

Institute of Software Technology

University of Stuttgart
Universitätsstraße 38
D-70569 Stuttgart

Fachstudie Nr. 204

Evaluation of Analysis and Visualization Tools for Performance Data

Alexander Miller, Dominik Lekar

Course of Study: Softwaretechnik
Examiner: Prof. Dr. Lars Grunske
Supervisor: Dr.-Ing. André van Hoorn

Commenced: 02.06.2014

Completed: 02.12.2014

CR-Classification: C.4

Abstract

Observing and improving the performance of an application is an important task, since it will enhance the user experience and lower the running costs. Instead of doing this task each time manually, the market offers a wide choice of tools which allow the user to analyse and visualize performance data automatically.

The purpose of this study is to evaluate such tools and compare them with each other. As these tools differ in various aspects, the evaluation has to cover both technical and nominal criteria such as supported databases and operating systems, license properties, states of development, range of support, and the given underlying conditions. Additionally each tool will be tested for its capability in recreating report examples, provided by Capgemini Deutschland GmbH. This study aids in decision making by providing a comparison and helps the user to weigh up each individual aspect to one's personal needs. Through this study we try to publish a detailed comparison of current tools as well as a final recommendation based on our personal experience.

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Goals	1
1.3	Project-Specific Requirements	1
1.4	Document Structure	2
2	Research Procedure	3
2.1	Kick-Off Meeting	3
2.2	Data Analysis	3
2.3	Create Assessment Criteria	3
2.4	Create Report Examples	4
2.5	Market Overview and Tool Selection	4
2.6	Evaluation of Tools	4
2.7	End of Study	4
3	Market Overview	5
4	Evaluation Criteria	9
4.1	Criteria Catalogue	9
4.2	Exemplary Reports	12
5	Tool Evaluations	15
5.1	Excluded tools	15
5.2	SiSense	16
5.3	QlikView	25
5.4	Birst	33
5.5	Splunk	39
5.6	Tabular Synopsis	46
6	Final Recommendation	47

Introduction

1.1 Motivation

Volume increases, broadening the range of products, process changes and statutory requirements continuously lead to extensions and adjustments to a system. In general these changes lead to an increase in complexity and resource consumption. Therefore an early detection of anomalies and in terms of planning, a more precise prediction of the effects of a change in the consumption of resources, is important for a system's optimization.

1.2 Goals

The goal of this study is to evaluate tools which support the selection, processing and visualization of given data. The tools will especially facilitate the continuation of the consumption of statistics and the creation of ad-hoc analysis (selection, aggregation, visualization). The study will provide an overview of eligible tools, based on the captured requirements and technical conditions from an industrial settings provided by Capgemini Deutschland GmbH located in Stuttgart-Degerloch. The tools will be evaluated with a created list of criteria. Finally a reasonable recommendation for a tool (or possibly several alternatives) will be given.

1.3 Project-Specific Requirements

While the tools' various evaluations are to be undertaken with a universal approach in mind this study was nonetheless set under way to help facilitate a solution for a specific use case on the side of Capgemini. As of the writing of this text the analysis of Capgemini's products' performance data (for the specific project this study is meant to help with) is done in a somewhat ad-hoc manner, making use of manually created and collected SQL queries and Excel sheets. This study is meant to help find the right tool to simplify, accelerate and automate this process.

In order to facilitate our research we received an excerpt from Capgemini's database as well as exemplary descriptions of multiple reports the tested tools are to create from the

1. Introduction

data. We are unable to go into much detail explaining the exact structure and meaning of neither the data, nor the reports, as this data pertains to a major customer of Capgemini's who understandably wish to keep such information confidential. As such only this small, abstract description may be given: the data set contains multiple years' worth of data pertaining to multiple entities' performance under several metrics such as daily task completions or the amount of CPU-minutes necessary to complete said tasks.

As for the given report descriptions we have created mock-up reports (mostly via Microsoft Excel) to serve as a clear prescription of what functionality the tested tools are meant to provide. These example reports will be further expounded in chapter Evaluation Criteria.

A final note on the data shown in this study: as has already been mentioned Capgemini and their customers wish for the data we have been provided with to remain confidential. In order to comply with this restriction while still being able to show off exemplary charts we have made sure that any and all data displayed in later chapters has either been obfuscated or is being shown without any revealing context.

1.4 Document Structure

This document is structured as follows: After this introduction chapter 2 will provide an overview over the timetable this study followed as well as giving a short description of the work necessary to reach each specific milestone.

Chapter 3 shows a rough overview over the tools currently available on the market.

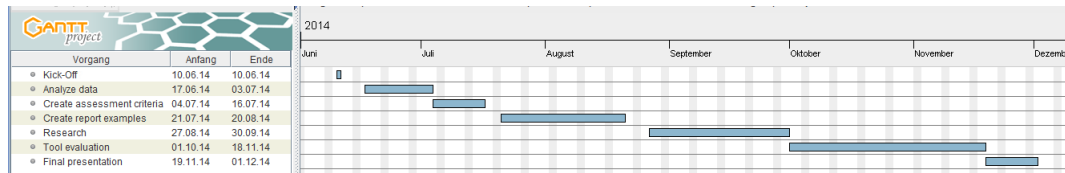
Chapter 4 contains a formalized list of criteria and their categories by which the various tools will be judged as well as examples of the types of reports the tools are supposed to be capable of generating.

Chapter 5 will provide a detailed analysis for each tool this study has investigated.

Chapter 6 will explain which tool(s) this study has found to be the best suited for the given tasks.

Research Procedure

This study was performed by a team of two members within a period of 6 months, starting on 02.06.14 and ending on 02.12.14. A meeting with the advisor was held every second week to discuss the current status of our work. From time to time these meetings were also used to present our currently finished milestones to representatives from Capgemini.



2.1 Kick-Off Meeting

The study started with the kick-off meeting in the office of Capgemini Deutschland GmbH in Stuttgart-Degerloch. After a short introduction of the company, we were made familiar with their project whose performance data needs to be analysed.

2.2 Data Analysis

After the kick-off meeting, Capgemini provided us an excerpt of their performance database in form of CSV files. We analysed the data scheme and imported the files into a locally run MySQL data base.

2.3 Create Assessment Criteria

In order to be able to formally compare tools we created an assessment criteria catalogue based on the statements of the representatives from Capgemini. Further important aspects were added for a more detailed comparison. The final criteria catalogue was presented and discussed in a meeting with the representatives of Capgemini.

2. Research Procedure

2.4 Create Report Examples

After creating the assessment criteria we created report examples based on a report file provided by Capgemini which contained a list of desired reports archetypes. These examples were to give a first impression of how the tools should be able to visualize the data and were discussed with the representatives of Capgemini.

2.5 Market Overview and Tool Selection

In this phase, available tools have been researched and listed. We started to gather basic information for each tool to categorize them and created a list of those tools suited for further research.

2.6 Evaluation of Tools

We installed each listed tool and tested it with the dataset provided by Capgemini. Each aspect of the criteria catalogue was tested and classified into one of the three possible outcomes.

2.7 End of Study

This study ended with the delivery of this document and the final presentation at Capgemini.

Market Overview

This chapter provides an overview of the currently available analysis and visualisation tools. In order to be suited, a tool at least needs to be able to recreate one of the reports. That means it has to be able to import the data, gain additional information by performing a query on it and display the result in any form of a chart.

Sisense

License: commercial
Webste: www.sisense.com



Qlikview

License: commercial
Webste: www.qlik.com



Birst

License: commercial, 30 day trial version
Webste: <http://www.birst.com/>



3. Market Overview

Splunk

License: commercial, free lite version
Webste: www.splunk.com



Spotfire

License: commercial, free trial
Webste: www.spotfire.tibco.com



SAS Visual Analytics

License: commercial, free demo version with predefined data
Webste: www.sas.com



Bime

License: commercial, free trial
Webste: www.bimeanalytics.com



CIWare

License: commercial, orderable demo
Webste: www.fortewares.com



InfoCaptor

License: commercial, free online demo
Webste: www.infocaptor.com



Tableau

License: commercial, free trial
Webste: www.tableausoftware.com



Logstash

License: open Source
Webste: www.logstash.net



Graylog2

License: open Source
Webste: www.graylog2.org



Evaluation Criteria

The tools tested in this study have all been evaluated based on the following criteria list. The criteria have been grouped under multiple umbrella subjects, meaning to emulate progression in program usage. Each criterion has further been given a three-step scale, the steps roughly evaluating to "bad", "acceptable" and "good", using the symbols "-", "~" and "+" respectively.

A tabular overview based on this catalogue, showing off all the tools' strengths and weaknesses at once, will also be provided. Of course it is not easy to fit the highly complex programs that are being tested into this three step scale. To gain a real understanding of the tools' various weaknesses and strengths reading the actual evaluation chapters is necessary.

