SDC
STORAGE DEVELOPER CONFERENCE
SNIA ■ SANTA CLARA, 2014

SPEC SFS 2014
An Under-the-Hood Review

Spencer Shepler
Nick Principe
Ken Cantrell

Today's Outline

- Intro / Contributions / Motivation
- SPEC SFS 2014 Framework
- SPEC SFS 2014 Reporting
- Workloads / Business Metrics
  - VDA
  - VDI
  - SWBUILD
  - Database

Tonight's BOF

- Drinks and Snacks
- Open discussion and additional detail…
- Tonight 6:00 PM – 7:00 PM
- Stevens Creek Room

SPEC
Standard Performance Evaluation Corporation

- **The Standard Performance Evaluation Corporation (SPEC)** is a non-profit corporation formed to establish, maintain and endorse a standardized set of relevant benchmarks that can be applied to the newest generation of high-performance computers. SPEC develops benchmark suites and also reviews and publishes submitted results from member organizations and other benchmark licensees
- www.spec.org

Disclaimer

- The SPEC SFS 2014 benchmark, as represented in this presentation, is *pre-release* software; the benchmark framework, workloads, and results and reporting structure are still under internal SPEC review and may change before final release of SPEC SFS 2014.

SPEC SFS 2014 Contributions

EMC	Tracing code, Validation, Testing, VDI Workload
Hitachi	Validation, Testing
Huawei	Validation, Testing
IBM	Reporting Tools, Validation, Testing, VDA Workload
Iozone.org	Source Code, Development, Testing, Validation, Binaries
Microsoft	Native Port to Windows, Validation, Testing, VDI Workload
NetApp	Validation, Testing, SWBUILD and DATABASE Workloads
Oracle	Validation, Testing, DATABASE Workload
Seagate	Validation, Testing

Motivation for SPEC SFS 2014

Motivation for SPEC SFS 2014

- SPEC SFS 2014 moves to Solution Benchmarking
  - Realistic, Solution-based workloads
    - DATABASE, SWBUILD, VDA, VDI
  - Workloads based on traces, like previous SFS 2008
    - Modern scenarios based on standard solutions
  - Benchmark measures application-level performance
    - Uses file system APIs at the *client*
  - Advanced measurement – quality of service
    - Ops and latency don't tell the whole story → business metrics
- Ability to measure broad range of products and configurations
  - Clients, Servers, Local File Systems, Networking Transports
    - All of these now contribute to measured performance of solution
  - Allowing for multi-tiered storage solutions

SPEC SFS 2014 Framework

SPEC SFS 2014 Framework

- Two components:
  - Load generator: netmist
    - Highly customizable, powerful, workload generator
    - SPEC SFS 2014 license includes full version
  - Wrappers: SfsManager
    - Provides ease of configuration
    - Coordinates running multiple load points (scaling)
    - Implements business metric logic
- Framework features:
  - Multi-client support is fundamental
  - Supports many operating systems and virtual machines
  - Protocol/file system agnostic
  - Definable workloads
- Full source code included with SPEC SFS 2014 benchmark

SPEC SFS 2014 Framework

- Benchmark execution phases
  - Validation
  - Initialization
  - Warmup
  - Measurement (Run)
  - Results
- This sequence of execution phases repeats for each requested load point

SPEC SFS 2014 Reporting

SPEC SFS 2014 Reporting
Publication of Results

- Prior to public disclosure, SPEC SFS 2014 results must be submitted for review by SPEC SFS subcommittee
  - Results are peer-reviewed for consistency and compliance with the SPEC SFS 2014 Run and Reporting Rules
    - Disclosure must be adequate for reproducibility
  - Accepted results are then published to the SPEC website
- Results can be Released publicly without prior committee review – however, if asked, full disclosure must be provided to SPEC

SPEC SFS 2014 Reporting
Run and Reporting Rules

- The SPEC SFS 2014 Run and Reporting Rules bound the measurement and configuration methodology
  - Primary goal of rules is to support SPEC's philosophy of fair and open benchmarking
  - Secondary goal is to ensure sufficient disclosure for reproducibility and comparability

