



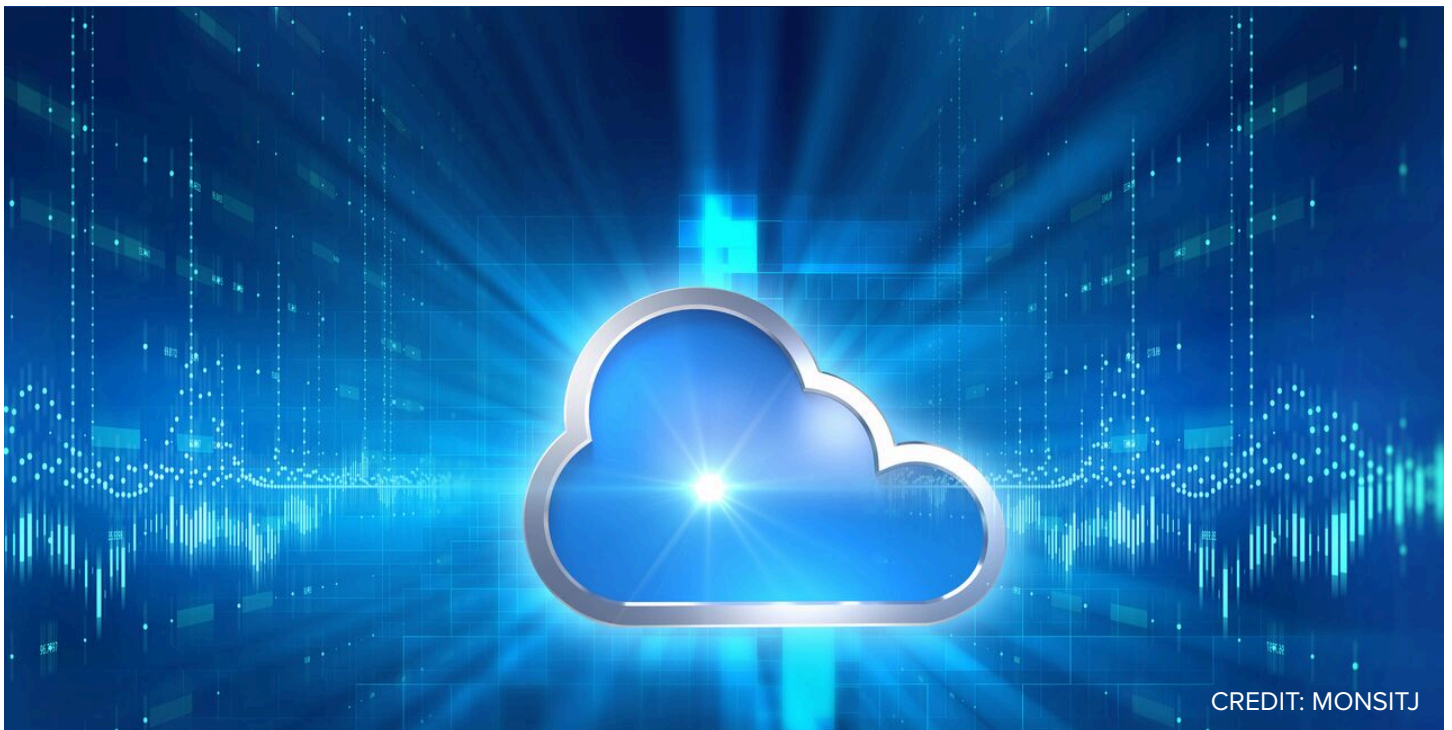
## REPORT

# Cloud Data Warehouse Performance Testing

*Product Profile and Evaluation: Amazon Redshift, Microsoft Azure SQL Data Warehouse, Google BigQuery, and Snowflake Data Warehouse*

WILLIAM MCKNIGHT

TOPIC: **DATA WAREHOUSE**



CREDIT: MONSITJ

SPONSORED BY MICROSOFT

# Cloud Data Warehouse Performance Testing

*Product Profile and Evaluation: Amazon Redshift, Microsoft Azure SQL Data Warehouse, Google BigQuery, and Snowflake Data Warehouse*

## TABLE OF CONTENTS

- 1** Cloud Analytics Platform Offerings
- 2** Field Test Setup
- 3** Field Test Results
- 4** Price Per Performance
- 5** Summary
- 6** In Closing
- 7** About the Analyst
- 8** About GigaOm Research
- 9** About Microsoft
- 10** Copyright

## 1 Cloud Analytics Platform Offerings

Big data analytics platforms load, store, and analyze volumes of data rapidly, providing timely insights to businesses. Data-driven organizations leverage this data, for example, for advanced analysis to market new promotions, as operational analytics to drive efficiency, or for predictive analytics to evaluate credit risk and detect fraud. Customers are leveraging a mix of relational analytical databases and data warehouses to gain insights.

This report focuses on relational analytical databases in the cloud, because deployments are at an all-time high and poised to expand dramatically. The cloud enables enterprises to differentiate and innovate with these database systems at a much quicker pace than was ever possible before. The cloud is a disruptive technology, offering elastic scalability vis-à-vis on-premises deployments, enabling faster server deployment and application development, and allowing for less costly storage. For these reasons and others, many companies have leveraged the cloud to maintain, or gain, momentum as a company.

This report outlines the results from a GigaOm Analytic Field Test derived from the industry standard TPC Benchmark™ DS (TPC-DS)<sup>1</sup> comparing Amazon Redshift, Azure SQL Data Warehouse, Google BigQuery, and Snowflake Data Warehouse — four relational analytical databases based on scale-out cloud data warehouses and columnar-based database architectures. Despite these similarities, there are some distinct differences between the four platforms.

### Amazon Redshift

Amazon Web Services Redshift was the first managed data warehouse cloud service and continues to maintain a high level of mindshare in this category. It does indeed tick all the table stakes boxes for a cloud analytic database.

Amazon Redshift Spectrum can create an external table to store raw data on Amazon S3. Redshift Spectrum has some future-proofing that a modern data engineering approach might utilize. In support of diverse data, Amazon Redshift has a few nice features like the `JSON_EXTRACT_PATH_TEXT` function for noSQL key value databases.

Amazon Redshift is a fit for organizations needing a data warehouse with little to no administrative overhead and a clear, consistent pricing model. Amazon Web Services supports most of the databases in this report, and then some.

### Azure SQL Data Warehouse

Azure SQL Data Warehouse storage is separate from the compute Data Warehouse Unit (DWU). This enables Azure SQL Data Warehouse to scale columnar storage capacity and compute resources independently. This capability adjusts to varying workload demands, offering potential cost savings

when demand is low. Azure SQL Data Warehouse can also pause and resume compute billing, meaning only the storage is billed for during the paused time. Azure SQL Data Warehouse achieves a good balance in both configurability and simplicity, in a way that is both easy to administer and flexible in handling almost any usage pattern.

Azure SQL Data Warehouse is fully ANSI-SQL compliant and users familiar with SQL Server will be very comfortable using this environment.

Azure SQL Data Warehouse can export data to a local file the same way an on-premises SQL Server can, e.g., via the SQL Server Import and Export Wizard.

## Google BigQuery

Google BigQuery is a managed service with some interesting distinctions. Google abstracts the details of the underlying hardware, database, and all configurations.

BigQuery is a hands-off database without indexes or column constraints. Defragmentation and system tuning are not required. It is truly serverless. Google Cloud manages the servers in a fully hands-off manner to the customer, dynamically allocating storage and compute resources. The customer does not define nodes and capacity of the BigQuery instance. The provisioning of compute is particularly fast and seamless.

You pay for the amount of data you query and store. Customers can pre-purchase flat-rate computation “slots” or units in increments of \$10,000 per month per 500 compute units.

We simply consumed the results for this field test, but should we have been looking to do more with the data, such as exporting it in different formats, BigQuery has the capabilities to do so.

Also, although we did not time data ingest, this is an area of strength for BigQuery. BigQuery ingest does not impact the performance of queries, because separate compute slots are used for loading apart from the ones performing the SELECT statements for users. Ingest stats are auto-generated and data with an unknown schema can be loaded using schema auto-suggestion. There is no charge for ingest.

