

IN&OUT AG

**Oracle Performance Benchmark
HPE Intel RHEL und IBM Power8 AIX**

Andreas Zallmann
CEO, In&Out AG

Version: 1.00

Datum: 18.11.2016

Klassifikation: nicht klassifiziert

In&Out AG IT Consulting & Engineering

Seestrasse 353, CH-8038 Zürich
Phone +41 44 485 60 60
Fax +41 44 485 60 68

info@inout.ch, www.inout.ch

Vorbemerkung

Die vorliegende Studie wurde von HPE, Intel und RedHat mitfinanziert. Dabei wurden die Hälfte der Dienstleistungsaufwände von diesen Firmen übernommen. In&Out hat die andere Hälfte der Dienstleistungen kostenfrei erbracht und die Lizenzen für OraBench™ kostenfrei zur Verfügung gestellt

In&Out als unabhängiges Beratungsunternehmen versichert, dass die mitfinanzierenden Firmen keinen Einfluss auf die Messungen und die Inhalte dieser Studie genommen haben und diese unabhängig erstellt wurde. Es besteht keinerlei finanzielle Verbindung zwischen In&Out und den oben genannten Firmen.

Ausgangslage

Aktuell sind für zentrale Banking- oder Versicherungslösungen auf Oracle Datenbanken in mittleren und grossen Unternehmen sehr häufig IBM POWER Systeme mit dem Betriebssystem AIX im Einsatz. Im Bereich der Applikationsserver und WebServer haben sich im Rechenzentrum x86 basierte Systeme auf Linux (in der Regel Red Hat Enterprise Linux, RHEL) auf breiter Front durchgesetzt. Viele Kunden stehen vor der Entscheidung, auf welcher Plattform sie ihre zentralen Datenbanksysteme und Core Applikationen zukünftig betreiben wollen.

Neben Aspekten wie Kosten, Stabilität, Skalierbarkeit, Betrieb und weiteren spielt auch immer der Performanceaspekt eine wichtige Rolle. Dabei sind öffentlich verfügbare CPU Benchmarks wie SPECint nur bedingt aussagekräftig, da diese nur die reine CPU Performance messen, jedoch die Performance des gesamten Systems entscheidend ist. Ebenso sind Benchmarks nur bedingt geeignet, die in erster Linie darauf abzielen neue Performancerekorde zu erzielen, z.B. TPC-C oder SAPS. Diese Benchmarks werden mit nicht repräsentativem maximalen Hardwareeinsatz erreicht. Zudem sind in den Benchmark häufig keine aktuellen Systeme vertreten.

In&Out verfügt mit der OraBench™ Performance Benchmark Suite über einen repräsentativen und leistungsfähigen Oracle Benchmark, mit dem die Performance von Oracle Datenbank Systemen mit allen relevanten Kennzahlen auf verschiedenen Plattformen verglichen werden kann. In&Out hat die Oracle Performance mit OraBench™ auf aktuellen x86 Servern mit verschiedenen Virtualisierungstechnologien und aktuellen IBM Power8 Server gemessen. Die Firma HPE hat uns im europäischen Benchmarkcenter in Böblingen (D) die entsprechende Hardware und Umgebungen bereitgestellt. Die Power Systeme wurden in einem Kundenumfeld gemessen.

Dieses Dokument dient als Entscheidungsgrundlage für Architektur- und Herstellerentscheide für Oracle Plattformen, z.B. Avaloq, Finnova, Temenos T24, Adcubum Sirius, oder anderer Oracle-basierter Applikationen.

In&Out AG

Die Züricher Beratungs- und Engineering Firma In&Out (www.inout.ch) bietet herstellerunabhängige und neutrale Beratung in den Kernthemen Technologie & Plattformen, Cloud Computing sowie Data Center & Continuity Management an. In&Out verkauft keine Produkte von Fremdherstellern, pflegt aber Technologiepartnerschaften mit einer Vielzahl von führenden Herstellern, unter anderem HPE, IBM, RedHat, EMC, HDS, VMware, etc.

In&Out berät Ihre Kunden häufig bei Plattformentscheidungen und hat einen grossen Schwerpunkt im Bereich Performance Benchmarks. Die Oracle Benchmark Suite OraBench™ wird seit über 10 Jahren von der In&Out entwickelt und ist bei mehr als 50 Performance Benchmarks bei namhaften Kunden eingesetzt worden.

Über den Autor



Andreas Zallmann,
andreas.zallmann@inout.ch
In&Out AG, Seestrasse 353, 8038 Zürich
www.inout.ch

Andreas Zallmann hat Informatik an der Universität Karlsruhe studiert und ist seit dem Jahr 2000 bei der In&Out AG tätig. Er ist verantwortlich für den Geschäftsbereich Technology und ist seit dem 1.10.2016 CEO der In&Out AG. Er ist der Entwickler der In&Out Performance Benchmarking Tools IOgen™ (Storage IO Benchmarks), NETgen™ (Netzwerk Benchmarks) und CPUgen™ (CPU Benchmarks) und hat in den letzten Jahren sehr viele Kunden- und Hersteller-Benchmarks durchgeführt.

Die In&Out verfügt über jahrelange Praxis-Erfahrung in Architektur, Konzeption, Benchmarking und Tuning von Storage- und Systemplattformen insbesondere für Core Applikationen für Banken und Versicherungen.

Management Summary

Die CPU Leistung der heutigen HPE Server mit aktuellen Intel Prozessoren unter RHEL ist den heutigen Power 8 Systemen unter AIX inzwischen überlegen, dies zeigt sich insbesondere auch in den Oracle PL/SQL Benchmarks.

Die Tests mit sehr hohem IO Durchsatz sind nicht wirklich zwischen IBM und HPE Intel RHEL vergleichbar, da die IO Anbindung dieser Server im Test erheblich höher als die der IBM Systeme war. Erkennbar können aber auf allen HPE Intel RHEL Systemen sehr hohe Data Writes und Data Scan Raten erreicht werden, die mit 5 GB/s selbst für sehr hohe Anforderungen ausreichend sein dürften.

Auch in den weiteren Tests zeigt der HPE Intel RHEL Stack eine sehr gute Leistung, die in nahezu allen Tests über der Leistung der IBM Plattform lag. Bei einigen Tests, besonders stark bei relativ vielen parallelen kleinen IOs zeigt sich allerdings ein signifikanter Performance Impact durch die Virtualisierung. Generell zeigt hier VMware ESX einen geringeren Impact wie die Red Hat Virtualisierung RHV.

Abschliessend kann der Einsatz von Oracle Datenbanken auf x86 Systemen mit Red Hat Linux aus Performancegesichtspunkten empfohlen werden. Dies speziell auch vor dem Hintergrund, dass die Leistung pro Core höher liegt, aber nur die Hälfte der Lizenzkosten pro Intel Core (Lizenzfaktor 0.5) im Vergleich zu Power Cores (Lizenzfaktor 1.0) anfällt.

Bedacht werden muss, dass IBM Power Plattformen über eine exzellente Virtualisierung verfügen und sehr gut selbst mit sehr hohen CPU Auslastungen bis zu 80% und mit einer sehr grossen Zahl von virtuellen Systemen zurechtkommen. Auf x86 Systemen mit VMware oder RHV wird üblicherweise mit geringerer Auslastung gearbeitet. Allerdings liegen die Hardwarekosten der Intel Systeme um Faktoren niedriger als die der Power Systeme.

Das Ziel dieses Benchmarks ist ein Vergleich mit gleichem Ressourceneinsatz auf der Basis von 8 physischen Cores. Dabei wurden 2 verschiedene CPU's von Intel vermessen: Die E5 Familie (maximal 2 Sockets) und die E7 Familie (2 - 16 Sockets).

Benchmark Setup

Systemplattformen

Folgende Plattformen werden in den Vergleich einbezogen:

- IBM Power8 Midrange Systeme Power E870 auf IBM AIX
- HPE DL380 Gen9 auf RedHat Enterprise Linux (RHEL)
- HPE DL580 Gen9 auf RedHat Enterprise Linux (RHEL)
- Die HPE Server wurden jeweils als Bare Metal (BM) ohne Virtualisierung und mit VMware ESXi und RedHat Virtualisierung (RHV) gemessen

Dabei wurde auf einen Performance Benchmark mit dem HPE Superdome X (SDX) bewusst verzichtet, da dieser die gleichen Prozessoren verwendet wie der HPE DL580 Gen9 (E7 Familie), allerdings bis zu 16 Sockeln skaliert und dieser Test einen Vergleich mit nur 8 Cores (entspricht bei Intel zwei CPU Sockel) zum Ziel hatte.

Die genaue Test-Konfiguration ist der nachfolgenden Tabelle zu entnehmen:

	IBM Power 8	HPE DL580 Gen9	HPE DL380 Gen9
System	IBM E870 MME	Proliant DL580 Gen9	Proliant DL380 Gen9
CPU	8 IBM Power8 4.2 GHz bis 80 Cores	4 Intel Xeon E7-8893 v4 3.20 GHz, Boost 3.50 GHz 16 Cores	2 Intel Xeon E5-2637 v4 3.50 GHz, Boost 3.70 GHz 8 Cores
Maximale Konfiguration	8 Sockel, 80 Cores	4 Sockel, 16 Cores	2 Sockel, 8 Cores
Getestete Konfiguration	8 Cores	2 Sockel, 8 Cores	2 Sockel, 8 Cores
Virtuelle Prozessoren	8 VP, 4 EC	8 VP	8 VP
Threads	SMT4, 32 Threads	Hyperthreading 16 Threads	Hyperthreading 16 Threads
Memory	128 GB	128 GB	128 GB
OS	AIX 7.1	RHEL 7.2	RHEL 7.2
Storage	HDS G1000 SSD Tier 8 Luns	HPE 3PAR StorServ 8450 All Flash 16 Luns	HPE 3PAR StorServ 8450 All Flash 16 Luns
Storageanbindung	4 x 8 Gbit	8 x 16 Gbit	8 x 16 Gbit
Virtualisierung	PowerVM	Bare Metal VMware ESX6 U2 RHV 4.0	Bare Metal VMware ESX6 U2 RHV 4.0
Oracle	12.1.0.2.0	12.1.0.2.0	12.1.0.2.0

Tabelle 1 – Benchmark Setup

Die HPE DL580 Gen9 und DL380 Gen9 Systeme werden im Folgenden als DL580 und DL380 bezeichnet.

Alle Messungen wurden mit virtuellen oder physischen Systemen mit jeweils 8 physischen Cores und aktiviertem Multithreading durchgeführt (IBM 4-fach Simultaneous Multithreading SMT-4, Intel Xeon 2-fach Hyperthreading). Auf der IBM werden die Cores mit allen Threads als virtuelle Prozessoren definiert (also 8 VP = 32 Threads bei SMT-4) auf den x86 Servern werden bei ESX und RHV jeweils die Threads konfiguriert (bei 2-fach Hyperthreading also 16 Threads und 8 Cores). Die DL380 verfügt über genau 8 Cores, bei der DL580 mit 16 Cores wurde per VM Pinning sichergestellt, dass die VM nur auf 8 Cores mit 16 Threads läuft und nicht die verbleibenden anderen 8 Cores mit nutzt. Bei den Bare Metal Tests auf DL580 wurden für diese Tests zwei CPUs physisch entfernt, damit nur 8 Cores verwendet wurden.

Auf IBM Systemen der Power8 Klasse ist auch noch 2-fach oder 8-fach Multithreading verfügbar. Vergleichstests mit SMT-8 haben keine Performanceverbesserungen und in machen Bereichen Performanceverschlechterungen gezeigt, deshalb wurde der Test mit SMT-4 durchgeführt, da dieser den besten Throughput zeigt.

Dieser Benchmark vergleicht also die Performance mit 8 physischen Cores, unabhängig von der maximal möglichen Konfiguration. Es ist darauf hinzuweisen, dass pro IBM P8 Core eine Oracle Enterprise Edition Lizenz benötigt wird, pro Intel x86 Core jedoch nur 0.5 Oracle Lizenzen verrechnet. Bei gleicher Anzahl Cores werden somit auf x86 Servern nur halb so viele Oracle Lizenzen benötigt wie auf Power Systemen.

Die Leistungsfähigkeit der Prozessoren IBM Power 8 und Intel Xeon E5-2637 v4 (verbaut im DL380) liegt laut SPECint bei jeweils 60 SPECint_base_rate pro Core gleich auf. Die Intel CPU E7-8893 v4 liegt etwas darunter bei 53.5 SPECint_base_rate (Quelle: www.spec.org).

Die Storageanbindung ist bei den hier verglichenen Benchmarks für die HPE Systeme deutlich stärker ausgelegt als bei den IBM Systemen. Beim IBM Setup ist mit 4 x 8 GBit/s der maximale theoretische Durchsatz 4 GB/s, beim HPE Setup mit 8 x 16 Gbps beträgt er theoretisch bis zu 16 GB/s. Wir werden im folgenden Papier darauf hinweisen, sofern diese unterschiedliche Storageanbindung eine Rolle spielt.