SPEC SFS 2014 Reporting Run and Reporting Rules Highlights

- There is no Uniform Access Rule
- The WARMUP time may be set to between 5 minutes and 1 week for a publishable run
- There is no requirement to reinitialize file systems before a publishable run
  - However, detailed documentation of actions taken since system (re)initialization is required
- Single workload may be run or reported
  - No requirement that all or more than one be reported at the same time

SPEC SFS 2014 Reporting
No “newfs” Requirement

- Re-initializing the storage under the file system may not be possible or realistic
  - Cloud storage, complex tiered storage
  - More than one file system in the storage hierarchy
- Must document procedures and steps taken since last re-initialization
  - Must be generally available and recommended for customers – no “benchmark specials”
  - Documentation/review allows for reproducibility
- Can be used to simulate “aged” systems
  - Especially in conjunction with long WARMUP

SPEC SFS 2014
Defining a Workload

Workloads and Business Metrics
Workload Definition

- Workloads are richly defined in SPEC SFS 2014
- Separate I/O size distributions for reads/writes
  - Each has 16 buckets; each bucket can be a range
  - Min I/O size: 1 byte; Max I/O size: size_t
- 22 file operations available to define workload
  - “Data”
    - Read/write ops: sequential, random, whole file, memory mapped
    - Read-modify-write, copyfile, append
  - “Metadata”
    - POSIX file ops: mkdir, stat, rename, chmod, etc.

Workloads and Business Metrics
Workload Definition

- Three parameters to control write behavior
  - % Write commits, % O_DIRECT, % O_SYNC
    - Equivalents to O_DIRECT/O_SYNC are used on platforms that do not support O_DIRECT/O_SYNC
- Other parameters to change workload and dataset behavior, such as
  - % Geometric – certain files will be accessed more
  - % Compress – compressibility of the dataset
- The dataset produced by SPEC SFS 2014 is not designed to be dedupable

Workloads and Business Metrics
Business Metric Definition

- A business metric is a unit of workload, made of:
  - One or more component workloads
  - Execution parameters
  - Success criteria (thresholds)
- Why business metrics?
  - Simulating real-world workloads
  - Reporting results in real-world language
  - Success criteria attach more meaning to results than just a load level: quality

Workloads and Business Metrics
Business Metric Scaling

- The definition of a single business metric is fixed
  - Discrete and independent units of workload
- Load scaling is achieved by adding additional business metrics
  - As load increases, so does
    - Proc count
    - Dataset size
  - The oprate of each proc is constant, however!

Workloads and Business Metrics
Business Metric Success Criteria

- Business metric success criteria (thresholds)
  - Global oprate threshold monitors the average oprate of all procs
  - Proc oprate threshold monitors the oprate of all procs
    - Any single proc exceeding the threshold invalidates that load point
  - Achieved oprate must be >= x% of defined

Workloads and Business Metrics
Business Metric Success Criteria

- Business metric success criteria (thresholds)
  - Workload variance threshold monitors ratio of global achieved oprates between all component workloads
    - This ratio must be within +/- x%, as defined in the threshold
    - Example: DATABASE has a 5:1 oprate ratio between the DB_TABLE and DB_LOG component workloads
      - Ratio of achieved oprates must be within +/- 5% of 5:1

Workloads and Business Metrics
Business Metric Success Criteria

- Business metric success criteria (thresholds)
  - With these success criteria, a business metric demands a certain quality of service at all load points
  - If a success criteria is not met for a requested load point, that point is marked INVALID
  - An INVALID data point does not stop the benchmark run, but is not publishable

Workloads and Business Metrics Benchmark Results

- There are two principal measures of performance in SPEC SFS 2014
  - Business Metrics
  - Overall Response Time
- Achieved Oprate and Total KBps will be included in official publications as well
- The sfssum file produced during a benchmark run contains all this info and more

Workloads and Business Metrics Benchmark Results

- Disclosure of results must include the summary result
  - Maximum achieved business metrics and overall response time of the entire benchmark run
    - Specific format in Run and Reporting Rules
- The full disclosure report is published on the SPEC website
  - Visual: Business Metrics vs. Response Time

