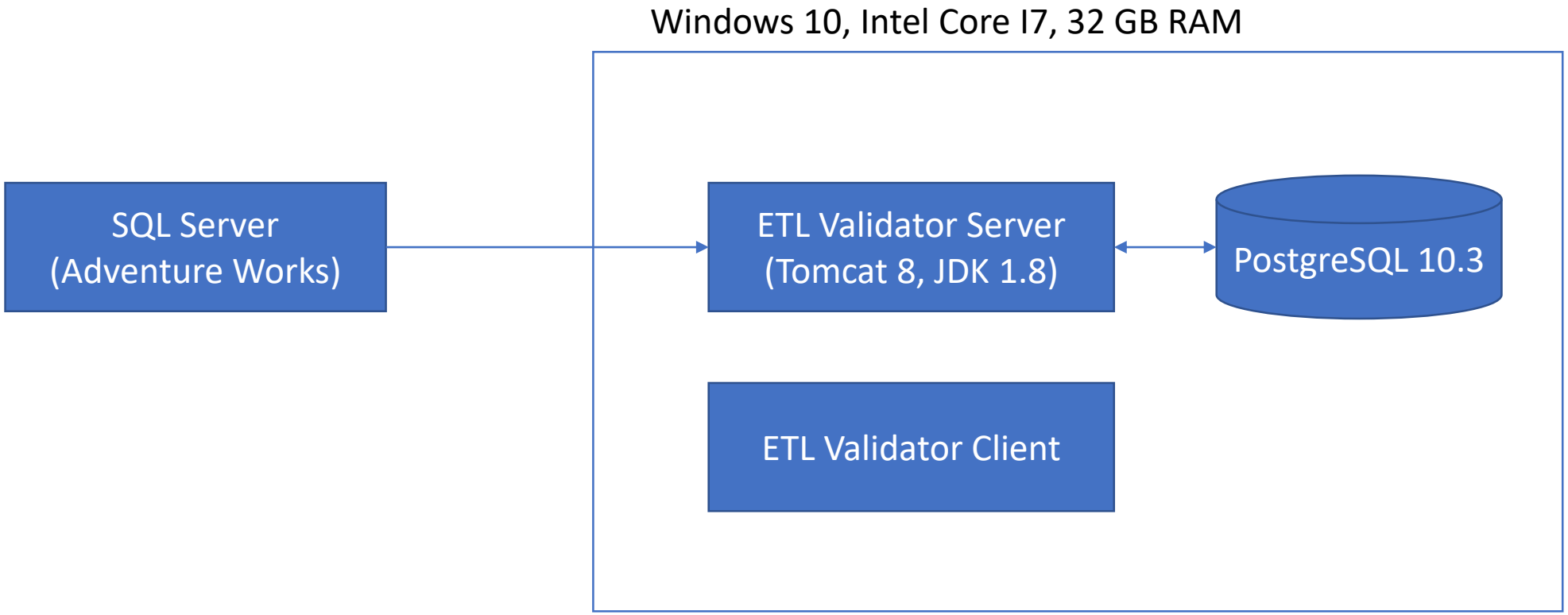
ETL Validator Performance Benchmark

Details of test run to compare ~20 million rows

Disclaimer

Actual results may vary, based on a broad range of implementation-specific factors, such as transaction mix, hardware platform, network parameters, and database size. Datagaps does not warrant or guarantee that customers will obtain the same or similar results, even if they use the same or similar equipment and/or software applications. Datagaps does not warrant, endorse, or guarantee any performance of any products, any results desired or achieved, or any statements made within this document.