4.1 Criteria Catalogue

Nominal Criteria

- Development Status
 - Development has (effectively) halted
 - ~ Slow development (sporadic patches, no new features)
 - + Active development, regular releases and patches
- License Cost
 - Expensive, temporary license
 - ~ Reasonably priced license/needs to be bought just once
 - + (Effectively) free
- Product Support
 - (Almost) None
 - ~ Support only for a fee
 - + Free support
- Available Documentation

4. Evaluation Criteria

- (Almost) None
- ~ Incomplete/disorganized/wrong documentation
- + Extensive and well organized documentation
- Difference between Commercial and Trial Version
 - Large difference in functionality/many features missing
 - ~ Differences are mostly peripheral
 - + Trial version is (effectively) feature-complete

Technical Criteria

- Supported Operating Systems
 - Requires a very specific version of an operating system
 - ~ Single-platform
 - + Multi-platform
- Supported Import Data Formats
 - Requires a very specific file format
 - ~ Supports common file types and data bases
 - + Supports multiple file formats which can freely be mixed

Data Visualization

- Chart Type Diversity
 - Only few/basic charts
 - ~ Good chart selection
 - + Many/Highly customizable chart types
- Chart Data Selection
 - None/chart must be remade
 - ~ Predefined Filters
 - + Predefined but also customizable filters
- Chart Malleability
 - Changing chart settings is a laborious process
 - ~ Most changes can be made with a few clicks
 - + Important parameters may also be switched instantly

Data Aggregation

- Ease of Data Selection
 - Selecting data is complex and highly unintuitive
 - ~ Data selection takes practice but is quickly understood
 - + Data selection has (almost) no learning curve
- Scope of Offered Functions
 - Too few to cover all use cases
 - ~ Large set of common functions for different data types
 - + Formulas are also highly customizable.
- Versatility of Formula Application
 - Formulas only ever apply to the entire data set
 - ~ Predefined limits and predicates may be set
 - + Limits and Predicates may be customized

Data Reporting

- Supported Export Data Formats
 - Cannot export data
 - ~ Supports only few/simple export variants
 - + Supports different export variants, both graphical and textual
- Reporting Automation
 - None
 - ~ Reports may be automatically created based on events/a schedule
 - + Reports may also be automatically distributed (e.g. via email)
- Report Sharing
 - None
 - ~ Dashboards may be shared freely
 - + Read/Write permission may also be set

4. Evaluation Criteria

4.2 Exemplary Reports

Next to the above catalogue Capgemini has provided us with a list of descriptions of various report types. These descriptions directly corresponded to the type of reports Capgemini wish to create through the tools tested in this study. In order to help us better understand these requirements we created a series of mockup reports using these descriptions and the real performance data we received. These mockups would then aid us in our evaluations - the sections on data visualization and aggregation in later chapters are directly based on the experience of trying to recreate these examples.

Some of these mockups looked as follows:

Figure 4.1 shows a comparison of various average values for some data set. The exact values are: a performance metric's value on some given date d , the metric's average value over the last year counting from d , as well as the last year's average when only counting those dates with the same day of the week as d , that is only counting Mondays/Tuesdays etc. Naturally creating this setup, from selecting the date(s) to applying averaging aggregations should be as simple and straightforward as possible.

Figure 4.2 shows a comparison of multiple growth rates over the the course of a year. Both the time frame and the data sets whose growth is shown should be quickly changeable.

Figure 4.3 shows both an absolute and a relative growth rate. The date used as the zero point should be freely selectable.

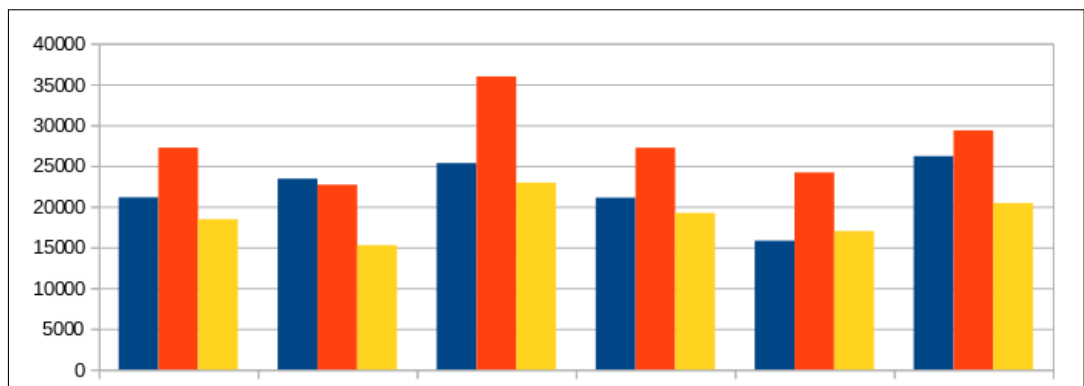


Figure 4.1. Comparison of averages

4.2. Exemplary Reports

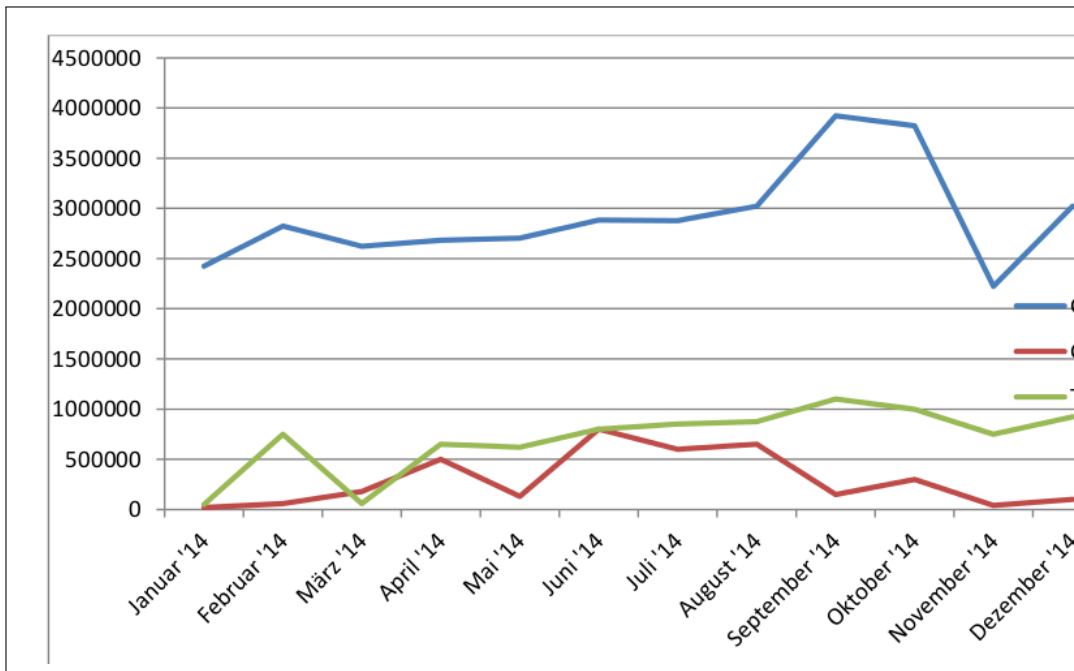


Figure 4.2. Comparison of various growth rates

4. Evaluation Criteria

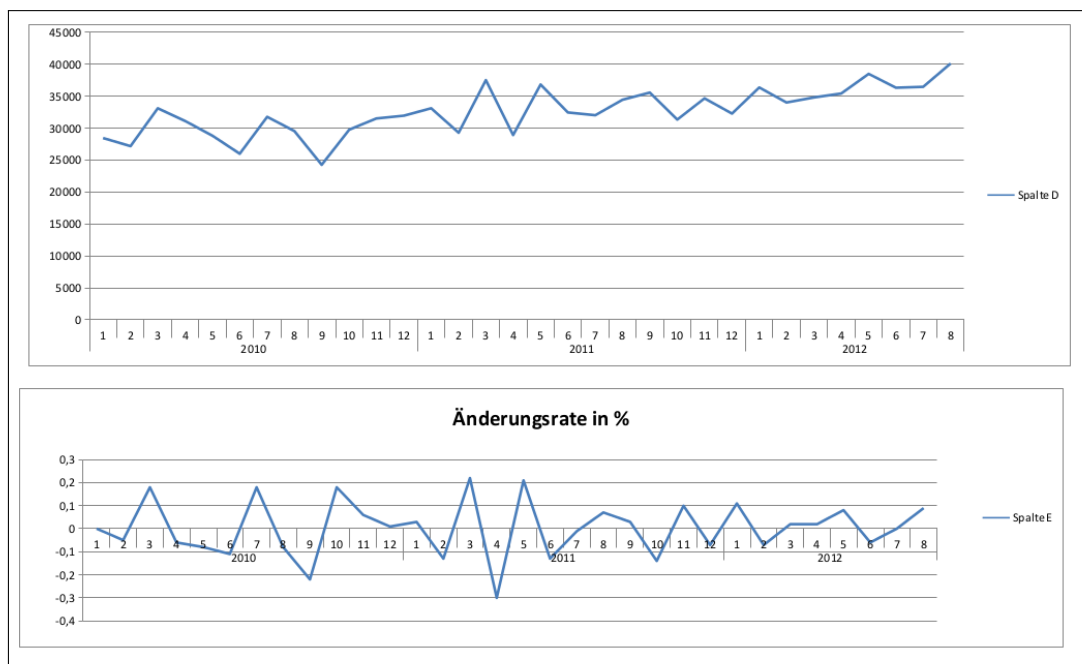


Figure 4.3. Absolute and relative growth

Tool Evaluations

This chapter will contain the individual evaluations of the previously introduced tools.

5.1 Excluded tools

Graylog2 has been excluded because of its lack of supported input file formats. Only syslogs and GELF (Graylog Extended Log Format) files can be imported. While searching for a free alternative for Graylog2 we found a tool called Logstash, which works in a similar fashion to Graylog2. Both run in combination with Elasticsearch and MongoDB, except that Logstash also allows the user to import CSV files. However, Logstash also had to be excluded due its difficult and unreliable installation procedure. Even when following alternative installation guides found on the internet, apart from the official documentation, we still didn't manage to get Logstash running properly. While browsing the internet, we also noticed, based on the issues from other users, that Logstash is very sensitive when it comes to updates: updating each one of the three components that make up Logstash can easily lead to incompatibility issues between the components and the whole system may cease working correctly.