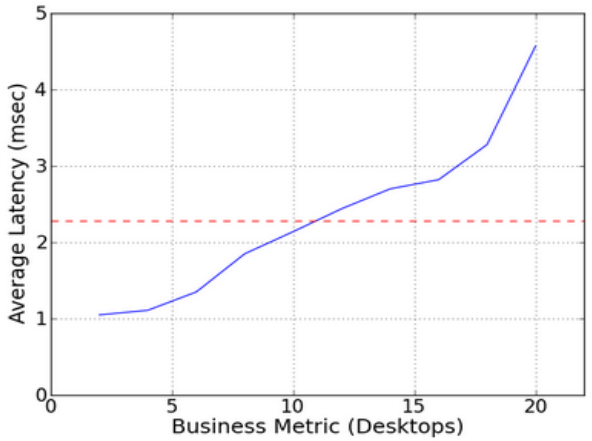
Workloads and Business Metrics Benchmark Results

SPEC SFS®2014 Result

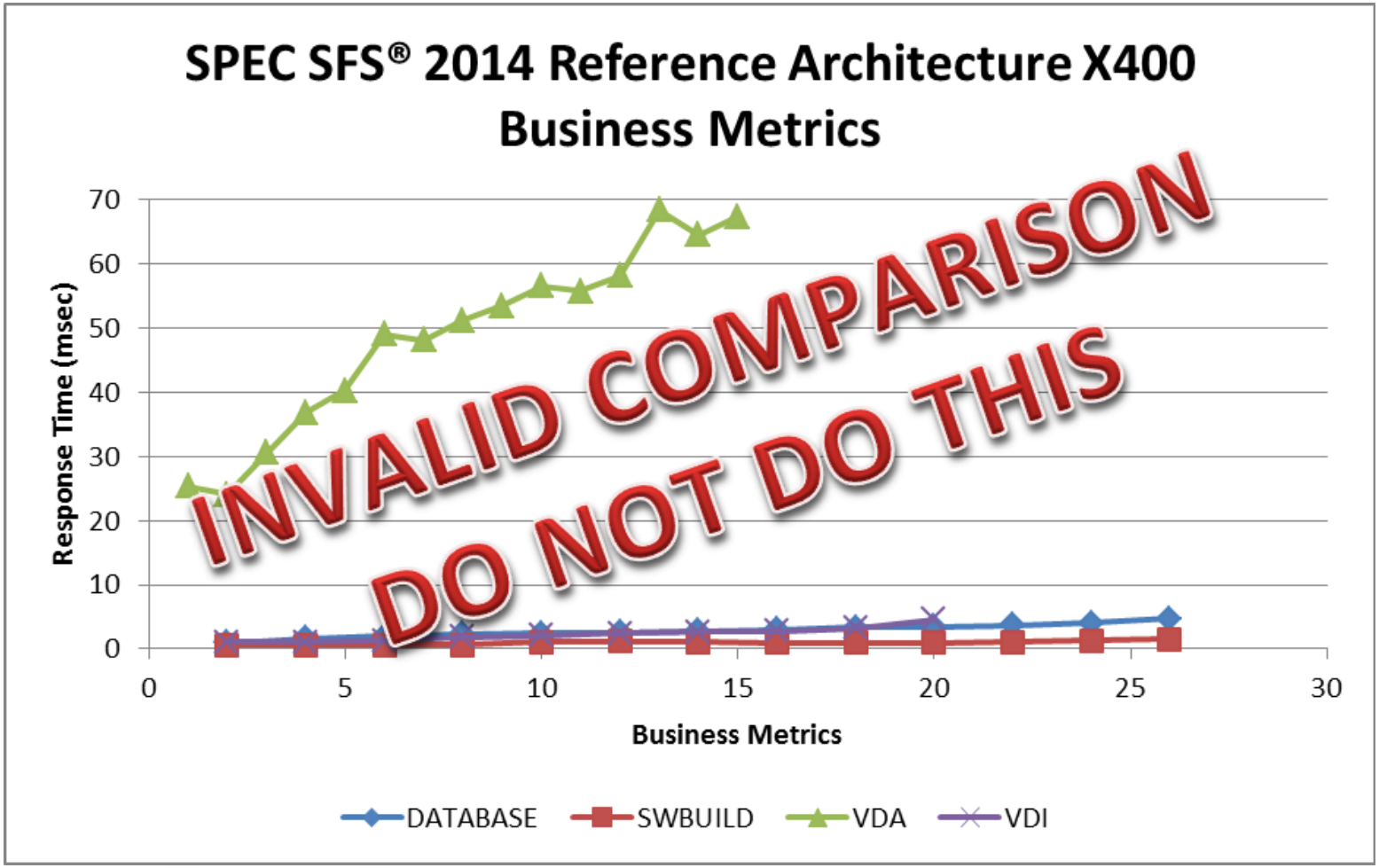
Acme Servers (fictional) Cloud Cluster 1000 (fictional)	SPEC SFS2014_vdi = 20 Desktops Overall Response Time = 2.28 msec

