

Kap. 7: DB/TP Benchmarks

- Einleitung
 - Anforderungen an Benchmarks
- OLTP-Benchmarks
 - TPC-C und TPC-E
- OLAP-Benchmarks
 - TPC-H, TPC-DS, TPC-DI
- Big Data Benchmarks
 - TPCx-HS, TPCx-BB, TPCx-IOT



Benchmarks

- Leistungsbewertung eines DBMS / Transaktionssystems / Big Data Systems für standardisierte Last und Ausführungsbedingungen
- unterschiedliche Einsatzbereiche / Lasten führen zu unterschiedlichen Benchmarks
 - Online Transaction Processing (OLTP)
 - Online Analytical Processing (OLAP, decision support)
 - Mischlasten
 - spezielle Anforderungen: Datenintegration, Graphanalysen ...
- unterschiedliche Benchmarks nach Datenmodell
 - relationale DBS (Wisconsin-Benchmark, TPC-Benchmarks)
 - OODBS (OO7)
 - XML (Xmark, Xmach-1)
 - NoSQL- / Graphdaten



Anforderungen an geeignete Benchmarks*

■ Domain-spezifische Benchmarks

- kein geeignetes Leistungsmaß für alle Anwendungsklassen möglich
- spezielle Benchmarks für techn./wissenschaftliche Anwendungen, DB-Anwendungen, etc.

■ Relevanz

- Berücksichtigung “typischer” Operationen des jeweiligen Anwendungsbereichs
- Messen der maximalen Leistung
- Berücksichtigung der Systemkosten (Kosteneffektivität)

■ Portierbarkeit

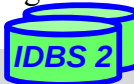
- Übertragbarkeit auf verschiedene Systemplattformen
- Vergleichbarkeit

■ Skalierbarkeit

- Anwendbarkeit auf kleine und große Computersysteme
- Übertragbarkeit auf verteilte/parallele Systeme

■ Einfachheit / Verständlichkeit

* J. Gray (ed.): The Benchmark Handbook for Database and Transaction Processing Systems. 2nd ed., Morgan Kaufmann, 1993. <http://jimgray.azurewebsites.net/benchmarkhandbook/toc.htm>



TPC-Benchmarks

■ Herstellergremium zur Standardisierung von DB-Benchmarks (www.tpc.org)

TPC Transaction Processing Performance Council

- Gründung 1988
- erste Benchmarks für Kontenbuchung (“Debit-Credit”): TPC-A, TPC-B (1989/90)

■ besondere Merkmale

- Leistung eines Gesamt-Systems wird bewertet
- Bewertung der Kosteneffektivität (Kosten / Leistung)
- skalierbare Konfigurationen
- verbindliche Richtlinien zur Durchführung und Dokumentation (Auditing; Full Disclosure Reports)
- Ausschluss von “Benchmark Specials” innerhalb von DBMS etc.

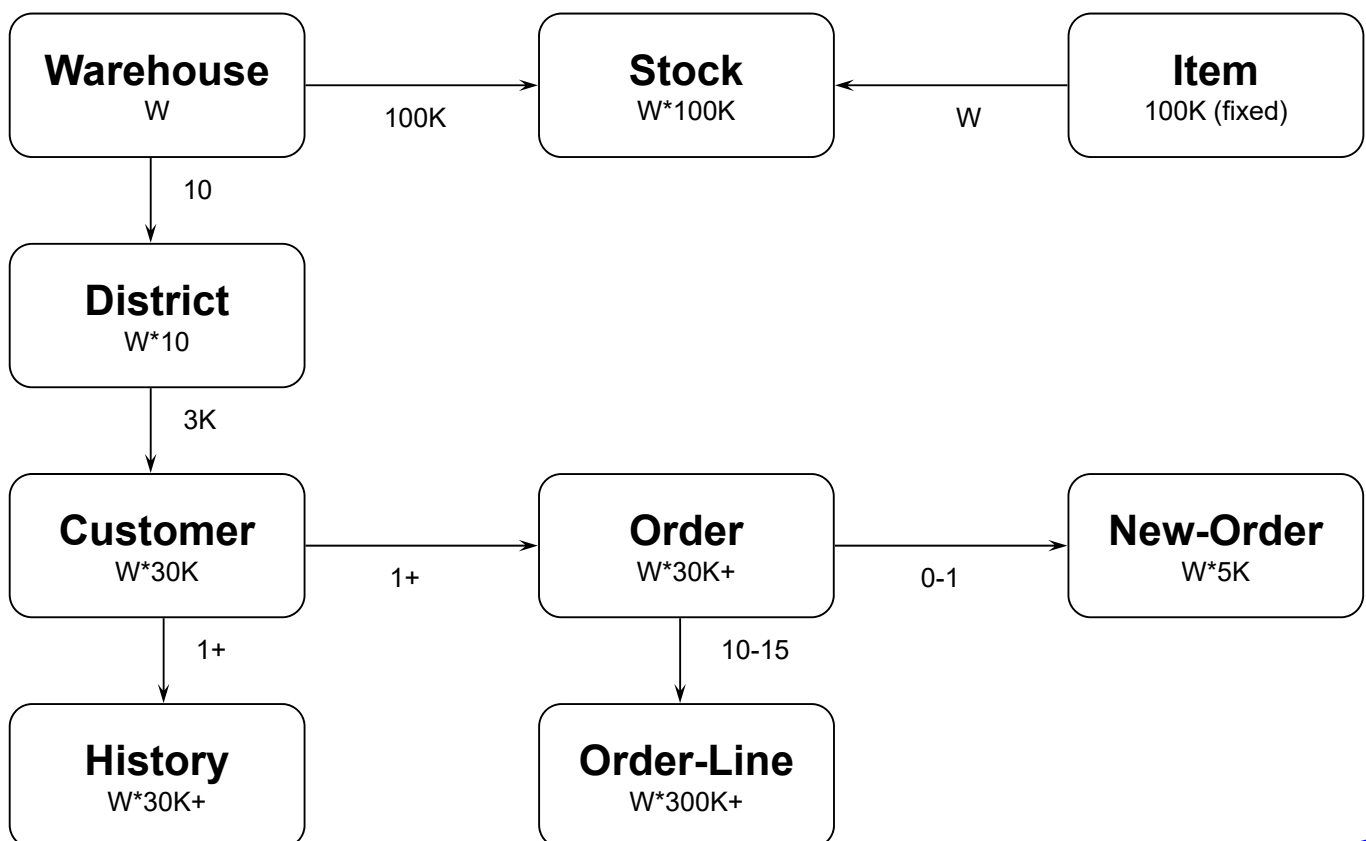
■ umfangreiche *Enterprise-Benchmarks* und schneller realisierbare *Express-Benchmarks* (TPCx-...)



