

BW-EML SAP Standard Application Benchmark

Heiko Gerwens and Tobias Kutning^(✉)

SAP SE, Walldorf, Germany
tobas.kutning@sap.com

Abstract. The focus of this presentation is on the latest addition to the BW SAP Standard Application Benchmarks, the SAP Business Warehouse Enhanced Mixed Load (BW-EML) benchmark. The benchmark was developed as a modern successor to the previous SAP Business Warehouse benchmarks. With near real-time and ad hoc reporting capabilities on big data volumes the BW-EML benchmarks matches the demands of modern business warehouse customers. The development of the benchmark faced the challenge of two contradicting goals. On the one hand the reproducibility of benchmark results is a key requirement. On the other hand the variability in the query workload was necessary to reflect the requirements for ad hoc reporting. The presentation will give an insight to how these conflicting goals could be reached with the BW-EML benchmark.

1 SAP Standard Application Benchmarks

SAP SE is the world's leading provider of business software¹. It delivers a comprehensive range of software products and services to its customers: Companies from all types of industries, ranging from small businesses to large, multinational enterprises engaged in global markets.

For more than 20 years now SAP and its hardware and technology partners have developed and run benchmarks to test the performance and scalability of both hardware and the SAP solutions running on the hardware.

The first benchmark was certified in 1995. Since then the SAP Standard Application Benchmarks have become some of the most important benchmarks in the industry. Especially the SAP SD standard application benchmark² can be named as an example.

The goal of the SAP Standard Application Benchmarks is to represent SAP business applications as realistic as possible. A close to real life workload is one of the key elements of SAP benchmarks.

The performance of SAP components and business scenarios is assessed by the benchmarks and at the same time input for the sizing procedures is generated. Performance in SAP Standard Application Benchmarks is determined by throughput numbers and system response times. The throughput numbers are defined in business

¹ <http://www.sap.com/about.html>, <http://www.softwaretop100.org/enterprise-software-top-10>.

² http://global1.sap.com/campaigns/benchmark/appbm_sd.epx.

application terms. For the SAP SD benchmark this would be for example fully processed order line items per hour. This business application throughput is then mapped onto the hardware resource consumption for example CPU and memory. The unit for the measurement of CPU power is SAP Application Performance Standard (SAPS)³. SAPS is a hardware independent measurement unit for the processing power of any system configuration. 100 SAPS is defined as 2,000 fully processed order lines items per hour which equals 2,000 postings or 2,400 SAP transactions.

By ensuring that the benchmark results are free of artifacts, the availability of the tested hardware and software combinations for productive use by customers SAP and its partners ensure that the benchmarks are highly business relevant for customers. The benchmarking methods are monitored and approved by the SAP Benchmark Council, which consists of SAP and its hard- and software partners.

Customers, SAP partners and also SAP profit from the SAP Standard Application Benchmarks. For customers the benchmarks can serve as a proof of concept illustrating the scalability and manageability of large SAP installations from both hard- and software point of view. Customers are also able to compare different hard- and software combinations and releases using the benchmark data.

Partners of SAP are enabled to optimize their technology for SAP applications with the help of the SAP Standard Application Benchmarks. The benchmarks are also used to prove the scalability of hardware components and of course the marketing departments do use the benchmark results to support sales.

For SAP the main benefits of the SAP Standard Application benchmarks are quality assurance, analysis of system configurations and parameter settings and the verification of sizing recommendations.⁴

2 Benchmark Simulation with the SAP Benchmark Tools

The SAP benchmark toolkit consists of collection of Perl scripts and C programs and is available for Windows, Linux and UNIX. In addition there are pre-configured SAP system environments available containing all necessary business data to execute the different benchmarks.

In the area of scripting the benchmark tools do allow the recording and playback of DIAG (the protocol used by the SAP frontend software SAP GUI) and HTTP protocol user interactions. The benchmark scripts can simulate a configurable number of system users executing a pre-defined series of transactions in the SAP system. The tools also take care of collecting all the necessary monitoring data from the collection tools built into the SAP System and also directly from the operating system. To check that the benchmark runs were free of errors the benchmark tools also collect the data returned for each system interaction of every user.

The Multi-User Simulation in SAP Standard Application Benchmarks can be divided in 3 sections Fig. 1 – the ramp up phase, the high load phase and the ramp-down

³ <http://global1.sap.com/campaigns/benchmark/measuring.epx>.

⁴ http://global1.sap.com/campaigns/benchmark/bob_overview.epx.