5. Tool Evaluations

5.2 SiSense

Introduction

SiSense is a business intelligence tool produced by a company of the same name. Its customers include big names like ebay and NASA while its website boasts that SiSense "allows non-technical users to analyse 100 times more data at 10 times the speed of current in-memory solutions, all while supporting thousands of queries."

Nominal Criteria

SiSense is actively developed commercial software, a free trial version had to be used for the purpose of this study. It is unclear which, if any features, might be missing compared to the paid-for product, but judging from the features mentioned and explained in the publicly available documentation the trial version appears to be feature-complete. The same restriction also applies to pricing - no explicit cost is provided and must be directly requested.

The documentation is for the most part both extensive and well arranged, available in both text and video formats and offers a walkthrough of the steps of using SiSense, from the first installation, to creating proper dashboards.

As for support a public forum is available. However it does not seem to see much use - the number of threads in most categories is in the low two digit area with many of these topics being official announcements and tutorials. We can not tell how long a waiting time a serious support request would entail.

Installation and Setup

Apart from the minor inconveniences of obligatory account creation and a trial license limited to only a few days (both of which are par for the course) the installation of the SiSense trial version is a very simple task, consisting of downloading and running an executable file and clicking "Next" in the installation screen. The program is ready to be used once this is done. Additional non-trivial post-hoc tuning to access various file formats, as is the case for e.g. Splunk and databases, is not necessary.

Operating systems supported by SiSense are: Windows 7 and up or Windows Server 2008 and up. Either way it must be a 64-bit operating system.

Accessing one's data to convert it to the format used by SiSense is an equally streamlined process. SiSense does, in fact, install two different applications, the first of them, the so called "SiSense Elasticube Manager", is what is used to convert data from different sources into SiSense's native Elasticube format. This program, when started for the first time, will lead the user step by step through the process of importing their data, supplying

just enough helpful hints that the data import is likely to succeed even on the first try, which has proven to be the exception to the rule.

For data input SiSense accepts various data formats including Excel and CSV files, data base connections such as MySQL and PostgreSQL and input from Google Analytics or Hadoop Hive.

Additional setup to connect to the database is, as already mentioned, not needed. Entering the database login credentials was the only work necessary, at least as far as the locally run MySQL database used in this study is concerned. The data import itself is particularly powerful in that it is heavily customizable in regards to how and what to import. It is possible to choose specifically which tables and columns to import (which is something many other tools offer as well). Once that choice is made additional work on the data may yet be done: One may opt to use foreign key references based on those present in the database and/or based on identical column names (Qlikview for example only offers the latter) or one can create or delete custom foreign key references on the fly. Table columns can also be renamed, altered or deleted, new ones may be created on the fly via custom SQL queries.

All this is happening based on a graphical interface where the data to be imported is shown in a table diagram (see figure 5.1), with lines connecting various table columns to signify a foreign key relationship. This interface makes it quite easy to prepare and manipulate the data import and quite difficult to actually lose track of this process.

In general the data import facilities offered by SiSense are on of the best among the programs tested in this study, and they are also the one area SiSense is able to score the most points.

Once data import has been completed the actual SiSense main piece will automatically open in-browser, and the analysis and visualization of the imported data can begin.

Data Visualization

Similar to Splunk, SiSense divides user data among multiple dashboards. Each dashboard is based on the data supplied by one Elasticube and consists of multiple widgets, that is, the actual user-made charts, as well as optional filters for the data the widgets are based on. As far as widget placement goes one is not afforded the same amount of freedom as in Qlikview, which allows the user to arbitrarily change placement and size of its charts as one would do with open windows on one's desktop. The system employed by SiSense is more comparable to a tiling window manager - starting from the first widget taking as much space as it can a part of the screen is divided either vertically or horizontally to make room for new widgets.

5. Tool Evaluations

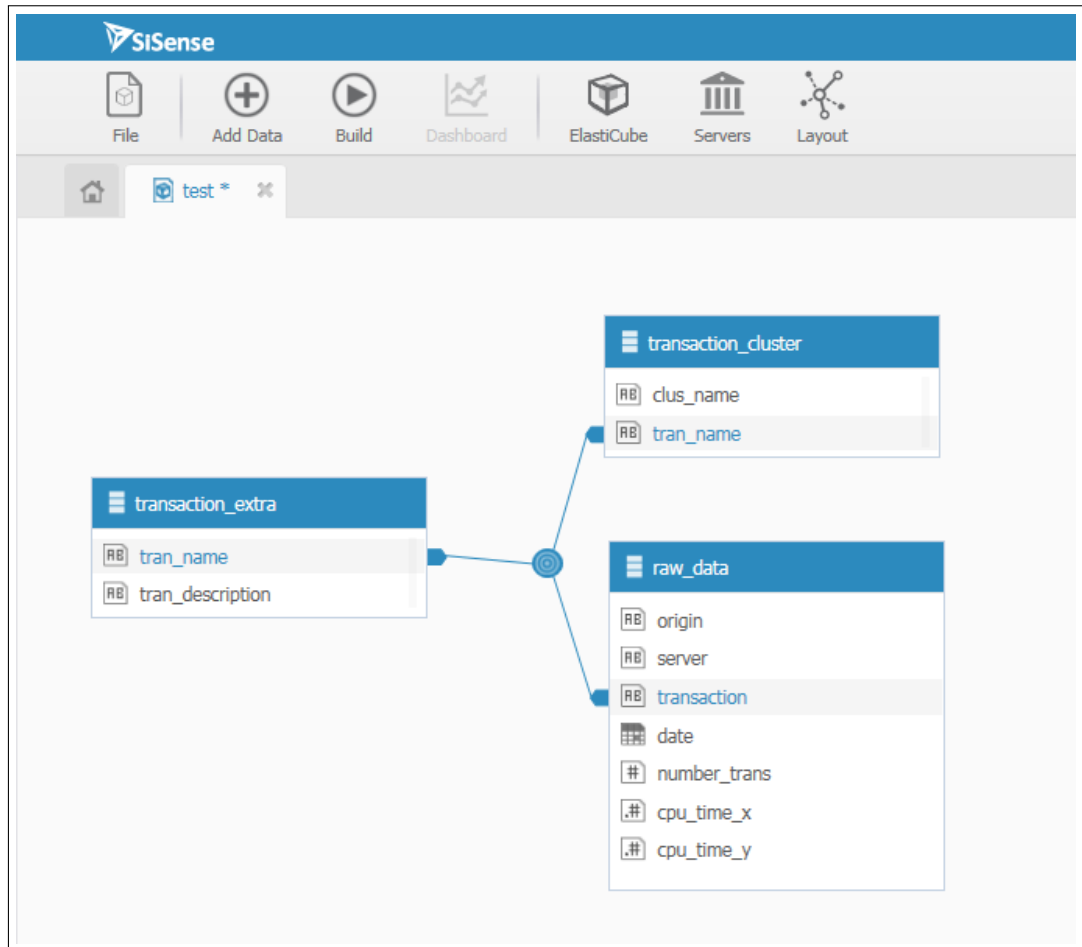


Figure 5.1. [SiSense] Graphical display of the data to be imported

As far as chart type diversity goes SiSense offers a reasonably sized assortment. The user has the choice between the following chart types:

- Indicator number
- Bar chart
- Column chart
- Line chart
- Area chart

- Pie chart
- Scatter chart
- Polar chart
- Table
- Scatter & area map

SiSense affords its charts a good amount of visual customizability, neither impressing, nor disappointing:

Charts may change in size and position. Content, sizes and visualization of the legend and various axes and data labels may be changed. The colors may be individually chosen either from a palette or directly via their RGB values. The formatting of values may be adapted on whether the values represent whole numbers/percentages/currencies/dates etc. The maximal number of x-values can be limited. The y-axis may be either logarithmic or absolute. The maximal y-value and the distance between milestones on the y-axis may also be chosen.

Some chart types come with their own custom settings like line-smoothing in a line chart or lining up vs. stacking columns & size of and gap between columns in a column diagram.

Data filters come in two varieties: on the one hand there are filters belonging only to one specific widget. These filter out data from this widget and this widget alone and are only accessible when editing said widget. On the other hand there are dashboard-wide filters. Filtering data for all widgets in the dashboard and also always available on the right side of the screen (widget filters are only accessible when specifically customizing a single widget). Dashboard filters will do one of two things: they will either fully remove deselected data from a widget, or merely gray out the deselected parts as shown in figure 5.2. The exact effects may be chosen for each widget, however some chart types will only use one of the two modes.

Chart types may also be restricted in the data they may contain, some cannot break down data a certain way, some do not make use of specific values. As such, changing which chart is displayed in a widget on the fly is oftentimes not possible. Other than having to specifically change a widget to be editable a partial reconstruction of the widget's data may be necessary.

Additionally a widget's own filters will always overwrite the dashboard filters. Both filter types existing for the same data may lead to unexpected results as the dashboard filter will seemingly not work correctly.