TPC-C

- 1992 eingeführter OLTP-Benchmark
- 9 Tabellen (Bestellverwaltung von Warenhäusern)
- 5 Transaktionstypen
 - New-order (45%)
 - Payment (43%), Delivery (4%) , Order-status (4%), Stock-level(4%)
- DB skaliert proportional zum Durchsatz
- hohe Lokalität
- Metriken
 - Durchsatz von NewOrder-Transaktionen pro Minute (**tpmC**)
 - Price/performance (**\$/tpmC**)

TPC-C DB-Schema



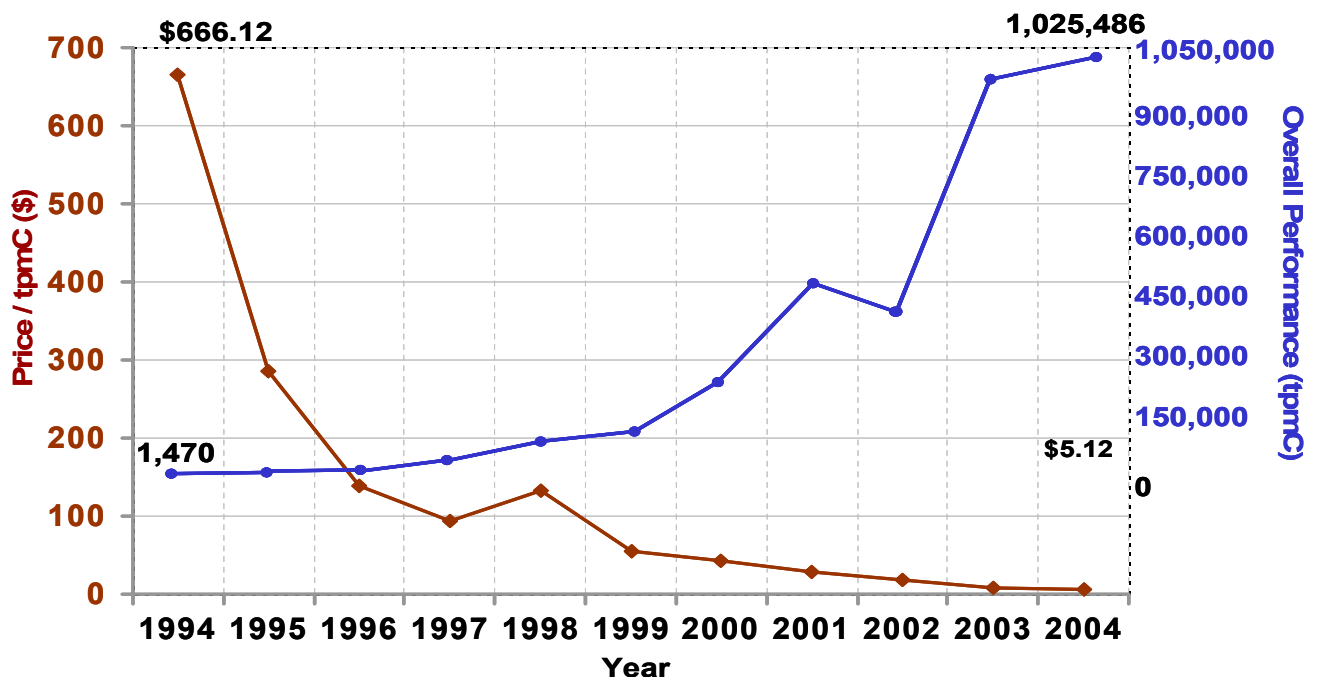
Transaktionstyp New Order (TPC-C)

```
BEGIN WORK { Beginn der Transaktion }  
SELECT ... FROM CUSTOMER  
    WHERE c_w_id = :w_no AND c_d_id = :d_no AND c_id = :cust_no  
SELECT ... FROM WAREHOUSE WHERE w_id = :w_no  
SELECT ... FROM DISTRICT (* -> next_o_id *)  
    WHERE d_w_id = :w_no AND d_id = :d_no  
UPDATE DISTRICT SET d_next_o_id := :next_o_id + 1  
    WHERE d_w_id = :w_no AND d_id = :d_no  
INSERT INTO NEW_ORDER ...  
INSERT INTO ORDERS ...  
pro Artikel (im Mittel 10) werden folgende Anweisungen ausgeführt:  
SELECT ... FROM ITEM WHERE ...  
SELECT ... FROM STOCK WHERE ...  
UPDATE STOCK ...  
INSERT INTO ORDER-LINE ...  
COMMIT WORK { Ende der Transaktion }
```

- im Mittel 48 SQL-Anweisungen (BOT, 23 SELECT, 11 UPDATE, 12 INSERT, EOT)
- 1% der Transaktionen sollen zurückgesetzt werden