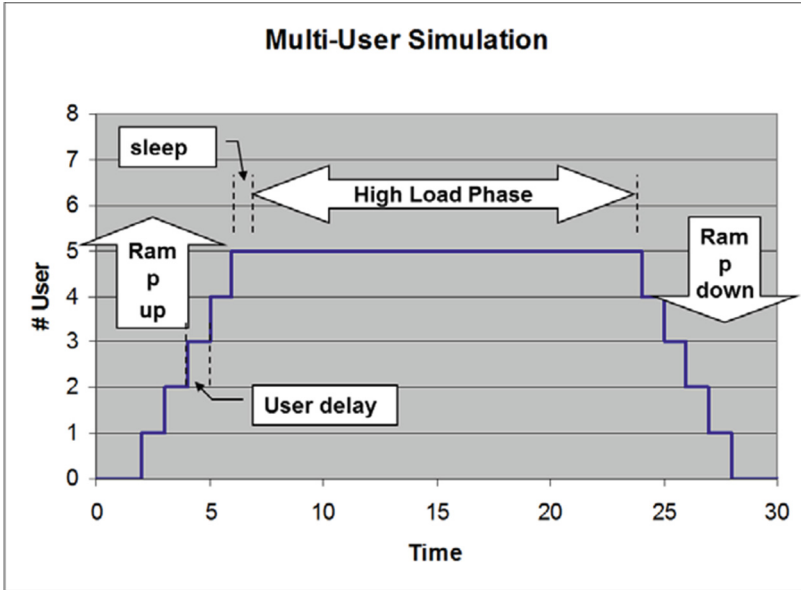


Fig. 1. Multi-user simulation in a SAP standard application benchmark

phase. In the ramp up phase all the configured users are logging into the system and start loops of their preconfigured steps of transactions. The speed with which users are logging onto the system can be adjusted using a user logon delay parameter. After all users are logged into the system the next phase called the high load phase starts. This is the phase where all configured users are logged into the system and perform the workload. For the SD benchmark the high load phase needs to be at least 15 min and for the BW-EML benchmark at least 1 h.⁵

3 SAP Business Warehouse Architecture

In order to understand the SAP terms used in the following discussion of the BW-EML benchmark let us take a look at some of the most important parts of the SAP BW architecture (Fig. 2).

3.1 InfoProvider

An InfoProvider is an object for which queries can be created or executed. InfoProviders are the objects or views that are relevant for reporting. The following types of InfoProviders⁶ used in the BW-EML benchmark.

⁵ http://global1.sap.com/campaigns/benchmark/bob_run.epx.

⁶ http://help.sap.com/saphelp_sem60/helpdata/en/4d/c3cd3a9ac2cc6ce10000000a114084/content.htm?frameset=/en/8d/2b4e3cb7f4d83ee10000000a114084/frameset.htm¤t_toc=/en/a3/fe1140d72dc442e10000000a1550b0/plain.htm&node_id=85&show_children=false.