## Snowflake Data Warehouse

As a cloud-only, fully managed solution, Snowflake has a clear separation between compute and storage. For Snowflake on AWS, which is what we used for the queries, data is stored in AWS S3 and is cached when queries are executed to bring the data in closer proximity to compute resources. Snowflake essentially offers two configuration “levers” — the size of the warehouse cluster and how many clusters are permitted to spin up to handle concurrency. Snowflake scales by cluster server count in powers of 2 (i.e., 1, 2, 4, 8, 16, and so on). If enabled, Snowflake will spin up additional clusters to handle multi-user concurrent query workloads. Snowflake would then automatically spin the

additional clusters down once demand has passed. If not enabled, it will place paused queries in a queue until resources free up.

In our estimation, Snowflake performance largely hinges on cache. If Snowflake must access Amazon S3 for additional data, query execution slows dramatically. Historically, Snowflake has not disclosed the server class or size of its AWS EC2 instances. Thus, one cannot calculate beforehand what configuration would be most optimal for workloads.

Snowflake supports an ANSI-compliant form of SQL. However, we experienced syntax differences when converting existing queries to Snowflake. Snowflake supports data export to S3 (on AWS), though it is possible to use the Snowflake command line interface results to direct data to a file in a Linux shell.

Table 1. Platform Summary

	Azure SQL Data Warehouse	Snowflake Data Warehouse	Google BigQuery	Amazon Redshift
First Released	1989 (SQL Server)			
	2005 (Analytics Platform System)	2014	2010	2014
	2016 (Azure SQL Data Warehouse)			
Current Version	10.0.10106.0	3.13	March 1, 2019 release	1.0.5833
SQL	Transact-SQL	Snowflake SQL	Ansi-2011 Compliant	PostgreSQL 8

<sup>1</sup>More can be learned about the TPC-DS benchmark at <http://www.tpc.org/tpcds/>.

## 2 Field Test Setup

The benchmark equivalency was established by price across the four platforms. Microsoft DW15000C costs \$183.86/hour to run. 3X-Large Snowflake @ \$3.00/node costs \$192/hour at their Enterprise support rate. (However, in the Price-Performance calculation below, we used the lowest, limited support \$2.00/node price for Snowflake.) Thirty nodes of Redshift dc2.8xlarge @ \$4.80/node costs \$144.00/hour to run. For Google, we used the cost-per-hour from BigQuery flat rate, which currently is \$55.00/hour.

### Field Test Data

The data sets used in the benchmark were a workload derived from the well-recognized industry standard [TPC Benchmark™ DS \(TPC-DS\)](#). The parameter values for the queries used across all vendors are informed by the TPC Benchmark™ DS (TPC-DS)<sup>2</sup> spec validation queries. This field test is not an official TPC benchmark.

From [tpc.org](http://tpc.org): “The TPCDS is a decision support benchmark that models several generally applicable aspects of a decision support system, including queries and data maintenance. The benchmark provides a representative evaluation of performance as a general purpose decision support system... The purpose of TPC benchmarks is to provide relevant, objective performance data to industry users. TPC-DS Version 2 enables emerging technologies, such as Big Data systems, to execute the benchmark.”

The queries were executed using the following setup, environment, standards, and configurations. The data model consists of 24 tables — 7 fact tables and 17 dimensions. To give an idea of the data volumes used in our field test, the following table gives row counts of fact tables in the database when loaded with 1TB and 10TB of GigaOm Analytic Field Test data:

Table 2. Database Row Count given 1TB and 10TB

GigaOm Analytic Field Test Table	Scale Factor 30,000 30TB Row Count
Catalog Returns	4,319,925,093
Catalog Sales	43,200,404,822
Inventory	1,627,857,000
Store Returns	8,639,952,111
Store Sales	86,399,341,874
Web Returns	2,160,007,345
Web Sales	21,600,036,511

## Cluster Environments

Our benchmark included four (4) different cluster environments:

- **Azure SQL Data Warehouse** DW15000C (30 compute nodes)
- **Amazon Redshift** dc2.8xlarge (30 compute nodes)
- **Snowflake Computing** 3X-Large (64 compute nodes)
- **Google BigQuery**, for which the underlying architecture is unknown and the environment is scaled automatically.

## Queries

The GigaOm Analytic Field Test is a fair representation of enterprise query needs. The GigaOm Analytic Field Test suite has 99 queries — four of which have two parts (14, 23, 24, and 39), for a total of 103 queries. The queries used for the tests were compliant with the standards set out by the TPC Benchmark™ DS (TPC-DS) specifications<sup>3</sup> and included only minor query modifications as set out by section 4.2.3 of the TPC-DS specification document. For example, minor query modifications included vendor-specific syntax for date expressions. Also in the specification, some queries require row limits and, thus, vendor-specific syntax was used (e.g., TOP, FIRST, LIMIT, and so forth) as allowed by section 4.2.4 of the TPC-DS specification.

Although concurrency was not tested in the benchmark, Azure SQL Data Warehouse supports [128 concurrent](#) queries. BigQuery supports a maximum [concurrency of 50](#) per project, but this is not a hard limit. Many customers use hundreds or thousands of projects, so they get effective concurrency that is much higher. Snowflake's maximum concurrency is difficult to calculate because it is a function of the number of queries, the submitted queries' execution plan, the size of the warehouse, and the maximum number of multi-cluster settings. In our experience, we saw an X-Large (16 node) Snowflake warehouse run 6 concurrent simple scan queries (SELECT with a single column filter WHERE clause) before starting to queue. Your results may vary.

<sup>2</sup> More can be learned about the TPC-DS benchmark at <http://www.tpc.org/tpcds/>.

<sup>3</sup> The TPC Benchmark™ DS (TPC-DS) specification we used was found at [http://www.tpc.org/tpc\\_documents\\_current\\_versions/pdf/tpc-ds\\_v2.1.0.pdf](http://www.tpc.org/tpc_documents_current_versions/pdf/tpc-ds_v2.1.0.pdf).

## 3 Field Test Results

This section analyzes the query results from the fastest of the three runs of the GigaOm Analytic Field Test queries (derived from the TPC-DS) described above. The primary metric used was the aggregate total of the best execution times for each query. Three power runs were completed. Each of the 99 queries was executed three times in order (1, 2, 3) against each vendor cloud platform, and the fastest of the three times was used as the performance metric. These best times were then added together to gain the total aggregate execution time for the entire workload.

### Selected Individual Query Results

Since the GigaOm Analytic Field Test included 103 TPC-DS derived queries, it would be too exhaustive to provide charts and analysis for every single one. GigaOm selected 10 queries to highlight that we determined to be good representatives of common data warehouse queries and use cases, irrespective of competitive performance. The section below discusses the queries and presents the results of the four platforms tested.

