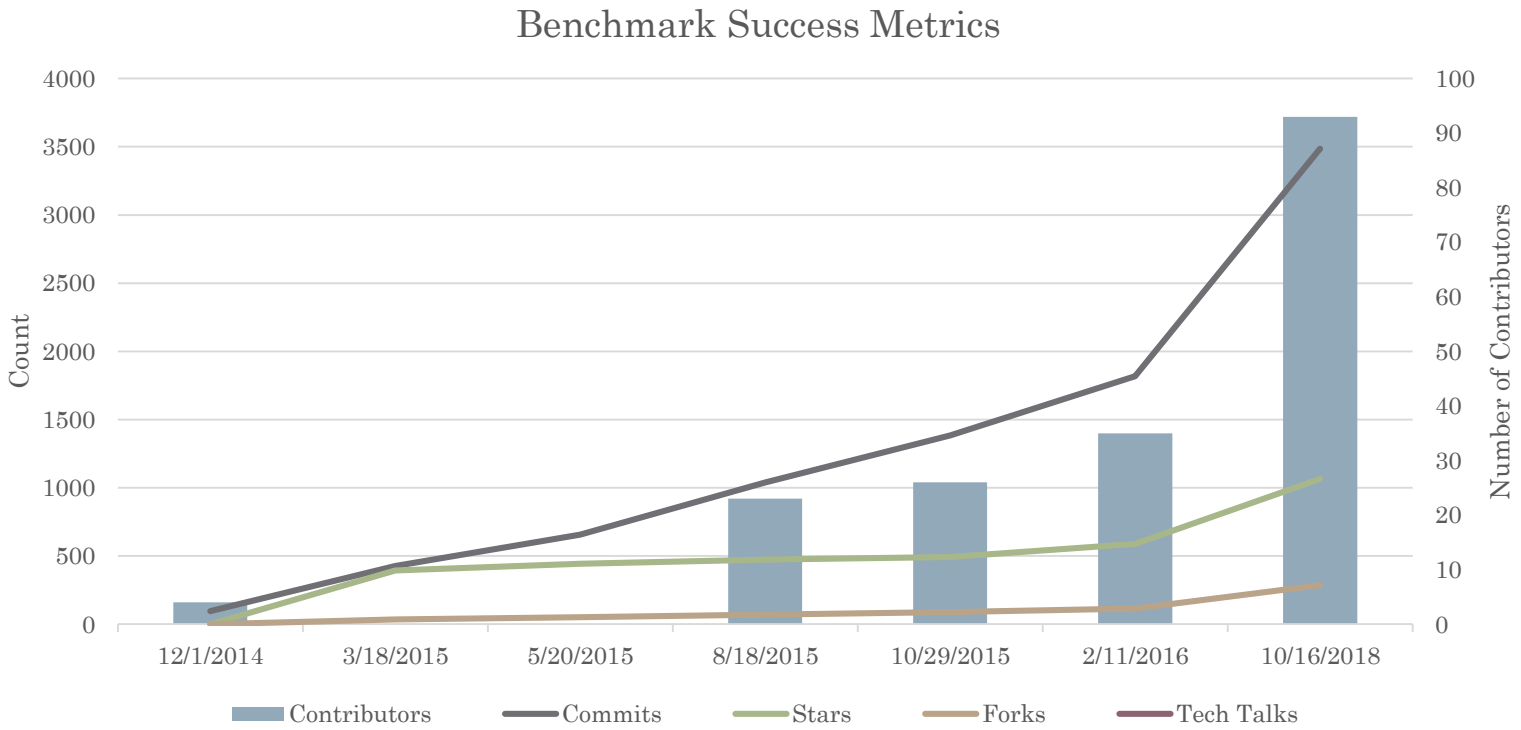
The community challenged in 2015

Measuring Success

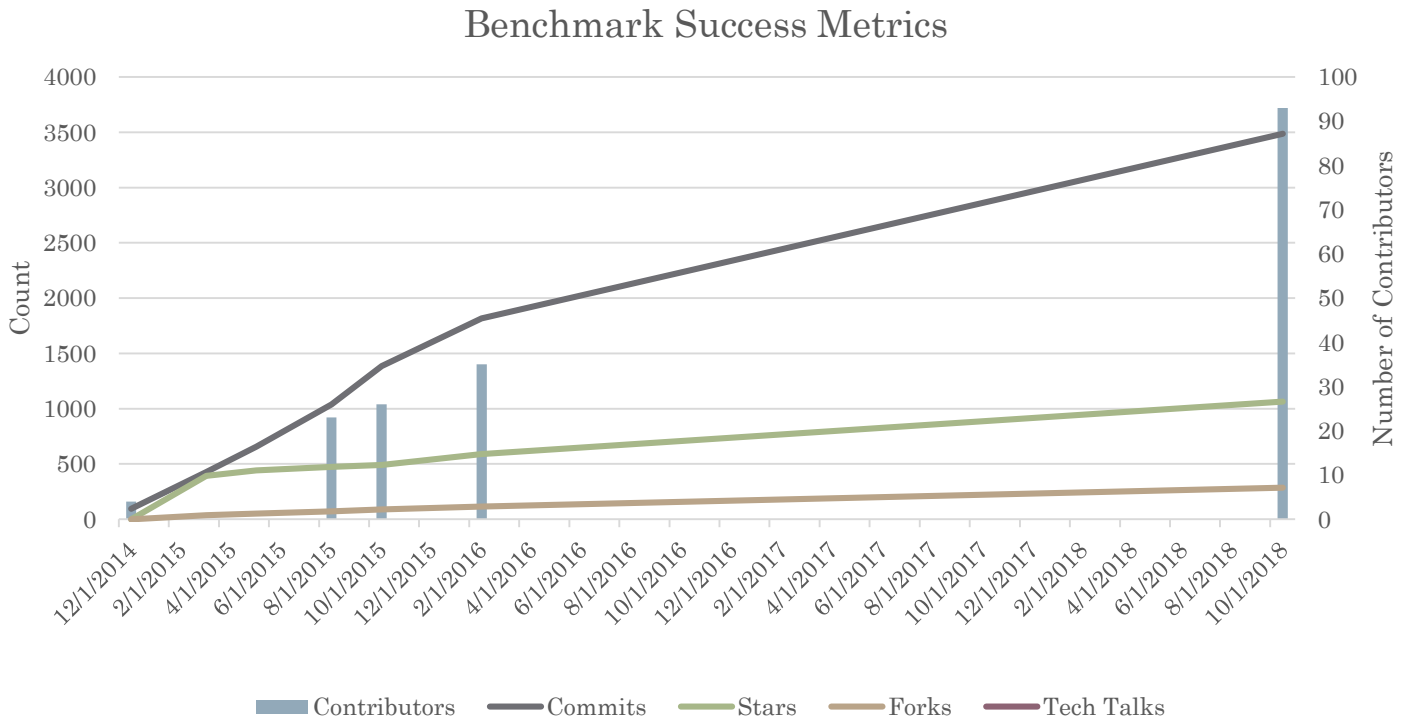
Success metrics

- Likes / Stars
- Shares
- Forks
- Pull requests
- Downloads
- New posts
- Community contributions

Perfkit Benchmarker Success



Perfkit Benchmarker Success



Anthony F. Voellm
Principal Engineer [Lead/Manager]
Microsoft Azure
Twitter: @p3rfguy
tvoellm@microsoft.com
Voellm@gmail.com

Bio

Anthony F. Voellm is currently a Principal Leader in Azure Cosmo DB leading Performance and Security. Anthony has a wide range of experience from kernel and database engines to graphics and automated image and map extraction from satellite images. Anthony is an avid inventor with 7 technology patents issued. In his current role at Microsoft he is building a planetary scale database. Prior to joining Microsoft in May 2016, Anthony held roles at Google leading the Cloud Performance Team and Security Teams, roles at Microsoft leading the Microsoft Windows Reliability, Security, and Privacy test team working on Windows8; Microsoft Hyper-V Performance Team; and SQL Server Performance team. He has also been a developer on the Windows Filesystem, SQL Server Engine, SGI and IRIX networking teams. Anthony has taught performance testing to over 2000 people worldwide and given dozens of informative talks