Performance

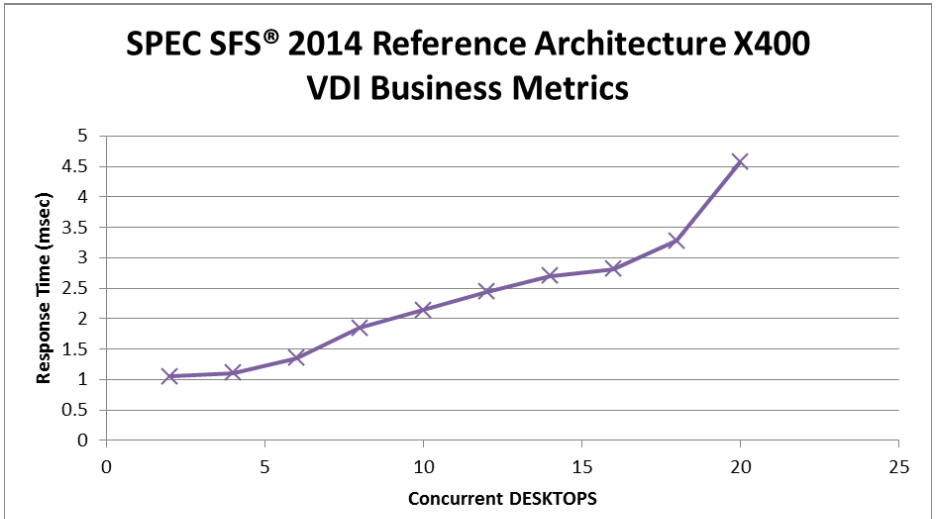
Business Metric (Desktops)	Average Latency (msec)	Desktops Ops/Sec	Desktops MB/Sec
2	1.1	400	5
4	1.1	800	11
6	1.4	1200	17
8	1.9	1600	23
10	2.1	2000	29
12	2.4	2400	35
14	2.7	2800	41
16	2.8	3200	46
18	3.3	3600	52
20	4.6	3999	58