Topology


Windows 10, Intel Core I7, 32 GB RAM
SQL Server
(Adventure Works)
ETL Validator Server
(Tomcat 8, JDK 1.8)
PostgreSQL 10.3
ETL Validator Client

Task Manager

File Options View

Processes | Performance | App history | Startup | Users | Details | Services

CPU
3% 1.15 GHz

Memory
13.2/31.8 GB (42%)

Disk 0 (C: D:)
1%

Disk 1 (H:)
0%

Disk 2 (E: F: G:)
0%

Ethernet
S: 200 R: 16.0 Kbps

Ethernet
S: 0 R: 0 Kbps

Ethernet
S: 0 R: 0 Kbps

Ethernet
Not connected

CPU

Intel(R) Core(TM) i7-4770 CPU @ 3.40GHz

% Utilization over 60 seconds

100%

Utilization	Speed		
3%	1.15 GHz		
Processes	Threads	Handles	
315	5569	210966	
Up time			
10:08:20:59			

Base speed:	3.40 GHz
Sockets:	1
Cores:	4
Logical processors:	8
Virtualization:	Enabled
L1 cache:	256 KB
L2 cache:	1.0 MB
L3 cache:	8.0 MB

Test Machine Configuration

Fewer details | Open Resource Monitor

Type here to search

22:24
08-07-2018

Datagaps - The data testing company

datagaps ETL Validator 3.4.7.20180519

Utilities Logs Help admin Status :

Connections

Data Source + | Workschema + | Switch to Tree View

HOME
DATA SOURCES
DATA MODELS
TEST CASES
TEST PLANS
TEST RUNS
PARAMETERS
SCHEDULER
WORK AREA

New Data Source

Name : MSSQL-72-AdW

Data Source Type : Sql Server Prerequisite Steps

Authentication : SQL Server Authentication

User Name : sa

Password : *********

Connection type : Auto Custom

Host Name : 192.168.6.72

Port : 1433

Database : AdventureWorks2014 Retrieve Databases

Connected successfully

Connection Timeout : 30.00 (Seconds)

Test Save Cancel

SQL Server Data Source

Type here to search

13:20
08-07-2018

Definition of data being compared

Source Column Name	Source Data Type
SalesOrderId	int
SalesOrderDetailId	int identity
CarrierTrackingNumber	nvarchar
OrderQty	smallint
ProductId	int
SpecialOfferId	int
UnitPrice	money
UnitPriceDiscount	money
LineTotal	numeric
rowguid	uniqueidentifier
ModifiedDate	datetime
BusinessEntityID	int
QuotaDate	datetime
SalesQuota	money

Datagaps - The data testing company

ETL Validator 3.4.7.20180519

Utilities Logs Help admin Status :

Query Compare TC_1088

Start | Queries | Mapping | Run

HOME
DATA SOURCES
DATA MODELS
TEST CASES
TEST PLANS
TEST RUNS
PARAMETERS
SCHEDULER
WORK AREA

Query Compare Test Case

Test Case Details

TEST CASE NAME* MSSQL_19MR_14C

TEST SUITE* TestSuite_Default

TEST FOLDER* TestFolder_Default

Click to add label..

Mark as Private

Click to add description..

Connection Details

TARGET DB CONNECTION* mssql72addw

SOURCE DB CONNECTION* mssql72addw

WORK SCHEMA* pgws

Query Compare Test Case was used for the test

Description

Compare results from two database queries

Helps compare data from queries on a single Source and Target database. It provides a visual query builder interface to select the columns to compare and generates the query automatically based on the data model. Supports specifying of SQL expression while comparing data between the source and target.

This Test Case is typically used to verify the data moved from the source to target database.

Pre-requisite: For using the visual query builder a data model is required for the selected connections.

Viewlet Tutorial

Save

Type here to search

13:26 08-07-2018

Datagaps - The data testing company

ETL Validator 3.4.7.20180519

Utilities Logs Help admin Status :

MSSQL_19MR_14C

Start | Queries | Mapping | Run

TARGET QUERY (use alias for functions)

```
select $[p_limit]
SSO.SalesOrderId,SSO.SalesOrderDetailId,SSO.CarrierTrackingNumber,SSO.OrderQty,SSO.ProductId,SSO.SpecialOfferId,SSO
.UnitPrice,SSO.UnitPriceDiscount,SSO.LineTotal,SSO.rowguid,SSO.ModifiedDate,SSP.BusinessEntityID,SSP.QuotaDate,SSP.
SalesQuota from  Sales.SalesOrderDetail SSO,Sales.SalesPersonQuotaHistory SSP
```