5. Tool Evaluations

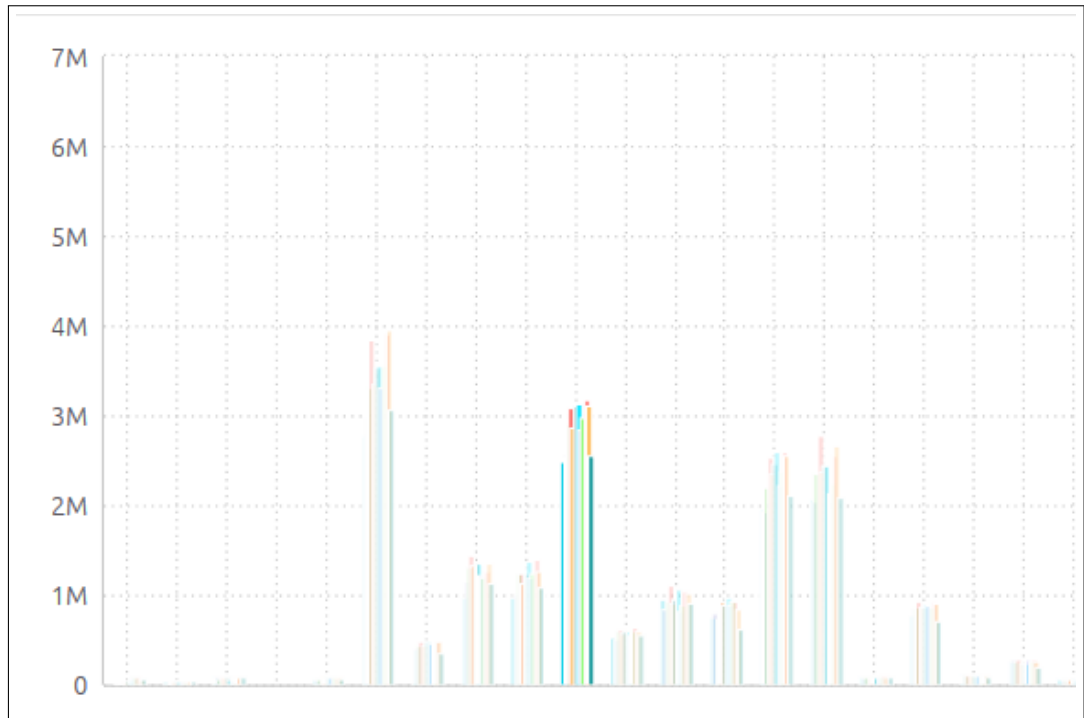


Figure 5.2. [SiSense] A filter graying out data instead of removing it from the chart

Data Aggregation

In SiSense data selection takes the usual approach: selection of x-values, of y-values and the way the data is then further broken down. One may, for example, use a list of servers as x-values, their performance data as y-values and break all it down based on time period (like the daily workload). Alternatively one can use the days as data for the x-axis and break down the visualization by server. Some diagram types will display correctly either way, others will need to have their data selected in a specific way to display as one would expect.

Practiced users will quickly come to understand and know how to circumvent this issue, however this does make it impossible to switch out some chart types on the fly, even if they do appear compatible and use the same data. For example a column diagram may correctly display data with one column for each month over the course of a year. Switching over to a line diagram what happens is that the y-values that used to belong to each column will now be displayed as multiple y-values all belonging to just one x-value. As figure 5.3 shows the resulting diagram is, of course, complete non-sense:

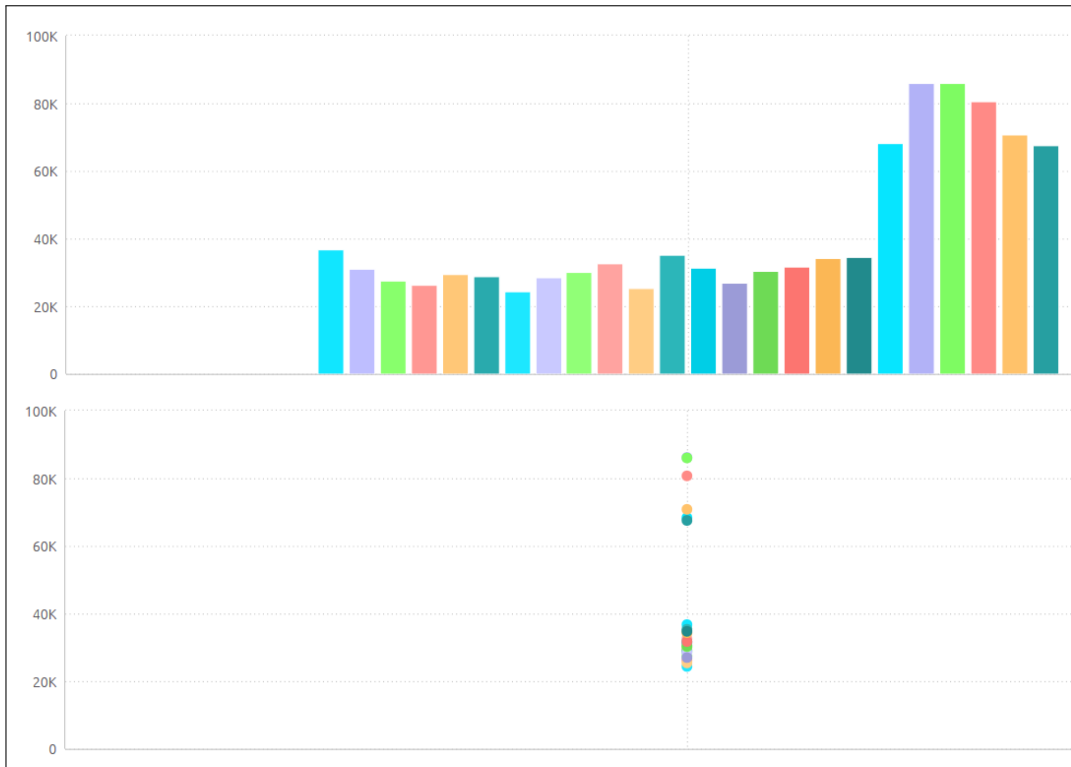


Figure 5.3. [SiSense] Both charts show the same data, only the chart type was changed

When selecting data filtering may be applied from the get go, using the same kind of filter that may be used on widgets or on the dashboard as described in the preceding chapter. Multiple filtering options are available: for most data enumerations filtering is based on selecting values one wants to filter from a list, or based on some type of ranking, like ignoring all values greater/smaller x , or selecting only the x smallest/biggest values. Additional data types are given their on filter options, e.g., dates may additionally be filtered based on years/months/days or specific time durations like "the last two days/weeks/months".

The filter system offers plenty of useful options from the very start, but quickly loses steam the farther one tries to go. There exists no real way for the user to define their own set of filtering operations, many use-cases are thus bound to be missed. To make an example: we have been unable to reproduce one of the filters found in one of our example reports: choosing some date d and only selecting data which was collected on a date with the same day of the week as d , that is fetching only the data that is collected on Mondays/Tuesdays etc. Needless to say such a filter would be quite useful for any sort of

5. Tool Evaluations

data gathering which involves different loads and values based on the specific time/ day of the week.

There does also exist an "advanced mode" in the filter (shown in figure 5.4), however it is completely based on manually editing a JSON file without any options to chose from or any other kind of help. Judging from the complete lack of publicly available documentation of this feature it appears far more likely that this mode merely contains a few obscure options most users will never need to touch, rather than any novel and advanced functionality.

Whatever issues have so far been presented in regards to chart utilization can be easily surmounted after only a few hours of practice, it is this lack of a freely configurable filtering utility which marks the first obvious failure on the part of SiSense.

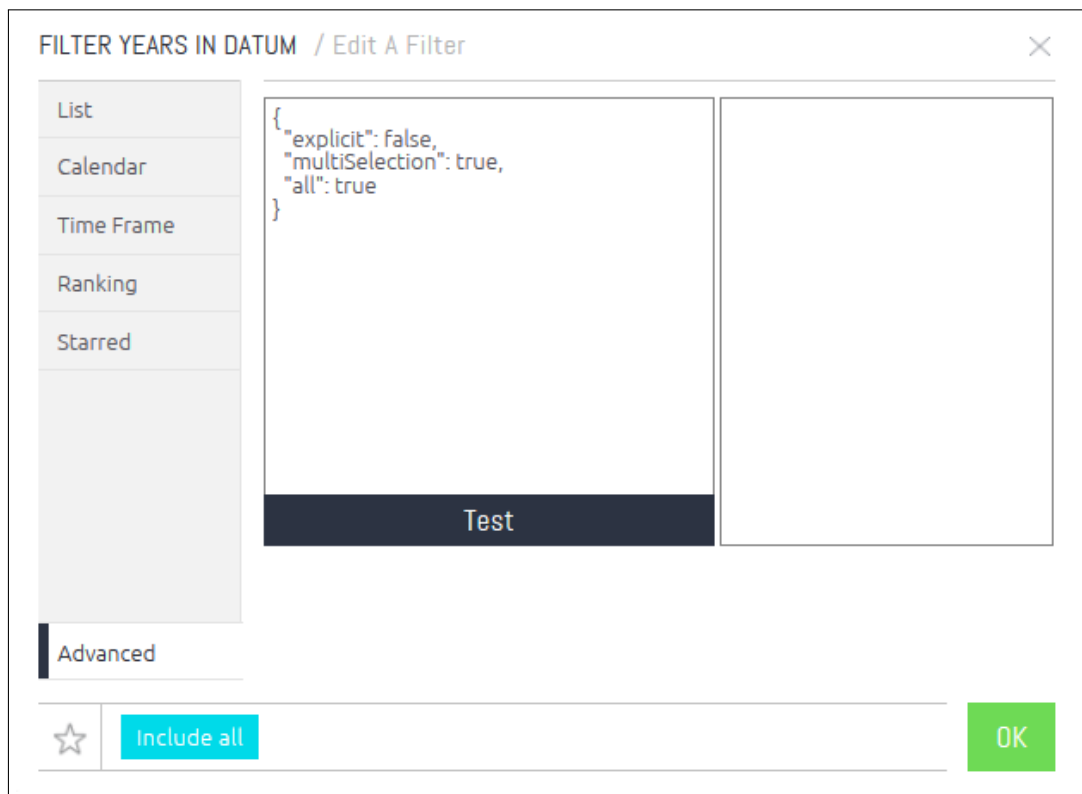


Figure 5.4. [SiSense] Advanced filter options

The pattern set by the filter functionality continues on into the choices of functions to be applied on the data. The amount of choices and the diversity offered is in fact quite

satisfactory. Next to the trivial main-stays like sum, average, min and max, there also exist various advanced predefined aggregations like standard deviation, growth rates, hourly/-monthly/yearly differences and averages, just to name a few. Most of these formulas may be written down and applied in the syntactically simple form of $f(x)$, hence using SiSense's function editor is oftentimes as simple as selecting the desired item out of a list. The breadth of the offered functionality is still only a fraction of the repertoire of something like Qlikview, nonetheless creative use of the provided formulas should be sufficient for many use-cases.