Benchmarking Tool

Die Benchmarks wurde mit dem In&Out Benchmarking Tool OraBench™ 6.9 durchgeführt. OraBench™ kann verschiedene Oracle Key Performance Indikatoren vollautomatisch messen und ermittelt die Transaktionsleistung, IO Kennzahlen, CPU und Memorybelastung und diverse weitere wichtige Kennzahlen auf dem Benchmarking Server.

Benchmarking Profil

Der Benchmark wurde mit der OraBench Datenbankgrösse L (LARGE) durchgeführt. Die wesentlichen Parameter sind der folgenden Tabelle zu entnehmen:

Parameter	N	S	M	L	VL	XL
Bezeichnung	Notebook	Small	Medium	Large	Very Large	Extra Large
# Cores	1-4	4-8	8-16	16-32	32-64	> 64
RAM	4-8 GB	8-64 GB	32-128 GB	64-256 GB	128-512 GB	> 512 GB
Database Size	64 GB	128 GB	256 GB	512 GB	1 TB	2 TB
SGA Total	1 GB	4 GB	16 GB	64 GB	256 GB	> 256 GB
db_cache_size	0.6 GB	3 GB	12 GB	42 GB	168 GB	
shared_pool_size	128 MB	256 MB	512 MB	1 GB	2 GB	
large_pool_size	128 MB	256 MB	512 MB	1 GB	2 GB	
log buffer	2 MB	8 MB	16 MB	32 MB	64 MB	
pga_aggregated_target	256 MB	1 GB	4 GB	8 GB	> 8 GB	
job_queue_processes	32	64	128	256	512	1024
parallel_max_servers	64	128	256	512	1024	2048
db_writers	1	2	4	8	16	>= 16
processes	128	256	512	1024	2048	4096

Tabelle 2 – OraBench Datenbankgrößen

Die folgende Tabelle fasst die 8 Testklassen von OraBench (T100-T800) zusammen. Pro Testklasse ist beschrieben, welche typischen Database Operationen diese umfasst, wie viele Tests in dieser Testklasse implementiert sind, was die Key Performance Indikatoren sind und welche Ressourcen besonders belastet werden.

Test Class	Real World Database Operations	# Tests	Key Performance Indicators	Ressource
T100	PL / SQL Tests	9	CPU Speed CPU Throughput	CPU
T200	Data Write	1	IO Throughput	Storage
T300	Data Load (Sequential)	7	Load Rate	CPU / Storage

Test Class	Real World Database Operations	# Tests	Key Performance Indicators	Ressource
T400	Data Scan (Sequential)	8	IO Throughput Scan Rate	CPU / Storage
T500	Data Aggregation	5	Processing Time	CPU / Storage
T600	Data Select (Random)	12	Transaction Service Time Transaction Throughput	CPU / Storage
T700	Data Update (Random)	6	Transaction Service Time Transaction Throughput IO Throughput	CPU / Storage
T800	Data Join	2	Transaction Service Time IO Throughput	CPU / Storage

Tabelle 3 – OraBench Tests

Insgesamt wurden auf allen gemessenen Systemen alle 50 Tests mit aufsteigender Parallelität durchgeführt. Der Test-setup seitens OraBench war für alle Systeme identisch. Während der Tests wurden die relevanten Oracle Parameter in AWR Reports dokumentiert und die Systembelastung ermittelt. Für die Darstellung in diesem Whitepaper wurden aus Platzgründen besonders wichtige und repräsentative Tests für die Darstellung ausgewählt.

Einige Tests sind mit * oder ** gekennzeichnet. Dabei handelt es sich um optionale Tests, die nicht zwingend durchgeführt werden müssen.

Erläuterung zu den Grafiken

Pro Grafik sind jeweils sieben Kurven dargestellt:

- Blau: IBM P8 / AIX
- Grün: HPE DL380 G9 / RHEL (3 Kurven)
- Gelb: HPE DL580 G9 / RHEL (3 Kurven)

Bei den HPE x86 Servern sind jeweils drei Kurven dargestellt:

- Durchgezogene Linie: RHEL auf Bare Metal (ohne Virtualisierung)
- Gestrichelte Linie: RHEL auf VMware ESXi
- Gepunktete Linie: RHEL auf RHV (Red Hat Enterprise Virtualisation)

Auf der X-Achse sind die Parallelitäten von 1 aufsteigend aufgeführt, auf der Y-Achse die jeweiligen Messwerte. Dies sind in der Regel kOps/s, also 1000 Operationen pro Sekunde. Ein Wert von 2'500 kOps bedeutet somit 2.5 Mio. Operationen pro Sekunde. Bei einzelnen Testreihen wird auch der IO Durchsatz in MB/s angegeben. Grundsätzlich ist ein hoher Wert "besser".

Benchmarks auf Oracle Datenbanken sind Benchmarks auf einem hochkomplexen System. Die Datenbank führt zu bestimmten Zeitpunkten, die sich nicht vollständig beeinflussen lassen, bestimmte Operationen durch, z.B. Redo Log File Switches; viele Operationen finden asynchron im Hintergrund statt und das Cache Verhalten ist nicht komplett deterministisch. Deshalb kann es bei einzelnen Tests durchaus zu gewissen Schwankungen kommen. Um diese Einflüsse und die Gesamtpformance zu erkennen, werden die Tests in aller Regel mit aufsteigender Parallelität gefahren und die Entwicklung der Performance mit zunehmender Parallelität beurteilt.

Erläuterung zu den Heatmaps

Die Heatmaps werden in zwei Ausprägungen dargestellt:

- Speed Heatmap zeigt die beste Single Thread Performance bei Parallelität 1
- Throughput Heatmap zeigt den besten Throughput bei beliebiger Parallelität

Dabei wird der jeweils beste Wert immer in dunkelgrün dargestellt, die jeweils schlechteren Werte werden über hellgrün, gelb, orange und rot abgestuft dargestellt. Dabei wird für rot immer der Wert 0 verwendet. Die Einfärbung erfolgt mit einer Standard Farbskala von Microsoft Excel.

Ergebnisse T100 PL/SQL Tests

Stellvertretend für die T100 PL/SQL Tests stellen wir die drei wichtigsten Tests aus dieser Messreihe vor. Die anderen Tests zeigen durchweg vergleichbare Ergebnisse. Die PL/SQL Tests sind sehr stark von der CPU Performance abhängig.

T132 PL/SQL Arithmetic Mix Number

Dieser Test führt eine Mischung von arithmetischen Operationen auf dem Oracle Datentyp NUMBER durch. Dieser Integer Datentyp kodiert jeweils eine Ziffer in ein Halbbyte (4 Bit) und rechnet auf diesen Halbbytes. Die Rechenweise ist grob vergleichbar mit einer schriftlichen Addition. Der Datentyp Number ist deshalb vergleichen mit einer direkten arithmetischen Operation auf einer Integer oder Floating Point Unit relativ langsam. Der Datentyp NUMBER wird häufig verwendet, da er in den Oracle Standard Libraries implementiert ist und nicht plattformspezifisch implementiert ist. Damit ist eine maximale Portabilität gewährleistet.

Innerhalb eines Durchlaufs werden 38 Operationen durchgeführt: 1 Random Generation (2.6%), 8 Additionen (21%), 7 Subtraktionen (18.5%), 4 Multiplikationen (10.5%), 3 Divisionen (7.9%), 6 ABS Funktionen (15.8%), 3 SQRT Funktionen (7.9%), 3 POWER Funktionen (7.9%), und 3 LN Funktionen (7.9%). Bei steigender Parallelität kann die Anzahl der Prozessoren und Threads optimal genutzt werden, da verschiedene arithmetische Operationen gleichzeitig ausgeführt werden.

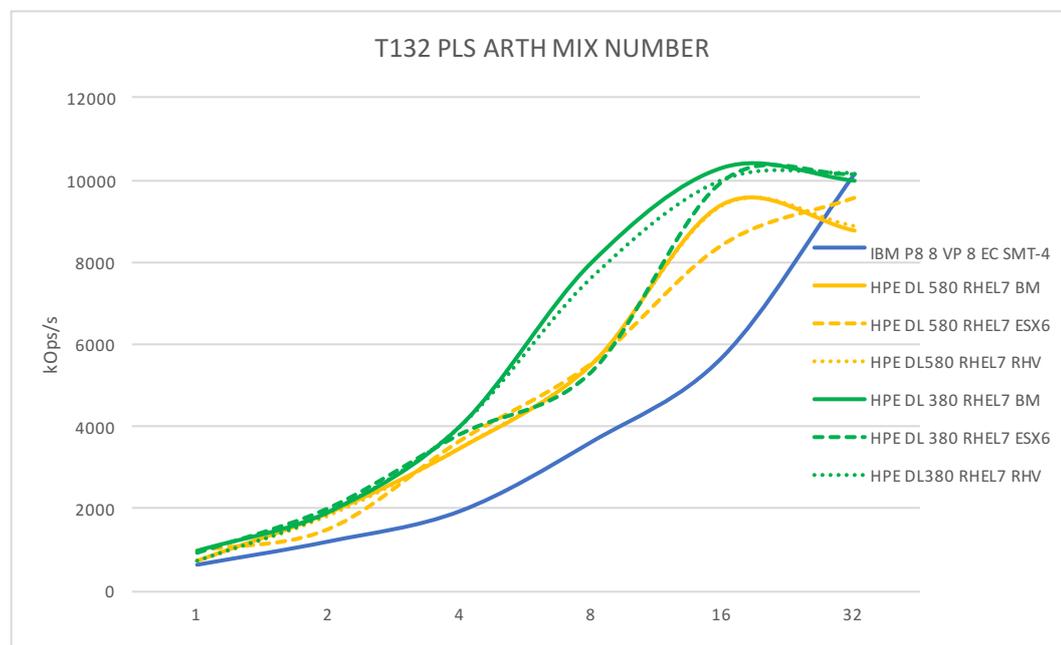


Abbildung 1 - T132 PL/SQL Arithmetic Mix Number

Die Ergebnisse der Tests zeigen, dass mit zunehmender Parallelität die Verarbeitungsleistung kontinuierlich bis zur maximalen Anzahl der verfügbaren Threads steigt. Bei IBM Power sind dies 32 Threads (8 Cores x 4 Threads) bei den x86 Systemen jeweils 16 Threads (8 Cores x 2 Threads). Bei diesem Test wird die Leistungsfähigkeit des Multithreadings ausgenutzt.

Erkennbar ist, dass bei gleicher Parallelität die x86 Systeme signifikant schneller sind als die Power Systeme. Die Single Thread Performance bei 1 Prozess ist bereits bis zu 40% höher, bei der optimalen Parallelität für die x86 Systeme von 16 ist der Performancevorteil ca. 80%. Das IBM System skaliert allerdings danach noch linear bis Parallelität 32 weiter, während die x86 Systemen mit nur 16 Threads keinen Leistungszuwachs mehr erzielen. Die Gesamtpformance pro Core ist für alle Systeme vergleichbar, die Power Prozessoren benötigen dafür allerdings die doppelte Anzahl von Prozessen.

Es gibt keine signifikanten Performanceeinbußen der verschiedenen Virtualisierungen auf x86 im Vergleich zu einem Bare Metal Setup. Grundsätzlich ist (wie bei allen CPU-lastigen Tests) der DL380 Server mit dem etwas höher getaktetem E5-2637 Prozessor leicht im Vorteil.

Die CPU Auslastung bei maximaler Parallelität liegt bei allen Tests bei 90-100%. Bei den x86 Systemen ist die CPU Belastung bereits bei 16 Prozessen so stark, bei den IBM Systemen aufgrund der höheren Anzahl Threads erst bei knapp 50%.

T133 PL/SQL Arithmetic Mix Float

Dieser Test führt eine Mischung von arithmetischen Operationen auf dem Oracle Datentyp FLOAT durch. Dieser Floating Point Datentyp wird in der Regel durch direkte arithmetische Operationen auf den Floating Point Units der CPU gerechnet. Er ist deshalb deutlich schneller als der Datentyp NUMBER und erreicht fast den 8-fachen Durchsatz (80 kOps statt 10 kOps).

Innerhalb eines Durchlaufs werden wie beim Datentyp NUMBER 38 Operationen in der gleichen Verteilung durchgeführt. Bei steigender Parallelität kann die Anzahl der Prozessoren und Threads optimal genutzt werden, da verschiedene arithmetische Operationen gleichzeitig ausgeführt werden.