#### *Long Running Queries*

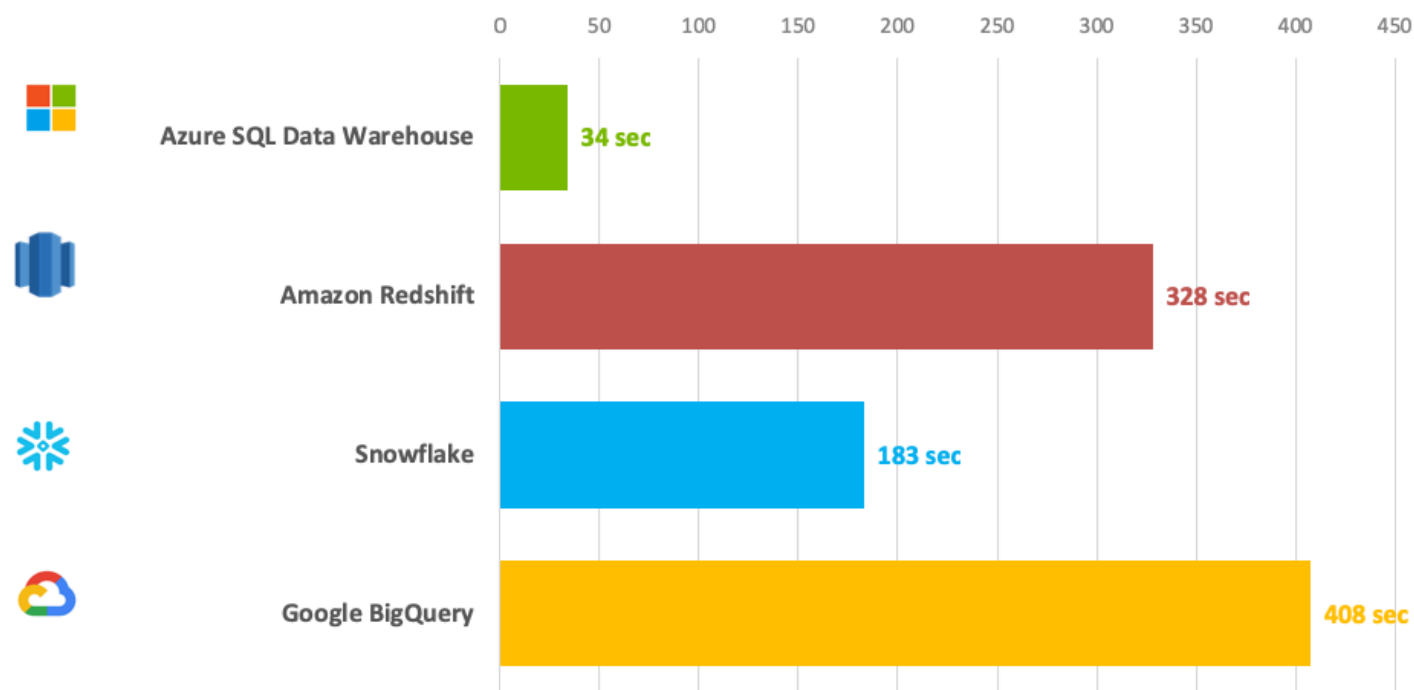
##### *Query 14a*

Query 14a is a lookup of sales by item, with breakouts of channels and brands. This information is valuable to marketing and sales leadership, and could be used behind interactive business intelligence (BI) dashboards driven by several parameters in the query, including date, brand, and category. If you have ever performed this SQL query in a modern organization, where there is a table for anything that can be determined to be discrete, you know the calculation is very complex, combining catalog, web, and store sales.

Query 14a is 1 of 2 queries in the GigaOm Analytic Field Test set that feature an intersect function and there are 2 intersects in the query. It is also one of the longer running queries in the set, so the performance is important.



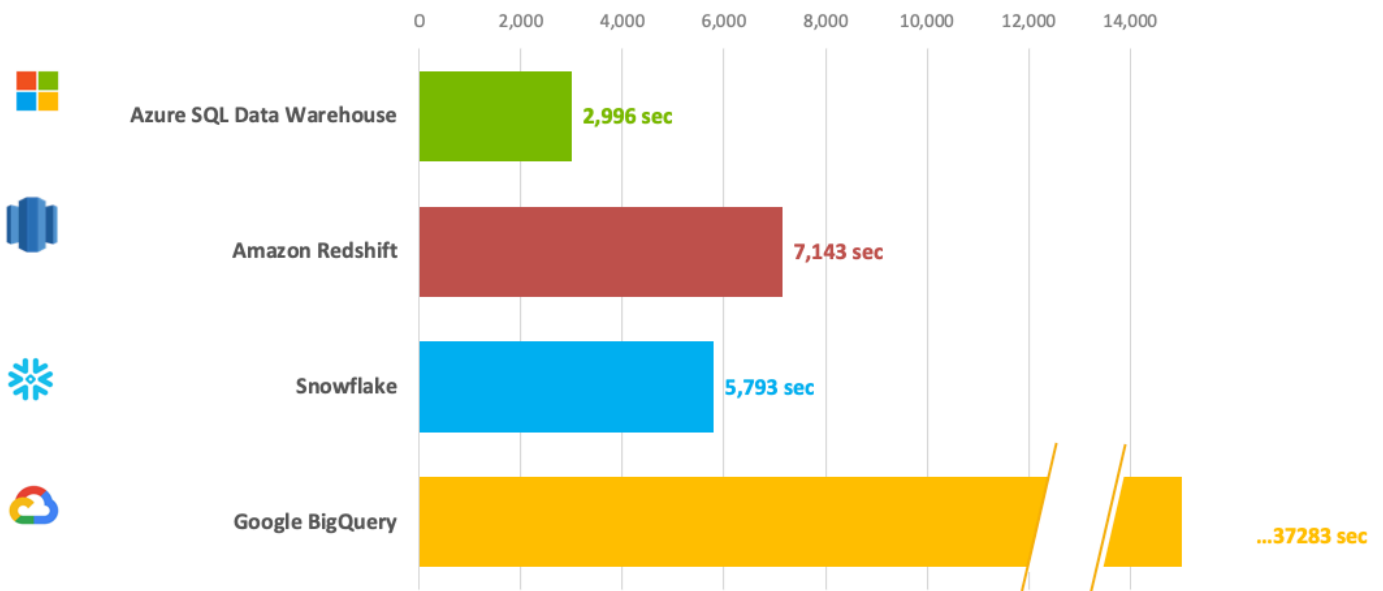
GigaOM Analytic Field Test Query 14 Execution Times



Query 80

Query 80 reports sales, net profit, and returns in all channels for a window of time for items with prices larger than a certain amount that are not promoted on television. Results are aggregated by channel and “means” within the channel (such as the store for store sales, the page of the catalog for catalog sales, and the actual web site for web sales). This is a helpful query for establishing the extended value of each channel.

Query 80 is complex to code and is also a bottom 10% query in terms of usual performance time.



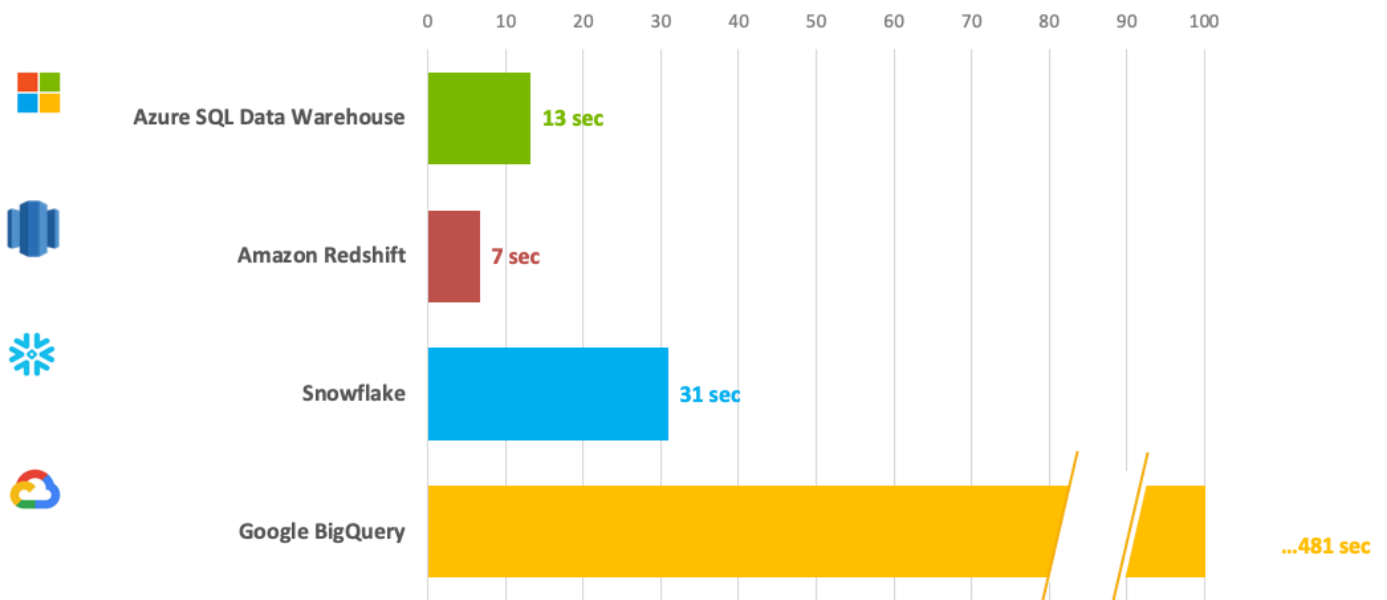
## Frequently Run Queries

### Query 6

Query 6 returns high value and volume buyers by state, which is analogous to queries that are routinely performed in the enterprise today where support, service, and marketing functions are being tiered based on customer value. Also, Query 6 triangulates item and pricing data, so this query benefits pricing teams within marketing groups, targeting future pricing that will be attractive to customers in various states and regions.

Query 6 is one of the faster GigaOm Analytic Field Test performing queries, but that does not take away from its importance. Query 6 is a simplistic 5-table join, with one correlated subquery, but is of sufficient complexity that performance can start becoming an issue in an enterprise.