Query on Adventure Works Dataset was used for the test

SOURCE QUERY (use alias for functions)

```
select $[p_limit]
SSO.SalesOrderId,SSO.SalesOrderDetailId,SSO.CarrierTrackingNumber,SSO.OrderQty,SSO.ProductId,SSO.SpecialOfferId,SS
O.UnitPrice,SSO.UnitPriceDiscount,SSO.LineTotal,SSO.rowguid,SSO.ModifiedDate,SSP.BusinessEntityID,SSP.QuotaDate,SS
P.SalesQuota from  Sales.SalesOrderDetail SSO,Sales.SalesPersonQuotaHistory SSP
```

Row Count : 19774671
Size : 3 GB

RESULTS (Shows the first 100 rows)

Sales	S	CarrierTracki		P	Spe	UnitPrice	UnitPr	LineTotal	rowguid	Modified	Busine	QuotaDate
43659	1	4911-403C-98	1	776	1	2024.9940	0.0000	2024.994000	B207C96D-D9E6-4(	2011-05-31 0	274	2011-05-31 00
43659	2	4911-403C-98	3	777	1	2024.9940	0.0000	6074.982000	7ABB600D-1E77-41	2011-05-31 0	274	2011-05-31 00
43659	3	4911-403C-98	1	778	1	2024.9940	0.0000	2024.994000	475CF8C6-49F6-48	2011-05-31 0	274	2011-05-31 00
43659	4	4911-403C-98	1	771	1	2039.9940	0.0000	2039.994000	04C4DE91-5815-45	2011-05-31 0	274	2011-05-31 00

RESULTS (Shows the fi

cnt
19774671

Query Builder Save

HOME
DATA SOURCES
DATA MODELS
TEST CASES
TEST PLANS
TEST RUNS
PARAMETERS
SCHEDULER
WORK AREA

Type here to search

13:27 08-07-2018

Datagaps - The data testing company

ETL Validator 3.4.7.20180519

Utilities Logs Help admin Status :

MSSQL_19MR_14C

Start | Queries | Mapping | Run

Join	Target Column Name	Target Data Type	Source Column Expression	Source Alias	Source Data Type
	salesorderid	numeric	salesorderid	salesorderid	numeric
	salesorderdetailid	numeric	salesorderdetailid	salesorderdetailid	numeric
	carriertrackingnumber	character varying	carriertrackingnumber	carriertrackingnumber	character varying
	orderqty	numeric	orderqty	orderqty	numeric
	productid	numeric	productid	productid	numeric
	specialofferid	numeric	specialofferid	specialofferid	numeric
	unitprice	numeric	unitprice	unitprice	numeric
	unitpricediscount	numeric	unitpricediscount	unitpricediscount	numeric
	linetotal	numeric	linetotal	linetotal	numeric
	rowguid	character varying	rowguid	rowguid	character varying
	modifieddate	timestamp without time zone	modifieddate	modifieddate	timestamp without time zone
	businessentityid	numeric	businessentityid	businessentityid	numeric
	quotadate	timestamp without time zone	quotadate	quotadate	timestamp without time zone
	salesquota	numeric	salesquota	salesquota	numeric

Column Mapping between Source and Target. There are a total of 14 columns that are being compared

Remap Columns by Name

HOME | DATA SOURCES | DATA MODELS | TEST CASES | TEST PLANS | TEST RUNS | PARAMETERS | SCHEDULER | WORK AREA

Type here to search

13:27 08-07-2018

Datagaps - The data testing company

ETL Validator 3.4.7.20180519

Utilities Logs Help admin Status :