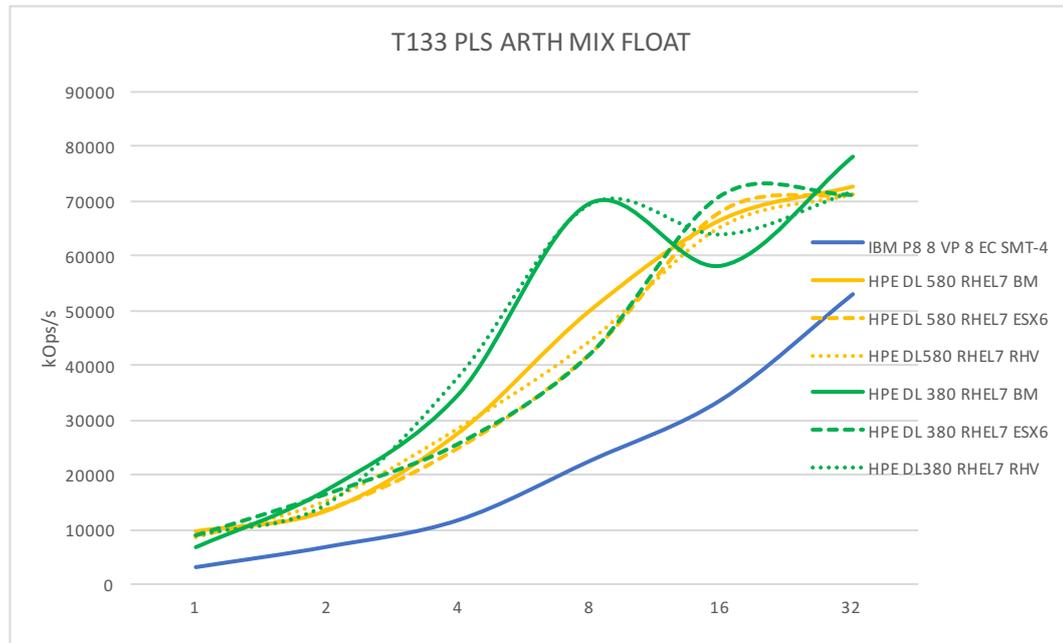


Abbildung 2 - T133 PL/SQL Arithmetic Mix Float

Die Ergebnisse der Tests zeigen, dass mit zunehmender Parallelität die Verarbeitungsleistung kontinuierlich bis zur maximalen Anzahl der verfügbaren Threads steigt. Bei IBM Power sind dies 32 Threads (8 Cores x 4 Threads) bei den x86 Systemen jeweils 16 Threads (8 Cores x 2 Threads). Bei diesem Test wird die Leistungsfähigkeit des Multithreadings ausgenutzt.

Erkennbar ist, dass bei gleicher Parallelität die x86 Systeme signifikant schneller sind als die Power Systeme. Die Single Thread Performance bei 1 Prozess ist bereits bis zu 100% höher, bei der optimalen Parallelität für die x86 Systeme von 16 ist der Performancevorteil ebenfalls 100%. Das IBM System skaliert allerdings danach noch bis Parallelität 32 weiter, während die x86 Systemen mit nur 16 Threads keinen Leistungszuwachs mehr erzielen. Die Gesamtperformance auf den x86 Systemen ist dennoch gut 30% höher.

Es gibt keine signifikanten Performanceeinbußen der verschiedenen Virtualisierungen auf x86 im Vergleich zu einem Bare Metal Setup. Grundsätzlich ist (wie bei allen CPU-lastigen Tests) der DL380 Server mit etwas höher getaktetem E5-2637 Prozessor leicht im Vorteil.

Die CPU Auslastung bei maximaler Parallelität liegt bei allen Tests bei 90-100%. Bei den x86 Systemen ist die CPU Belastung bereits bei 16 Prozessen so stark, bei den IBM Systemen aufgrund der höheren Anzahl Threads erst bei knapp 50%.

T141 PL/SQL String

Dieser Tests führt eine Mischung von String Operationen auf dem Oracle Datentyp VARCHAR2 (32) durch. Innerhalb eines Durchlaufs werden 29 Operationen durchgeführt: 1 Random String Generation (3.4%), 6 INSTR (20.7%), 9 CONCAT (31.0%), 3 LENGTH (10.3%), 3 SUBSTR (10.3%), 1 CHR (3.4%), 5 UPPER (17.2%), 1 MOD (3.4%). Bei steigender Parallelität kann die Anzahl der Prozessoren und Threads optimal genutzt werden, da verschiedene String- Operationen gleichzeitig ausgeführt werden.

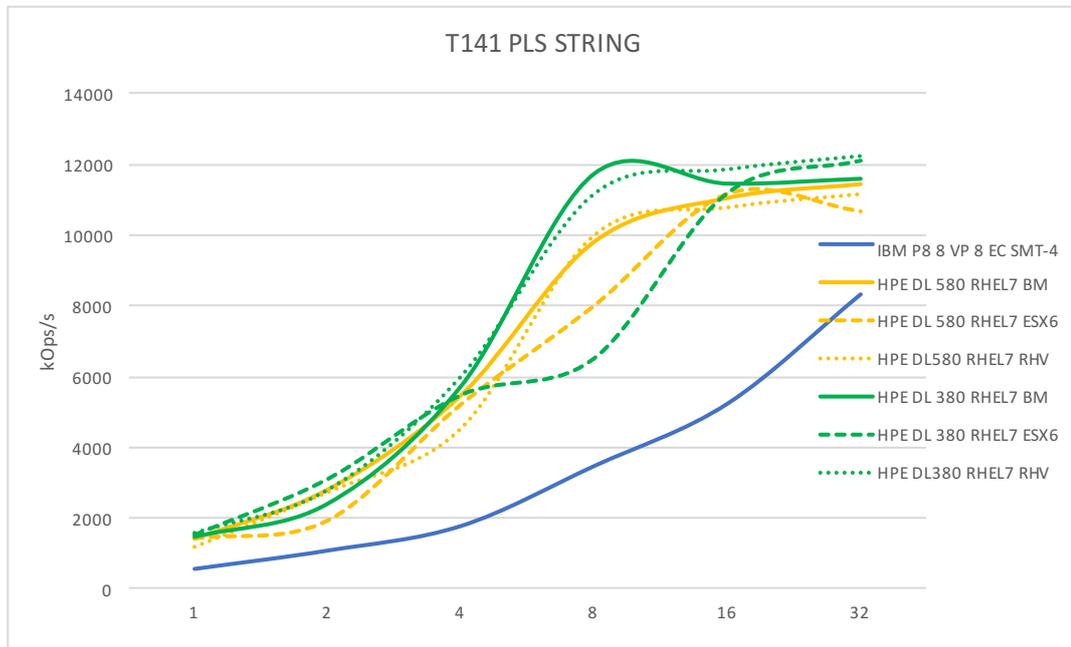


Abbildung 3 - T141 PL/SQL String

Die Ergebnisse der Tests zeigen, dass mit zunehmender Parallelität die Verarbeitungsleistung kontinuierlich bis zur maximalen Anzahl der verfügbaren Threads steigt. Bei IBM Power sind dies 32 Threads (8 Cores x 4 Threads) bei den x86 Systemen jeweils 16 Threads (8 Cores x 2 Threads). Allerdings wird bei den x86 Systemen in der Regel bereits bei 8 Prozessen, die maximale Verarbeitungsleistung erreicht, die Auswirkungen des Multithreadings kommen bei diesem Test auf x86 nicht stark zum tragen.

Erkennbar ist, dass bei gleicher Parallelität die x86 Systeme signifikant schneller sind als die Power Systeme. Die Single Thread Performance bei 1 Prozess ist bereits bis zu 100% höher, bei der optimalen Parallelität für die x86 Systeme von 8 ist der Performancevorteil ebenfalls 100%. Das IBM System skaliert allerdings danach noch bis Parallelität 32 weiter, während die x86 Systemen ab 16 parallelen Prozessen keinen signifikanten Leistungszuwachs mehr erzielen. Die Gesamtperformance auf den x86 Systemen ist dennoch gut 35% höher.

Es gibt keine signifikanten Performanceeinbußen der verschiedenen Virtualisierungen auf x86 im Vergleich zu einem Bare Metal Setup. Grundsätzlich ist (wie bei allen CPU-lastigen Tests) der DL380 Server mit etwas höher getaktetem E5-2637 Prozessor leicht im Vorteil.

Die CPU Auslastung bei maximaler Parallelität liegt bei allen Tests bei 90-100%. Bei den x86 Systemen ist die CPU Belastung bereits bei 16 Prozessen so stark, bei den IBM Systemen aufgrund der höheren Anzahl Threads erst bei knapp 50%.

Zusammenfassung T100 PL/SQL Tests

Die T100 Tests sind sehr CPU intensive Tests. Der Einfluss der Virtualisierung ist jeweils sehr gering.

In der folgenden Tabelle wird die Single Thread Performance mit Parallelität 1 (=Speed) für die Plattformen dargestellt. Man erkennt, dass mehrheitlich der DL380 mit dem 3.5 GHz Prozessor am schnellsten ist, dicht gefolgt vom DL580 mit dem 3.2 GHz Prozessor. In der Single Thread Performance ist der IBM P8 trotz der schnellsten Taktung von 4.2 GHz deutlich weniger leistungsstark.

Tests	Description	IBM P8	DL580 BM	DL580 VM	DL580 RV	DL380 BM	DL380 VM	DL380 R
T111	PL SQL Loop	10204	18519	22727	20833	21739	22727	22727
T121	PL SQL Random Number	176	417	455	463	510	472	521
T122	PL SQL Random String generation	22	53	58	53	61	63	63
T131	PL SQL arithmetic mix, integer	556	952	833	952	769	1000	1000
T132*	PL SQL arithmetic mix, number	645	741	667	909	952	1000	952
T133*	PL SQL arithmetic mix, float	3175	9524	6061	7143	6897	9524	9524
T134*	PL SQL arithmetic mix, double	3846	6061	10000	6250	10000	10000	6667
T141*	PL SQL string mix	561	1395	1395	1429	1463	1579	1579
T151*	PL SQL fibonacci number n=39	32	20	20	20	22	20	24

Tabelle 4 - T100 PL/SQL Speed Heatmap (höhere Werte = besser, ausser bei T151)

In der folgenden Tabelle wird der maximale Durchsatz bei jeweils 8 physischen Cores (= Throughput) dargestellt. Es zeigt sich ein ähnliches Bild wie im Speed Test, allerdings ist der P8 in einigen Tests (T131-133) auf Augenhöhe und in den anderen Tests zumindest deutlich näher an den mit den HPE X86 RHEL Ergebnissen. Hier kommt der Vorteil des SMT-4 des IBM P8 Servers im Vergleich zum 2-fach Hyperthreading der Intel CPUs zum tragen. Insgesamt zeigen aber die Intel basierten Server immer noch einen deutlich höheren Durchsatz bei den meisten Tests.

Tests	Description	IBM P8	DL580 BM	DL580 VM	DL580 RV	DL380 BM	DL380 VM	DL380 R
T111	PL SQL Loop	153846	172043	170213	166667	179775	181818	181818
T121	PL SQL Random Number	2649	3828	3687	3604	3922	4000	3941
T122	PL SQL Random String generation	337	451	456	438	487	478	479
T131	PL SQL arithmetic mix, integer	10000	9697	9697	9143	10323	10159	10159
T132*	PL SQL arithmetic mix, number	10159	9412	9412	8312	10323	10323	9846
T133*	PL SQL arithmetic mix, float	52893	21	33	28	29	21	21
T134*	PL SQL arithmetic mix, double	57658	76190	71111	72727	82051	80000	72727
T141*	PL SQL string mix	8348	11429	11163	10847	11707	12229	12000

Tabelle 5 - T100 PL/SQL Throughput Heatmap

Ergebnisse T200 Data Write

Dieser Test legt neue Datenbank Files mit einer Größe von jeweils 8 GB in Oracle an. Zunächst wird 1 File zu 8 GB, dann 2, 4, 8 und 16 Files parallel angelegt, insgesamt somit 31 Dateien mit insgesamt 248 GB.

T211 Data Write

Dieser Test ermittelt die Sequential Write Performance von Oracle. Sequentielle Writes kommen in Oracle z.B. bei Redo Log Writer Process LGWR, RMAN Backup und Restore, Export, Data Pump, Redo Log Archiver ARCH, Austauschen von Sortierbereichen im TEMP Tablespace und dem Recovery Writer RVWR vor und ist ein wichtiger Indikator für die Effizienz des gesamten IO Stacks.