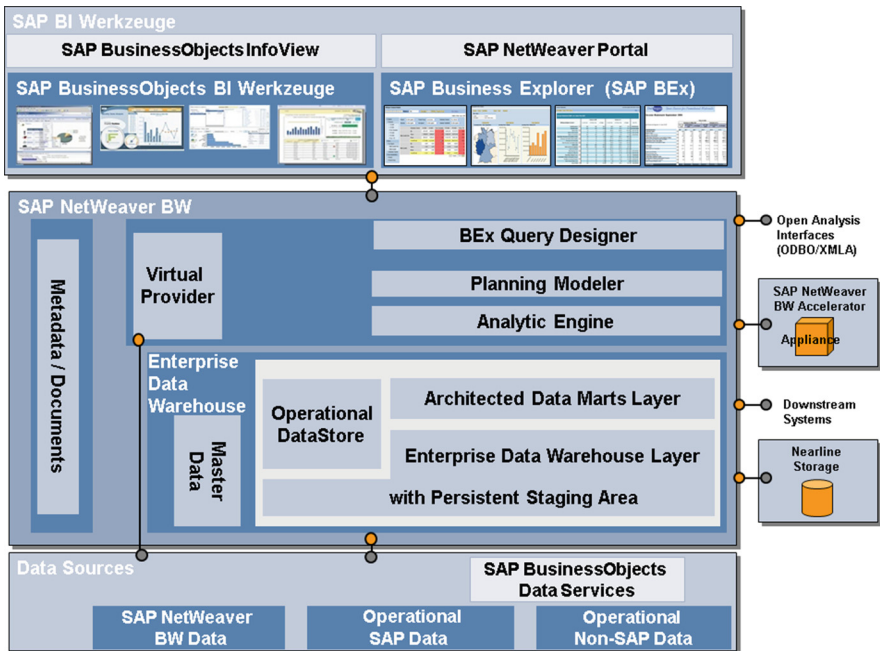


Fig. 2. SAP BW architecture overview (source: http://help.sap.com/erp2005_ehp_07/helpdata/en/47/5fa4468d0268b4e10000000a42189b/content.htm?frameset=en/46/8c6361e4c70ad3e10000000a11466f/frameset.htm)

3.2 InfoCube

An InfoCube is a fact table and its associated dimension tables in the star schema.

3.3 DSO

DSO is the short version of DataStore Object.

A DataStore object serves as a storage location for consolidated and cleansed transaction data or master data on a document (atomic) level. This data can be evaluated using queries.

A DataStore object contains key fields (such as document number, document item) and data fields that, in addition to key figures, can also contain character fields (such as order status, customer).

In contrast to InfoCubes, the data in DataStore objects is stored in transparent, flat database tables. Fact tables or dimension tables are not created.

The DSO has typically been used as the store for incoming data which is further processed into the reporting InfoCubes. Therefore DSOs are the core building element of a layered scalable warehousing architecture (LSA). However with new technology many reporting scenarios can be now directly implemented on top of DSOs with good

performance making InfoCubes completely obsolete in such cases (see also below in chapter “TCO reduction”).

3.4 Multi-cube/Multi-provider

A multi-provider is a union of basic InfoProviders. The multi-provider itself does not contain any data; rather, data reside in the basic InfoCubes or DSOs. To a user, the multi-provider is similar to a basic InfoProvider. When creating a query, the user can select characteristics and key figures from different basic InfoProviders.⁷

4 The BW-EML Standard Application Benchmark

Among the SAP Standard Application Benchmarks the topic of analyzing large data volumes is also present for a long time now. Customers interested in information about performance and scalability of their SAP Business Warehouse implementations, initially turned to the BW benchmark where first results have been published 2003. This benchmark, like all SAP Standard Application Benchmarks was designed to represent relevant, real-life scenarios involving various SAP business applications to help customers find appropriate hardware configurations. Like all SAP standard application benchmark the execution just requires the complete SAP BW software stack and an additional benchmark package to run this benchmark. Because of the abstraction of the SAP stack of the DB provider and hardware this benchmark is open for many platforms. The first version of the BW benchmark consists of two separate phases one for uploading data and a separate phase for query processing. Both benchmark phases had individual metrics. For the load phase the throughput in terms of number of rows loaded per hour either to Info Cubes or Operational Data Store (ODS) was published, whereas for the query phase the number of query navigation steps per hour was mentioned. While initially this benchmark provided valuable results the need of more instant availability of business data in many use cases was recognized. In 2008 this lead to the next step of the evolution of SAP’s benchmarks for the SAP Business Warehouse, the mixed load (BI-MXL) benchmark, which integrates the data upload and the query work load into a single benchmark phase. In this benchmark only the query navigation steps per hour were published. The data upload ran just as an additional background work load. The amount of data being uploaded during the benchmark phase of 1 h – 1/1000 of the original data in the system – represents the capability of the system to manage the near real time upload of data from the source system while handling heavy query work load.

With the changing demands of business decision making, like the instant availability of the latest business data and ad hoc query capability, the aspects that customers looked at when considering which configuration was the right fit changed as well. There was a clear need for a new benchmark type that covered these new

⁷ <http://wiki.scn.sap.com/wiki/display/BI/BW+Glossary,+Definitions+and+Explanations>.

demands. SAP decided 2012 to develop a new set of metrics: the BW enhanced mixed load (BW-EML) benchmark.⁸

Today, when it comes to decision making systems, customers are looking for ones that support the following:

4.1 Near Real-Time Reporting Capabilities

To make informed decisions in a fast-paced world, the ability to get instant information from analytical business applications is crucial. Not only do businesses require quick access to information, they also need this information to include up-to-the minute details. Smart meter analytics and trade promotion management are just two examples of business processes that rely on near real-time reporting. To reflect this need we increased the frequency of the data upload compared to the BI MXL benchmark. Instead of having a total of three upload phases - one every 20 min, we increased this to a total of 12 delta uploads - one every 5 min.