Yet the weakness of SiSense continues to be found in the act of data *selection*. The data filter allowed to select data only based on conditions already predefined by SiSense. Such a choice does not exist for application of aggregations at all. It is not at all possible to say "Apply formula f on some data set x only for data points which fulfill predicate p ". Or to make a concrete example: it is not possible to define a function which looks at the average server load only on weekends, unless all working days are removed from the entire time data set beforehand. But then what if one wants to compare the average workloads on weekdays vs. week ends? The standard filter does not include the ability to filter dates based on the day of the week, selecting or removing all weekends from a year's worth of data would have to be done by manually picking out the individual dates from a list. If a workaround even exists for this situation it is unlikely to be easy to implement and use.

Finally it is also not possible to directly create new data sets from existing data. Other tools make it possible to create a new column by, for example, applying a filter to a data set. This data would then be added to the existing data base and could be freely used throughout the dashboard. In SiSense this is only possible if data is incorporated in some widget. You cannot keep the associated data after a widget was deleted if that data was not originally derived directly from the Elasticube.

Ultimately the great ease of using SiSense's data aggregation facilities comes at the cost of being too weak to handle advanced use-cases. It is only a question of whether the advantages of the former outweigh the disadvantages of the latter.

It is also worth mentioning that according to the SiSense website SiSense is using special technology to bypass the operating system and load data directly into the CPU cache, which is supposed to lead to a dramatic speed boost. The veracity of this claim could absolutely not be verified. SiSense did not enjoy a performance advantage over other tools tested in this study and was even visibly slower than some. In fact doing calculations on large data sets (around one million entries) would oftentimes render the browser non-responsive and issue a warning that "a script has stopped working" before this option was permanently turned off.

5. Tool Evaluations

Data Reporting

As was already mentioned in the section about Installation and Setup SiSense combines various external data sources into its own native format, the so called "SiSense Elasticube". Once the Elasticube is created changes to its initial data sources are no longer registered, but may be acquired by rebuilding an Elasticube, either manually, or on an automated schedule. This is also where SiSense's automatic reporting ability comes into play. Rebuilding an Elasticube may be coupled with automatically distributing the dashboards constructed from this Elasticube via email.

An automatic reporting schedule independent of Elasticube construction may also be used. Additionally the amount of emails sent may be limited.

Dashboards may also be shared among multiple users much in the same way multiple people would share a Dropbox folder.

An individual widget may also be exported into a CSV file. Hereby only the rows and columns relevant to the individual widget will be exported.

Final Thoughts

"Good is the enemy of great" is a maxim which readily applies in the case of SiSense. It easily scores many points on the front of installation and data preparation, offers good data reporting support, does not require much practice to become useful and offers some great functions right from the get-go. Yet, despite being off to such a good start, SiSense fails to achieve actual greatness.

It is SiSense's weak filtering mechanism and lack of ability to freely customize the minutiae of data selection and formula application which renders it ill-equipped to deal with many use-cases and ultimately forces us to come to the conclusion that SiSense is just not powerful enough a solution for the kind of problems this study is meant to solve.

5.3 QlikView

Introduction

QlikView is a commercial software developed by the formerly Swedish company QlikTech. QlikTech in fact offers two different products which would be suitable for this study: Qlik Sense and QlikView. Since the former is a yet untested newcomer, released only around July 2014, and also quite similar to QlikView, we have decided to forego Qlik Sense and to concentrate on the tried and true QlikView.

Nominal Criteria

QlikView is currently under active development. The free version of QlikView is just as powerful as the fully licensed product, we have been unable to find any information on missing or abridged features.

The documentation is something of a mixed bag. On the one hand it is quite extensive and seemingly well organized, available both in the form of text and videos. On the other hand it is not always up-to-date, especially the exact process of allowing QlikView to read data from a MySQL database required extensive Google searches and crude clicking around through the menus until the proper settings were found through dumb luck. The official documentation had proven itself rather unreliable, at least as far as this specific facet is concerned.

On the bright side QlikView has something many other tools do not have: a thriving community which doubles as a support system and source of unofficial documentation. A Java programmer, when he encounters a problem, need not directly contact Oracle, and may instead turn towards thousands of his peers to find help online. The QlikView community functions in much the same way (though of course on a much smaller scale) to the point that the QlikView community forum even looks just like StackOverflow, but with a different color scheme. This gets rid of the need to leaf through tomes of documentation in the need to find some feature or explanation that may not even fully apply to whatever problem one is currently trying to solve. Instead community support is able to offer precise and relevant answers and even complete solutions - the latter only if one does not mind publicly offering a (obfuscated) subset of one's data.

QlikTech additionally offer free training courses for QlikView based around different topics and delivery systems. We have received many an invitation to QlikView Webinars over the course of this study.

License pricing is rather on the expensive side with prices ranging from 1010€ per named user for a Named User License, 10500€ per concurrent user for a Concurrent License and a grand 26250€ per server for an Enterprise Edition Server License.

5. Tool Evaluations

Ultimately one has to take the good along with the bad. Luckily the bad parts do not have any long term consequences while the good parts do.

Installation and Setup

QlikView's installation follows the usual routine: create an account, download and run the executable file, click "next" a few times and you are done.

Various versions of Windows are supported by QlikView, from Windows 8 going as far back as Windows XP, for both 32 and 64 bit systems.

QlikView accepts various data formats as input ranging from Excel, XML and CSV files and various databases to more obscure formats like DataStax. However importing data from a MySQL database source required the download of a special ODBC Connector, which we were only able to find after doing some searching of our own.

Only once this is out of the way will QlikView gain the capacity to connect to a MySQL database and make use of the data stored therein. Like with other tools it is possible to select exactly which tables and table columns will and will not be used.

All these actions do not have an immediate effect. What actually happens is that QlikView generates lines of a scripting language of its own design, This script has to be explicitly executed by the user, only then will QlikView connect to the database and make all the necessary queries. This makes it possible to mix data from various sources, or further adapt the imported data, be it simple renaming of fields or extraction of additional columns.

```
1 SET ThousandSep='.';
2 SET DecimalSep=',';
3 SET MoneyThousandSep='.';
4 SET MoneyDecimalSep=',';
5 SET MoneyFormat='#.##0,00 €;-#.##0,00 €';
6 SET TimeFormat='hh:mm:ss';
7 SET DateFormat='DD.MM.YYYY';
8 SET TimestampFormat='DD.MM.YYYY hh:mm:ss[.fff]';
9 SET MonthNames='Jan;Feb;Mrz;Apr;Mai;Jun;Jul;Aug;Sep;Okt;Nov;Dez';
10 SET DayNames='Mo;Di;Mi;Do;Fr;Sa;So';
11
12 ODBC CONNECT TO [Fachstudie Source];
13 SQL SELECT number_trans as Transactions,
14        srv_name as Server,
15        date as Date,
16        weekday(date) as Day_Of_Week
17 FROM fachstudiedb.`srv_transactions`;
18
```

Figure 5.5. [QlikView] Exemplary data import script

The script's syntax is generally simple and importing data from a database allows the direct use of SQL queries, as such learning how to manipulate QlikView's import scripts will not be much of a problem. An example of one such import script is provided in figure 5.5.

Unfortunately it is not possible for QlikView to directly assign foreign key relationships when importing data. QlikView will only recognize such associations if the columns the data is stored in all share the same name. The database schema must be properly formatted in advance. Alternatively the columns may simply be renamed in QlikView's import script.

As such QlikView's data import facility may well be summarized as being somewhat slow to get going but more than powerful enough to get the job done.

Data Visualization

Upon starting out QlikView will present the user with the choice which of their data to import into list boxes. These lists are the go-to object in QlikView when it comes to selecting which data to work with. For example selecting an entry from the "Countries" list will also highlight all the entries in the "Customers" list which are located in the selected country. Likewise it will cause an update and rebuild of all charts using the selected data. This will only work if QlikView knows which columns refer to the same data set, hence why aptly naming the columns is a crucial step of the setup process.

QlikView supports a reasonably sized set of different chart types and widgets (collectively referred to as "Objects" by QlikView). The different chart types offered by QlikView are:

- Bar chart
- Column chart
- Line chart
- Radar chart
- Gauge chart
- Mekko chart
- Scatter chart
- Grid chart
- Pie Chart
- Funnel chart

5. Tool Evaluations

- Block chart
- Table chart

It is possible to assign multiple chart types to a single object, these charts may then be instantaneously switched with only one single mouse click. Though of course not all chart types will properly look or work based on a specific selection of data. This is however not a fault of QlikView but rather the result of a general incompatibility among the various charts.