MSSQL_19MR_14C

Start | Queries | Mapping | Run

Test Case Run Details Run Duration : 00:00:01

Execution Started

Source (mssql72addw) RUNNING

Rows Loaded : 3400000 of 19774671

Run Duration : 00:00:49

Throughput : 66096 rows/sec

Source data load. Note that throughput of 66096 rows per second

Target (mssql72addw) QUEUED

Rows Loaded : 0

Run Duration : 00:00:01

Throughput : 0 rows/sec

Data Comparison

Run Duration : 00:00:01

Running... Run Abort Run Date: 08-07-2018 13:27:51 Remove Refresh

HOME DATA SOURCES DATA MODELS TEST CASES TEST PLANS TEST RUNS PARAMETERS SCHEDULER WORK AREA

Type here to search 13:28 08-07-2018

Datagaps - The data testing company

ETL Validator 3.4.7.20180519

Utilities Logs Help admin Status :

MSSQL_19MR_14C

Start | Queries | Mapping | Run

HOME
DATA SOURCES
DATA MODELS
TEST CASES
TEST PLANS
TEST RUNS
PARAMETERS
SCHEDULER
WORK AREA

Test Case Run Details Run Duration : 00:00:01

Execution Started

Source (mssql72addw) RUNNING

Rows Loaded : 9200000 of 19774671

Run Duration : 00:02:14

Throughput : 68391 rows/sec

Target (mssql72addw) QUEUED

Rows Loaded : 0

Run Duration : 00:00:01

Throughput : 0 rows/sec

Data Comparison

Run Duration : 00:00:01

Source data load. Note that throughput of 68391 rows per second

Task Manager

File Options View

Processes | Performance | App history | Startup | Users | Details | Services

Name	Status	24% CPU	39% Memory	3% Disk	56% Network
Photos (2)		0%	470.6 MB	0 MB/s	0 Mbps
Commons Daemon Service R...		10.0%	329.1 MB	0 MB/s	80.3 Mbps
About ETL/Data Validator		0.2%	224.4 MB	0 MB/s	0 Mbps
Java(TM) Platform SE binary		0%	179.9 MB	0 MB/s	0 Mbps
Antimalware Service Executable		3.2%	145.0 MB	0.1 MB/s	0 Mbps
Commons Daemon Service Run...		0%	130.4 MB	0 MB/s	0 Mbps
Cortana (2)		0%	130.2 MB	0 MB/s	0 Mbps
Windows Explorer (6)		0%	69.5 MB	0 MB/s	0 Mbps
Service Host: Remote Desktop S...		0.1%	63.8 MB	0 MB/s	0.1 Mbps
pgAdmin 4 Desktop Runtime (3...		0%	47.1 MB	0 MB/s	0 Mbps
Desktop Window Manager		0%	45.8 MB	0 MB/s	0 Mbps
Resource and Performance Mo...		0%	41.6 MB	0 MB/s	0 Mbps
Windows Shell Experience Host		0%	40.7 MB	0 MB/s	0 Mbps
µTorrent (32 bit)		0%	30.8 MB	0 MB/s	0 Mbps
Service Host: Diagnostic Policy ...		0%	29.4 MB	0 MB/s	0 Mbps