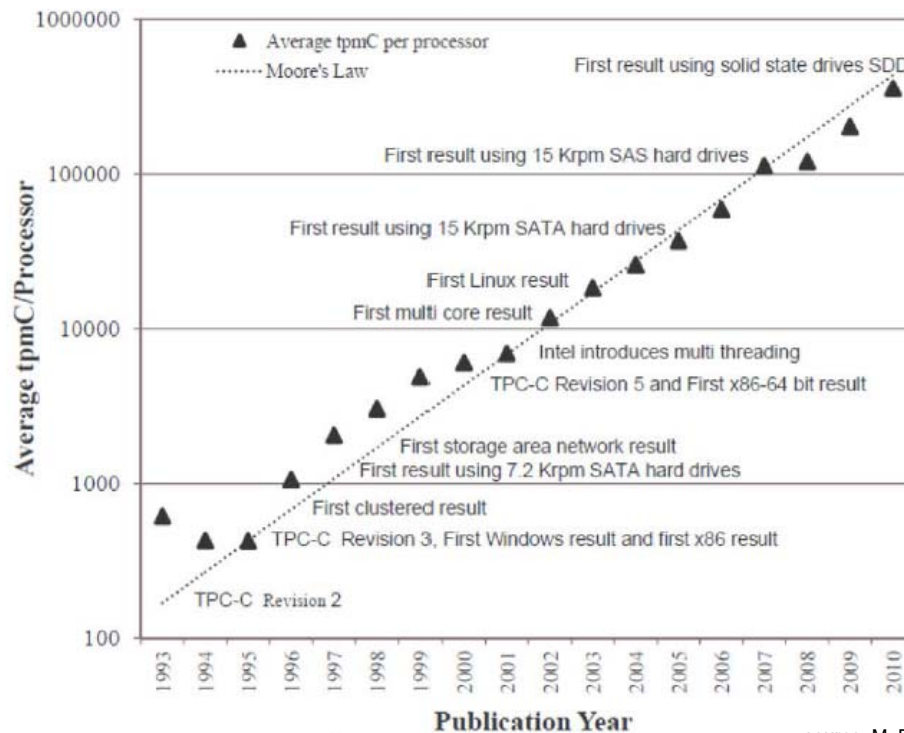


TPC-C: Entwicklung 1994-2004



Entwicklung relativer TPC-C Performance

Transaction performance vs. Moore's Law, Milestones, 1993 to 2010



source: M. Poess, WDBD 2012



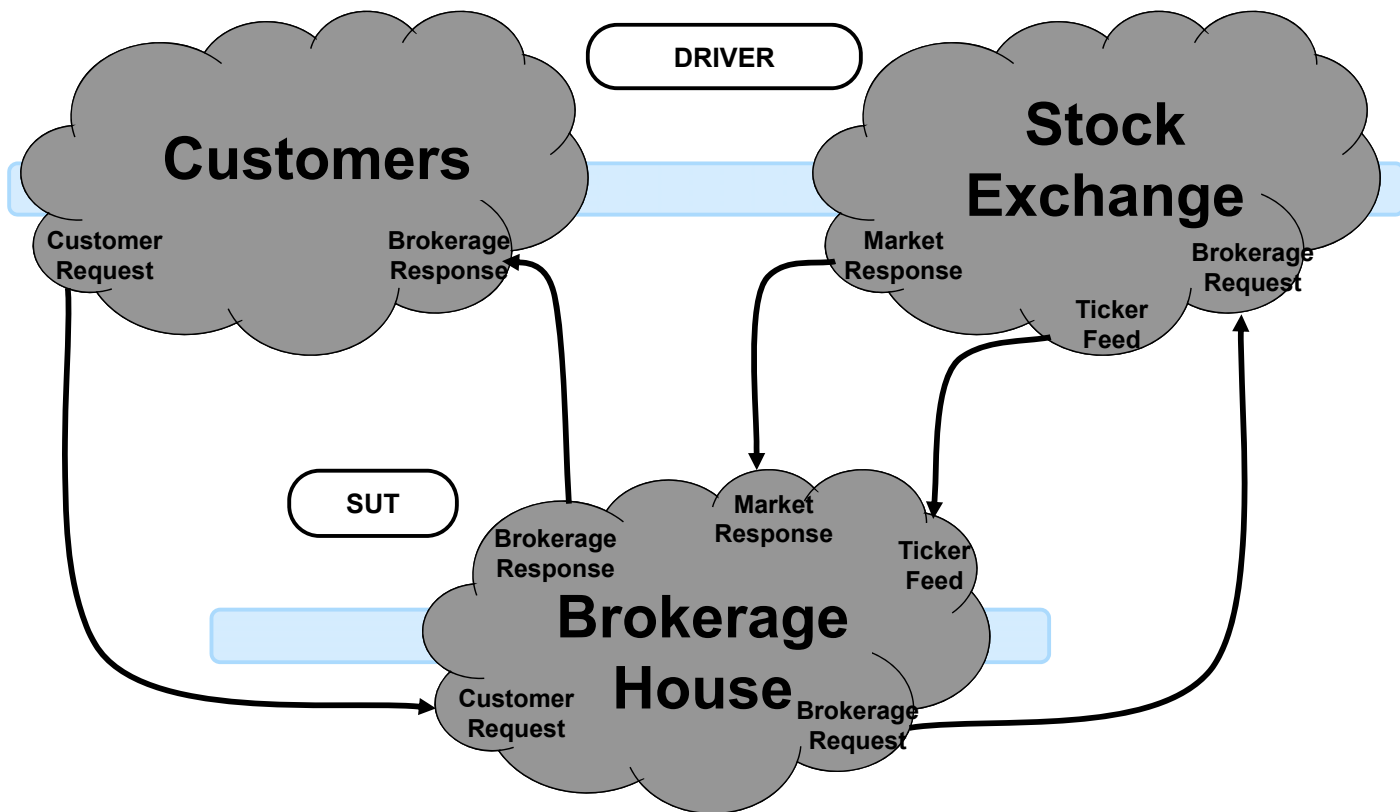
TPC-C Top-Durchsatz (Juni 2013/2018)

Rank	Company	System	Performance (tpmC)	Price/tpmC	Watts/KtpmC	System Availability	Database	Operating System	TP Monitor	Date Submitted	Cluster
1	ORACLE	SPARC SuperCluster with T3-4 Servers	30,249,688	1.01 USD	NR	06/01/11	Oracle Database 11g R2 Enterprise Edition w/RAC w/Partitioning	Oracle Solaris 10 09/10	Tuxedo CFS-R	12/02/10	Y
2	IBM	IBM Power 780 Server Model 9179-MHB	10,366,254	1.38 USD	NR	10/13/10	IBM DB2 9.7	AIX Version 6.1	Microsoft COM+	08/17/10	Y
3	ORACLE	SPARC T5-8 Server	8,552,523	.55 USD	NR	09/25/13	Oracle 11g Release 2 Enterprise Edition with Oracle Partitioning	Oracle Solaris 11.1	Oracle Tuxedo CFSR	03/26/13	N

Sponsor	System	Performance (TpmC)	Price/TpmC	System Availability	Date Submitted	DB Software Name
	KTNF-KR580S1	152,330	1680.00 KRW	11/20/2018	11/19/2018	Goldilocks v3.1 Standard Edition
	Jet-speed HHA2212	139,909	1730.00 KRW	5/9/2017	5/8/2017	Goldilocks v3.1 Standard Edition



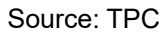
TPC-E: Business Model – Financial Market



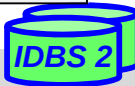
TPC-E

- 2007 verabschiedet
- wesentlich komplexer als TPC-C
 - 33 statt 9 Tabellen
 - 188 statt 92 Spalten
 - referentielle Integrität einzuhalten
 - obligatorische RAID-Datenspeicherung
 - pseudo-reale Daten (z.B. für Kundennamen)
- 10 Transaktionstypen
 - Consumer-to-Business: Trade-Order, Trade-Result, Trade-Update, ...
 - Business-to-Business: Broker Volume, Market-Watch, ..
- Metriken
 - Durchsatz von TradeResult-Transaktionen pro Sekunde (**tpsE**)
 - Price/performance (**\$/tpsE**)

TPC-E Database – Mile High View



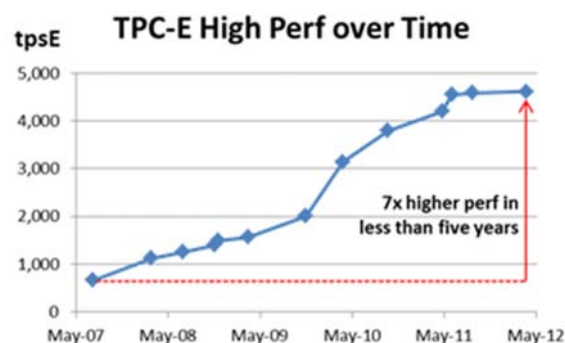
7 - 15



TPC-E Top-Durchsatz (2018)

Sponsor	System	Performance (TpsE)	Price/TpsE	System Availability	Date Submitted	DB Software Name
	Lenovo ThinkSystem SR950	11,357	98.83 USD	11/6/2017	11/1/2017	Microsoft SQL Server 2017 Enterprise Edition
	Lenovo System x3850 X6	9,068	139.85 USD	7/31/2016	5/31/2016	Microsoft SQL Server 2016 Enterprise Edition
	FUJITSU Server PRIMERGY RX4770 M3	8,796	116.62 USD	7/31/2016	7/12/2016	Microsoft SQL Server 2016 Enterprise Edition
	Lenovo ThinkSystem SR650	6,779	92.49 USD	9/10/2018	9/5/2018	Microsoft SQL Server 2017 Enterprise Edition
	FUJITSU Server PRIMERGY RX2540 M4	6,606	92.85 USD	3/31/2018	3/31/2018	Microsoft SQL Server 2017 Enterprise Edition