## GigaOM Analytic Field Test Query 6 Execution Times

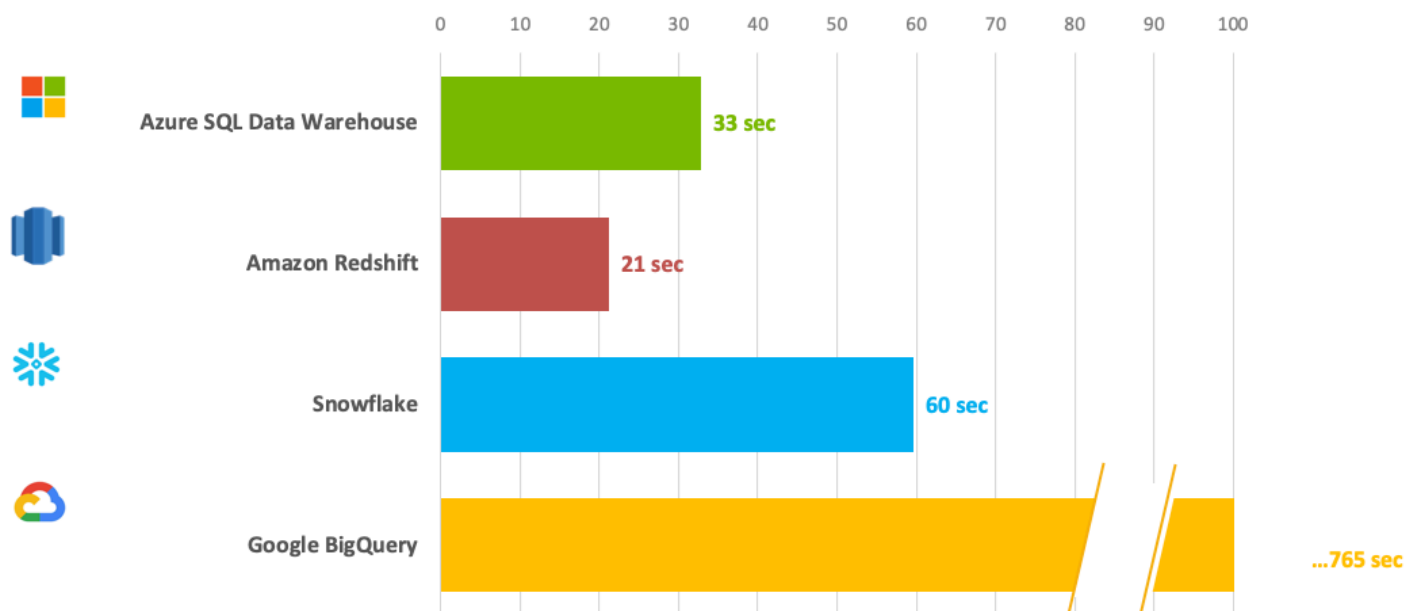


## Query 29

Query 29 retrieves how many items were sold in stores, which were returned in a time window, and which were purchased through the catalog in a time window. It shows the quantity with results grouped by item and store. This query could drive sales BI reports and dashboards from mid-level district managers down to store managers to make predictions for future inventory levels, stocking, and purchasing needs.

In terms of complexity, Query 29 is a 6-table join with date joined into the core tables 3 times. It is a mid-tier performer, but it is really important to see how well the database performs the joins.

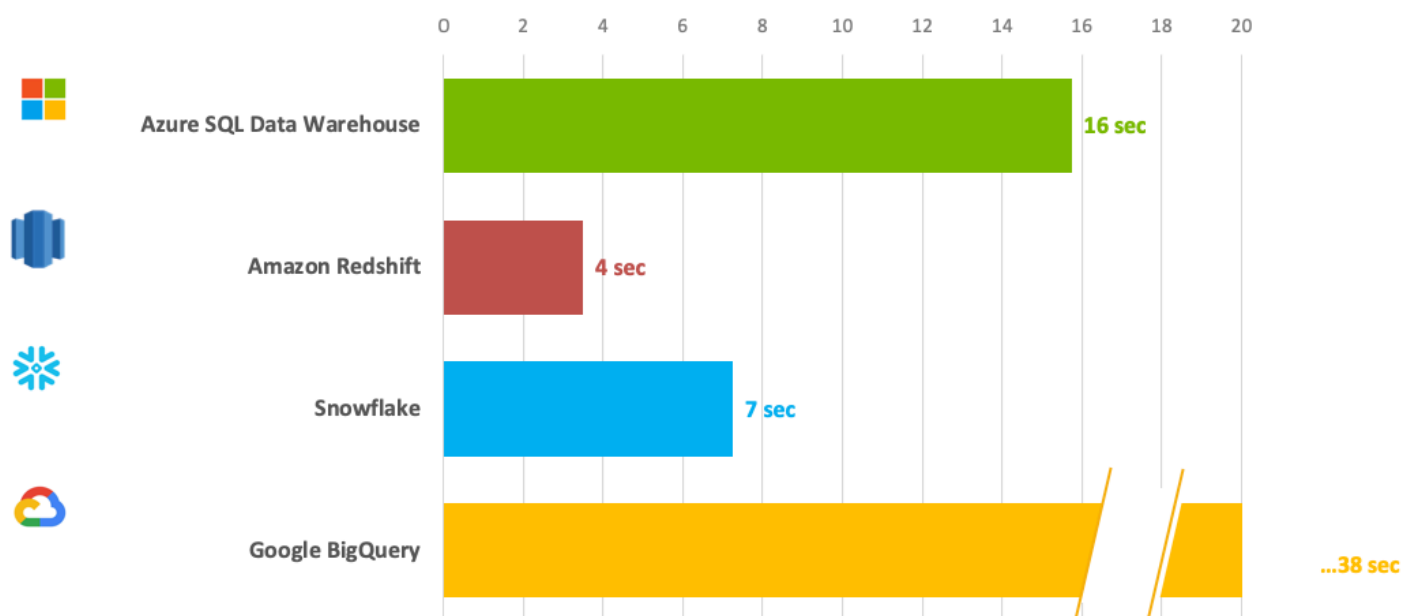
## GigaOM Analytic Field Test Query 29 Execution Times

*Query 30*

Query 30 identifies customers who have returned a percentage more than the average for their state. Identifying outlying customers like this is important in data-driven organizations, who must utilize all data to get an advantage. Analyzing customers who exhibit outlier behavior, and their data, are precursors to machine learning algorithms to detect patterns including fraudulent returns or overly unsatisfactory products.

Query 30 has some complex subqueries, which are common in enterprise queries.

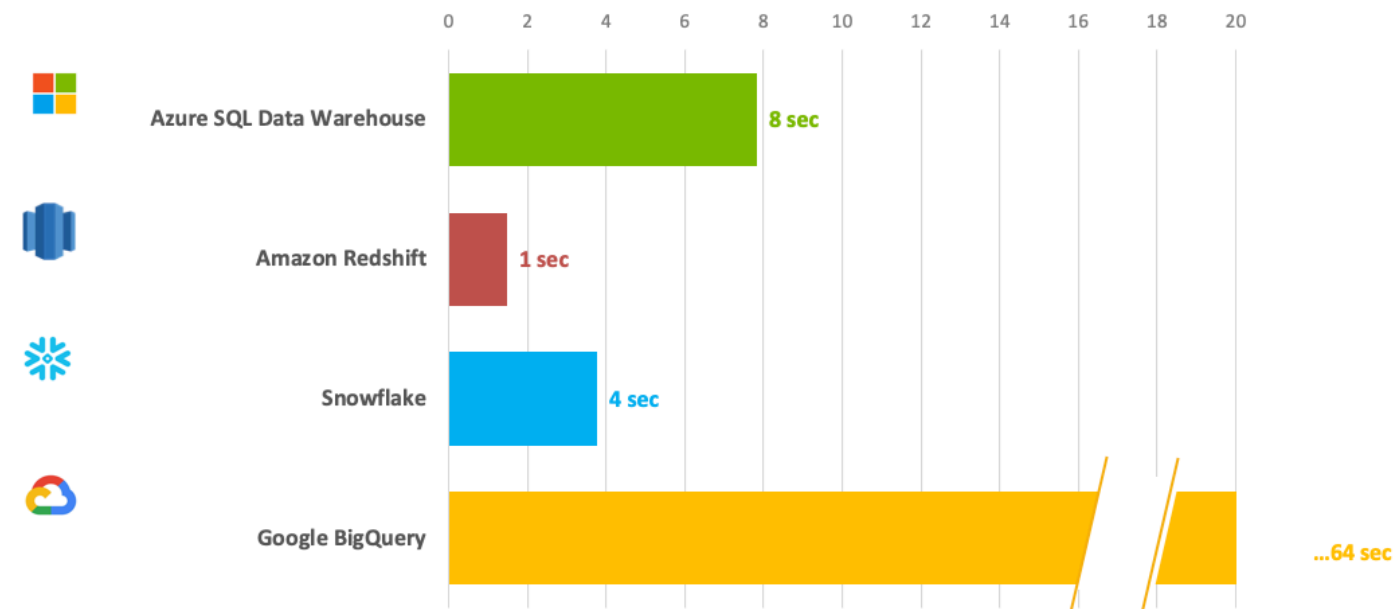
## GigaOM Analytic Field Test Query 30 Execution Times

*Query 39a*

Query 39a calculates the coefficient of variation and mean of every item and warehouse for two consecutive months. This moving average inventory analysis would be very useful to supply chain managers, in coordination with suppliers and procurement teams, to ensure the proper balance of inventory flows.