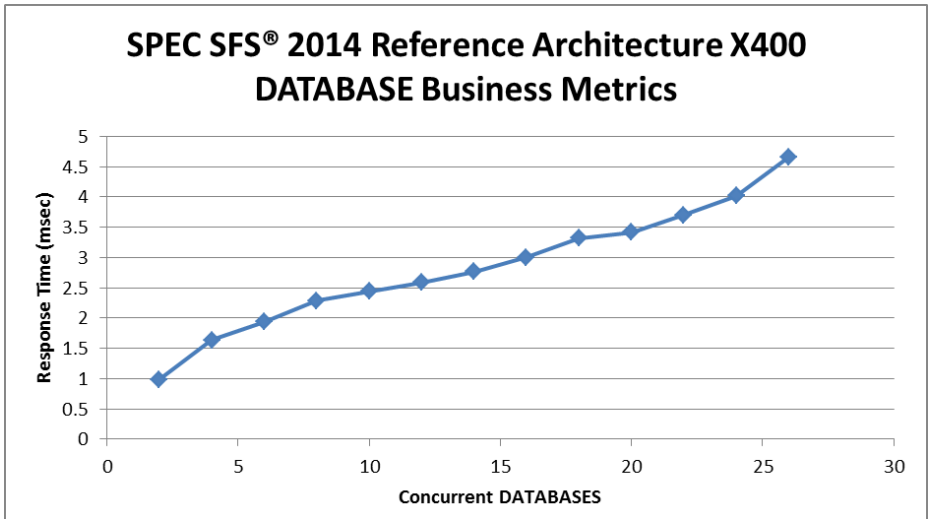


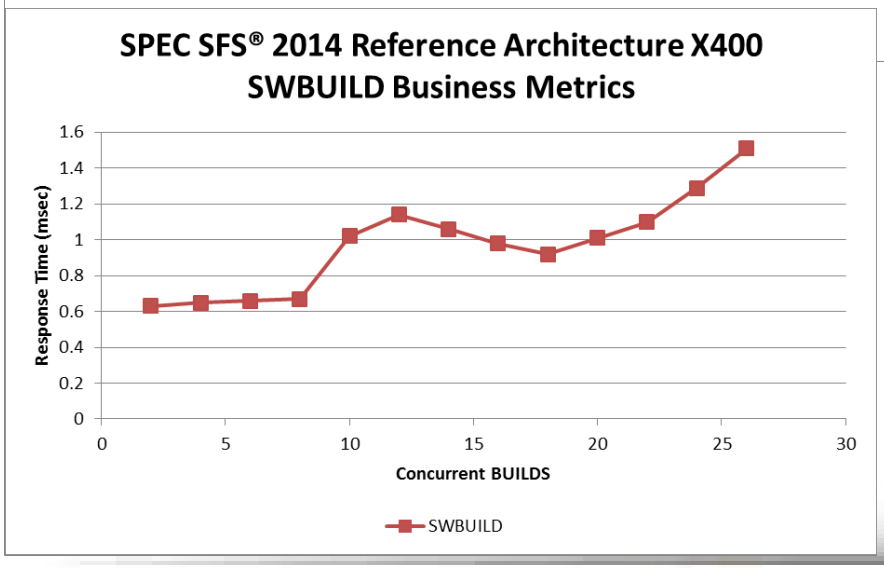
Workloads and Business Metrics Benchmark Results

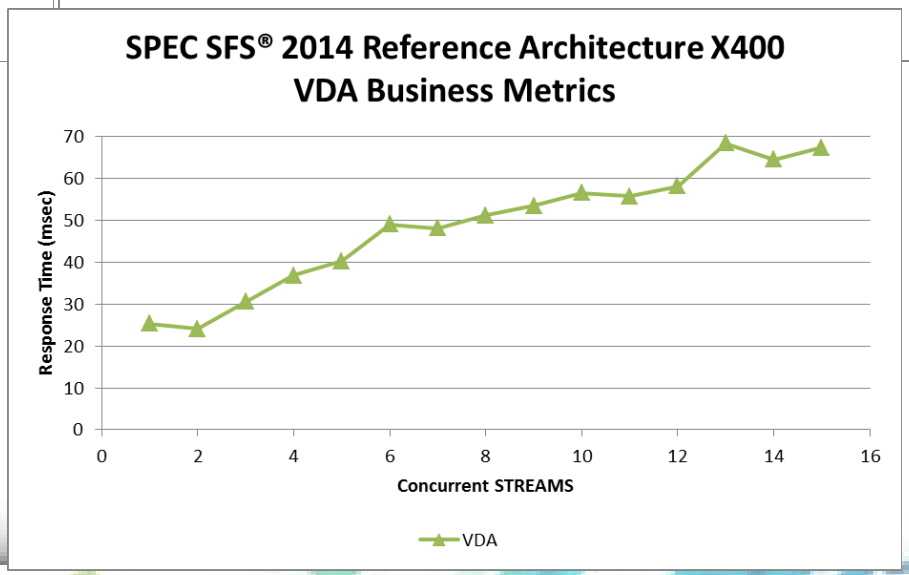


Workloads and Business Metrics Benchmark Results









Workloads and Business Metrics
Overall Response Time

- Overall response time is calculated differently in SPEC SFS 2014
  - Still the area under the curve divided by the maximum achieved business metric
  - Origin point (0,0) is no longer assumed
    - First point used in the calculation is the first achieved result
    - No longer seems appropriate to assume the curve will be a certain shape