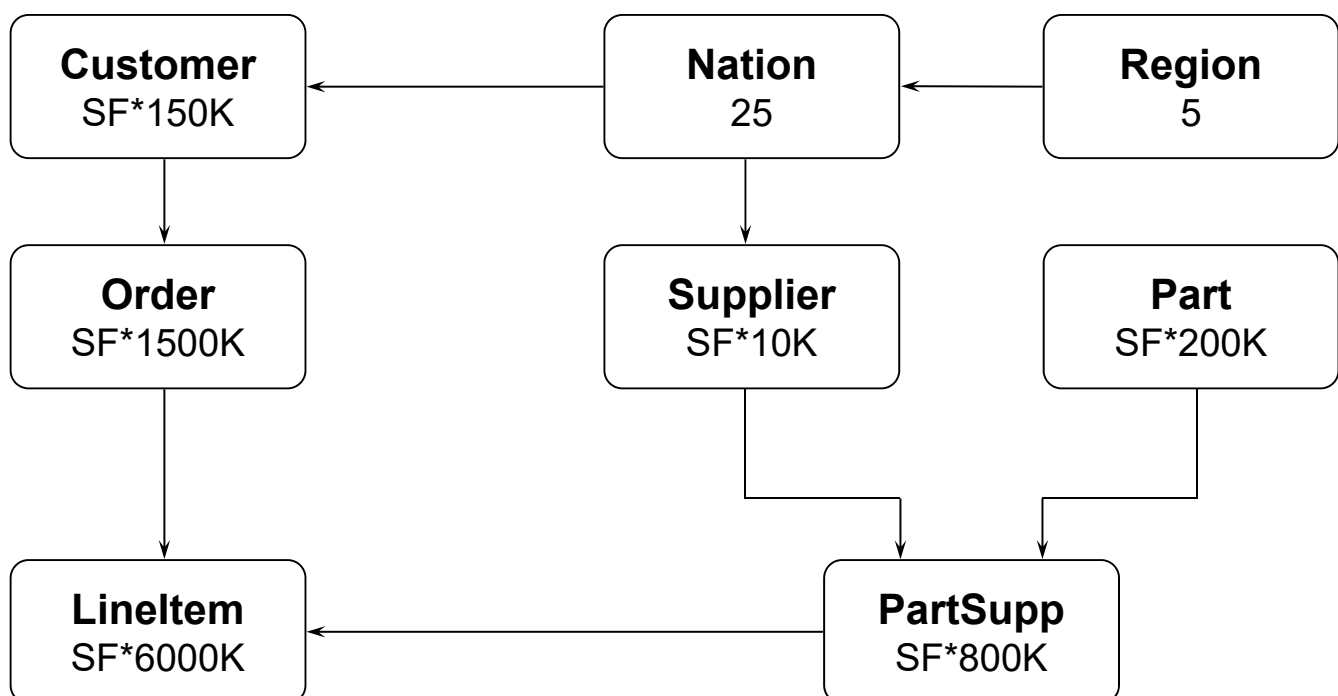
Performance-Entwicklung für SQL-Server:



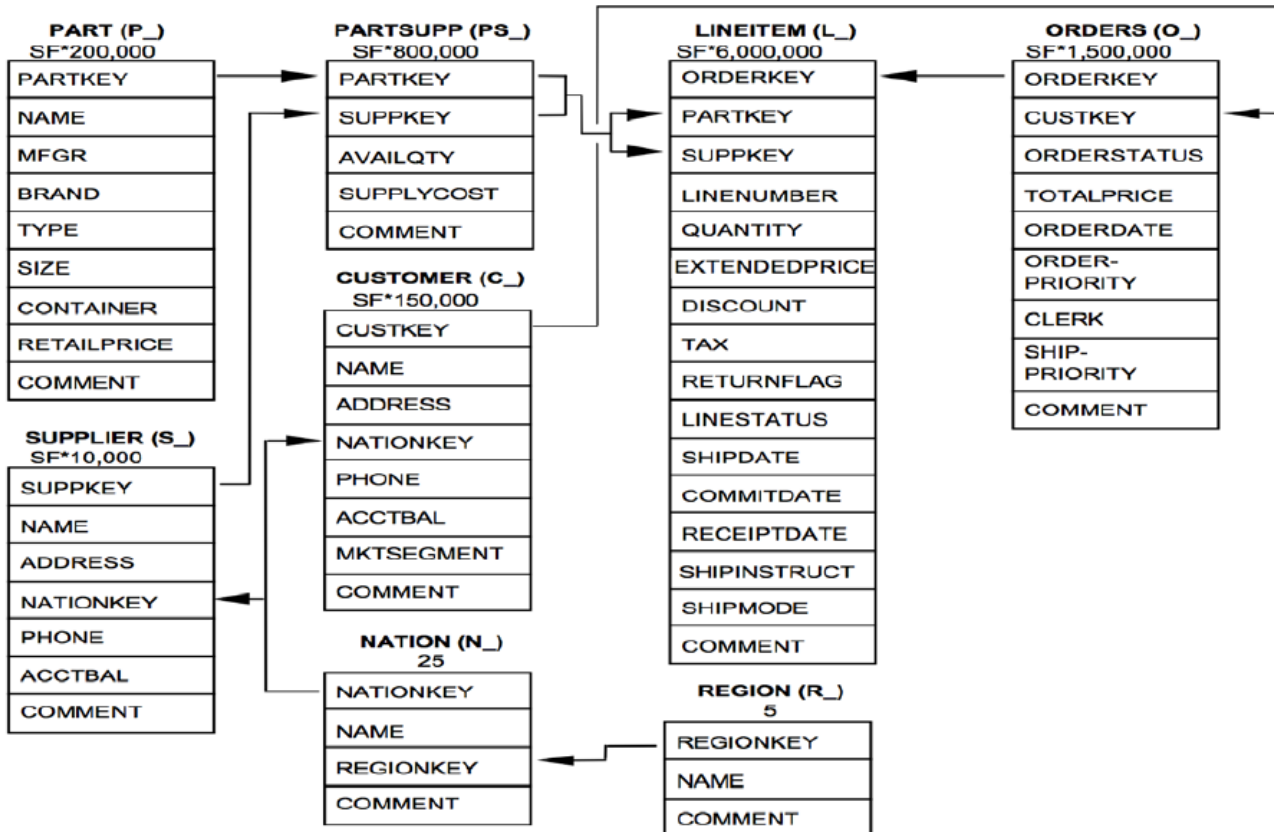
TPC-H

- Benchmark für Decision Support (Ad-Hoc-Queries)
 - 1999 eingeführt
- unterschiedliche DB-Größen (100 GB – 100 TB) über Skalierungsfaktor
- 22 Query-Typen, 2 Update-Operationen
- Zeitbedarf für Einbenutzer- und Mehrbenutzerausführung (single/multiple streams)
- Metriken
 - Query-per-Hour (QphH@Size)
 - \$/QphH@Size

TPC-H Schema



TPC-H Schema (detailliert)