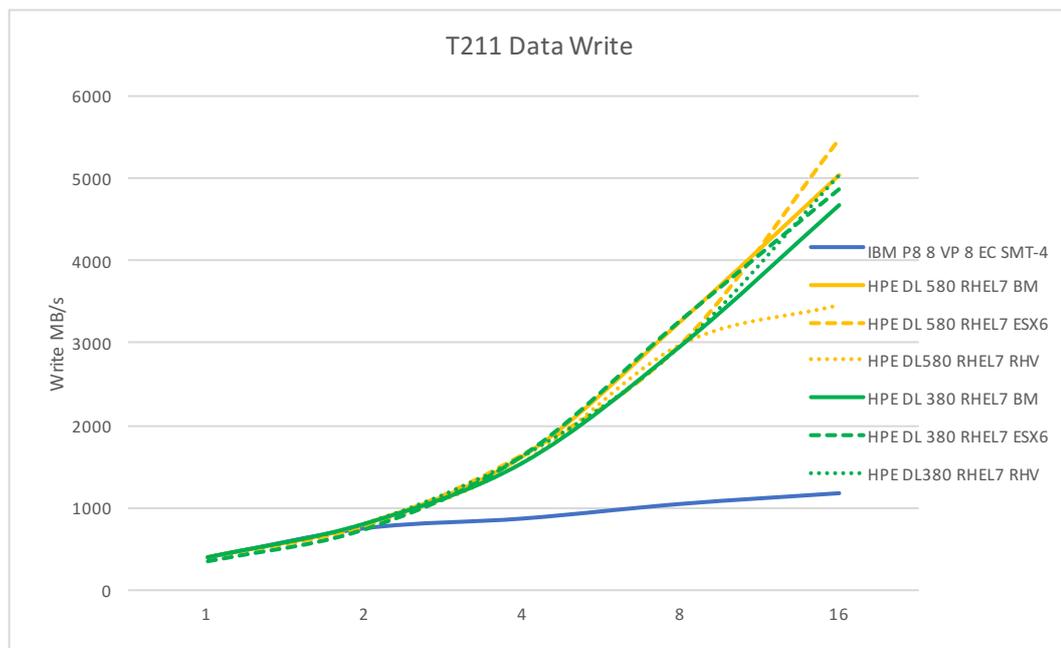


Abbildung 4 - T211 Data Write

Die Ergebnisse des Tests zeigen, dass mit zunehmender Parallelität die Sequential Write Performance ansteigt. Bei den HPE Intel RHEL Systemen steigt diese quasi linear bis zu einer Bandbreite von 5 GB/s, die Oracle effektiv auf den Storage schreiben kann, bei den IBM Power System werden nur gut 1 GB/s erreicht. Dabei muss allerdings darauf hingewiesen werden, dass die Storageanbindung des IBM Systems mit 4 x 8 Gbit (maximale theoretische Bandbreite 4 GB/s) deutlich geringer war als bei den HPE Systemen mit 8 x 16 Gbit (maximale theoretische Bandbreite 16 GB/s). Ein direkter Vergleich der Data Write Performance ist damit nicht zulässig, alle Systeme holen ca. 30% der theoretischen Storagebandbreite ab.

Die Verarbeitungsleistung könnte bei allen Systemen sicher noch deutlich steigen, wenn noch höhere Parallelitäten zum Einsatz kommen würden. Die CPU Leistung ist auf der IBM im Bereich von unter 5% bei den HPE Systemen bei vierfacher Write Leistung bei 5-13%.

Die wesentliche Erkenntnis dieses Tests ist, dass sämtliche HPE Intel RHEL Plattformen mit 4-5 GB/s in Oracle Daten schreiben können und dies auch mit aktivierter Virtualisierung. Das sind beeindruckende Zahlen die in der Realität selten notwendig sind.

Ergebnisse T300 Data Load

Die T300 Tests simulieren das Einfügen von grossen Datenmengen in eine Oracle Datenbank. Dies geschieht über das Kopieren einer Tabelle auf zwei grundsätzliche Arten:

- T310 und T320: Conventional Load:
Commit nach jeweils 2 Rows, letzter Test nach jeweils 10 Rows, typisch für OLTP Umgebungen
- T330: Bulk Load: Commit jeweils am Ende des Kopiervorganges, typisch für DWH Load

Weiterhin wird zwischen der Art der Tabellen unterschieden:

- T310: Kleine Tabelle, non-partioned
- T320: Kleine Tabelle, list-partioned
- T330: Grosse Tabelle, range-list partioned

Zusätzlich werden ausserdem verschiedene Speicherstrukturen getestet:

- Heap organized, Test 311, 321, 331 (uncompressed) und 332 (compressed)
- Index organized, Test 312, 322
- Hash-Cluster organized, Test 313

T321 Data Load Conventional Heap Organized, List-Partitioned

Der Test führt einen konventionellen Insert durch, bei dem jeweils nach 2 Rows ein Commit erfolgt. Dies wird mit einer Parallelität von 1-8 durchgeführt. Der Test mit Parallelität 1 wird zweimal durchgeführt, damit die Anzahl der eingefügten Zeilen stimmt. Der letzte Test mit Parallelität 8 wird mit einem Commit nach jeweils 10 Rows wiederholt. Dadurch vervielfacht sich der Durchsatz.

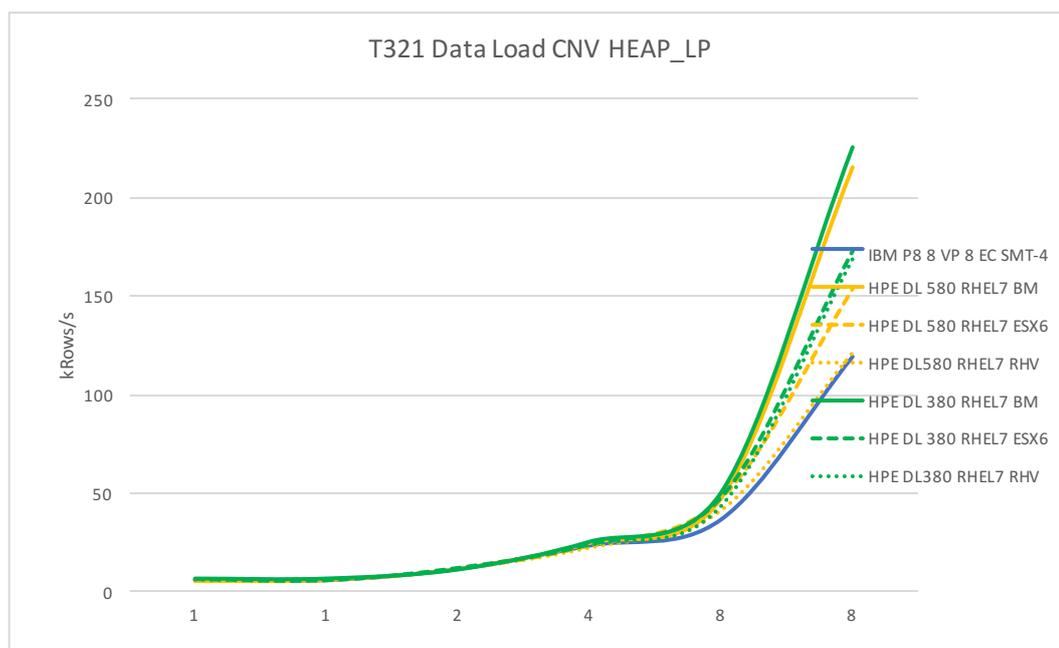


Abbildung 5 - T321 Data Load Conventional heap-organized, list-partioned table

Zu Beginn mit 2 Rows pro Commit zeigen sich zwischen allen Plattformen keine signifikanten Unterschiede. In dieser Phase liegt die Limitierung intern in der Oracle Datenbank, die extrem vielen Commits zu verarbeiten. Erst beim Wechsel auf 10 Rows pro Commit kommen die Unterschiede zum Tragen. Am schnellsten sind hier die beiden HPE Bare Metal Systeme ohne Virtualisierung. Die Virtualisierung per VMware ESXi und RHV erreichen etwas geringere Maximalwerte. Die IBM Werte liegen am unteren Rand dieser Bandbreite.

Bei allen Systemen kann hier nicht die maximale Anzahl von Threads ausgereizt werden, dies trifft die IBM Plattform mit 32 Threads deutlich stärker als die HPE Systeme mit 16 Threads. Demzufolge liegt die CPU Auslastung auf den HPE Systemen bei 50%, auf dem IBM System nur bei 25%. Es ist zu erwarten, dass bei höherer Parallelität das IBM System beim Durchsatz aufholen würde. Die IO Anbindung spielt bei diesem Test keine wesentliche Rolle, dafür sind die maximalen Bandbreiten mit 500 MB/s zu gering.

T331 Data Load Bulk Heap Organized, Rangelist-Partitioned

Es wird ein Bulk Insert mit einer Parallelität von 1-16 durchgeführt, mit jeweils einem Commit zum Abschluss. Der Test mit Parallelität 1 wird am Anfang zweimal durchgeführt, damit die Gesamtzahl der eingefügten Zeilen den Vorgaben entspricht.

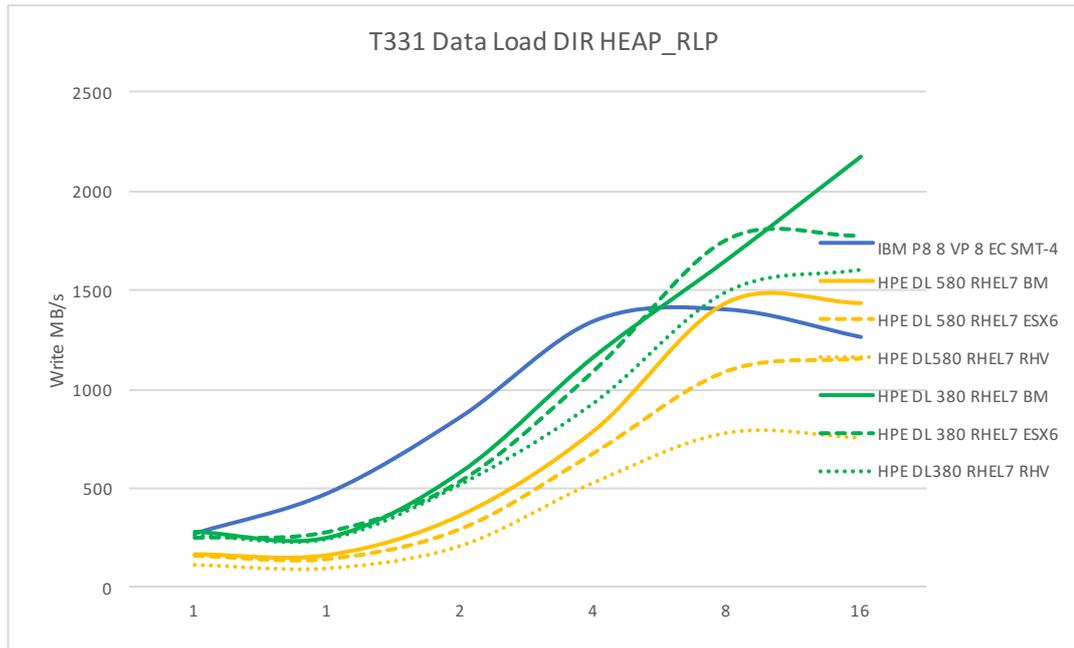


Abbildung 6 - T331 Data Load Bulk heap-organized, rangelist-partioned table

Bei den HPE Intel RHEL Systeme skaliert dieser Test bis zu einer Parallelität von 8 recht gut, bei 16 parallelen Prozessen zeigt sich keine oder nur noch eine geringe Steigerung des Durchsatzes. Generell sieht man, dass die Bare Metal Variante sehr gut funktioniert und die Virtualisierung einen schlechteren Durchsatz aufweist. Nicht ganz erklärbar sind an dieser Stelle die deutlich schlechteren Ergebnisse des DL580 Servers, die in diesem Ausmass nicht mit der etwas langsamer getakteten CPU erklärt werden können. Eine Analyse der AWR Reports hat einen deutlich höheren Anteil am Wait Event "Log buffer space" gezeigt. Der LGWR Prozess kann die Daten nicht schnell genug wegschreiben. Die Werte konnten bei einer Wiederholung der Test reproduziert werden, eine weitere Analyse war in der verfügbaren Zeit leider nicht möglich.

Die IBM Systeme erreichen hier einen deutlich schlechteren Durchsatz, der allerdings partiell durch die weniger leistungsfähige Storageanbindung verursacht wird. In diesem Test werden bis zu 2 GB/s IO verursacht. Die CPU Auslastung liegt auf den HPE Intel RHEL Systemen mit sehr hohem Durchsatz bei bis zu 75%, auf den IBM Systemen können nur maximal 15% der CPUs ausgelastet werden.

T332 Data Load Bulk Compressed Heap Organized, Rangelist-Partitioned

Es wird ein Bulk Insert mit einer Parallelität von 1-16 durchgeführt, mit jeweils einem Commit zum Abschluss. Der Test mit Parallelität 1 wird am Anfang zweimal durchgeführt, damit die Gesamtzahl der eingefügten Zeilen den Vorgaben entspricht.