Query 39a has a join, subquery, and case. Other than the calculations, it is a straightforward query, nicely representative of common enterprise queries.

GigaOM Analytic Field Test Query 39 Execution Times

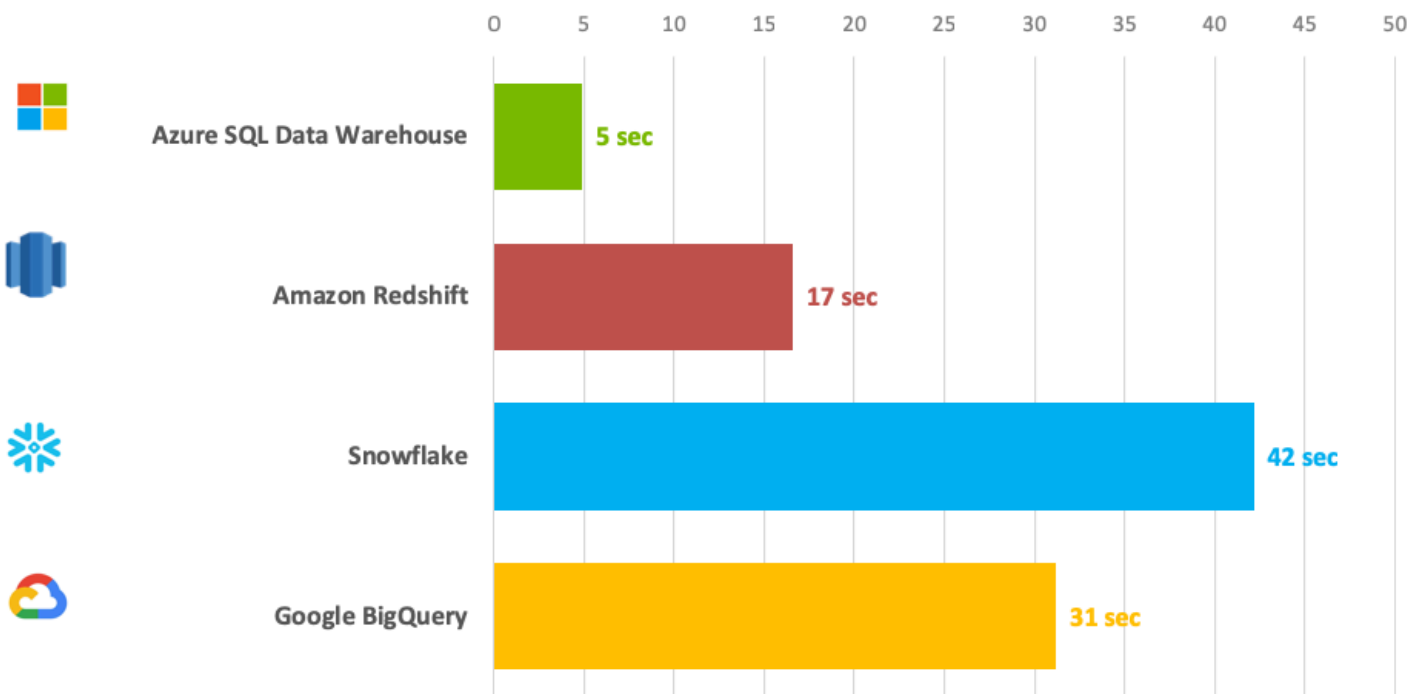


Query 44

Query 44 is looking for the best and worst performing items, a common query for an enterprise, but it is measured by net profit which adds complexity (as anyone who has ever attempted this calculation in a modern enterprise knows). These insights are critical for profit-and-loss financial analysts, as well as actuarial departments, who must forecast the best performing items and how current sales might impact the future bottom line and an upcoming stock market earnings report.

An enterprise would want to note the performance of nested subqueries, which this query features, and is why we find it is worth attention.

GigaOM Analytic Field Test Query 44 Execution Times

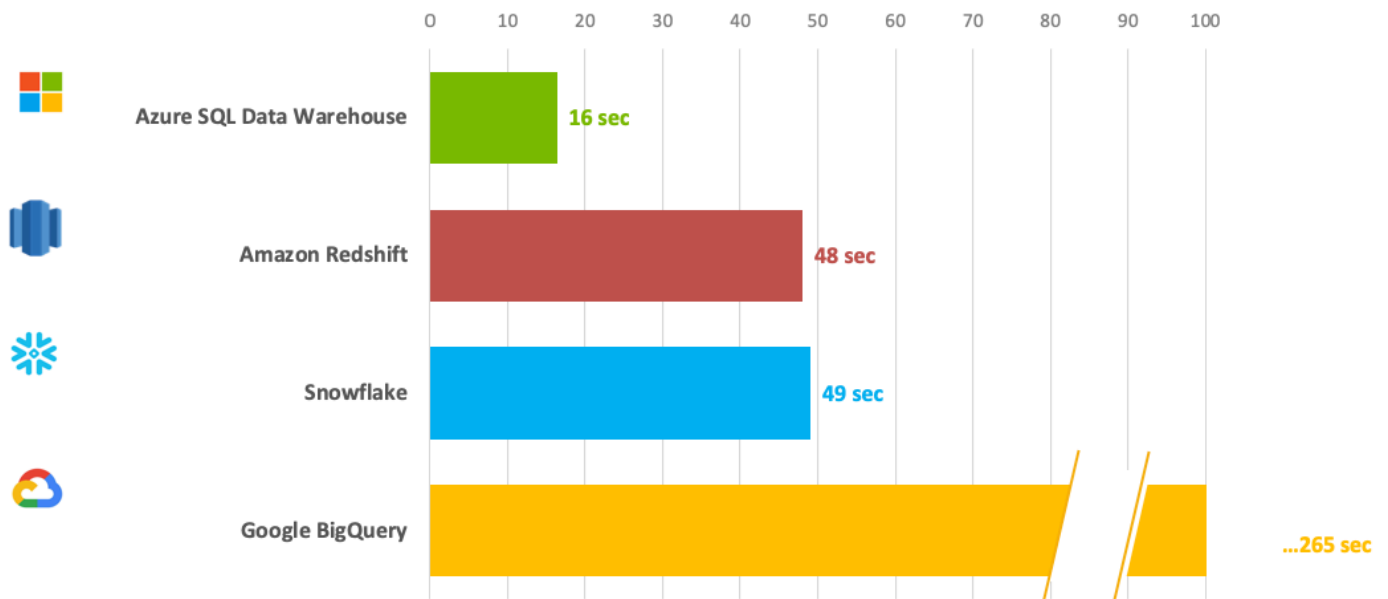


Query 47

Query 47 is looking for monthly sales that deviate by a certain percentage from the average monthly sales for the year, sorted by deviation and store. Performance management or performance-based incentive managers would use this information to track the changes over time, and store or sales managers might use it to measure the impact of new initiatives versus current efforts.

Query 47 is a trending and ranking query that is similar to what many enterprises are doing in determining points of attention that are based on business condition change. It is also a mid-tier query, where an enterprise would care about its performance.