Beispiel-Query Q7 (Volume Shipping)

```

/* TPC_H Query 7 - Volume Shipping */
SELECT SUPP_NATION, CUST_NATION, L_YEAR, SUM(VOLUME) AS REVENUE
FROM ( SELECT N1.N_NAME AS SUPP_NATION
      , N2.N_NAME AS CUST_NATION, datepart(yy, L_SHIPDATE) AS L_YEAR
      , L_EXTENDEDPRICE*(1-L_DISCOUNT) AS VOLUME
FROM SUPPLIER, LINEITEM, ORDERS, CUSTOMER, NATION N1, NATION N2
WHERE S_SUPPKEY = L_SUPPKEY AND O_ORDERKEY = L_ORDERKEY
      AND C_CUSTKEY = O_CUSTKEY AND S_NATIONKEY = N1.N_NATIONKEY
      AND C_NATIONKEY = N2.N_NATIONKEY
      AND ((N1.N_NAME = 'FRANCE' AND N2.N_NAME = 'GERMANY')
          OR (N1.N_NAME = 'GERMANY' AND N2.N_NAME = 'FRANCE'))
      AND L_SHIPDATE BETWEEN '1995-01-01' AND '1996-12-31' ) AS SHIPPING
GROUP BY SUPP_NATION, CUST_NATION, L_YEAR
ORDER BY SUPP_NATION, CUST_NATION, L_YEAR
    
```

Zur Unterstützung bei der Verhandlung über neue Lieferverträge soll der Wert der zwischen Frankreich und Deutschland transportierten Güter festgestellt werden. Dazu werden jeweils die rabattierten Einnahmen in den Jahren 1995 und 1996 berechnet, die aus Auftragspositionen resultieren, bei denen der Lieferant aus dem einen, und der Kunde aus dem anderen Land stammt (also vier Ergebnistupel).

TPC-H: Ergebnisse 2017

1,000 GB Results

Rank	Company	System	QphH	Price/QphH	Watts/KQphH	System Availability	Database	Operating System	Date Submitted	Cluster
1		Dell PowerEdge R720xd using EXASolution 5.0	5,246,338	.14 USD	NR	09/24/14	EXASOL EXASolution 5.0	EXASOL EXACluster OS 5.0	09/23/14	Y
2		HPE ProLiant DL380 Gen9	717,101	.61 USD	NR	10/19/17	Microsoft SQL Server 2017 Enterprise Edition	Red Hat Enterprise Linux Server 7.3	04/17/17	N
3		HPE ProLiant DL380 Gen9	678,492	.64 USD	NR	07/31/16	Microsoft SQL Server 2016 Enterprise Edition	Microsoft Windows Server 2012 R2 Standard Edition	03/24/16	N
4		Cisco UCS C460 M4 Server	588,831	.97 USD	NR	12/16/14	Microsoft SQL Server 2014 Enterprise Edition	Microsoft Windows Server 2012 R2 Standard	12/15/14	N

30,000 GB Results

Rank	Company	System	QphH	Price/QphH	Watts/KQphH	System Availability	Database	Operating System	Date Submitted	Cluster
1		Dell PowerEdge R720xd using EXASolution 5.0	11,223,614	.23 USD	NR	09/24/14	EXASOL EXASolution 5.0	EXASOL EXACluster OS 5.0	09/23/14	Y
2		Lenovo System x3950 X6	1,056,164	2.04 USD	NR	09/30/16	Microsoft SQL Server 2016 Enterprise Edition	Microsoft Windows Server 2016 Standard Edition	04/28/16	N

100,000 GB Results

Rank	Company	System	QphH	Price/QphH	Watts/KQphH	System Availability	Database	Operating System	Date Submitted	Cluster
1		Dell PowerEdge R720xd using EXASolution 5.0	11,612,395	.37 USD	NR	09/24/14	EXASOL EXASolution 5.0	EXASOL EXACluster OS 5.0	09/23/14	Y



TPC-H: Ergebnisse Jan. 2019

Sponsor	System	Scale Factor	Performance (QphH)	Price/QphH	System Availability	Date Submitted	DB Software Name
	KTNF KR580S1	100	24,037	18739.00 KRW	12/20/2018	12/19/2018	Goldilocks v3.1 Standard Edition
	HPE ProLiant DL380 Gen10	1,000	1,009,065	0.47 USD	5/21/2018	11/17/2017	Microsoft SQL Server 2017 Enterprise Edition
	HPE ProLiant DL580 Gen9 Action Vector 5.0	3,000	2,140,307	0.38 USD	7/31/2016	6/2/2016	Action Vector 5.0
	Cisco UCS C240 M5	3,000	1,029,593	0.47 USD	6/12/2018	6/12/2018	Microsoft SQL Server 2017 Enterprise Edition
	HP ProLiant DL580 Gen10	10,000	1,479,748	0.95 USD	10/31/2017	10/31/2017	Microsoft SQL Server 2017 Enterprise Edition
	Lenovo System x3950 X6	30,000	1,056,164	2.04 USD	9/30/2016	4/28/2016	Microsoft SQL Server 2016 Enterprise Edition

Ergebnisse mit unterschiedlichen Skalierungsfaktoren sind nicht direkt vergleichbar



A New Benchmark With Real Business Impact Using SQL Server 2017 on Linux and the Power of NVMe™.

Another TPC-H Benchmark™ with SQL Server 2017 on Linux by Western Digital? Yes! Today we are pleased to announce that Western Digital and Cisco set the first 3,000GB world record using a two-socket server and the first 3,000GB world record using SQL Server 2017 on Linux. This TPC-H benchmark is our first official TPC benchmark using the HGST Ultrastar® SN200 Series PCIe NVMe SSD.

SQL Server 2017 on Linux – Showing Real World Benefits

For many it was a big surprise when Microsoft announced that SQL Server would run on Linux. Yet with the multitude of changes the IT world is undergoing, choice and flexibility have become key goals and enablers. SQL Server has undergone many transformations, and SQL Server 2017 represents a major step in bringing powerful tools to wherever its users are and the prevalence of open source tools.

Our new benchmark shows the real-world benefits of this shift, and is enabled by the latest demonstration of hardware innovations from Western Digital and Cisco.

Breaking Down the Metrics

While delivering less than 1% difference in TPC-H throughput and 4% difference in the TPC-H Composite (QphH) metrics, Cisco's C240 M5 Rack Server with HGST Ultrastar SN200 Series PCIe SSD delivers on par performance when compared to the previous Cisco 3,000GB SQL Server TPC-H world record while:

- Reducing CPU power consumption by 38%
- Reducing physical core count by 23%
- Reducing total system DRAM by 50%
- Reducing rack density by 50%

The table provides a side-by-side comparison of the previous 3,000GB TPC-H benchmark using SQL Server 2016 and Cisco's C460 M4 Rack Server to the new 3,000GB TPC-H world record on SQL Server 2017 and Cisco's C240 M5 Rack Server.