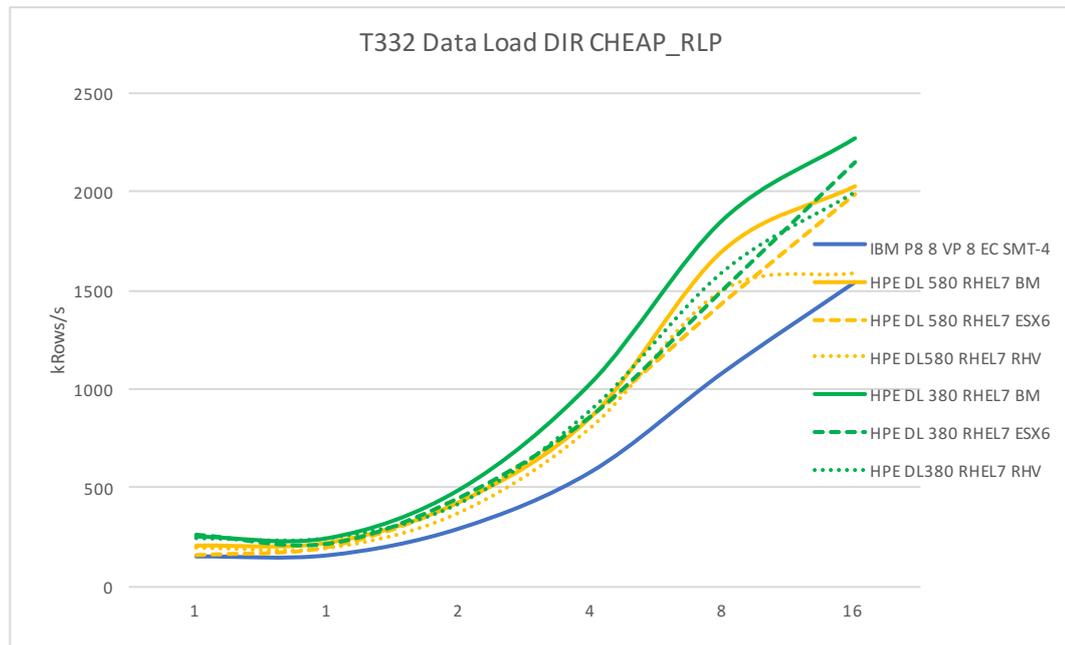


Abbildung 7 - T332 Data Load Bulk compressed heap-organized, rangelist-partioned table

Im Gegensatz zu den T331 Tests skalieren hier alle HPE Intel RHEL Systeme bis zu einer Parallelität von 16 Prozessen und auch bei den virtualisierten Umgebungen zeigt sich kein signifikanter Performanceimpact (ausser bei RHV auf der DL580)

Die IBM Systeme erreichen hier einen etwas schlechteren Durchsatz. Im Vergleich zum Test T331 muss aufgrund des komprimierten Heap Organized Indexes sehr viel weniger auf die Disk geschrieben werden, sodass der Engpass nicht hier zu suchen ist. Generell ist die IBM Plattform vermutlich aufgrund der 4-fach Threads in der Lage über die Parallelität 16 weiter zu skalieren.

Zusammenfassung T300 Data Load Tests

Die T300 Load Tests belasten vor allem im Bereich der Conventional Loads (T310 / T320) die internen Mechanismen der Datenbank sehr stark, vor allem den Log Writer. Die T300 Bulk Tests haben einen sehr deutlichen IO Anteil.

In der folgenden Tabelle wird die Single Thread Performance mit Parallelität 1 (=Speed) für die Plattformen dargestellt. Man erkennt, dass es im Bereich der Conventional Loads keine signifikanten Unterschiede gibt, bei den Bulk Loads sind die DL380 Systeme jeweils am schnellsten, gefolgt von den DL580 Systemen und der IBM Power Plattform.

Tests	Description	IBM P8	DL580 BM	DL580 VM	DL580 RH	DL380 BM	DL380 VM	DL380 RH
T311	Data Load conventional, heap non partitioning	6.25	5.95	5.95	6.25	5.68	6.25	5.68
T312*	Data Load conventional, iot non portioning	6.25	5.68	5.68	5.68	6.25	5.68	5.95
T313*	Data Load conventional, hash cluster non portioning	5.95	6.25	3.38	5.68	5.95	5.95	5.95
T321**	Data Load conventional, heap list-partioning	6.25	5.68	6.25	5.95	6.25	6.25	6.25
T322**	Data Load conventional, iot list-partioning	5.95	6.25	5.43	5.68	6.25	5.95	5.68
T331	Data Load bulk direct, uncompressed	210.53	363.64	210.53	326.53	615.38	615.38	615.38
T332*	Data Load bulk, direct, compressed	150.94	210.53	207.79	210.53	253.97	258.06	228.57

Tabelle 6 - T300 Data Load Speed Heatmap

In der folgenden Tabelle wird maximale Durchsatz bei jeweils 8 physischen Cores (= Throughput) dargestellt. Der beste Durchsatz wird auf den DL380 Systemen erreicht, während die DL580 Systeme insgesamt einen geringeren Throughput aufweisen, jeweils mit einem spürbaren Einfluss der Virtualisierung. Die Performance der IBM Plattform ist hier in aller Regel schlechter.

Tests	Description	IBM P8	DL580 BM	DL580 VM	DL580 RH	DL380 BM	DL380 VM	DL380 RH
T311	Data Load conventional, heap non partitioning	88.05	85.89	59.32	62.78	166.67	160.92	140
T312*	Data Load conventional, iot non portioning	64.81	37.84	28.28	25.74	95.89	79.55	63.93
T313*	Data Load conventional, hash cluster non portioning	51.47	35.81	26.92	26.67	89.74	75.68	67.31
T321**	Data Load conventional, heap list-partioning	119.66	215.38	186.67	123.89	225.81	229.51	184.21
T322**	Data Load conventional, iot list-partioning	120.69	218.75	186.67	101.45	218.75	208.96	170.73
T331	Data Load bulk direct, uncompressed	1142.86	2245.61	1729.73	1620.25	3240.51	2612.24	2327.27
T332*	Data Load bulk, direct, compressed	1542.17	2031.75	1969.23	1802.82	2265.49	2169.49	1882.35

Tabelle 7 - T300 Data Load Throughput Heatmap

Ergebnisse T400 Data Load

Die Tests im Bereich T400 sind quasi das lesende Gegenstück zu den schreibenden T200 Tests. Sie simulieren einen sequentiellen Scan über grosse Oracle Tabellen und sind sehr storagelastig. Data Scan Operationen kommen in Oracle sehr häufig vor (z.B. bei Table Scans, Index Scans oder Cluster Scans) und sind eine der Schlüsselfaktoren für eine gute Oracle Performance.

T426 Data Scan

Dieser Test ist ein aufschlussreicher Test für die Sequential Read Performance von Oracle. Dabei wird ein non-cached Full Table Scan mit einer steigenden Anzahl von Prozessen durchgeführt. Dieser Test ist ein wichtiger Indikator für die Effizienz des gesamten IO Stacks. Pro IO Request wurden dabei 64 8 KB Blöcke gelesen, also 512 KB pro IO Request.

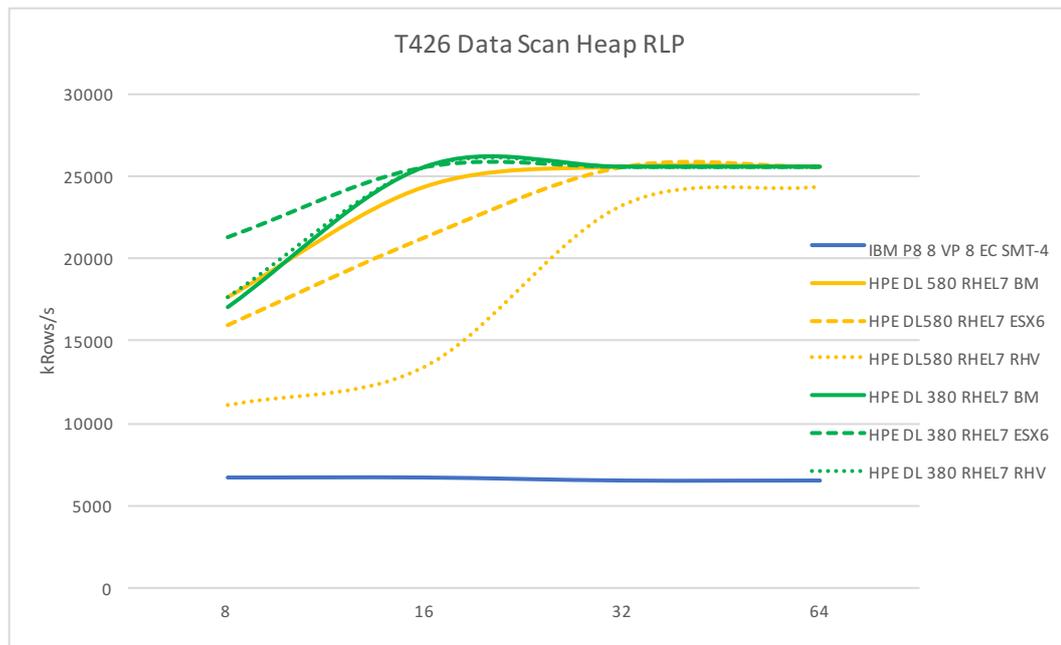


Abbildung 8 - T426 Data Scan

Die Ergebnisse des Tests zeigen, dass mit zunehmender Parallelität die Scan Rate sehr schnell auf das Maximum von knapp 6 GB/s für die HPE Intel RHEL Systeme steigt und auf der IBM P8 Plattform nur ca. 1.6 GB/s erreicht werden. Dabei muss allerdings darauf hingewiesen werden, dass die Storageanbindung des IBM Systems mit 4 x 8 Gbit (maximale theoretische Bandbreite 4 GB/s) deutlich geringer war als bei den HPE Systemen mit 8 x 16 Gbit (maximale theoretische Bandbreite 16 GB/s). Ein direkter Vergleich der Data Write Performance ist damit nicht zulässig, relativ gesehen holen alle Systeme ca. 35% der theoretischen Storagebandbreite ab.

Die CPU Leistung ist auf dem IBM System im Bereich von 6-7% bei den HPE Systemen bei vierfacher Read Leistung bei 30-60%. Dieser Test ist im wesentlichen IO-Bound.

Die wesentliche Erkenntnis dieses Tests ist, dass sämtliche Red Hat / HPE Plattformen mit fast 6 GB/s in Oracle Daten scannen können und dies auch mit aktivierter Virtualisierung. Das sind beeindruckende Zahlen, die in der Realität oft einen entscheidenden Performancevorteil ausmachen können.

Ergebnisse T500 Data Aggregation

Die Tests im Bereich T500 erzeugen Aggregationen, wie sie in verschiedenen Datenbankanwendungen sowohl im OLTP wie im DWH Bereich vorkommen, z.B. Sortieren, Gruppieren, Auswertungen wie AVG, MAX, MIN, etc.).

T512 Data Aggregation Create BTree (with Logging)

Bei diesem Test wird ein BTree Index mit einer steigenden Anzahl von Prozessen erzeugt. Dieser Test ist ein wichtiger Indikator für die Effizienz des gesamten IO Stacks. Pro IO Request wurden dabei 512 KB gelesen.

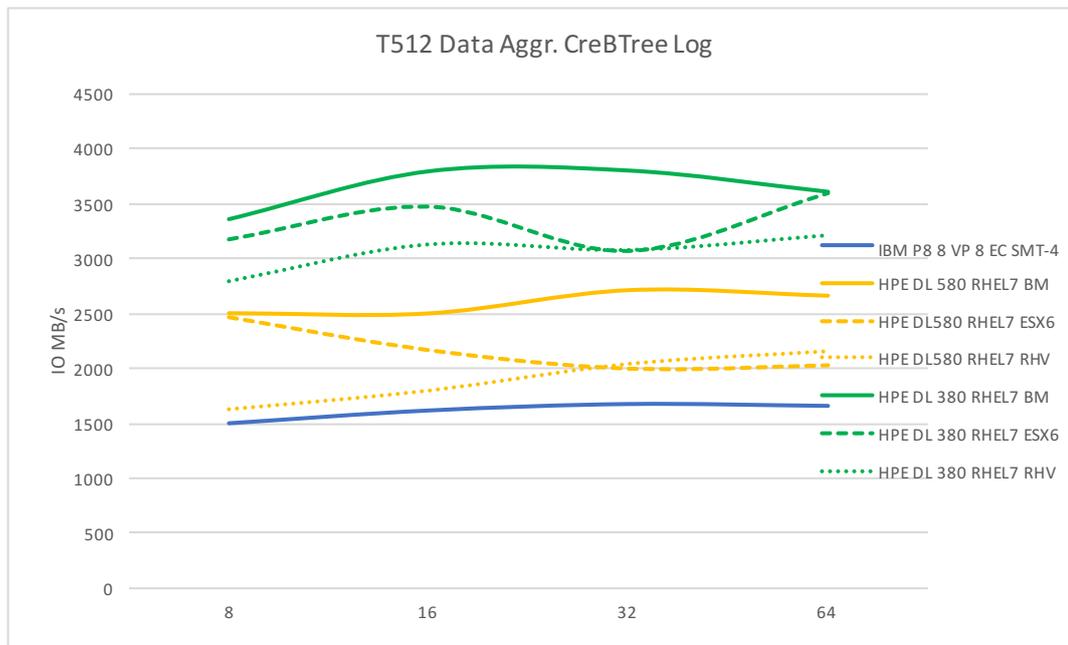


Abbildung 9 - T512 Data Aggregation Create BTree Index (with Logging)

Die Ergebnisse des Tests zeigen dass bereits bei geringer Parallelität von 8 Prozessen bis zu 4 GB/s IO Durchsatz erzeugt wird und dieser mit zunehmender Parallelität nicht weiter ansteigt. Auf dem IBM P8 System können nur ca. 1.6 GB/s Durchsatz erreicht werden. Dabei muss allerdings darauf hingewiesen werden, dass die Storageanbindung des IBM Systems mit 4 x 8 GBit (maximale theoretische Bandbreite 4 GB/s) deutlich geringer war als bei den HPE Systemen mit 8 x 16 Gbit (maximale theoretische Bandbreite 16 GB/s). Ein direkter Vergleich der Data Write Performance ist damit nicht zulässig. Auffällig ist, dass die virtualisierten Systeme deutlich weniger IO Throughput liefern und die DL580 wiederum schlechtere Ergebnisse als die DL380 liefert.

Die CPU Leistung liegt auf dem IBM System bei 20% und auf den HPE Systemen bei höherer IO Leistung bei 60-80%.