4.2 Ad Hoc Reporting Capabilities

Data volumes in enterprise data warehouses have grown significantly over the past few years due to increased complexity of business data models and the level of detail captured in data. Sheer data volumes and the demand for unforeseeable navigation patterns make it impossible to use standard techniques like pre-aggregation of data for speeding up query response times. Modern analytical applications must allow users to navigate instantly through these huge amounts of data by providing extensive slicing and dicing functionality. The new benchmark reflects this need by using randomized sets of input parameters for the initial report execution and randomized drill down and filtering dimensions during query navigation steps.

Another alternative approach to show the ad hoc reporting capabilities would be to use more static queries and to combine this with disallowing any technique to pre-calculate and store final or intermediate query results. We decided explicitly not to use such an approach for several reasons.

Excluding specific technologies always has the big problem to define exactly what has to be excluded. There would always be the tendency to invent technologies which achieve similar results, but circumvent the definitions of things that are excluded.

Technology advances can lead to the fact, that the excluded technologies can still be applied in a beneficial way for the given use case. This might lead to an artificial constraint, so that not the best possible technology is used.

Verification of the benchmark results would need a procedure to verify, if a technique has been used which is not allowed. This would imply a deep knowledge of the used technologies which sometimes can even collide with the interest of the technology provider to protect their IP.

⁸ http://global1.sap.com/campaigns/benchmark/appbm_bweml.epx.

Bottom line a benchmark should always test the capabilities of a system from the system usage point of view. How and with which technology a specific capability is achieved should not be important.

4.3 TCO Reduction

Since data warehouses can contain hundreds of terabytes of data, it is crucial to minimize data redundancy, while at the same time maintaining layered data models. With SAP NetWeaver Business Warehouse 7.30, it is possible to run reports directly on DataStore Objects (DSOs), which helps reduce TCO by saving precious storage space. DSOs are the core building element of a layered scalable warehousing architecture (LSA). Since reports can now analyze data in DSOs as fast as in multidimensional InfoCube data structures, InfoCubes have become completely obsolete in many reporting scenarios. To prove this capability we implemented the same set of queries with the same flexibility of query navigation steps on both – Info Cubes and DSOs.

4.4 The Main Differences Between BI-MXL and BW-EML

The Loading of delta requests happens every 20 min in BI-MXL. In BW-EML this timeframe was reduced to 5 min.

In BI-MXL there are pre-defined navigation paths and drilldowns only with the same characteristics set in display. In the BW-EML on the other hand ad hoc reporting requirements are verified by using randomized navigation paths and changes in the displayed characteristics.

All benchmark queries in the BI-MXL benchmark are defined on InfoCubes. The BW-EML benchmark uses DataStore Objects (flat table) and InfoCubes (star schema like data model) for reporting for a reduction of TCO.

The single metric of query navigation steps as well as the total data volume for the delta upload was kept like in the BW-MXL.

4.5 The BW-EML Benchmark in Detail

The BW-EML benchmark, like its predecessor, simulates a mix of multi-user reporting workload and the loading of delta data into the database simultaneously with user queries. Let's drill down further into the details of the benchmark definition.

4.6 Data Model

Even though in many cases InfoCubes do not provide a major benefit we like to insure that the database being benchmarked can efficiently use both InfoCubes and DSOs for reporting. Therefore the BW-EML benchmark's data model consists of three InfoCubes and seven DSOs, each of which contain the data of one specific year. The three InfoCubes contain the same data (from the last three years) as the corresponding DSOs.

Both types of objects consist of the same set of fields. The InfoCubes come with a full set of 16 dimensions, which comprise a total of 63 characteristics, with cardinalities of up to one million different values and one complex hierarchy. To simulate typical customer data models, the InfoCube is made up of 30 different key figures, including those that require exception aggregation. In the data model of the DSOs, the high cardinality characteristics are defined as key members, while other characteristics are modeled as part of the data members.

4.7 Scaling of Data Volumes

To test hardware configurations of various sizes, the BW-EML benchmark can be executed with different data volumes. The smallest configuration defined in the benchmark rules starts with an initial load of 500 million records (50 million records for each InfoCube and DSO). The records are loaded from ASCII flat files with a total record length of 873 bytes each. Larger volume configurations of the EML Benchmark include initial load volumes of one billion, two billion or even more records.