Fewer details End task

Source data load. Note that network throughput of 80.3 Mbps

Running... Run Abort Run Date: 08-07-2018 13:27:51 Remove Refresh

Type here to search 13:30 08-07-2018

Task Manager

File Options View

Processes | Performance | App history | Startup | Users | Details | Services

Name	Status	20% CPU	41% Memory	2% Disk	56% Network
pgbouncer.exe (32 bit)		0%	0.3 MB	0 MB/s	0 Mbps
PostgreSQL Server		0%	2.3 MB	0 MB/s	0 Mbps
PostgreSQL Server		0%	2.3 MB	0 MB/s	0 Mbps
PostgreSQL Server		0%	2.6 MB	0 MB/s	0 Mbps
PostgreSQL Server		0%	2.6 MB	0 MB/s	0 Mbps
PostgreSQL Server		0%	2.3 MB	0 MB/s	0 Mbps
PostgreSQL Server		0%	5.1 MB	0 MB/s	0 Mbps
PostgreSQL Server		7.5%	3.6 MB	10.5 MB/s	0 Mbps
PostgreSQL Server		0%	2.6 MB	0 MB/s	0 Mbps
PostgreSQL Server		0%	2.0 MB	0 MB/s	0 Mbps
PostgreSQL Server		0%	2.0 MB	0 MB/s	0 Mbps
PostgreSQL Server		0%	2.0 MB	0 MB/s	0 Mbps
PostgreSQL Server		0%	1.9 MB	0 MB/s	0 Mbps
PostgreSQL Server		0%	1.9 MB	0 MB/s	0 Mbps
PostgreSQL Server		0%	1.9 MB	0 MB/s	0 Mbps
PostgreSQL Server		0%	1.9 MB	0 MB/s	0 Mbps
PostgreSQL Server		0%	1.9 MB	0 MB/s	0 Mbps
PostgreSQL Server		0%	1.9 MB	0 MB/s	0 Mbps
PostgreSQL Server		0%	1.9 MB	0 MB/s	0 Mbps
PostgreSQL Server		0%	1.9 MB	0 MB/s	0 Mbps
PostgreSQL Server		0%	1.9 MB	0 MB/s	0 Mbps
PostgreSQL Server		0%	1.8 MB	0 MB/s	0 Mbps
PostgreSQL Server		0%	1.8 MB	0 MB/s	0 Mbps
PostgreSQL Server		0%	1.4 MB	0 MB/s	0 Mbps
PostgreSQL Server		0%	1.4 MB	0 MB/s	0 Mbps
PostgreSQL Server		0.2%	1.3 MB	21.4 MB/s	0 Mbps
PostgreSQL Server		0.1%	1.2 MB	0 MB/s	0 Mbps
PostgreSQL Server		0%	1.2 MB	0 MB/s	0 Mbps
PostgreSQL Server		0.1%	1.2 MB	8.6 MB/s	0 Mbps
PostgreSQL Server		0%	0.9 MB	0 MB/s	0 Mbps
PostgreSQL Server		0%	0.8 MB	0 MB/s	0 Mbps

Disk writes by PostgreSQL process : 10.5 MB/s

Fewer details

End task

Type here to search

13:31
08-07-2018

Datagaps - The data testing company

datagaps ETL Validator 3.4.7.20180519

Utilities Logs Help admin Status :

MSSQL_19MR_14C

Start | Run

HOME
DATA SOURCES
DATA MODELS
TEST CASES
TEST PLANS
TEST RUNS
PARAMETERS
SCHEDULER
WORK AREA

Test Case Run Details Run Duration : 00:00:01

Execution Started

Source (mssql72addw) SUCCESS

Rows Loaded : 19774671 of 19774671

Run Duration : 00:04:47

Throughput : 68844 rows/sec

Target (mssql72addw) SUCCESS

Rows Loaded : 19774671 of 19774671

Run Duration : 00:04:49

Throughput : 68408 rows/sec

Data Comparison

Run Duration : 12:29:27

Running...