## GigaOM Analytic Field Test Query 47 Execution Times

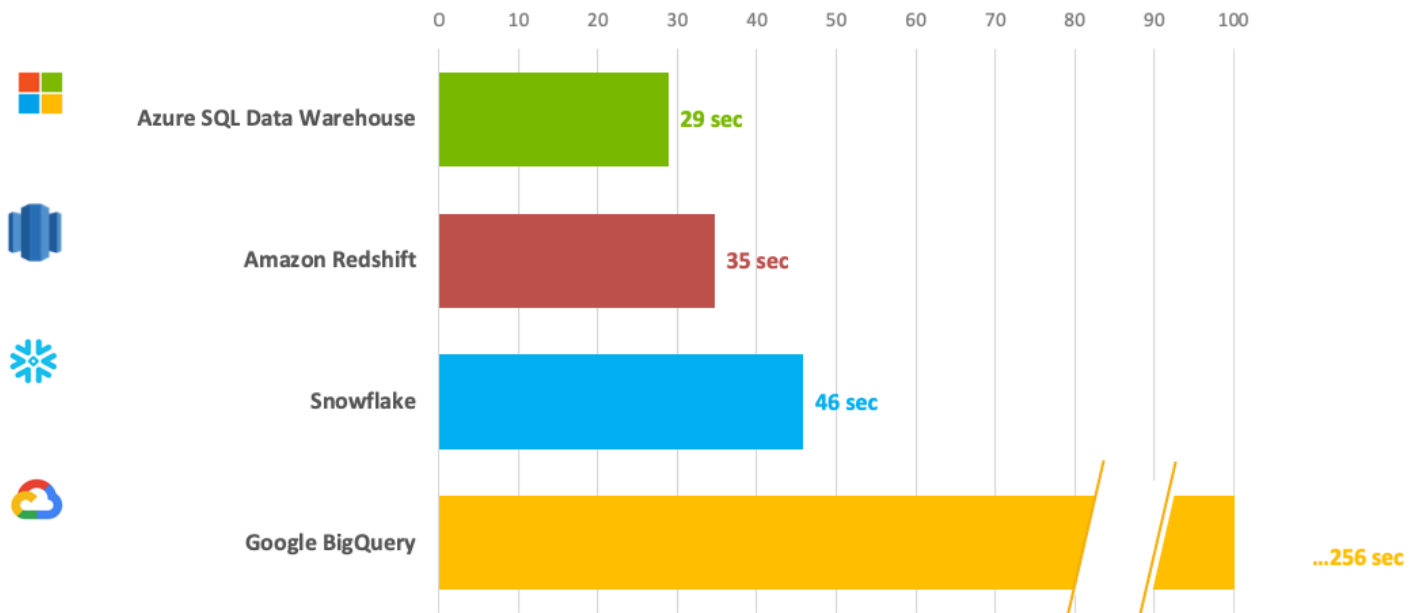
*Query 93*

Query 93 looks at customer purchases with returns backed out, where the returns are looked up in ticketing tables. This is an operational query that could be used by return departments to track returned items.

Query 93 is a simple query to code and is a very common query run in enterprises today.



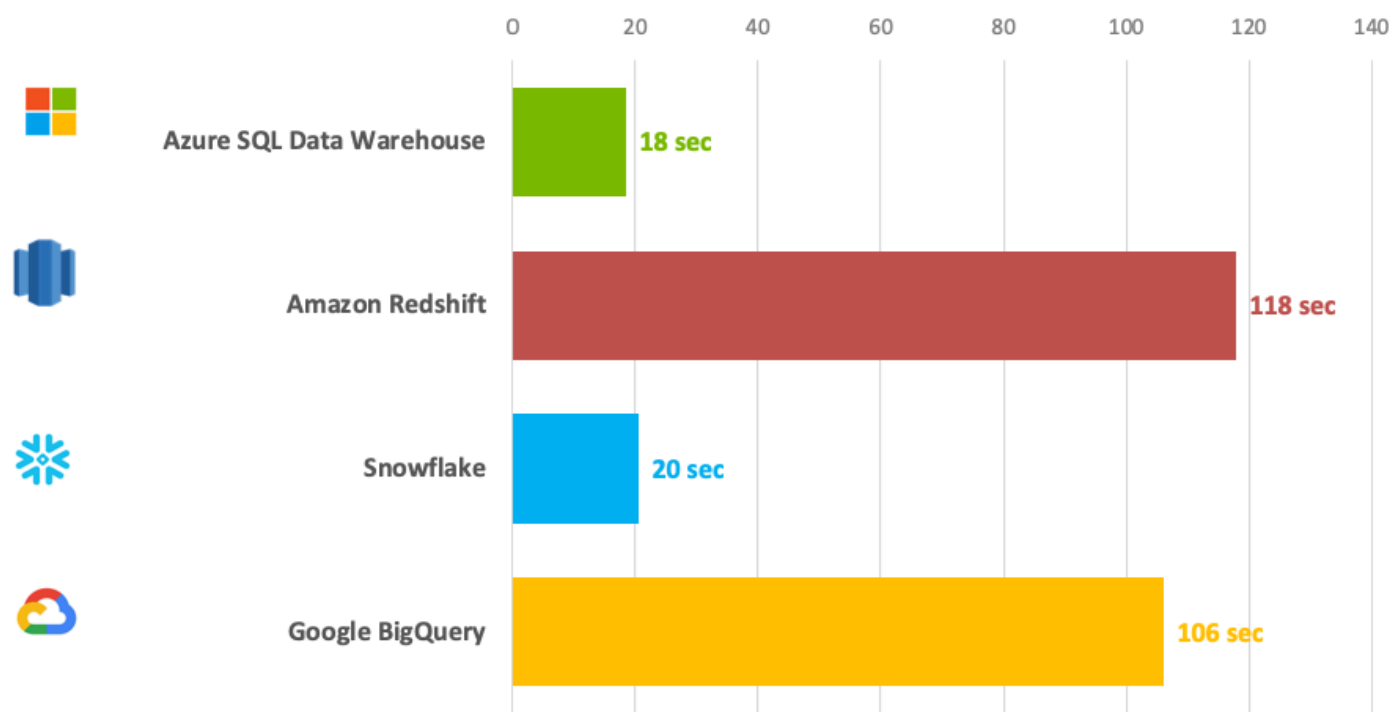
## GigaOM Analytic Field Test Query 93 Execution Times



## Query 94

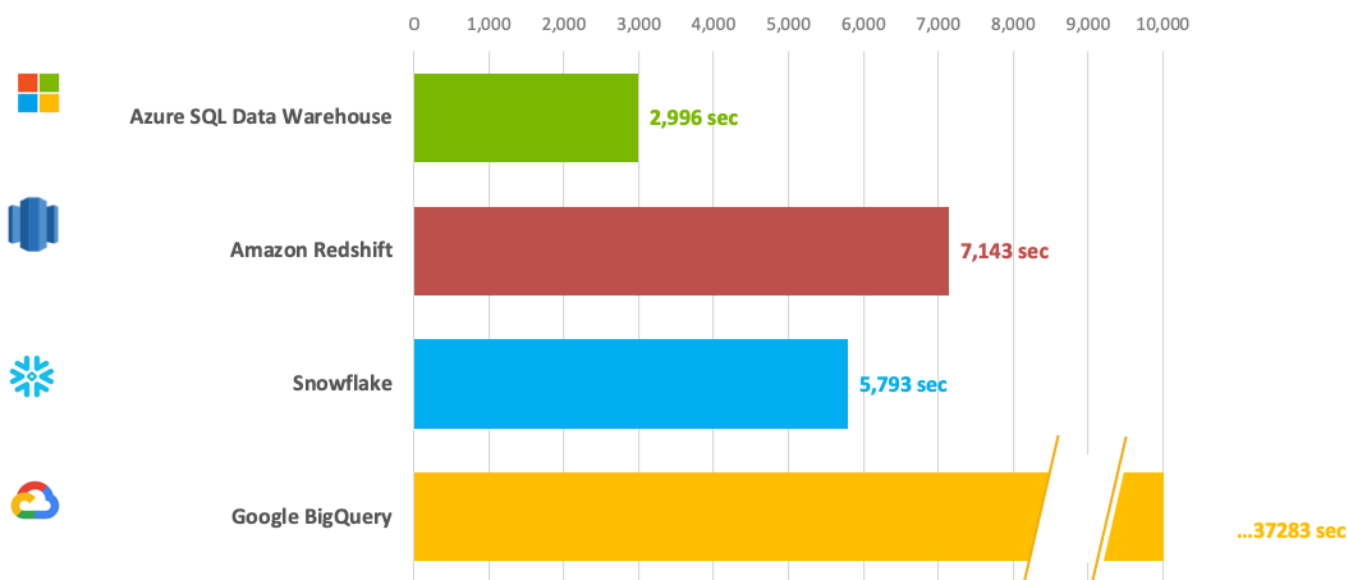
Query 94 is calculating order counts. Calculating order counts for a specific use is usually more complex than a simple count. The count includes web sales, total shipping cost, and net profit in a time window for customers in a given state from a named website for non-returned orders shipped from more than one warehouse. For companies offering free shipping incentives on their e-commerce sites, this type of analysis would be valuable in triangulating pricing and sales data to ensure the free shipping offer is both cost effective and incentivizing for customers.