on software fundamentals. In addition to computer interests his passions lie in growing engineers, building things, and doing anything outdoors. Anthony holds a Master of Science from George Washington University, BA in Physics and a BS in Computer Science and Mathematics from the University of Vermont.

References

- http://www.roylongbottom.org.uk/whetstone.htm
- http://www.roylongbottom.org.uk/whetstone.pdf
- http://en.wikipedia.org/wiki/Graphics_processing_unit
- http://en.wikipedia.org/wiki/Benchmark_(computing)
- ftp://reports.stanford.edu/pub/cstr/reports/cs/tr/86/1117/CS-TR-86-1117.pdf
- http://en.wikipedia.org/wiki/Graphics_processing_unit
- http://en.wikipedia.org/wiki/Beowulf_cluster
- http://www.spec.org
- http://www.eembc.org/
- http://www.netlib.org/benchmark/hpl/
- http://www.opengl.org/resources/benchmarks/
- http://www.intel.com/pressroom/archive/releases/2007/20070417gloc1.htm
- http://www.spec.org/virt_sc2010/
- http://en.wikipedia.org/wiki/Frame_rate
- http://techblog.netflix.com/2011/11/benchmarking-cassandra-scalability-on.html
- http://www.clusters4all.com/tech/scale.html
- http://en.wikipedia.org/wiki/10BASE-T
- http://en.wikipedia.org/wiki/No_sql
- http://en.wikipedia.org/wiki/Memcache
- http://en.wikipedia.org/wiki/100_Gigabit_Ethernet
- http://bertrandmeyer.com/2011/06/20/concurrent-programming-is-easy/