In addition to these charts there are various other objects which may be included: One may add decorative line or arrow objects which further explain the chart and help orient the user. One may add search boxes which can help find a specific piece of data from a large data set, thus relieving the need to manually scroll through long lists. One may add date sliders or calendar objects to, for example, easily select a specific day's date out of multiple years of data. One may add input boxes and buttons, the latter of which may trigger a variety of different actions as shown in figure 5.6. Additionally users may be restricted from various actions such as moving, adding or removing objects right up to making the whole document read-only.

Data selection may also be quickly customized. It is very much possible to define data selection objects which select data based only on certain criteria. Filtering which data (it is possible) to select for aggregation in QlikView is not a simple, plug-and-play process the way it is in SiSense, but QlikView more than makes up for it with its sheer power. Unlike SiSense, QlikView is extremely malleable, and allows the user to churn their data through an arbitrary amount of filters and predicates of various forms. As an example take the use case of filtering out all weekends from a year's worth of data. SiSense was not able to do it (unless you count removing dates by hand). QlikView however is perfectly capable of selecting dates based on their day of week and would have easily allowed to further specify selection by, for example, restricting dates to some specific year or month or only selecting those where some other column value associated with that date is within some given bound.

Combine QlikView's ability to define a custom set of button, input box and field objects with its strong customizable data filters and a QlikView document may well be made automatized and fool-proof to similar degree as to what is possible with Excel documents.

QlikView continues to be versatile even on the cosmetic front. All of its objects may be freely moved around and resized, the layout is thus not restricted to any kind of *n-rows and m-columns* scheme as is the case with most other tools.

Objects may take on different styles, colors and shapes. The amount of settings which exists for charts, as shown in figure 5.7, is of downright overwhelming size (which is quite often the case with QlikView). Almost everything may be changed in one way from another,

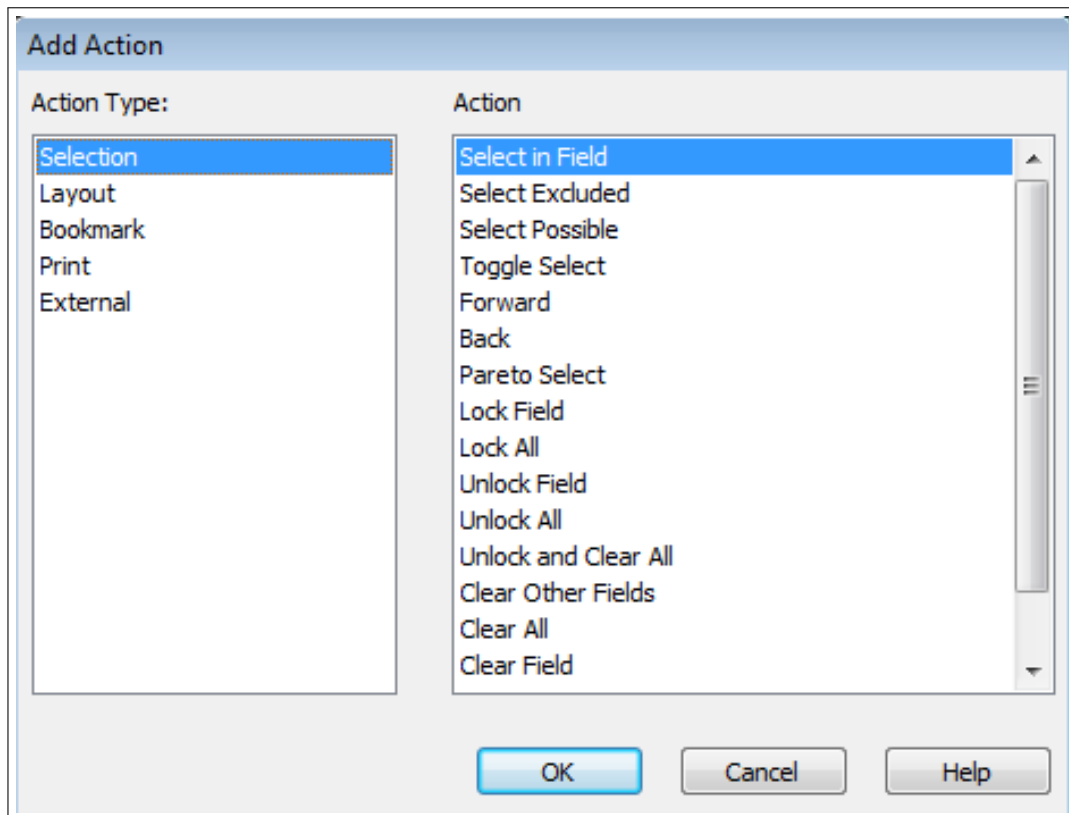


Figure 5.6. [QlikView] Actions a button push may trigger

from the used colors, to position and setup of the legend and the formatting and values. Individual chart types come with their own adjustments ranging from line thickness to switching between 2D and 3D display.

Data Aggregation

Starting out with data aggregation in QlikView is not quite as easy as it is for some other tools, you first need to find your way around multiple tabs, option panes and checkboxes before the position of the data choosing facilities becomes fully apparent. Once this is over new charts can be setup quickly and efficiently. The order in which one chooses to define charts' values (revenue per month per location vs revenue per location per month) may lead to some unexpected and unwanted results, but this difference, too, is quickly understood.

5. Tool Evaluations

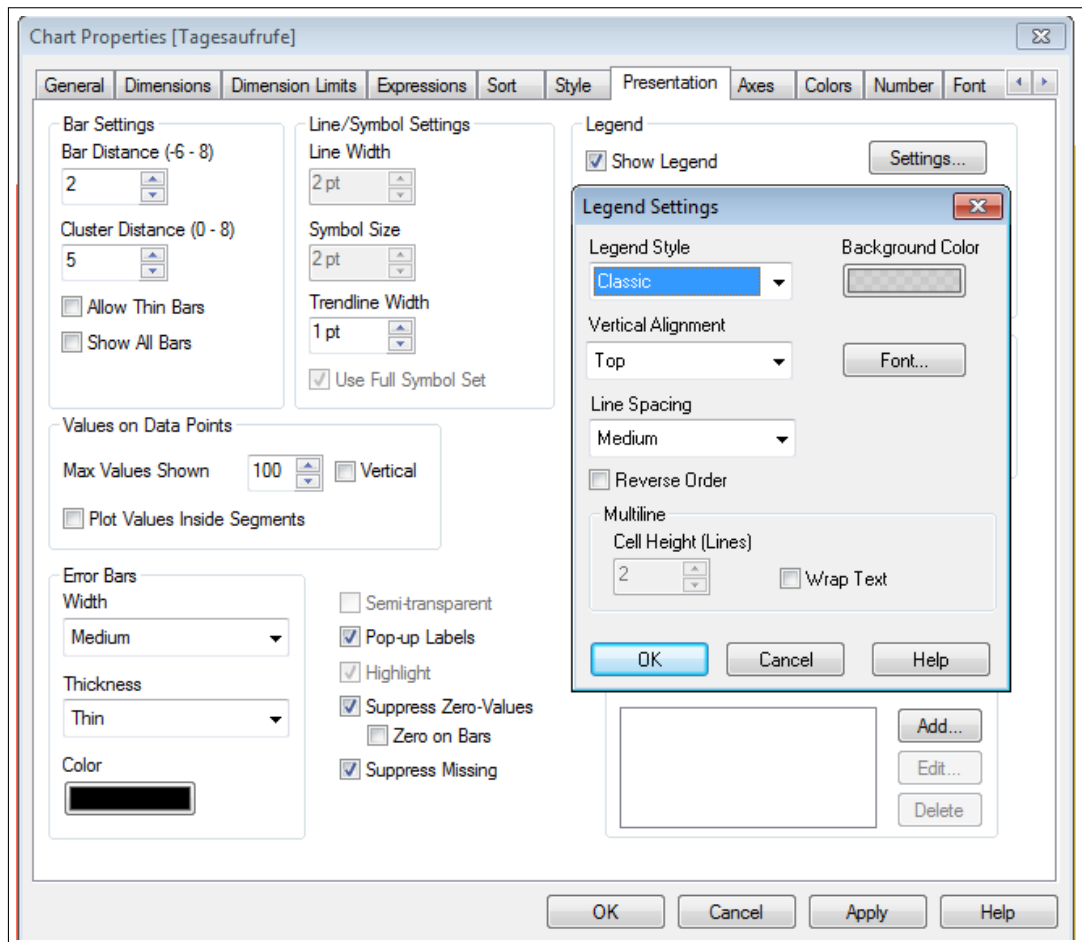


Figure 5.7. [QlikView] Overview of style settings

Quick understanding, however, is restricted to handling vanilla data. Transforming and filtering data is where QlikView shows its incredibly sharp claws. The amount of available functions, distributed among multiple categories and data types, literally numbers in the hundreds. Selecting the right one for the job, or even knowing it exists in the first place, is no easy task and will require plenty of experience.

The flood of information does not stop at just selecting a function, knowing how to apply it can be an equally daunting task as the function syntax is far from trivial and is in fact quite ugly and difficult to read and understand as figure 5.8 shows. Yet as complicated QlikView may be - it is just as powerful. QlikView's many functions come with many

additional selectors and properties and as such may be applied to exactly whatever data in exactly whatever way is needed. For other tools it was sometimes a question of whether they are capable of a specific use-case at all, with QlikView it is always just a question of "how?".