Ergebnisse T600 Data Select

Die T600 Tests simulieren den zufälligen Zugriff auf einzelne Oracle Datenbank Blöcke (Random Select). Diese Operationen sind sehr typisch für OLTP Anwendungen, kommen aber auch in Big Data oder DWH Applikationen vor. Grundsätzlich wird unterschieden zwischen den folgenden Tests:

- T610: Zugriff über primary key mit geringer Selektivität (geringe Treffermenge)
- T620: Zugriff über secondary key mit hoher Selektivität (hohe Treffermenge)

Dabei werden folgende 6 Tests jeweils mit primary und secondary key Zugriff durchgeführt:

- T6x1: heap organized, non-partitioned, **cached**
- T6x2: index organized, non-partitioned, **cached**
- T6x3: hash cluster organized, non-partitioned, **cached**
- T6x4: heap organized, list partitioned, **cached**
- T6x5: index organized, list-partitioned, **cached**
- T6x6: heap organized, rangelist-partitioned, **non-cached**

Zu beachten ist, dass die ersten fünf Tests jeweils alle cached sind, also auf einer sehr kleinen Tabelle arbeiten, die relativ schnell in die SGA geladen wird. Hier spielt die Storageanbindung keine wesentliche Rolle, sondern vor allem die Verarbeitungsleistung des DB Servers. Die Tests T616 und T626 sind non-cached Tests auf einer Tabelle, die nicht in die SGA passt. Bei diesen Tests spielt insbesondere die Storgelatenzy eine grosse Rolle für die Verarbeitung und es werden insbesondere im Test T626 mit grosser Treffermenge sehr viele IOs generiert.

T611 Data Select Primary Key Heap Organized, Non-Partitioned, Cached

Stellvertretend für die cached T61x Tests wird hier der Test 611 dargestellt. Die anderen Tests T612-T615 zeigen vergleichbare Resultate. Die Tests führen Random Selects über den Primary Key durch, mit einer sehr kleinen Treffermenge (1 Row).

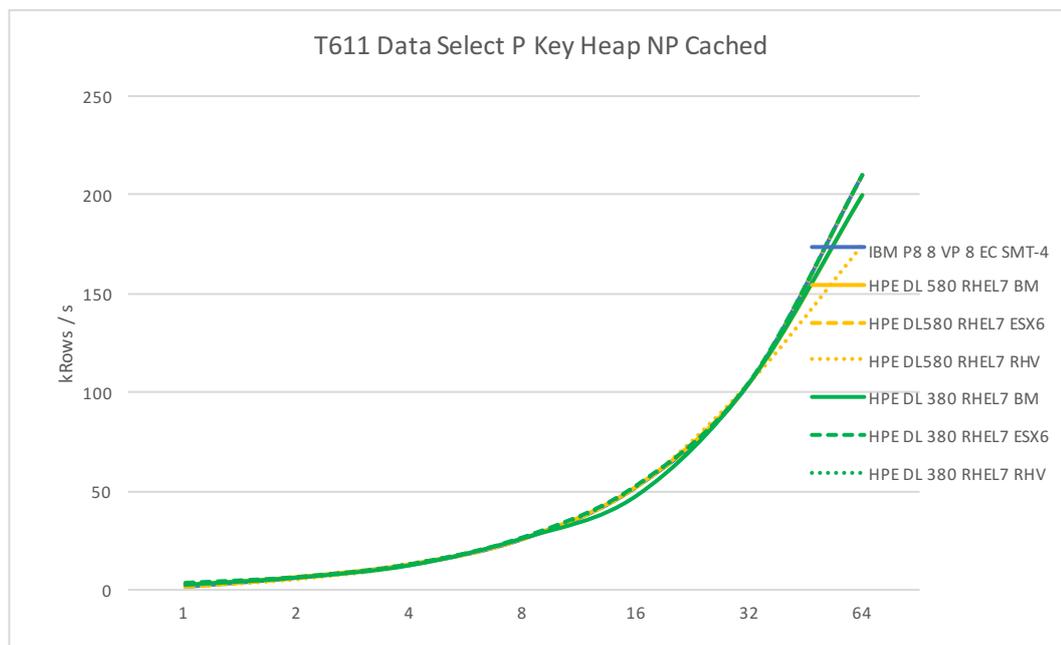


Abbildung 10 - T611 Data Select Primary Key Heap Organized, Non Partitioned, Cached

Die Ergebnisse sind für alle Plattformen nahezu identisch (die Kurven überlappen sich teilweise exakt), es wird jeweils eine Verarbeitungsleistung von ca. 200'000 Rows/s erreicht. Der Durchsatz steigt kontinuierlich bis zur maximal gewählten Parallelität von 64 Prozessen an. Mit 64 Prozessen lag die CPU Belastung beim IBM Server bei 40%, bei den HPE Intel RHEL Servern bei 50%. Die maximale Verarbeitungsleistung könnte mit zusätzlicher Parallelisierung noch weiter gesteigert werden. Die IO Leistung spielt bei den Cached Tests keine wesentliche Rolle.

T616 Data Select Primary Key Heap Organized, Rangelist-Partitioned, Non-Cached

Der Test T616 führt Random Selects auf einer sehr grossen Tabelle durch, dadurch können die Resultate nicht im grösseren Stil gecached werden. Die Treffermenge der Selects ist jeweils nur eine Row.

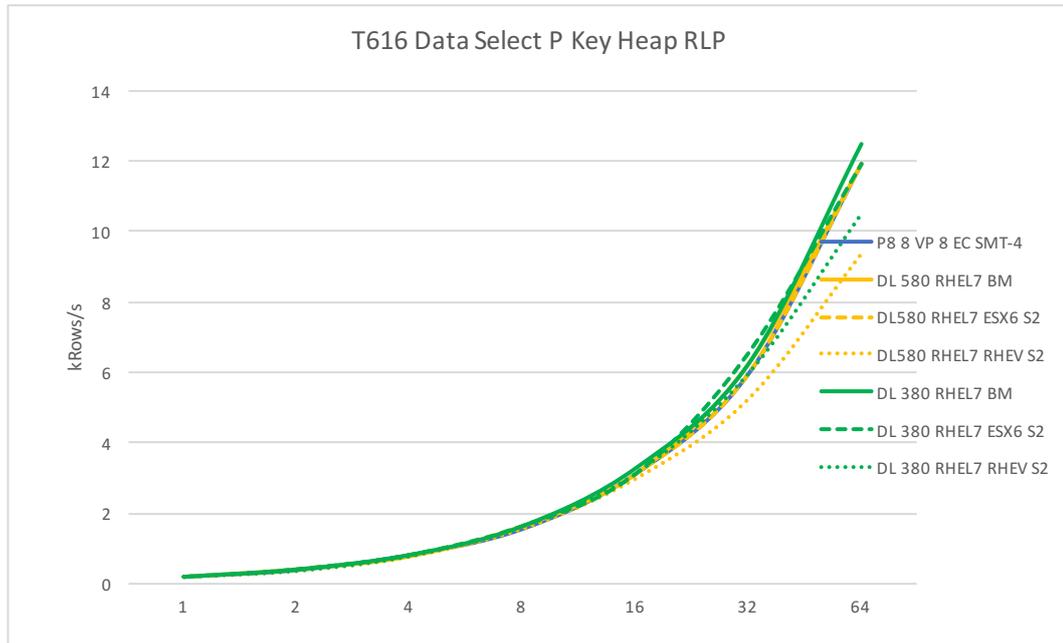


Abbildung 11 - T616 Data Select Primary Key Heap Organized, Rangelist-Partitioned, Non-Cached

Die Ergebnisse sind für alle Plattformen sehr ähnlich (die Kurven überlappen sich teilweise exakt), es wird jeweils eine Verarbeitungsleistung von ca. 9-12'000 Rows/s erreicht. Der Durchsatz steigt kontinuierlich bis zur maximal gewählten Parallelität von 64 Prozessen an.

Die Verarbeitungsleistung auf der non-cached Tabelle liegt mit ca. 10'000 Rows/s um Faktor 20 niedriger als bei der Tabelle im Cache mit 200'000 Rows/s. Hier erkennt man auch das Beschleunigungspotential von Datenbankzugriffen im Cache verglichen mit Datenbankzugriffen auf dem Storage (selbst wenn dieser als ultraschneller SSD Storage zur Verfügung steht).

Die CPU Utilization lag bei allen Systemen sehr niedrig (< 20%), in diesem Test spielt vor allem die Storage Latency eine wesentliche Rolle. Diese lag bei allen getesteten Konfiguration gemäss AWR Report bei ca. 0.5 ms.

T621 Data Select Secondary Key Heap Organized, Non-Partitioned, Cached

Stellvertretend für die T62x Tests wird der Test 621 hier dargestellt. Die Tests T622-T625 zeigen vergleichbare Resultate. Die Tests führen Random Selects über den Secondary Key durch, die eine grössere Treffermenge liefern.

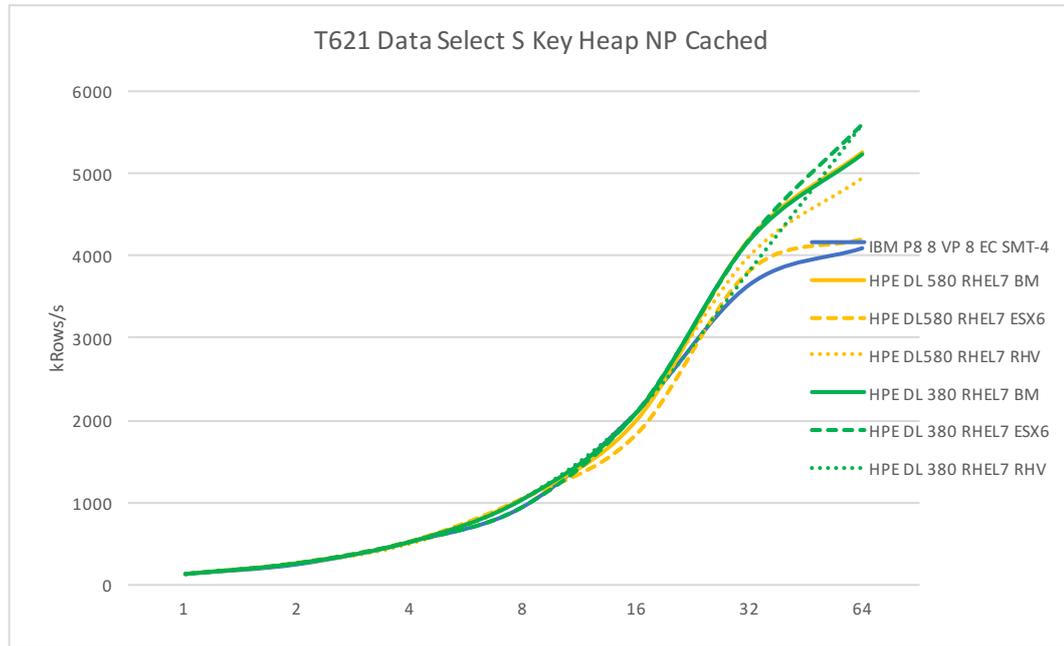


Abbildung 12 - T621 Data Select Secondary Key Heap Organized, Non Partitioned, Cached

Die Ergebnisse sind für alle Plattformen ähnlich, es wird jeweils eine Verarbeitungsleistung von 4 bis 5.6 Mio. Rows pro Sekunde erreicht. Der Durchsatz steigt kontinuierlich bis zur maximal gewählten Parallelität von 64 Prozessoren an. Bedingt durch die grössere Treffermenge ist dieser Test sehr CPU intensiv. Bei den HPE Intel RHEL Systemen lag die CPU Belastung mit 64 Prozessoren bereits bei fast 100%, beim IBM System bei 75%. Die Art der Virtualisierung auf den HPE Intel RHEL Servern spielt keine entscheidende Rolle. Die IO Leistung spielt Tests keine wesentliche Rolle.

T626 Data Select Secondary Key Heap Organized, Rangelist-Partitioned, Non-Cached

Der Test T626 führt Random Selects auf einer sehr grossen Tabelle durch, dadurch können die Resultate nicht gecached werden. Die Treffermenge der Selects ist jeweils gross und somit spielt die IO Leistung eine grosse Rolle.