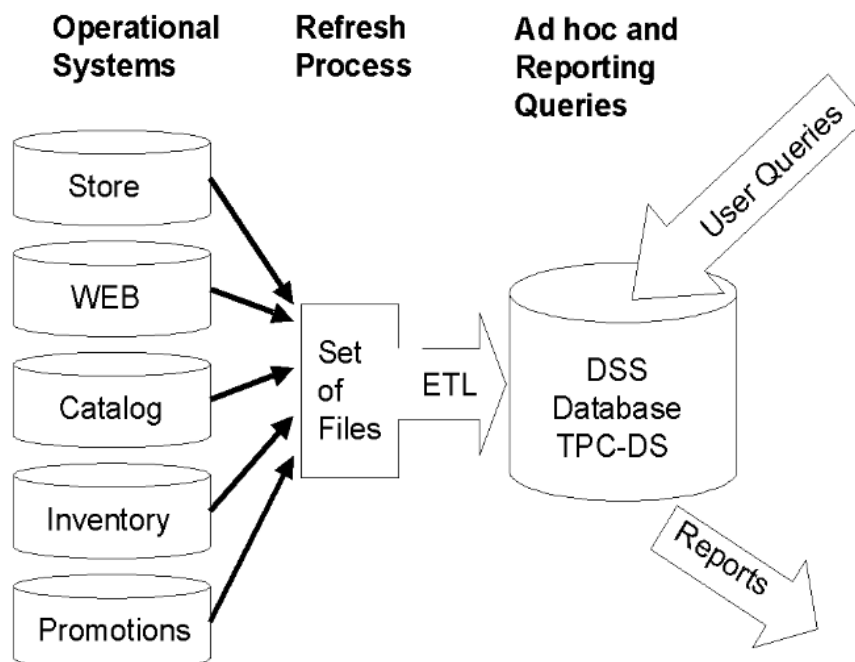
	Previous 3TB C460	Current 3TB C240	Difference
Sockets	4	2	50%
Total CPU Power Consumption	660W	410W	38%
Rack Units	4	2	50%
Cores	72	56	23%
Frequency	2.5	2.5	0%
Total GHz	180	140	23%
RAM	3000	1500	50%
Power (QppH)	1,488,163	1,371,925	8%
TPC-H Throughput (QthH)	770,803	772,683	<1%
TPC-H Composite (QphH)	1,071,018	1,029,593	4%
System Hardware Cost	\$145,315.00	\$98,404.93	32%
SQL Server Licensing	\$485,010.00	\$377,230.00	23%
Total Cost	\$630,325.00	\$481,900.98	24%
Price/Performance	\$0.60	\$0.47	22%

<https://blog.westerndigital.com/sql-server-on-linux-3tb-tpch/>



TPC-DS

- seit 2012
- komplexere Verhältnisse als mit TPC-H



TPC-DS (2)

■ Merkmale des Data-Warehouses

- mehrere (7) Faktentabellen mit gemeinsamen Dimensionen (Snowflake Schemas)
- 24 Tabellen mit durchschnittlich 18 Attributen
- 99 Queries + 12 Änderungsoperationen: Ad-hoc-Queries, Reporting, OLAP (Drill-Down/Up), Data Mining
- ungleichmäßige Datenverteilung: lineare Skalierung für Faktentabellen, nicht-lineare Skalierung für Dimensionstabellen
- ETL-artige Änderungen

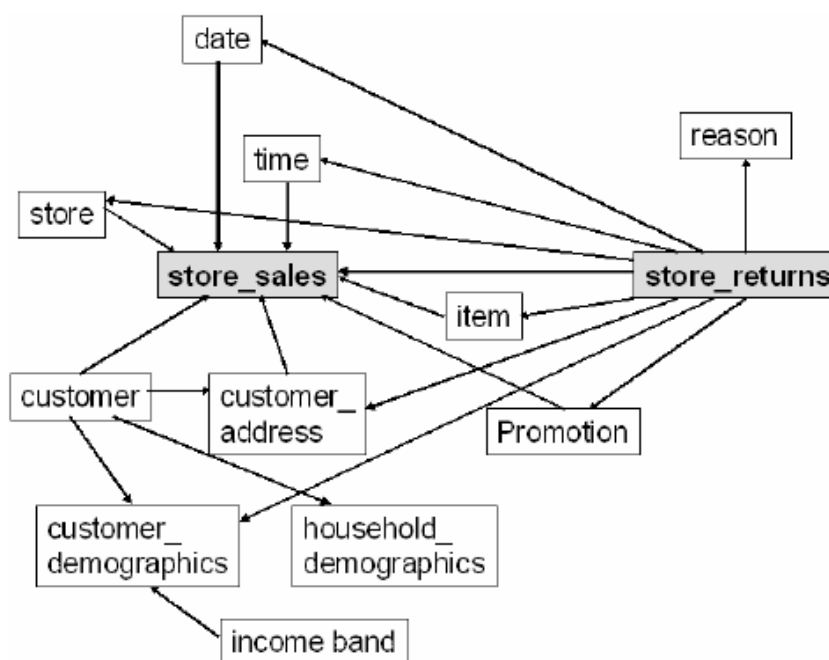
■ Skalierung von 100 GB bis 100 TB

■ Sequenz von 4 Tests: Database Loading, Power test, Query throughput tests 1 und 2

■ Metriken:

- Query-Durchsatz QphDS@SF und Kosten $\$ / \text{QphDS@SF}$
- sekundäre Angaben: Ladezeit, Dauer (elapsed time) der anderen Phasen
- ...

TPC-DS Teilschema (Store Sales)



■ daneben Tabellen für Produktkatalog und Web Sales

TPC-DS Ergebnisse

TPC-DS Advanced Sorting Results

Sponsor	System	Scale Factor	Performance (QphDS)	Price/QphDS	System Availability	Date Submitted	Withdrawn	DB Software Name
	Cisco UCS Integrated Infrastructure for Big Data	10,000	1,580,649	0.64 USD	3/5/2018	3/5/2018		Transwarp Data Hub v5.1

TPC-DI

■ Datenintegrations/ETL-Benchmark

- seit 2014

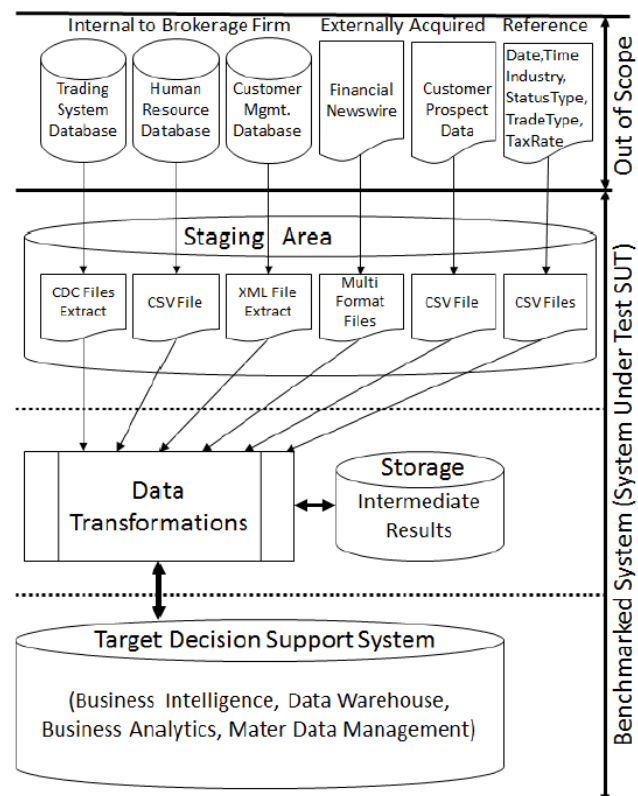
■ Merkmale

- mehrere heterogene Datenquellen
- Laden und inkrementelle Anpassung von Fakten- und Dimensionstabellen
- Mix von Datentransformationen zur Datenvalidierung, Konvertierung, komplexen Aggregationen etc.