Task Manager

File Options View

Processes | Performance | App history | Startup | Users | Details | Services

Name	Status	30% CPU	75% Memory	1% Disk	0% Network
Oracle MTS Recovery Service		0%	0.3 MB	0 MB/s	0 Mbps
oravssw.exe		0%	0.2 MB	0 MB/s	0 Mbps
pg_ctl - starts/stops/restarts the...		0%	1.1 MB	0 MB/s	0 Mbps
pgAdmin 4 Desktop Runtime (3...		0%	47.1 MB	0 MB/s	0 Mbps
PostgreSQL Server		0%	2.2 MB	0 MB/s	0 Mbps
PostgreSQL Server		0%	5.1 MB	0 MB/s	0 Mbps
PostgreSQL Server		15.7%	8,550.0 MB	0 MB/s	0 Mbps
PostgreSQL Server		0%	2.9 MB	0 MB/s	0 Mbps
PostgreSQL Server		0%	2.0 MB	0 MB/s	0 Mbps
PostgreSQL Server		0%	2.0 MB	0 MB/s	0 Mbps
PostgreSQL Server		0%	1.8 MB	0 MB/s	0 Mbps
PostgreSQL Server		0%	1.8 MB	0 MB/s	0 Mbps
PostgreSQL Server		0%	1.4 MB	0 MB/s	0 Mbps
PostgreSQL Server		0%	1.4 MB	0 MB/s	0 Mbps
PostgreSQL Server		0%	1.3 MB	3.0 MB/s	0 Mbps
PostgreSQL Server		0%	1.2 MB	0 MB/s	0 Mbps
PostgreSQL Server		0%	1.2 MB	0 MB/s	0 Mbps
PostgreSQL Server		0%	1.2 MB	0 MB/s	0 Mbps
RDP Clipboard Monitor		0.2%	2.1 MB	0 MB/s	0 Mbps
Runtime Broker		0%	4.8 MB	0 MB/s	0 Mbps
Runtime Broker		0%	4.7 MB	0 MB/s	0 Mbps
Runtime Broker		0%	2.7 MB	0 MB/s	0 Mbps
Settings		0%	0.4 MB	0 MB/s	0 Mbps
Spooler SubSystem App		0%	6.0 MB	0 MB/s	0 Mbps
Stereo Vision Control Panel API ...		0%	0.9 MB	0 MB/s	0 Mbps
System Guard Runtime Monitor...		0%	3.2 MB	0 MB/s	0 Mbps
TeamViewer 13 (32 bit)		0%	2.0 MB	0 MB/s	0 Mbps
Usermode Font Driver Host		0%	1.6 MB	0 MB/s	0 Mbps
Usermode Font Driver Host		0%	0.3 MB	0 MB/s	0 Mbps
Video Application (2)		0%	0.9 MB	0 MB/s	0 Mbps
VMware Authorization Service (		0.1%	1.8 MB	0 MB/s	0 Mbps

Fewer details End task

In-memory data comparison. Note the use of 8.5 GB RAM

Type here to search

13:38 08-07-2018

Datagaps - The data testing company

ETL Validator 3.4.7.20180519

Utilities Logs Help admin Status :

MSSQL_19MR_14C

Start | Queries | Mapping | Run

Run Summary SUCCESS

All 19774671 rows matched

UnMatched Matched

View Report in Browser

SHOW EXPORT DIFFERENCES

SHOW QUERIES

HIDE RUN DURATIONS

Total Duration : 00:14:54
Source Data Load : 00:04:47
Target Data Load : 00:04:49

SHOW DATA SOURCES

UNMATCHED	MATCHED	SOURCE	TARGET
0 (0%)	19774671 (100%)	19774671	19774671
(unique mismatches)	(unique matches)	Duplicates: 0 (0%)	Duplicates: 0 (0%)
		Unique Keys: 19774671 (100%)	Unique Keys: 19774671 (100%)

Data Differences | Join Based Column Differences | Row Differences

No Unmatched Data Found

Source Data Load : 4 mins 47 secs
Target Data Load : 4 mins 49 secs
Data Comparison : 5 mins 18 secs
Total Run Time : 14 mins 54 secs

Run | Abort | Run Date: 08-07-2018 13:27:51 | Remove | Refresh

HOME | DATA SOURCES | DATA MODELS | TEST CASES | TEST PLANS | TEST RUNS | PARAMETERS | SCHEDULER | WORK AREA

Type here to search

13:43 08-07-2018

Recommendation for Memory

Size of data being compared	Minimum RAM (Recommended)
1 GB	8 GB
2 GB	16 GB
4 GB	24 GB
8 GB	32 GB
16 GB	64 GB
32 GB	128 GB

Recommendation for Cores (Processor)

Number of test cases in parallel	Number of Cores (Recommended)
1	4
2	6
4	8
8	16
16	32