SPEC SFS 2014 Workloads

SPEC SFS 2014 Workloads
Video Data Acquisition (VDA)

- Simulates acquisition of data from a temporally volatile source (surveillance, big data ingest)
  - Metric: Concurrent STREAMS
  - Workload derived from IBM Redbooks
- Two component workloads, 9:1 oprate ratio
  - VDA1, data stream
    - ~36 Mb/sec sequential writes (upper range of HD video)
  - VDA2, companion applications/user access
    - 89% read, 2% read-modify-write, 9% metadata

SPEC SFS 2014 Workloads
Video Data Acquisition (VDA)

- VDA2 workload ensures that quality of data ingestion is maintained despite other activity
  - Starvation of reads or writes will be detected by success criteria violation
    - Per-proc oprate: >= 75% of requested
    - Overall oprate: >= 95% of requested
    - Component workload variance: <= 5% of defined

SPEC SFS 2014 Workloads
Virtual Desktop Infrastructure (VDI)

- Simulates the workload generated by a hypervisor to support a heavy steady-state knowledge worker workload
  - Workload derived from traces of ESXi, Hyper-V, and Xen environments
  - Metric: concurrent DESKTOPS
- One component workload, 2 procs per desktop
  - Data-heavy workload: 1% metadata ops

SPEC SFS 2014 Workloads
Virtual Desktop Infrastructure (VDI)

- Simulates steady-state VDI workload
  - Does not include boot storm or login storm
- All writes use Direct I/O
- Dataset consists of compressible (50%) large files (500MB)
- Dataset is not dedupable – simulates a VDI scenario using Full Clones

SPEC SFS 2014 Workloads
Software Build (SWBUILD)

- Simulates large software project compilation or build phase of an EDA workflow
  - Workload derived from traces taken during software build activity and ClearCase documentation
  - Metric: concurrent BUILDS
- One component workload, 5 procs per build
  - Metadata-heavy: 87% metadata ops

SPEC SFS 2014 Workloads
Software Build (SWBUILD)

- Reads and writes are done on a whole file
  - Average file size is a Gaussian distribution centered at 16 KiB, ~573,000 files per build
  - Files are highly compressible (80%)
- This workload has the most potential to be cached/modified by the load generating clients
  - Also most likely to introduce/measure a bottleneck on load generating clients vs. storage solution