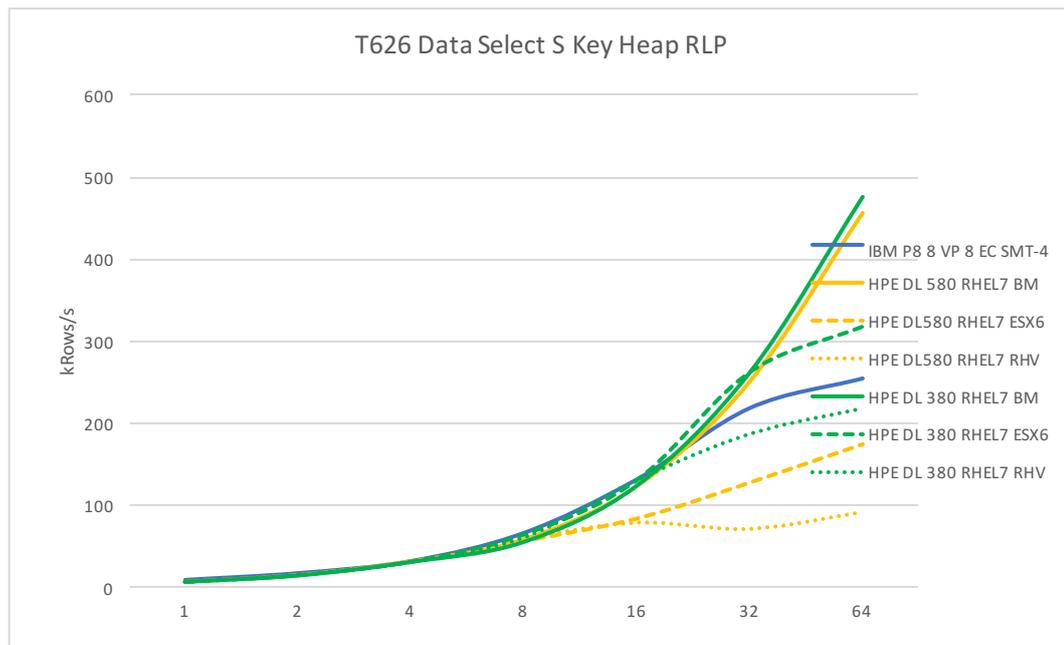


Abbildung 13 - T626 Data Select Secondary Key Heap Organized, Rangelist-Partitioned, Non-Cached

Die Ergebnisse unterscheiden sich bei diesen Tests relativ deutlich. Insbesondere sieht man einen deutlichen Einfluss der Virtualisierung auf den HPE Intel RHEL Systemen, die offensichtlich sehr viel kleine IOs nicht so effizient und parallel handhaben können. Dies insbesondere auf dem Setup der DL580 Systeme. Die Bare Metal Ergebnisse der x86 Server sind hingegen exzellent und zeigen den doppelten Durchsatz der IBM Plattform.

Die Auswertung der Ergebnisse zeigt, dass die CPU Belastung bei den HPE Systemen bis zu 80% beträgt (für Bare Metal) und für die virtualisierten Umgebungen mit geringerem Durchsatz logischerweise deutlich tiefer ist. Die Analyse der Servicetimes zeigt, dass diese direkt mit dem Durchsatz korrelieren. Bei den Bare Metal Systemen liegt die durchschnittliche Servicetime von User IOs bei unter 10ms, während diese bei sich auf dem DL380 mit RHV auf 20 ms verdoppelt und beim DL580 mit RHV sogar auf 46ms mehr als vervierfacht. Die IBM Power Plattform liegt mit einer Servicezeit von 17ms im Mittelfeld.

Die Verarbeitungsleistung auf der non-cached Tabelle liegt mit maximal 500'000 Rows/s um Faktor 10 niedriger als bei einer Tabelle im Cache mit 5 Mio. Rows/s. Hier erkennt man das Beschleunigungspotential von Datenbankzugriffen im Cache verglichen mit Datenbankzugriffen auf dem Storage (selbst wenn dieser als ultraschneller SSD Storage zur Verfügung steht).

Zusammenfassung T600 Data Select Tests

In der folgenden Tabelle wird die Single Thread Performance mit Parallelität 1 (=Speed) für die Plattformen dargestellt. Bei den meisten Tests gibt es keine signifikanten Unterschiede bei einem einzelnen Thread. Insbesondere auch die non-cached Tests 616 und 626 zeigen mit 1 Thread identische Ergebnisse.

Tests	Description	IBM P8	DL580 BM	DL580 VM	DL580 RV	DL380 BM	DL380 VM	DL380 RV
T611	Data select, random via primary key, heap non partitioning, cached	1.93	2.52	1.93	1.56	2.52	2.98	2.43
T612*	Data select, random via primary key, iot non partitioning, cached	3.28	3.28	3.28	3.28	3.28	3.12	3.28
T613*	Data select, random via primary key, hash cluster non partitioning, cached	2.05	2.11	2.85	1.93	2.62	2.98	2.11
T614*	Data select, random via primary key, heap list partitioning, cached	2.85	2.43	1.04	0.68	2.34	1.46	2.62
T615*	Data select, random via primary key, iot list partitioning, cached	3.28	3.28	3.28	3.28	3.28	3.28	3.28
T616	Data select, random via primary key, non cached	0.01	0.01	0.01	0.01	0.01	0.01	0.01
T621*	Data select, random via secondary key, non-partionied, cached	131	131	125	131	131	131	131
T622*	Data select, random via secondary key, iot non partitioning, cached	131	125	105	105	131	131	131
T624*	Data select, random via secondary key, heap list partitioning, cached	0.82	0.71	0.82	0.82	0.82	0.78	0.82
T625*	Data select, random via secondary key, iot list partitioning, cached	131	114	131	131	131	125	131
T626	Data select, random via secondary key, non cached	8.36	8.25	8.22	8.06	8.34	8.19	8.06

Tabelle 8 - T600 Data Load Select Heatmap

In der folgenden Tabelle wird maximale Durchsatz bei jeweils 8 physischen Cores (= Throughput) dargestellt. Bei den cached Tests 611-615 und 621-625 zeigen sich nur relativ geringe Unterschiede. Bei den non-cached Tests T616 und T626 zeigen sich hingegen ein spürbarer Einfluss der Virtualisierung. Die IBM Systeme liegen bei diesen Tests im Mittelfeld.

Tests	Description	IBM P8	DL580 BM	DL580 VM	DL580 RV	DL380 BM	DL380 VM	DL380 RV
T611	Data select, random via primary key, heap non partitioning, cached	210	200	210	175	200	210	210
T612*	Data select, random via primary key, iot non partitioning, cached	200	210	210	210	210	210	210
T613*	Data select, random via primary key, hash cluster non partitioning, cached	200	210	210	210	210	210	210
T614*	Data select, random via primary key, heap list partitioning, cached	210	210	210	210	210	210	210
T615*	Data select, random via primary key, iot list partitioning, cached	210	210	200	210	210	210	210
T616	Data select, random via primary key, non cached	11.92	11.92	11.92	9.36	12.48	11.92	10.49
T621*	Data select, random via secondary key, non-partionied, cached	4'092	5'243	4'195	4'935	5'243	5'592	5'592
T622*	Data select, random via secondary key, iot non partitioning, cached	3'495	3'355	2'943	3'165	2'996	3'570	3'647
T624*	Data select, random via secondary key, heap list partitioning, cached	4'794	5'084	4'661	4'935	5'992	5'785	5'786
T625*	Data select, random via secondary key, iot list partitioning, cached	4'194	4'302	3'813	3'995	4'415	4'660	3'814
T626	Data select, random via secondary key, non cached	255	456	174	92	477	318	219

Tabelle 9 - T600 Data Select Throughput Heatmap

Ergebnisse T700 Data Update

Die T700 Tests simulieren den Update einzelner Datenbank Elemente (Random Update). Diese Operationen sind sehr typisch für OLTP Anwendungen, kommen aber auch in Big Data oder DWH Applikationen vor.

Dabei werden folgende 6 Tests jeweils mit primary key Zugriff und einer geringen Selektivität (1 row) durchgeführt:

- T711: heap organized, non-partitioned, **cached**
- T712: index organized, non-partitioned, **cached**
- T713: hash cluster organized, non-partitioned, **cached**
- T714: heap organized, list partitioned, **cached**
- T715: index organized, list-partitioned, **cached**
- T716: heap organized, rangelist-partitioned, **non-cached**

Zu beachten ist, dass die ersten fünf Tests jeweils alle cached sind, also auf einer sehr kleinen Tabelle arbeiten, die relativ schnell in die SGA geladen wird. Hier spielt die Storageanbindung keine wesentliche Rolle, sondern vor allem die Verarbeitungsleistung des DB Servers. Der Test T716 ist ein non-cached Test auf einer Tabelle, die nicht in die SGA passt. Bei diesem Test spielt insbesondere die Storgelatency eine grosse Rolle für die Verarbeitung und es werden auch viele asynchrone Blockupdates und Redolog Einträge generiert.

T711 Data Update Primary Key Heap Organized, Non-Partitioned, Cached

Stellvertretend für die T71x Tests wird der Test 711 dargestellt. Die Tests T712-T715 zeigen vergleichbare Resultate. Die Tests führen Random Updates über den Primary Key durch, die eine sehr kleine Treffermenge(1 Row) liefern.

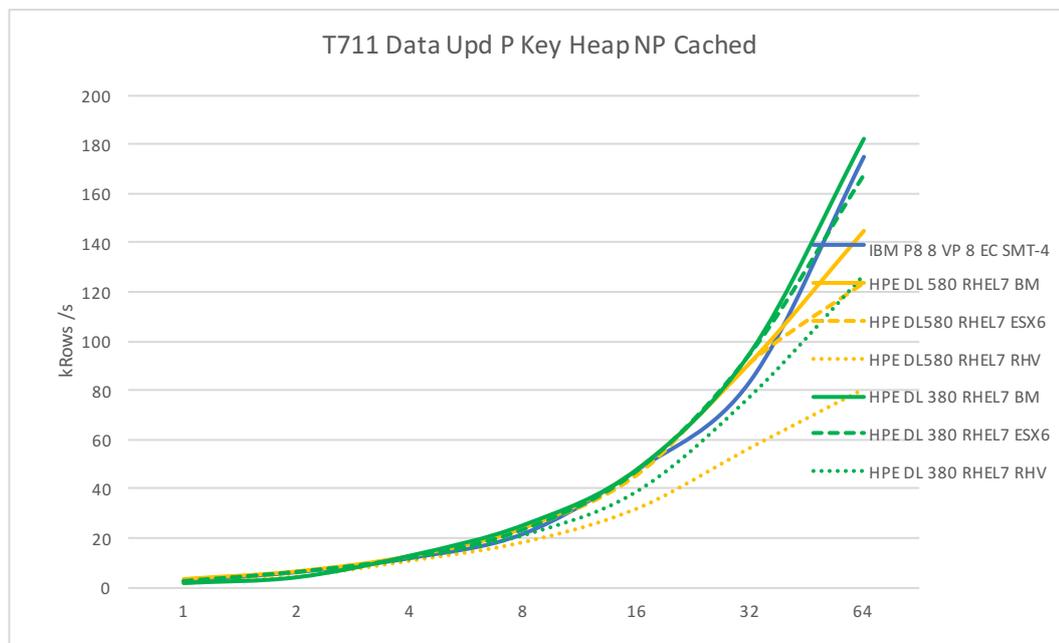


Abbildung 14 - T711 Data Update Primary Key Heap Organized, Non Partitioned, Cached

Die Ergebnisse unterscheiden sich bei diesen Tests relativ deutlich. Insbesondere sieht man einen deutlichen Einfluss der RedHat Virtualisierung RHV auf den HP Intel RHEL Systemen, die offensichtlich sehr viel kleine IOs nicht so effizient und parallel handhaben können. Dies insbesondere auf dem Setup der DL580 Systeme. Die Bare Metal Ergebnisse der x86 Server sind gleichwertig mit dem Durchsatz der IBM P8 Plattform.

Die Auswertung der Ergebnisse zeigt, dass die CPU Belastung im Maximum bei 80% liegt (für die Bare Metal) und für die virtualisierten Umgebungen mit geringerem Durchsatz entsprechend tiefer ist. Ansonsten liegt der wesentliche Unterschied bei den virtualisierten Umgebungen in höheren Wartezeiten für den Event "Log Buffer Space", d.h. die Redologs können bei diesen Systemen weniger schnell weggeschrieben werden, was den Durchsatz reduziert. Bei IBM taucht dieser Wait Event nicht auf, deshalb sind im Vergleich zu den sonstigen Resultaten die Ergebnisse bei diesem Test besonders gut.

Dieser Test ist durchaus storagelastig, da bis zu 20'000 IOPS und bis zu 600 MB/s geschrieben werden (Redologs). Die Anzahl der lesenden IOs ist wegen dem Caching entsprechend niedrig.

T716 Data Update Primary Key Heap Organized, Rangelist-Partitioned, Non-Cached

Der Test T716 führt Random Updates auf einer sehr grossen Tabelle durch, dadurch können die Resultate nicht im grösseren Stil gecached werden. Die Treffermenge der Selects ist jeweils nur eine Row.