In each of the mentioned configurations, the total number of records that need to be loaded, in addition to the initial load during the benchmark run, is one thousandth of the initial amount of records. The high load phase of the benchmark must run for at least one hour. During this time, the delta data must be loaded in intervals of five minutes. The same number of records must be loaded in each InfoCube and DSO.

4.8 Query Model

For the BW-EML benchmark, eight reports have been defined on two MultiProviders – one for the three InfoCubes, and another for the seven DSOs. The respective reports on both MultiProviders are identical. This leads to two sets of four reports each. The four reports are categorized as follows:

- Report Q001: Customer-based reporting
- Report Q002: Material-based reporting
- Report Q003: Sales area-based reporting
- Report Q004: Price comparison reporting

During the report execution data is randomly selected for one particular year, and by that implicitly picking the InfoCube or DSO that contains the data. Up to six further navigation steps – filter and drill down steps with randomly chosen filter criteria and drill down dimension - are performed within each report, each of which results in an individual query and database access point.

Although the first three reports share similar navigation patterns, the filter and drill-down operations are randomized to address the demand for ad hoc types of queries. To make sure that the benchmark accesses different partitions of data, random choice of filter characteristics and the corresponding parameter values are used. Additionally, a random choice of characteristics for drilldowns or other slice-and-dice operations

ensures that a huge number of different characteristic combinations are covered in a multi-user reporting scenario.

4.9 Multi-user Workload

A script controls the multi-user reporting workload in the BW-EML benchmark. The script starts for each user a simulated HTTP client which serves as a simulated frontend. By this not only the data base is tested but the complete backend stack including the SAP BW OLAP engine and the HTML rendering. With the script, the number of simulated users can be defined. By this the intended query throughput can be easily scaled through an increased number of simulated users. Each simulated user logs on to the system and then executes a sequence of four reports on InfoCubes and four reports on ODS objects including the ad hoc navigation steps consecutively, resulting in a total of 40 ad hoc navigation steps per user loop. After finishing all the predefined steps, the user logs off and then starts the next loop with a new log on. Simulated users are ramped up at a pace of one user logon per second. Once all configured users are running the high load phase starts and the benchmark control environment automatically starts a process chain that controls the delta load, which is scheduled every five minutes. The high load phase runs at least one hour. Within that hour each user runs multiple loops of the same sequence of 40 navigation steps. At the same time 12 delta uploads are started.

After a high load phase of at least one hour, the simulated users are ramped down, and the delta loads finish. A control program checks if the correct number of records has been uploaded during the benchmark run and if the uploaded records are visible in the report results. The essential key figure that is reported for a benchmark run is the number of ad hoc navigation steps per hour that the system executes successfully during the high load phase.

4.10 Verification of the Benchmark Results

The benchmark script environment automatically collects a large set of monitoring data available in the SAP system. This includes data like the number of table accesses to each individual table.

Specifically for the BW-EML (and BI-MXL) we collect statistical data for each individual query. This includes a total run time and the number of records processed from the Info Cubes or DSOs. This allows us to compare the query work load with previous runs to see if there are any unusual patterns visible. On top we implemented a test query which verifies after each delta upload, if the data has been loaded correctly and in time.

4.11 Key Challenge for the Design of the BW-EML Benchmark

One the one hand benchmarks need to provide highly reproducible results to be relevant, because otherwise results cannot be compared. On the other hand to reflect the

ad hoc query capabilities we decided to introduce a huge variability into the query work load, with a huge amount of possible different query navigations and drilldowns.

Two things are important to nevertheless achieve good reproducibility

1. Characteristics are grouped by their respective cardinalities, and only characteristics of the same cardinality are considered for a randomized operation. E.g. we have a group of high cardinality characteristics like customer and material with 1 million distinct values and groups of characteristics with 1000 like e.g. plant, with 100 like sales organization or with 10 distinct values like distribution channel. Even though the navigation path is chosen randomly, the data volume that needs to be processed is similar independent of the chosen navigation path.

Because of the data skew that is also included in the data generator the drill down to the most detailed level of data nevertheless leads to a significant different amount of processed data.

Data skew is introduced for characteristics which are obviously depended on each other in most cases. E.g. there is a different characteristic for “ship to party” and the “bill to party”. In real life in most cases they are identical. Here we simulated it, by setting the same value in 95 % of the generated records. Another type of data skew introduced was sparsity, so that only a small fraction of the records contained a different value and all other records either contained empty values or one specific default value.

This data skew led to the fact that the actual processed data varied from a few result records up to a couple of 100 thousand resulting records.