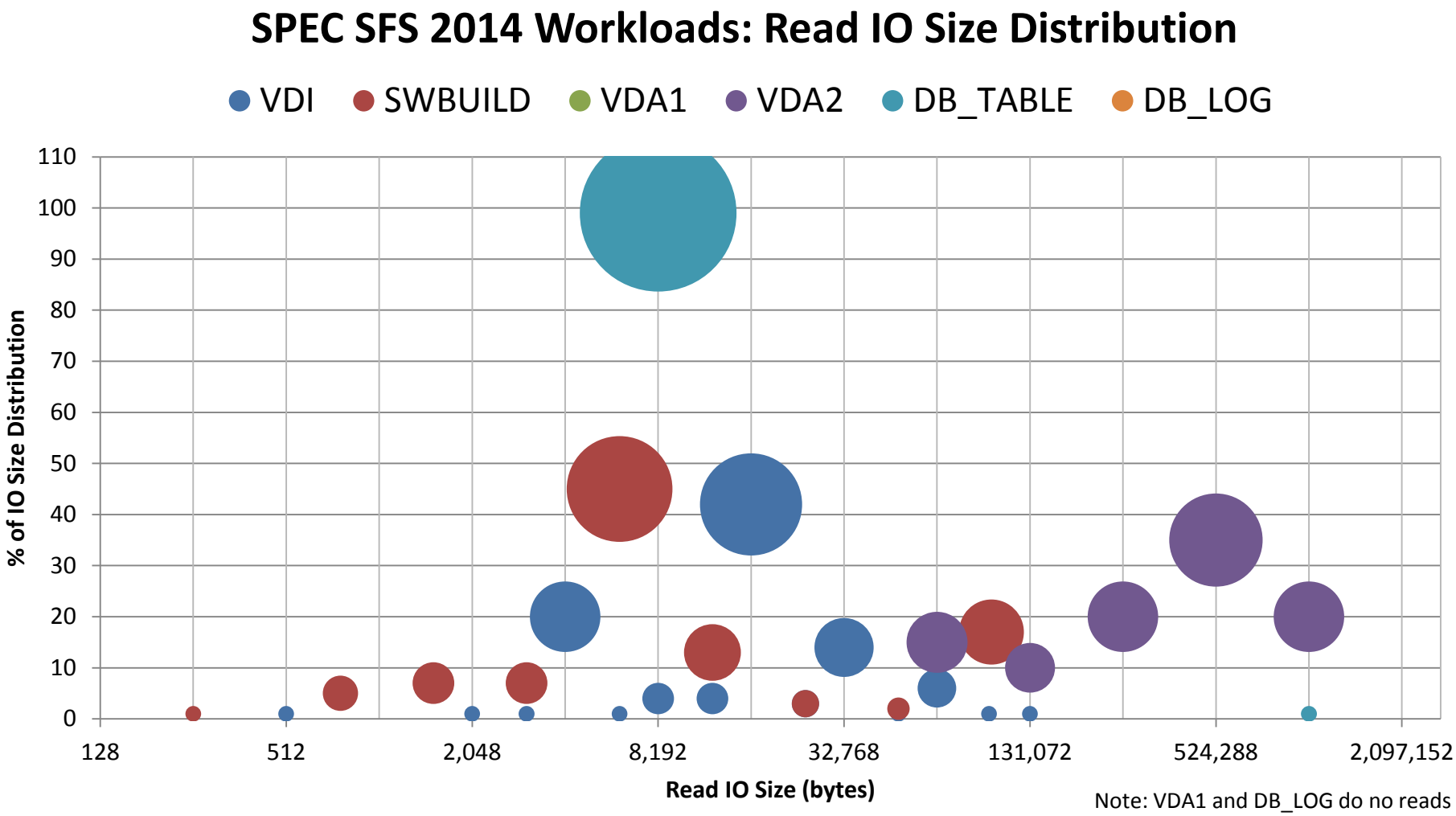
SPEC SFS 2014 Workloads DATABASE

- Simulates an OLTP database consolidation scenario
  - Workload derived from data from Oracle
  - Metric: concurrent DATABASES
- Two component workloads, 5:1 oprate ratio
  - DB_TABLE
    - Random reads and writes
  - DB_LOG
    - Mostly sequential writes