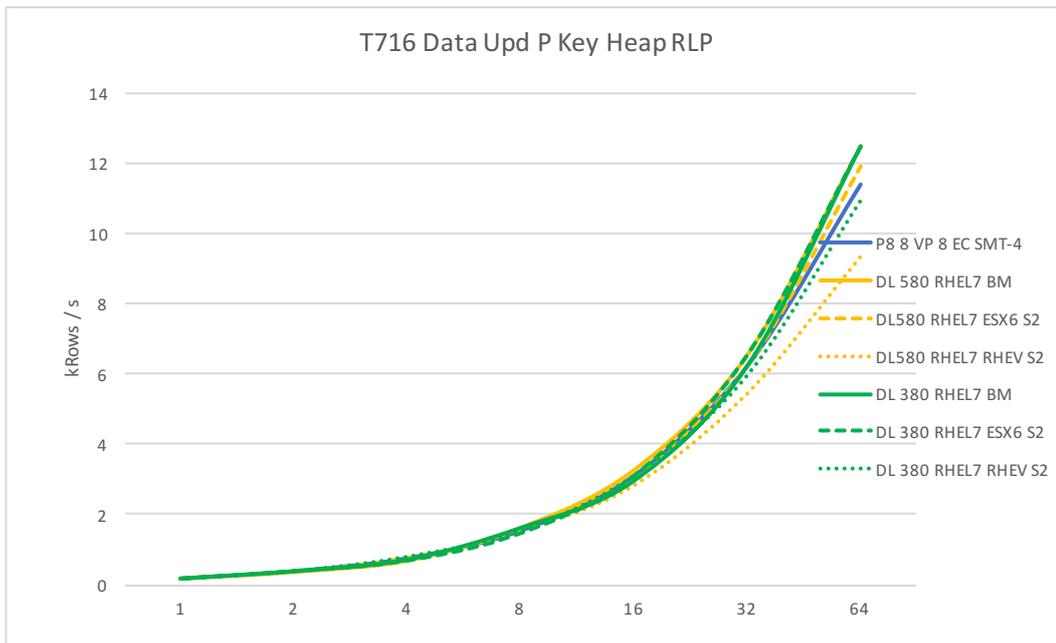


Abbildung 15 - T716 Data Update Primary Key Heap Organized, Rangelist-Partitioned, Non-Cached

Die Ergebnisse sind für alle Plattformen sehr ähnlich (die Kurven überlappen sich teilweise exakt), es wird jeweils eine Verarbeitungsleistung von ca. 9-12'000 Rows/s erreicht. Die beiden mit RHV virtualisierten Umgebungen haben einen minimal geringeren Durchsatz. Der Durchsatz steigt kontinuierlich bis zur maximal gewählten Parallelität von 64 Prozessen an.

Die Verarbeitungsleistung auf der non-cached Tabelle liegt mit ca. 10'000 Rows/s fast um Faktor 20 niedriger als bei der kleinen Tabelle im Cache mit 180'000 Rows/s. Hier erkennt man auch das Beschleunigungspotential von Datenbankzugriffen im Cache verglichen mit Datenbankzugriffen auf dem Storage (selbst wenn dieser als ultraschneller SSD Storage zur Verfügung steht).

Die CPU Utilization lag bei allen Systemen sehr niedrig (< 20%), in diesem Test spielt vor allem die Storage Latency eine wesentliche Rolle. Diese lag bei allen getesteten Konfiguration gemäss AWR Report bei ca. 0.5 ms.

Zusammenfassung T700 Data Update Tests

In der folgenden Tabelle wird die Single Thread Performance mit Parallelität 1 (=Speed) für die Plattformen dargestellt. Beim non-cached Test 716 gibt es keinerlei Unterschiede. Bei den anderen Tests sind in der Regel die virtualisierten Umgebungen etwas langsamer als die Bare Metal Umgebungen. Die IBM P8 Plattform ist bei diesen Tests auf Augenhöhe mit den HPE Intel RHEL Servern.

Tests	Description	IBM P8	DL580 BM	DL580 VM	DL580 RV	DL380 BM	DL380 VM	DL380 RV
T711	Data update, random via primary key, heap non partitioned, cached	2.62	3.12	1.72	1.37	1.29	2.85	2.26
T712*	Data update, random via primary key, iop non partitioned, cached	1.26	1.46	1.31	0.91	2.98	1.87	1.52
T713*	Data update, random via primary key, hash custer non partitioned, cached	2.62	2.26	2.98	2.11	1.99	2.18	1.93
T714*	Data update, random via primary key, heap list-partionined, cached	1.04	0.95	0.9	0.64	2.62	1.26	1.06
T715*	Data update, random via primary key, iot list-partionined, cached	1.31	1.42	1.34	1.04	2.11	1.99	1.68
T716	Data update, random via primary key, range-list partionined, non cached	0.2	0.2	0.2	0.2	0.2	0.2	0.2

Tabelle 10 - T700 Data Load Select Heatmap

In der folgenden Tabelle wird maximale Durchsatz bei jeweils 8 physischen Cores (= Througput) dargestellt. Beim non-cached Test 716 zeigen sich keine Unterschiede. Bei den anderen Tests sind in der Regel die virtualisierten Umgebungen etwas langsamer als die Bare Metal Umgebungen. Die IBM ist bei diesen Tests auf Augenhöhe mit den HPE Intel RHEL Servern.

Tests	Description	IBM P8	DL580 BM	DL580 VM	DL580 RV	DL380 BM	DL380 VM	DL380 RV
T711	Data update, random via primary key, heap non partitioned, cached	175	145	123	81	182	168	127
T712*	Data update, random via primary key, iop non partitioned, cached	168	140	123	82	161	145	108
T713*	Data update, random via primary key, hash custer non partitioned, cached	175	155	145	84	191	175	131
T714*	Data update, random via primary key, heap list-partionined, cached	168	155	131	84	182	145	127
T715*	Data update, random via primary key, iot list-partionined, cached	120	117	108	84	155	150	98
T716	Data update, random via primary key, range-list partionined, non cached	11.4	12.48	11.92	9.36	12.48	12.48	10.92

Tabelle 11 - T700 Data Select Throughput Heatmap

Ergebnisse T800

Die Tests im Bereich T800 führen Data Joins in einer Nested Loop aus.

T816 Data Join Nested Loop Heap Organized Rangelist Partitioned Non-Cached

Bei diesem Test wird ein Data Join mit einer steigenden Anzahl von Prozessen erzeugt.

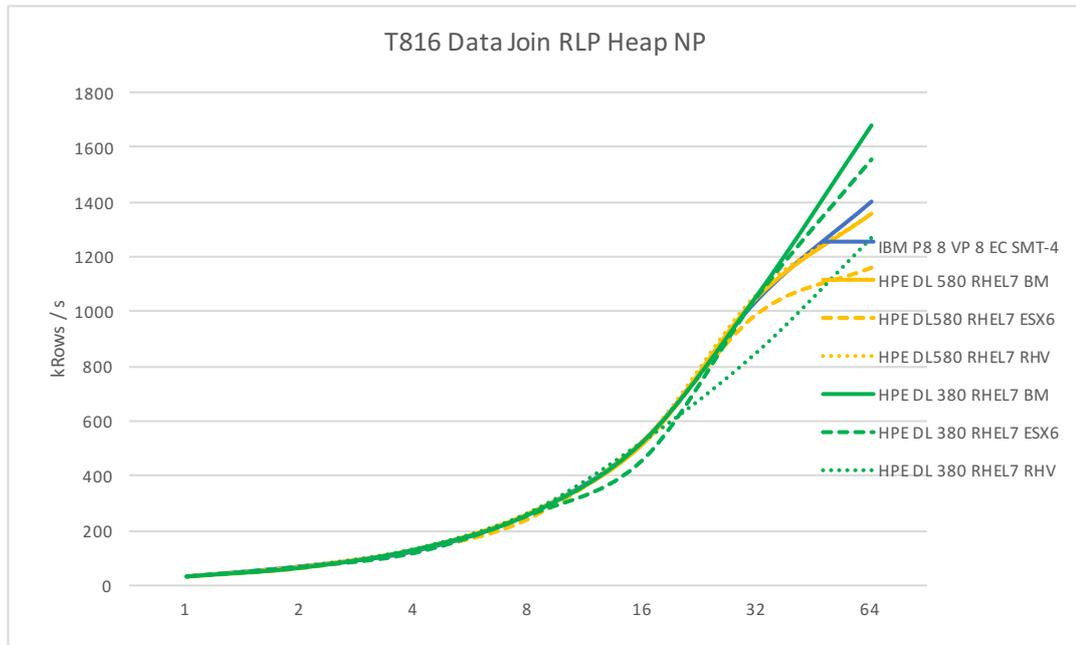


Abbildung 16 - T816 Data Join Nested Loop Heap Organized Rangelist Partitioned Non-Cached

Die Ergebnisse sind auf allen getesteten Plattformen relativ ähnlich, insbesondere bei einer geringen Parallelität bis 16 Prozessen. Bei hohen Parallelitäten werden die Systeme sehr stark ausgelastet (CPU Belastung > 90%) und auch die IO Leistung ist mit über 40'000 Random Reads pro Sekunde relativ hoch.

Zusammenfassung

Bei den CPU-lastigen T100 Tests zeigt sich ein deutlicher Vorteil für die HPE Intel RHEL Systeme. Dies gilt besonders für die Single Thread Performance (Speed) aber auch für den gesamten Throughput. Bei gleicher Anzahl von Threads sind die HPE Intel RHEL Systeme deutlich schneller, die IBM Plattform benötigt generell eine höhere Parallelität um ihre maximale Leistung zu erreichen. Der Einfluss der Virtualisierung bei den CPU-lastigen Tests ist nicht signifikant, auch virtualisiert wird die gleiche Leistung erreicht wie ohne Virtualisierung (Bare Metal). Der im DL380 verbaute E5-2637 v4 Prozessor mit 3.5 GHz Taktfrequenz ist erkennbar schneller als der im DL580 benutzte E7-8893 v4 Prozessor mit 3.2 GHz Takt. Bei diesem Test mit nur 8 physischen Cores weist demzufolge das DL380 System den höheren Durchsatz aus, beim DL580 wurde aufgrund der Vergleichbarkeit nur die Hälfte der verfügbaren 16 Cores benutzt.

Die storagelastigen T200 und T400 Tests sind aufgrund der unterschiedlichen Storageanbindung von IBM und HPE nicht wirklich zu vergleichen. Es zeigt sich jedoch sowohl bei sequentiellen Schreiben wie auch beim sequentiellen Lesen auf allen HPE Plattformen eine gleichbleibend exzellente IO Rate von bis zu 5 GB/s. Diese Werte werden auch in den virtualisierten Umgebungen erreicht.

Bei Load Szenarien (T300) und der Datenaggregation (T500) sind die HPE Intel RHEL Server durchweg besser als die Power Plattform, allerdings ist ein teilweise starker Performance Impact der Virtualisierung messbar.

Bei den Data Selects (T600) sind die Ergebnisse in der Regel vergleichbar, werden sie bei den Cached Tests von der Verarbeitungsleistung der Oracle Datenbank limitiert und bei den Non-Cached Tests von der Servicezeit der Storage-systeme. Ausnahme stellt der Test T626 dar, in dem sich ein sehr starker Impact der Virtualisierung zeigt, während die Bare Metal Resultate der HPE Intel RHEL Systeme deutlich besser als die IBM P8 Ergebnisse liegen.

Data Update (T700) und Data Join (T800) zeigen eine vergleichbare Performance auf allen Systemen, jedoch ebenfalls mit einem spürbaren Einfluss der Virtualisierung auf den HPE Intel RHEL Servern.

Fazit

Die CPU Leistung der heutigen HPE Server mit aktuellen Intel Prozessoren unter RHEL ist den heutigen Power 8 Systemen unter AIX inzwischen überlegen, dies zeigt sich insbesondere auch in den Oracle PL/SQL Benchmarks.

Die Tests mit sehr hohem IO Durchsatz sind nicht wirklich zwischen IBM und HPE Intel RHEL vergleichbar, da die IO Anbindung dieser Server im Test erheblich höher als die der IBM Systeme war. Erkennbar können aber auf allen HPE Intel RHEL Systemen sehr hohe Data Writes und Data Scan Raten erreicht werden, die mit 5 GB/s selbst für sehr hohe Anforderungen ausreichend sein dürften.

Auch in den weiteren Tests zeigt der HPE Intel RHEL Stack eine sehr gute Leistung, die in nahezu allen Tests über der Leistung der IBM Plattform lag. Bei einigen Tests, besonders stark bei relativ vielen parallelen kleinen IOs zeigt sich allerdings ein signifikanter Performance Impact durch die Virtualisierung. Generell zeigt hier VMware ESX einen geringeren Impact wie die Red Hat Virtualisierung RHV.

Abschliessend kann der Einsatz von Oracle Datenbanken auf x86 Systemen mit Red Hat Linux aus Performancegesichtspunkten empfohlen werden. Dies speziell auch vor dem Hintergrund, dass die Leistung pro Core höher liegt, aber nur die Hälfte der Lizenzkosten pro Intel Core (Lizenzfaktor 0.5) im Vergleich zu Power Cores (Lizenzfaktor 1.0) anfällt.

Bedacht werden muss, dass IBM Power Plattformen über eine exzellente Virtualisierung verfügen und sehr gut selbst mit sehr hohen CPU Auslastungen bis zu 80% und mit einer sehr grossen Zahl von virtuellen Systemen zurechtkommen. Auf x86 Systemen mit VMware oder RHV wird hingegen üblicherweise mit Auslastungen bis ca. 50% gearbeitet. Allerdings liegen die Hardwarekosten der Intel Systeme um Faktoren niedriger als die der Power Systeme.

Das Ziel dieses Benchmarks ist ein Vergleich auf der Basis von 8 physischen Cores. Dabei wurden 2 verschiedene CPU's von Intel vermessen: Die E5 Familie (maximal 2 Sockets) und die E7 Familie (2 - 16 Sockets).