■ Durchsatzmessungen

- für initiales Laden des Warehouses und
- zwei inkrementelle Update-Phasen

■ noch keine veröffentlichten Ergebnisse (Jan 2019)



Big Data Benchmarks

- komplexe Analyseprozesse und Systemumgebungen (Data Warehouse, Hadoop, Streaming-System, Kombinationen)
 - Anforderungen; Volume, Variety, Velocity, Veracity, Analysen
 - kaum Standards
 - meist kein ACID
- Benchmarking von Komponenten vs End-to-End
- Datengenerierung?
- zum Teil Fokussierung auf bestimmte Art von Daten
 - z.B. LinkBench für Graphdaten
- TPC Express for Hadoop Systems (TPCx-HS)
 - basiert auf *Terasort* zum Sortieren großer Datenmengen
- TPCx-BigBench (TPCx-BB)
 - basiert auf TPC-DS für Hadoop-basierte Big Data Analysen



TPCx-HS

- TPC Express for Hadoop Systems (TPCx-HS)
 - seit 2014
- basiert auf *Terasort* zum Sortieren großer Datenmengen
 - Zeit zum (externen) Sortieren von 1 TB Daten (Jim Gray et al)
 - andere Sortierwettbewerbe *MinuteSort* / *PennySort* (sortiere so viel wie möglich in einer Minute/für einen Penny)
- Datenmengen/Skalierungsfaktoren: 1 bis 10.000 TB (10 PB)
 - von Sätzen a 100B
- parallele Ausführung mit Hadoop (HDFS) und MapReduce
 - 2009: 173 Minuten für 100 TB auf Cluster mit 3452 Knoten zur Generierung, Sortierung und Validierung der Daten/Ergebnisse
- Messen der Elapsed Time (Durchsatz) und Kosteneffektivität
 - $HSpH@SF = 3600 * SF / T$ (stundenbezogene Durchsatzmetrik für Skalierungsfaktor SF; Elapsed Time T in Sekunden)
 - $\$/HSpH@SF$: Kosteneffektivität



Ergebnisse TPCx-HS (Juni 2017)

10 TB Results

Rank	Company	System	HSph	Price/HSph	Watts/KHSph	System Availability	Apache Hadoop Compatible Software	MR1/MR2	Operating System	Date Submitted	Nodes
1		Cisco UCS Integrated Infrastructure for Big Data	12.18	314,594.50 CNY	NR	06/15/17	Transwarp Data Hub (TDH) 4.6	2	Red Hat Enterprise Linux Server 6.7	06/15/17	17
2		Cisco UCS Integrated Infrastructure for Big Data	12.02	32,135.61 USD	NR	03/31/16	MapR Converged Community Edition Version 5.0	1	Red Hat Enterprise Linux Server 6.7	03/30/16	16
3		Cisco UCS Integrated Infrastructure for Big Data	11.56	37,066.53 USD	NR	03/23/16	MapR Converged Community Edition Version 5.0	1	Red Hat Enterprise Linux Server 6.5	03/22/16	16
4		Dell PowerEdge 730/730xd	9.07	38,101.22 USD	NR	10/19/15	Cloudera Distribution for Apache Hadoop (CDH) 5.4.2	1	Red Hat Enterprise Linux Server 6.5	10/16/15	13

100 TB Results

Rank	Company	System	HSph	Price/HSph	Watts/KHSph	System Availability	Apache Hadoop Compatible Software	MR1/MR2	Operating System	Date Submitted	Nodes
1		Cisco UCS Integrated Infrastructure for Big Data	22.26	37,839.54 USD	NR	07/07/16	IBM Open Platform (IBM IOP) 4.1	2	Red Hat Enterprise Linux Server 6.7	07/07/16	32
2		Cisco UCS Integrated Infrastructure for Big Data	21.99	39,193.64 USD	NR	10/26/15	Cloudera Distribution for Apache Hadoop (CDH) 5.3.2	1	Red Hat Enterprise Linux Server 6.5	10/23/15	32

300 TB Results

Rank	Company	System	HSph	Price/HSph	Watts/KHSph	System Availability	Apache Hadoop Compatible Software	MR1/MR2	Operating System	Date Submitted	Nodes
1		Cisco UCS Integrated Infrastructure with S3260	11.34	50,472.05 USD	NR	11/11/16	Cloudera Distribution for Apache Hadoop (CDH) 5.3.2	1	Red Hat Enterprise Linux Server 6.7	11/11/16	17

Beispiel 100 TB (SF 100) Cisco/IBM: Systemkosten 842.000 USD (32 Knoten mit 896 Cores),
Dauer für Generierung/Sortierung/Verifikation: 3.908/10.530/1.728 s., insgesamt 16.166 s = 4,5h.



Ergebnisse TPCx-HS (Jan 2019)

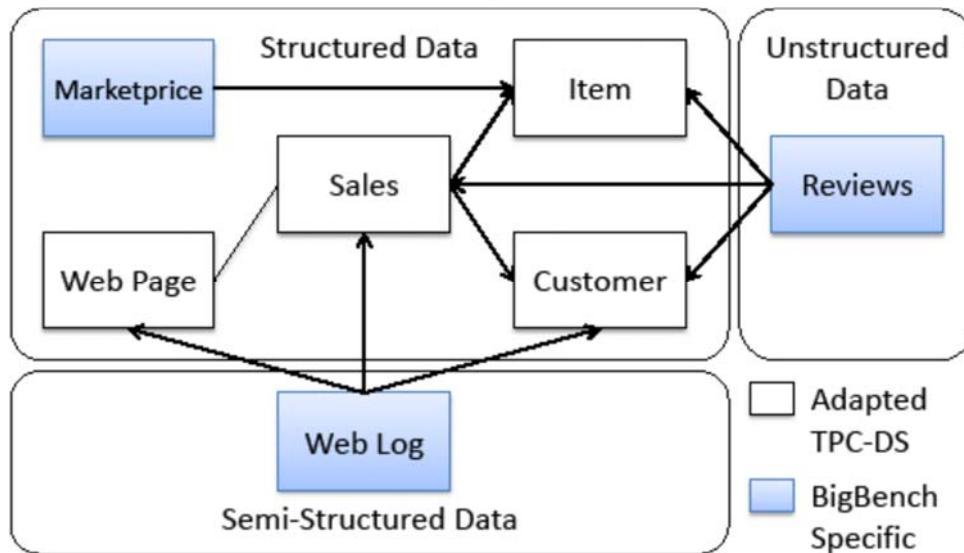
Sponsor	System	Scale Factor	Performance (HSph)	Price/HSph	System Availability	Date Submitted	Data Processing Software Platform	Framework
	Cisco UCS Integrated Infrastructure for Big Data	1	\$8.76	74896.47 USD	1/30/2018	1/30/2018	Cloudera Distribution for Apache Hadoop (CDH) 5.13.0	1
	Dell PowerEdge R6415	1	\$9.09	46871.62 USD	10/10/2018	10/10/2018	Cloudera Distribution for Apache Hadoop (CDH) 6.0	1
	Dell PowerEdge R6415	1	\$11.01	46107.91 USD	10/10/2018	10/10/2018	Cloudera Distribution for Apache Hadoop (CDH) 6.0	2
	Cisco UCS Integrated Infrastructure for Big Data	3	\$10.96	59862.78 USD	1/30/2018	1/30/2018	Cloudera Distribution for Apache Hadoop (CDH) 5.13.0	1
	Cisco UCS Integrated Infrastructure for Big Data	10	\$11.13	58948.43 USD	1/30/2018	1/30/2018	Cloudera Distribution for Apache Hadoop (CDH) 5.13.0	1
	Cisco UCS Integrated Infrastructure for Big Data	30	\$11.54	56854.08 USD	1/30/2018	1/30/2018	Cloudera Distribution for Apache Hadoop (CDH) 5.13.0	1