SPEC SFS 2014 Workloads DATABASE

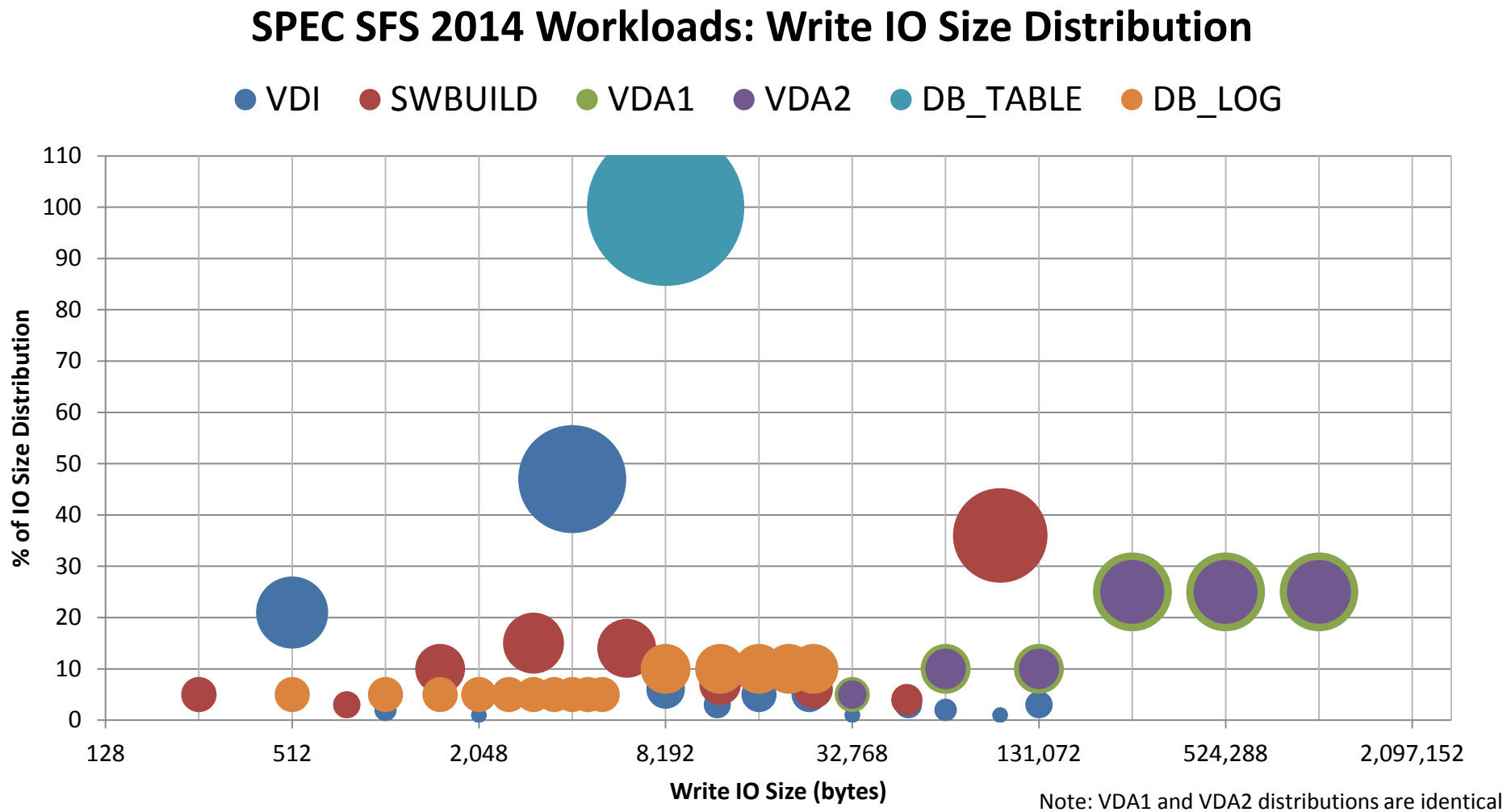
- All DB_TABLE threads for an individual business metric share the same dataset
  - Multiple threads working the same tables
- Workload simulates moving hot spots in the dataset
  - These hot spots move over time
- Solution under test must provide good quality of service to both table and log I/O
  - Maximum component workload variance is <= 5%

SPEC SFS 2014 Read Size Distribution



Note: VDA1 and DB_LOG do no reads

SPEC SFS 2014 Write Size Distribution



Note: VDA1 and VDA2 distributions are identical

SPEC SFS 2014 Workloads Summary

- SPEC SFS 2014 Workloads are richly-defined, realistic, solution-based workloads
- Results are measured in Business Metrics
  - Real-world language for real-world workloads
  - Quality of service is measured with success criteria
- Performance is measured at the application level
  - Performance of whole solution is measured
- Modern scenarios based on standard solutions
  - Workload definitions and source available to all SPEC SFS 2014 licensees
  - Open, transparent, and fair benchmarking

Future Investigations

- More Workloads
  - Windows Homefolders (a.k.a. FSCT)
  - HPC
  - Movie Production
  - Video Distribution
- Support More Storage APIs
  - Block Device
  - MPI-IO
  - HDFS
  - CDMI
- Energy Efficiency
  - Work with SNIA and the EPA. Energy Star standard
  - Power Measurement
- ***Source code continues to be provided for everything as SPEC maintains openness and transparency***

Conclusion

- Thank You!

- Please come to the BOF tonight for more open discussion and additional detail…
- 6:00 PM – 7:00 PM
- Stevens Creek Room