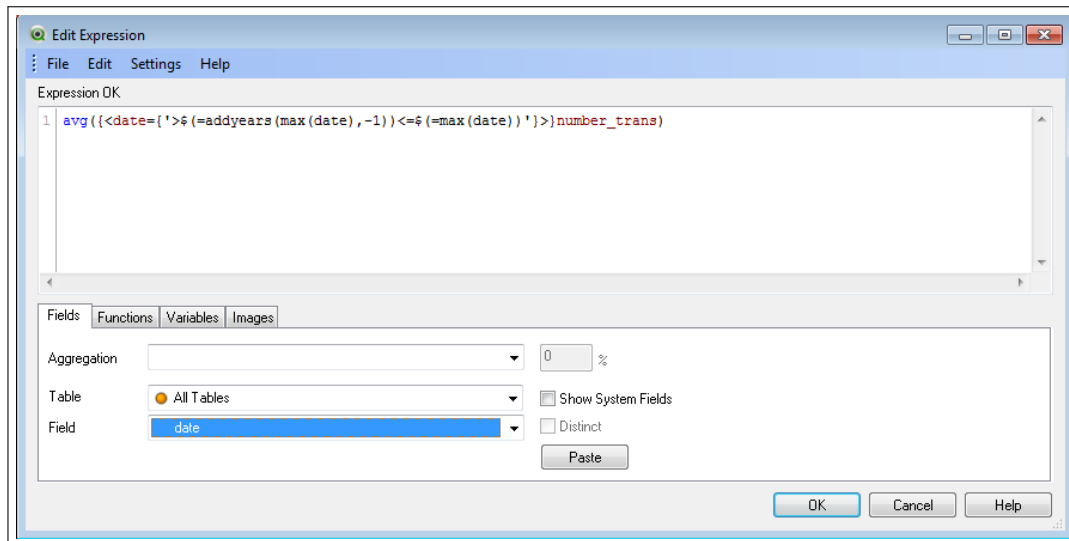


Figure 5.8. [QlikView] Typical function syntax

In summary it can be said that while QlikView may be a program that will take a long time to learn and an even longer time to master, but once this is accomplished the benefits are immense.

Data Reporting

So far QlikView has more or less been going from strength to strength. This section will put a dampener on this streak. While far from inadequate QlikView's reporting ability is definitely its weakest point.

Each object's specific data may be exported into an Excel file. Other than that reporting in QlikView is primarily geared for printing, which means that reports are always static. Report creation is no automatic process, instead single objects may be added to a report via drag-and-drop. Object placement and size within a report can then be freely changed just like in QlikView proper.

There is no automatic reporting, reports are always created by hand. On the positive side there does exist an *alarm* function which, based on whichever condition (the same

5. Tool Evaluations

functions as discussed in the preceding chapter may be applied), will automatically send out emails with a predefined message. While no proper report this feature may well be used to recognize anomalies and dangerous conditions in the data.

Data sharing, too, is not innately a part of QlikView. Sharing documents will thus require the use of the local network or a service like Dropbox.

One possible way to redress these issues to some degree would be the third-party program QV Mailer which would allow to automatically distribute QlikView reports via email.

Final Thoughts

QlikView is powerful, it is capable and it is even fast, too. A tangible delay was only ever felt when it needed to handle almost our entire test dataset - well over a million entries - simultaneously. But QlikView is also somewhat unwieldy and slow to get going. As such employing QlikView will require a certain amount of specialization and a good bit of persistence, its great power cannot be brought to bear in a quick-come-quick-go fashion. If this overhead is acceptable then QlikView is a strong candidate for being the best tool this study has looked at.

5.4 Birst

Introduction

Birst Inc. is an American company providing their business intelligence platform in the form of a cloud-based, Software-as-a-Service application.

As Birst is a commercial service a free trial version was used during this study.

Nominal Criteria

Birst is currently under active development. The difference between the trial and commercial version can, as usual, not be conclusively established because no such information is provided. The Terms Of Service do limit the amount of data stored to be no more than 100 GB and it does not appear to be possible to save data visualizations (more on this in the section on Data Aggregation), but otherwise the trial version appears feature-complete.

Same goes for the license cost. No explicit information is provided and the exact price will likely need to be negotiated.

Birst does feature a Support Portal boasting with 24x7 customer support and a large array of tutorials and FAQs, yet this is only open for paying customers. Public support and documentation however is, even after extensive searching, limited to maybe half a dozen (albeit useful) tutorial videos.

Installation and Setup

Being a cloud-based service used exclusively from within browser rather than an executable program Birst's installation process (as least as far as the trial version is concerned) is virtually non-existent. You sign up to create an account and log in and this is it - you are now free to use Birst.

This also renders Birst fully multi-platform, as any operating system you can use a browser on is supported.

Unfortunately, getting data into Birst did not turn out to be as easy. Quite the opposite, we have been unable to do it via SQL. Birst refused to connect to our locally run MySQL database, supposedly because it was lacking the proper SQL drivers. Despite our best efforts to remedy the issue we ultimately had to resort to uploading data as CSV files instead. This was also the first time that Birst's lack of public documentation and support became glaringly apparent. Scouring the web for help revealed either generic advertisements from Birst's own website or information about the completely unrelated SQL BIRT system.

Nonetheless after data has been uploaded into Birst's systems it can be freely looked

5. Tool Evaluations

at and further aggregated using Birst's custom formula facilities. Sadly this data customization lacks many formula options and the guided formula creation of Birst's dedicated visualization suite (which will be discussed in the next chapter). The exact details and reasons for this discrepancy are unclear.

Data Visualization

Birst's visualization suite may boast the accomplishment of being at the same time both powerful and simple. Creating and expanding a chart or selecting data is a simple matter of dragging and dropping a data column into the right category. Aided by a search function which can quickly find data points by their name charts may thus be created quickly and efficiently.

Birst offers multiple containers data is to be dropped into, as shown in figure 5.9. At first there are Measures and Categories, these correspond a chart's y- and x-Axis. Data may then be further broken down by another column adding a Color and even further by using what Birst calls a "Trellis" which would mean rendering the chart multiple types for entries inside yet another data column.

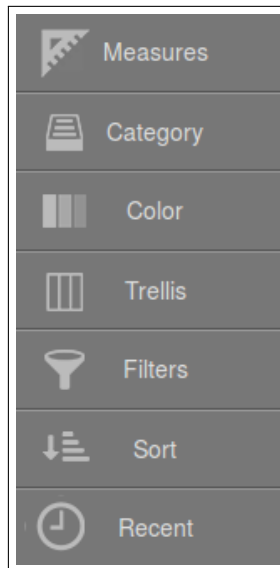


Figure 5.9. [Birst] Charts are built by dropping data into this column

If that is still not enough more data can be added in the form of "Shapes" and their various sizes, effectively turning the chart into a quickly customized scatter plot.

The latter option, as well as other further customizations, are offered in the form of a menu,

5.4. Birst

as shown in figure 5.10, that pops in automatically whenever some data is currently being dragged - selecting any of these, too, works via drag-and-dropping data into the desired shape. Additionally chart types and visualization options may be quickly switched around via a drop-down menu. In case of badly selected data for the current chart type (like a column chart being unable to do anything with data put into the Shape category) Birst will even inform the user which part of the data is usable and which ought to be removed.

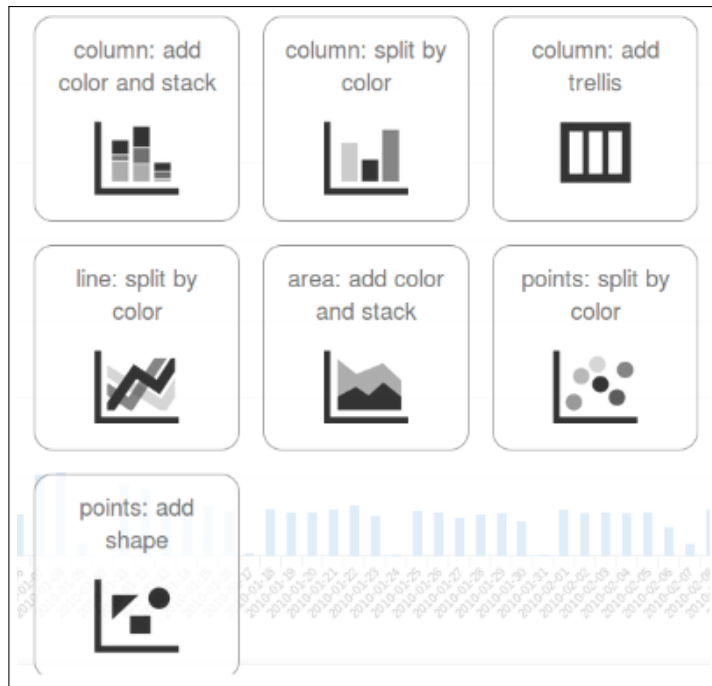


Figure 5.10. [Birst] Contextual chart customization help

In general Birst supports the following chart types:

- Column chart
- Bar chart
- Line chart
- Area chart
- Point chart
- Scatter chart

5. Tool Evaluations

- Bubble chart
- Pie chart
- Funnel/Pyramid chart
- Table chart

All these charts come with the usual set of generic and chart-specific customizations like choice of color, placement of the legend, line smoothing in a line chart, bar stacking in a bar chart, etc.

Data selection, too, is a quite smooth process. Birst's visualization suite comes with a good set of easy-to-use filters, especially where the creation of custom filters is concerned (which will be elaborated on in the next section). Predefined filters offer an array of logical conditions such as smaller/greater than, between, in, contains, is equal to etc. Not particularly impressive on its own this filtering facility receives a great boost in power thanks to Birst's handling of time data. For whatever dates are currently loaded Birst offers a great array of time related functions and predicates such as Day/Day Number/Day Of Week/Week Of Year/x Days Ago/x Months Ago etc. Thus the use-case of filtering out all weekends from a data set that SiSense was not able to do and QlikView needed a long-winded function to accomplish poses absolutely no problem for Birst.