Ergebnisse mit unterschiedlichen Skalierungsfaktoren sind nicht direkt vergleichbar



TPCx-BigBench

- seit 2016
 - basierend auf BigBench-Paper (Sigmod 2013) von Teradata, Oracle, Univ. Toronto
- Erweiterung von TPC-DS (Enterprise Data Warehouse)
 - semistrukturierte / unstrukturierte Daten: Weblogs, Product reviews



TPCx-BB (2)

- Arbeitslast: 30 Analysen (SQL bzw. Machine Learning)
- Anwendungsbereiche
 - Marketing (cross-selling, ...): 18 Queries
 - Merchandizing (Sortiment- und Preisoptimierung): 5 Queries
 - Betrieb (Produktrückgaben ...),
 - Supply Chain
- unterschiedliche Analysetechniken
 - OLAP mit HiveQL, SparkSQL etc
 - statistische Analysen (Korrelationsanalyse, Regression, Time Series)
 - Data Mining (Klassifikation, Clustering, Assoziationsregeln)
- Benchmark-Aspekte
 - Skalierungsfaktoren: 1, 1000, 3000, ... 1 Million (ca 1 GB – 1000 TB)
 - Messen der Ladezeiten, Powertest (Single User), Durchsatztest (Multi-User Betrieb)

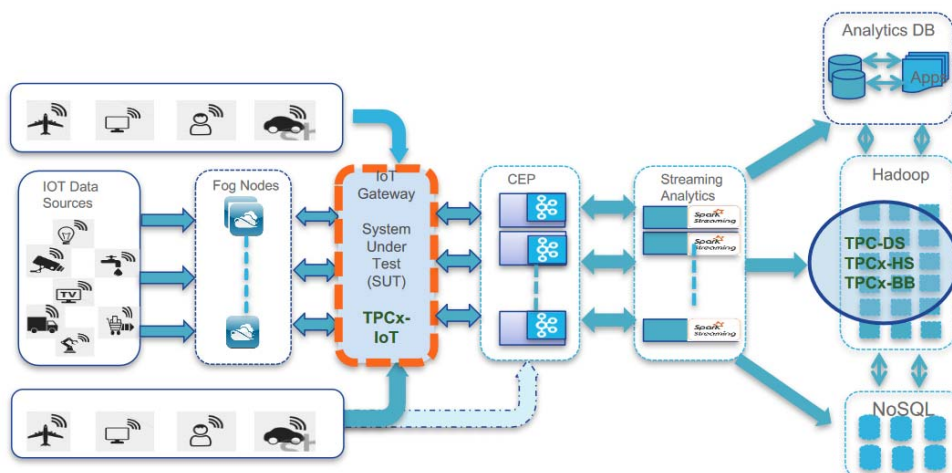
TPCx-BB Ergebnisse

Sponsor	System	Scale Factor	Performance (BBQpm)	Price/BBQpm	System Availability	Date Submitted	DBMS Software (Big Data Software Framework)
	Huawei FusionServer for Big Data	1000	366.69	1,077 USD	2/16/2017	12/29/2016	Cloudera for Apache Hadoop (CDH) 5.9.0
	Huawei FusionServer for Big Data	3000	611.31	646 USD	2/16/2017	12/29/2016	Cloudera for Apache Hadoop (CDH) 5.9.0
	Dell 14G R640/R740xd	10000	1,660.75	546 USD	3/22/2018	3/22/2018	Cloudera Distribution for Apache Hadoop (CDH) 5.13.1
	Sugon Cluster	30000	3,383.95	307 USD	1/22/2018	1/21/2018	Cloudera for Apache Hadoop (CDH) 5.11.1

Ergebnisse mit unterschiedlichen Skalierungsfaktoren sind nicht direkt vergleichbar

TPCx-IOT

- seit 2017
- Evaluierung von IOT-Gateway-Systemen
 - 200 unterschiedliche Sensoren
 - Stromverbrauch in verschiedenen Regionen
 - konstante Datengenerierung (JSON-Format) und Filterung/Aggregation durch Gateway
 - Metrik $IoTps = \text{Skalierungsfaktor } SF / T$ (Elapsed Time für Dateneinfügungen/Queries)
- Referenzimplementierungen für HBase und Couchbase verfügbar



TPCx-IOT Ergebnisse

Rank	Company	System	Performance (IoTps)	Price/IoTps	Watts/IoTps	System Availability	DBMS Software	OS	Processors / Cores / Threads	Date Submitted
1		Dell PowerEdge R6415	165,054.00	0.79 USD	NR	11/30/2018	Hbase 2.0.0 on CDH 6.0	Red Hat Enterprise Linux Server Release 7.5	4/96/192	11/30/2018
2		Cisco UCS Mini	142,493.85	0.94 USD	NR	11/20/2017	Hbase 1.2.1 on CDH 5.10.0	Red Hat Enterprise Linux Server Release 7.2	8/112/224	11/20/2017

Zusammenfassung

- Benchmarks spielen wichtige Rolle bei der Bewertung und Optimierung von DBS
- TPC-Benchmarks
 - meist Bewertung vollständiger Systeme
 - Berücksichtigung von Kosten / Energieverbrauch
 - relativ hoher Aufwand zur Durchführung (Auditing etc.), der für Express-Benchmarks begrenzt wird
 - neuere Benchmarks für Datenintegration und Hadoop/Big Data
- Beobachtungen
 - starke Verbesserungen in Performanz, Preis-Leistungsverhältnis aufgrund technologischer Fortschritte / Systemoptimierungen
 - eingeschränkte Realitätsnähe älterer Benchmarks
 - zunehmender Fokus auf BigData-Benchmarks