GigaOM Analytic Field Test Query 94 Execution Times



Aggregate Results

As mentioned, the best times of each platform’s three power runs were then added together to gain the total aggregate execution time for the entire workload. The following chart shows the overall performance of each platform for the given workload, in terms of total time it took to execute all 103 queries in the GigaOm Analytic Field Test.



The next table presents the complete set of the best results of all query executions for each platform. The spark chart for each row provides a visual reference for the fastest and slowest times for each query. The following is a legend of the colors and headings used:

– Fastest Time

– Slowest Time

**DW** – Azure SQL Data Warehouse

**RS** – Amazon Redshift

**SF** – Snowflake

**BQ** – Google BigQuery

Table 12. Results for All Queries

Query	SQL DW	Redshift	Snowflake	BigQuery	DW	RS	SF	BQ
1	6.580	6.189	5.253	34.285	—	—	—	—
2	42.587	10.818	38.983	161.081	—	—	—	—
3	3.360	5.118	12.559	35.864	—	—	—	—
4	129.220	151.706	138.531	1,346.933	—	—	—	—
5	32.990	110.544	59.957	340.166	—	—	—	—
6	13.237	6.617	30.881	480.618	—	—	—	—
7	14.056	13.934	59.343	66.102	—	—	—	—
8	16.120	7.058	18.829	41.122	—	—	—	—
9	21.163	61.230	127.016	88.784	—	—	—	—
10	15.423	64.724	26.863	59.197	—	—	—	—
11	72.667	71.718	83.539	806.247	—	—	—	—
12	5.220	2.318	6.101	19.462	—	—	—	—
13	13.567	21.465	84.656	164.975	—	—	—	—
14	33.963	328.102	183.009	407.651	—	—	—	—
14b	45.330	334.808	198.516	2,502.835	—	—	—	—
15	13.837	3.991	5.611	121.217	—	—	—	—
16	31.563	784.880	7.572	128.854	—	—	—	—
17	25.990	16.013	40.055	804.656	—	—	—	—
18	21.243	46.531	13.960	803.040	—	—	—	—
19	11.163	10.401	34.836	31.275	—	—	—	—
20	6.067	3.589	2.883	37.562	—	—	—	—
21	3.890	0.485	3.496	7.310	—	—	—	—
22	4.820	3.050	7.750	24.510	—	—	—	—
23	78.133	238.914	244.657	2,429.327	—	—	—	—
23b	74.137	238.130	247.550	2,086.158	—	—	—	—
24	19.190	46.293	85.579	542.550	—	—	—	—
24b	27.020	43.236	89.010	565.740	—	—	—	—
25	22.493	21.702	56.871	733.442	—	—	—	—
49	30.980	33.545	79.611	127.813	—	—	—	—
50	26.424	67.223	43.481	254.204	—	—	—	—
51	19.227	32.081	31.726	415.552	—	—	—	—
52	3.010	5.968	13.612	23.971	—	—	—	—
53	3.987	7.468	15.031	30.939	—	—	—	—
54	15.453	14.750	38.669	403.032	—	—	—	—
55	2.920	6.040	13.373	24.797	—	—	—	—
56	12.080	14.789	31.689	41.448	—	—	—	—
57	20.243	24.508	20.175	96.284	—	—	—	—
58	13.920	11.455	25.054	110.161	—	—	—	—
59	84.123	47.015	64.891	200.279	—	—	—	—
60	13.314	17.884	32.738	58.777	—	—	—	—
61	14.640	21.242	81.391	76.748	—	—	—	—
62	6.360	3.840	10.684	17.466	—	—	—	—
63	4.117	7.668	23.836	28.965	—	—	—	—
64	85.070	248.720	176.065	1,917.566	—	—	—	—
65	22.410	22.615	54.556	137.121	—	—	—	—
66	9.600	14.098	21.730	39.510	—	—	—	—
67	651.083	464.580	641.240	2,291.993	—	—	—	—
68	20.517	16.818	71.217	72.956	—	—	—	—
69	16.940	43.005	22.695	51.175	—	—	—	—
70	10.690	32.247	29.331	115.037	—	—	—	—
71	15.770	21.409	73.426	48.157	—	—	—	—
72	39.396	1,436.155	22.067	3,600.122	—	—	—	—
73	12.313	6.647	36.721	45.325	—	—	—	—
74	86.660	60.951	63.694	844.625	—	—	—	—
75	100.947	79.675	104.901	530.170	—	—	—	—
76	16.187	12.556	66.669	81.148	—	—	—	—

26	9.690	6.987	7.670	29.952	
27	9.800	44.325	57.998	129.586	
28	23.860	42.985	105.227	51.956	
29	32.887	21.237	59.693	764.899	
30	15.750	3.504	7.261	37.552	
31	32.827	38.903	40.461	98.771	
32	6.413	6.625	3.249	38.277	
33	14.300	16.534	46.862	135.034	
34	15.923	14.113	62.151	33.873	
35	26.960	180.582	20.444	125.728	
36	6.327	37.580	36.046	117.064	
37	3.453	1.763	7.597	20.802	
38	51.353	32.868	46.552	78.514	
39	7.837	1.480	3.779	64.299	
39b	9.774	1.586	5.822	66.600	
40	13.930	44.008	4.204	160.672	
41	2.110	0.386	0.414	1.218	
42	2.663	6.401	20.497	27.365	
43	11.994	7.696	14.561	25.394	
44	4.913	16.603	42.177	31.217	
45	14.867	3.669	10.644	93.772	
46	14.327	15.071	58.279	49.127	
47	16.357	48.060	49.092	265.334	
48	8.134	13.048	41.527	117.884	

77	16.440	21.846	57.154	227.059	
78	96.730	442.275	503.549	1,871.047	
79	16.473	20.711	74.259	55.035	
80	46.560	232.172	99.704	1,969.868	
81	20.970	5.074	8.323	40.865	
82	4.580	2.951	12.227	22.837	
83	11.747	2.228	4.381	14.643	
84	4.760	1.215	4.139	6.042	
85	18.257	13.314	14.097	173.184	
86	4.220	6.991	8.454	33.445	
87	55.556	34.499	40.164	1.266	
88	22.510	35.104	80.431	118.398	
89	6.520	8.791	27.493	33.241	
90	6.493	3.042	11.587	16.891	
91	7.980	1.258	2.796	11.414	
92	4.950	4.432	11.603	29.035	
93	28.863	34.647	45.841	255.616	
94	18.466	117.805	20.459	106.034	
95	28.330	72.557	88.040	2,944.355	
96	3.214	5.156	26.437	24.398	
97	36.087	27.399	39.216	119.807	
98	6.167	8.770	21.643	59.361	
99	10.240	6.475	14.694	27.683	
TOTAL	2,995.972	7,143.242	5,793.037	37,282.718	

## 4 Price Per Performance

The price-performance metric is dollars per query-hour (\$/query-hour). This is defined as the normalized cost of running the GigaOm Analytic Field Test workload on each of the cloud platforms. It was calculated by multiplying the best on-demand rate (expressed in dollars) offered by the cloud platform vendor (at the time of testing) times the number of computation nodes used in the cluster, and then dividing this amount by the aggregate total of the best execution times for each query (expressed in hours).

To determine pricing, each platform has different options, which include reserved instances, annual contracts, and prepay. However, reserved instance pricing can only be procured with annual commitments, and is most cost-effective when paid in full upfront. Long commitments are out of scope for this field test, so we chose the lowest on-demand rate. Buyers should evaluate all of the pricing options, not just the ones presented in this report.

For [Azure SQL Data Warehouse](#), you pay for compute resources as a function of time. The hourly rate for SQL Data Warehouse varies slightly by region. We chose the lowest found hourly rate of \$6.04 per node for the DW15000C service level.<sup>4</sup> Also, we added the separate storage charge to store the 14TB of data (automatically compressed down from 30TB) at a rate of \$0.19 per TB per hour.