Even with these significant differences in the execution of equivalent steps – so same step number in different loops – we nevertheless could achieve well reproducible results. Actually here the second effect helped us.

2. Even for the smallest benchmark the total number of more than 60,000 queries executed during the benchmark run leads to a sufficient averaging of the individual queries so that the final benchmark result is well reproducible. We never did an in depth statistical analysis, but looking at the results of run times and records processed of individual queries in detail it's obvious that for each navigation step only a limited number of variations of an individual step seemed to be possible.

5 Results of the Benchmarks

So far the following 9 results have been published for the SAP BW-EML standard application benchmark Fig. 3.^{9,10}

The result of a series of identical performance test runs using the BW-EML workload was that the variability between the different runs is less than 1 %. The results were measured internally at SAP on the SAP HANA database Fig. 4.

⁹ <http://global.sap.com/solutions/benchmark/bweml-results.htm>.

¹⁰ http://global.sap.com/campaigns/benchmark/appbm_cloud.epx.

Certifi- cation	Number of ini- tial records	Ad-hoc navigation steps per hour	Scale- out	Nu mber of DB Servers	Number of Applica- tion Serv- ers
2013020	500,000,000	66,900	No	1	1
2014001	500,000,000	113,390	No	1	2
2014013	500,000,000	137,510	No	1	2
2012023	1,000,000,000	65,990	No	1	1
2013027	1,000,000,000	129,930	No	1	1
2014021	2,000,000,000	111,850	No	1	2
2014009	2,000,000,000	126,980	No	1	2
2014014	2,000,000,000	177,590	Yes	5	3
2013037	3,000,000,000	128,650	Yes	4	2

Fig. 3. Overview SAP BW-EML standard application benchmark results

Run Num- ber	Date	Time	Ad-hoc navigation steps per hour
1	02/22/13	09:04:29	62008
2	02/22/13	10:09:36	62311
3	02/22/13	11:14:15	62059
4	02/22/13	12:19:05	62412
5	02/22/13	13:24:12	62363

Fig. 4. Table of performance test results BW-EML

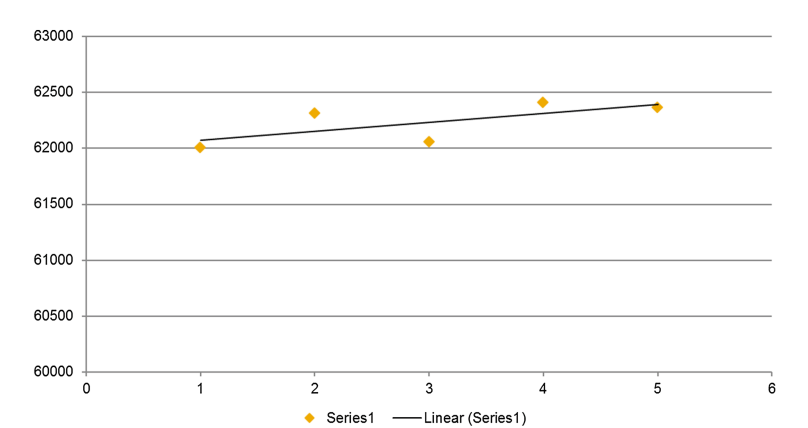


Fig. 5. Chart of performance test results BW-EML

What can be seen from the certified benchmark results and the internal testing performed so far is that the challenge of the benchmark was mastered. The BW-EML benchmark provides highly reproducible results and at the same time delivers the ad hoc query capabilities with a wide variability of the query work load and a huge amount of different query navigations and drill downs (Fig. 5).

6 Conclusion

Compared to earlier benchmarks the main difference is the usage of random query navigation steps and random choice of initial filter characteristics with a huge number of possible variations. This leads to a query work load where pre-calculated query results and aggregates are of limited value and so this benchmark stresses the ad hoc query capability to a large extend.

Also we have shown that it is possible to design a benchmark work load which on the one hand introduces a huge variability into the work load by choosing a data generator which generates random values including data skew, by using random query navigation steps and random query predicates and on the other hand achieve well reproducible results.

Big Data Benchmarking

5th International Workshop, WBDB 2014, Potsdam,
Germany, August 5-6- 2014, Revised Selected Papers
Rabl, T.; Sachs, K.; Poess, M.; Baru, C.; Hans-Arno, J.
(Eds.)

2015, X, 157 p. 54 illus., Softcover

ISBN: 978-3-319-20232-7