For [Amazon Redshift](#), you also pay for compute resources as a function of time. For price-performance, we used the lowest hourly rate we found for Redshift dc2.8xlarge instance type, \$4.80 per node.<sup>5</sup>

For [Snowflake](#), you pay for compute resources as a function of time — just like SQL Data Warehouse and Redshift. However, with Snowflake, you choose the hourly rate based on the enterprise features you need. For the lowest level of support, the Standard rate is \$2.00 per node, per hour. Features like multi-cluster (automatic scale out of additional cluster), enterprise-level security, and more support, are \$3.00 per node, per hour at the Enterprise level. There are higher levels. For price-performance, we used both Standard and Enterprise.<sup>6</sup>

With [Google BigQuery](#), the default option is to pay for bytes processed. During a single run of the GigaOm Analytic Field Test derived queries, we processed roughly 113TB of data. If we used the on-demand pricing, the \$5 per TB for BigQuery rate would have cost \$564. The second option is to pay a flat rate cost-per-hour. If we used the cost-per-hour from BigQuery flat rate, which was \$55 at the time of this report, the total workload, which ran for 10.4 hours, would have cost \$570. Most likely, anyone using large scale data warehousing would be on the flat-rate. We show both options.<sup>7</sup>

If you contiguously ran all 103 of these queries to completion of the set, the cost at an hourly basis is indicated in the chart below. Azure SQL Data Warehouse provided the best price-performance for the GigaOm analytic field test queries.

Price-Performance @ 30TB (\$ per Query, per Hour):

	SQL Data Warehouse	Redshift	Snowflake Standard	Snowflake Enterprise	BigQuery Flat Rate	BigQuery Per Byte
Instance Type	DW 15000C	dc2.8xlarge	3X-Large	3X-Large	Flat Rate	Per Byte
Nodes	30	30	64	64	N/A	N/A
Compute \$/node/hour	\$ 6.04	\$ 4.80	\$ 2.00	\$ 3.00		
<b>Compute \$/hour</b>	<b>\$ 181.20</b>	<b>\$ 144.00</b>	<b>\$ 128.00</b>	<b>\$ 192.00</b>	<b>\$ 55.00</b>	<b>\$ 5.00 / TB</b>
\$/hour storage	\$ 0.19					
Total Storage TB	14 TB					
<b>Total Storage \$/hour</b>	<b>\$ 2.66</b>					
<b>Price Basis: Total Execution Time (Sum of Best Times) or TB processed</b>	<b>2,996</b>	<b>7,143</b>	<b>5,793</b>	<b>5,793</b>	<b>37,283</b>	<b>113 TB</b>
<b>Price-Performance</b>	<b>\$ 153.01</b>	<b>\$ 285.73</b>	<b>\$ 205.97</b>	<b>\$ 308.96</b>	<b>\$ 569.60</b>	<b>\$ 1310.00</b>



<sup>4</sup>Azure SQL Data Warehouse pricing: <https://azure.microsoft.com/en-us/pricing/details/sql-data-warehouse/gen2/>

<sup>5</sup>Amazon Redshift pricing: <https://aws.amazon.com/redshift/pricing/>

<sup>6</sup>Snowflake pricing: <https://www.snowflake.com/pricing/>

<sup>7</sup>Google BigQuery pricing: <https://cloud.google.com/bigquery/pricing>

## 5 Summary

Cloud databases are a way for enterprises to avoid large capital expenditures, provision quickly, and provide performance at scale for advanced analytic queries. Relational databases with analytic capabilities continue to support the advanced analytic workloads of the organization with performance, scale, and concurrency. For the GigaOm Analytic Field Test, which contains a representative set of corporate-complex queries derived from the well-known TPC Benchmark™ DS (TPC-DS)<sup>8</sup> standard, Azure SQL Data Warehouse consistently outperformed the competition.

Overall, the benchmark results were insightful in revealing query execution performance and some of the differentiators between Azure SQL Data Warehouse, Snowflake Data Warehouse, Amazon Redshift, and Google BigQuery. Azure SQL Data Warehouse query response times on the 30TB GigaOm Analytic Field Test data set were twice as fast as Snowflake, approximately two and a half times the performance of Redshift, and 12 times faster than BigQuery overall.

In terms of price per performance, Azure SQL Data Warehouse ran the GigaOm Analytic Field Test queries 30% cheaper than Snowflake in terms of cost per query per hour, and was 1.9 times more cost-effective in terms of cost per query, per hour than Redshift. Azure SQL Data Warehouse also ran the GigaOm Analytic Field Test queries 3.7 times cheaper than Google BigQuery in terms of price per query, per hour.

Price and performance are critical points of interest when it comes to selecting an analytics platform, because they ultimately impact total cost of ownership, value, and user satisfaction. Our analysis reveals Azure SQL Data Warehouse to be very powerful and comparative in value.

<sup>8</sup>This was NOT an official TPC Benchmark™ DS (TPC-DS) benchmark. More can be learned about the TPC-DS benchmark at <http://www.tpc.org/tpcds/>.

## 6 In Closing

Price-performance is important, but it is only one criterion in selecting a data warehouse platform. There are numerous other factors to consider including ease of administration, integration with existing software and systems, workload management, user interface, scalability, vendor support, and reliability, among others.

This price-performance test is a single point-in-time check. It is our experience that performance changes over time and varies widely for different workloads. Also, a platform performance leader can hit up against the point of diminishing returns and viable contenders can quickly close the gap.

GigaOm runs all of its performance tests to strict ethical standards. The results of the report are the quantitative results of the application queries to the simulations described in the report. The report clearly defines the selected criteria and process used to establish the field test. The report also clearly states the data set sizes, the platforms, the queries, etc. used. The reader is left to determine for themselves how to utilize the information for their individual needs. The report does not make any claim regarding third-party certification, and presents the objective results received from the application of the process to the criteria as described in the report. The report strictly measures price-performance and does not purport to evaluate other factors that potential customers may find relevant when making a purchase decision.

This is a sponsored report. Microsoft chose the test, the Azure configuration, and the platforms evaluated. GigaOm chose the most compatible configurations for the other tested platforms, configured and ran the queries. Choosing compatible configurations is subject to judgment. We have attempted to describe our decisions in this paper.

In this writeup, all the information necessary is included to replicate this test. You are encouraged to compile your own representative queries, data sets, data sizes, and compatible configurations and test for yourself.



## 7 About the Analyst



An Ernst & Young Entrepreneur of the Year Finalist and frequent best practices judge, William is a former Fortune 50 technology executive and database engineer. He provides Enterprise clients with action plans, architectures, strategies, and technology tool selection to manage information.

William McKnight is an Analyst for Gigaom Research who takes corporate information and turns it into a bottom-line producing asset. He's worked with companies like Dong Energy, France Telecom, Pfizer, Samba Bank, ScotiaBank, Teva Pharmaceuticals and Verizon — Many of the Global 2000 — and many others. William focuses on delivering business value and solving business problems utilizing proven, streamlined approaches in information management.

He is a frequent international keynote speaker and trainer. William has taught at Santa Clara University, UC-Berkeley and UC-Santa Cruz.

## 8 About GigaOm Research

GigaOm provides technical, operational, and business advice for IT's strategic digital enterprise and business initiatives. Enterprise business leaders, CIOs, and technology organizations partner with GigaOm for practical, actionable, strategic, and visionary advice for modernizing and transforming their business. GigaOm's advice empowers enterprises to successfully compete in an increasingly complicated business atmosphere that requires a solid understanding of constantly changing customer demands.

GigaOm works directly with enterprises both inside and outside of the IT organization to apply proven research and methodologies designed to avoid pitfalls and roadblocks while balancing risk and innovation. Research methodologies include but are not limited to adoption and benchmarking surveys, use cases, interviews, ROI/TCO, market landscapes, strategic trends, and technical benchmarks. Our analysts possess 20+ years of experience advising a spectrum of clients from early adopters to mainstream enterprises.

GigaOm's perspective is that of the unbiased enterprise practitioner. Through this perspective, GigaOm connects with engaged and loyal subscribers on a deep and meaningful level. Find out more at [www.GigaOm.com/about](http://www.GigaOm.com/about)

## 9 About Microsoft

Microsoft offers Azure SQL Data Warehouse.

Learn more at <https://azure.microsoft.com/en-us/services/sql-data-warehouse/>.

© [Knowingly, Inc.](#) 2019. "Cloud Data Warehouse Performance Testing" is a trademark of [Knowingly, Inc.](#). For permission to reproduce this report, please contact [sales@gigaom.com](mailto:sales@gigaom.com).