As was already explained in a preceding section: Birst's trial version does come with one glaring flaw: the inability to either save or load any work done in its visualizer. As such it has not been possible to properly assess the process by which single charts (the visualizer only ever works on one single chart, unless one counts the Trellis) would go on to be united into a complete dashboard - we had to make due with Birst's pre-made example dashboards instead.

What can be said is that the dashboard interface is not quite as versatile as the visualizer as chart type and properties cannot be switched on the fly, which however does make sense as the dashboard is meant to show off finished reports. For data choice it is possible configure so called "Prompts" much the same way one creates a filter in order to chose which data may or may not be selected. These prompts would then make themselves available as toggleable drop down lists in the actual dashboard.

Data Aggregation

As already mentioned in the preceding section Birst provides a filter interface that even in its raw state can deal with use-cases other tools cannot. The same may be said about Birst's functions: what is predefined already offers a good choice and will be enough to deal with most challenges. However it is setting up one's own functions and filters where

Birst really gets to shine.

The process is easy and completely intuitive. If users know SQL syntax, they will immediately feel at home. What's more is that it is a guided process, at every step in the making of the formula Birst will keep track of what was entered, give warning about errors and make reasonable and syntactically correct suggestions on what to do next. The entire process features virtually no learning curve, all that's needed is actually knowing what data you want and then selecting the appropriate fields from Birst's drop-down menus in the fashion of select *A* where *B*, as shown in figure 5.11. Whatever data was the result of this selection is hence available to be used in the program in the same way as if it was a previously loaded data column, including further filtering or using it in more formulas.

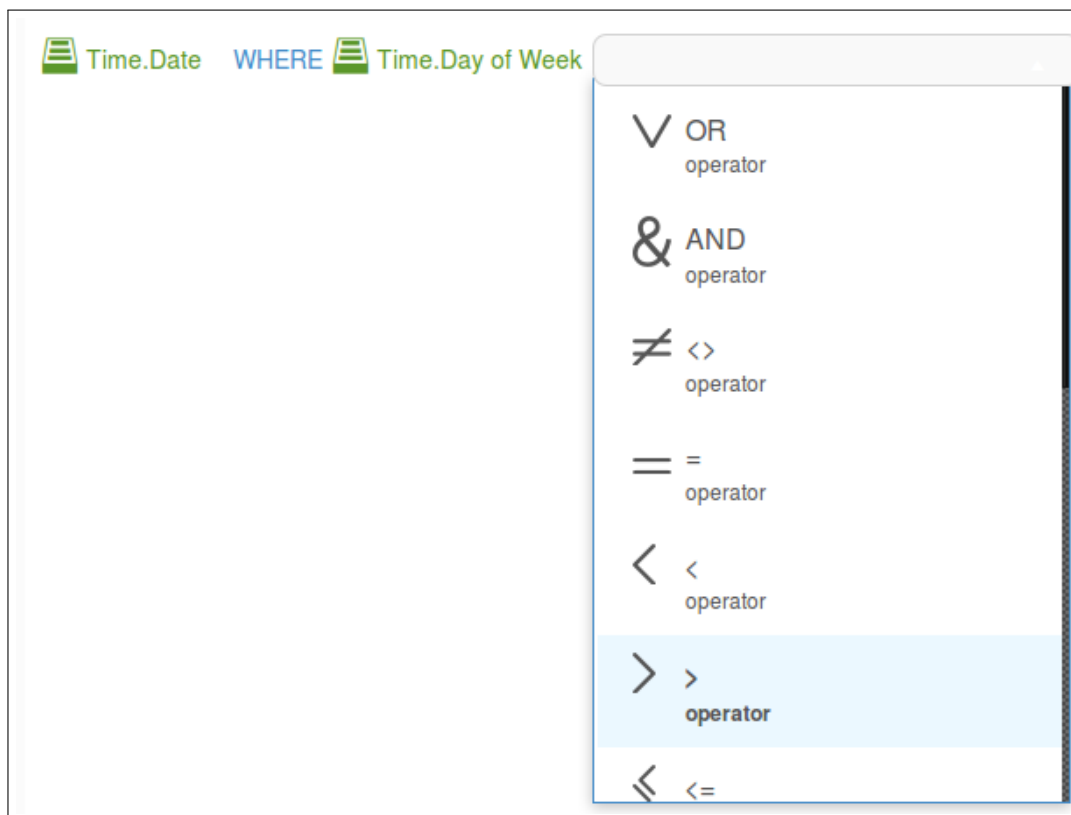


Figure 5.11. [Birst] Guided formula creation facility

Everything that so far has been said in this section applies solely to Birst's dedicated visualization suite. Birst also offers a report design facility, the major difference between the two, as far as we could tell, being that the former allows to define more than one chart

5. Tool Evaluations

at the same time - while simultaneously being slower and more cumbersome and having fewer visualization options and plain not offering any of the utilities like draw-and-drop data selection or guided formula creation that make Birst convenient and efficient in the first place. Whatever reasons this design decision may have had, we chose not to waste our time investigating such oddities and concentrated on the dedicated visualization aspect which, as the preceding sections have shown, more than makes up for its weaker twin.

Data Reporting

A report in Birst may be saved in one of two different ways. Privately - which means it will only be accessible to whoever created it, or publicly - which means it will be shared among all accounts who have been invited by a user with administrative rights. The same sharing mechanism also applies to complete dashboards.

Additionally reports (but surprisingly enough not complete dashboards) may be automatically generated and distributed. The distribution happens via email while the report may be sent as a CSV, HTML, Excel, PowerPoint, PDF or Word file on a daily, weekly or monthly basis.

Final Thoughts

Of all the tools investigated in this study Birst is definitely the oddest. Its various sections are noticeably shorter than those of other tools, because instead of lengthy explanations of how and why a use-case is not possible (or difficult) to achieve, Birst only features a quick list of things it *can* do. But when an issue does come up it presents a serious headache: The inability to save the charts we created prevented us from investigating Birst as thoroughly as we would have liked, while the presence of two visualization suites, one being much weaker than the other on nearly every front, calls into question the tool's entire design philosophy.

Yet even despite these weird issues Birst offers a power to convenience ratio that is absolutely unmatched by any other tool this study has looked at.

5.5 Splunk

Introduction

Splunk is a Log-, Monitoring- and Reporting-Tool developed by the American Company Splunk. According to their website, over 7,900 customers in 100 countries rely on Splunk. Splunk comes with 4 different Versions: Free, Enterprise, Global and Cloud. Splunk Free is the free version. Enterprise and Cloud are the commercial versions. Global is a support upgrade.

Nominal Criteria

The documentation of Splunk is extensive and well organized, both in form of text and videos. The big and active community of Splunk proved to be an important resource when facing a specific problem. Splunk also offers webinars and education courses. Some courses also have recommended prerequisites.

After having installed Splunk for the first time, it will run as a trial enterprise version. The trial duration is 60 days and 500MB of data per day to index. Afterwards the user has to convert to a perpetual Free license or purchase an Enterprise license.

Splunk Free allows the user to index up to 500MB of data per day. Also it is restricted in some functions. Some of the major drawbacks in our opinion are “scheduling and automated generation and delivery of reports and dashboards”, as well as, “monitor and alert for individual and correlated real-time events”.

Splunk Enterprise is the commercial version which offers full functionality, including all premium apps. Pricing is not determined by the amount of users, but by the amount of indexed data per day. There are two options for licensing Splunk Enterprise:

- Perpetual license: this includes the full functionality of Splunk Enterprise and starts with 4,500 for 1 GB/day, plus annual support fees
- Term license: this provides the option of paying a yearly fee instead of the one-time perpetual license fee. Term licenses start at 1,800 per year, which includes annual support fees

It also exist a volume discount: the license, for example for 1GB per day, leads to 4500/GB, compared to 2500/GB for the 10GB per day license. Splunk Cloud is a cloud-based version of Splunk Enterprise and also offers full functionality. Splunk Global is a support upgrade as shown in figure 5.12.

5. Tool Evaluations

Case Priority Levels

Splunk offers different response times and case handling based on case priority levels.

- P1 = A Production Splunk installation is completely inaccessible or the majority of its functionality is unusable.
- P2 = One or more important features of a Production Splunk installation has become unusable.
- P3 = Any other case.
- P4 = All enhancement requests.

Support Service Offerings

	Community	Enterprise	Global
Access to Splunk Documentation	✓	✓	✓
Access to Splunk Answers	✓	✓	✓
Live Product Roadmap & Input	✓	✓	✓
Online Case Submission	✓	✓	✓
Online Case Status		✓	✓
Guaranteed Response Times		✓	✓
Phone Support		✓ Phone Support during business hours except US holidays. After-hours Phone Support for Critical Issues.	✓ Phone Support during business hours except US holidays. After-hours Phone Support for Critical Issues.
Assigned Primary Support Contact			✓
Quarterly Account Status Reviews			✓

Enterprise and Global Services Agreements

	Response Time	Status Update	Fix or Workaround
P1	4 Hours	Daily	1 Business Day
P2	Next Business Day	Weekly	1 Week
P3	2 Business Days		Next Release
P4	2 Business Days		At Splunk's discretion

Figure 5.12. [Splunk] All levels of support service

Note that the community version equals Splunk Free. Once purchased the Enterprise edition, the support service will automatically upgrade to Enterprise as well.

Technical Criteria

Splunk can be installed on almost any modern operating system and supports a big variety of data. In term of this study we only tested importing files from the same Computer running windows 7. Importing data can be easily done using the web interface as shown in figure 5.13.

Splunk itself only supports to export Dashboards as a PDF file. However there are a lot of apps to add other file formats for export, such as Excel, CSV and databases.

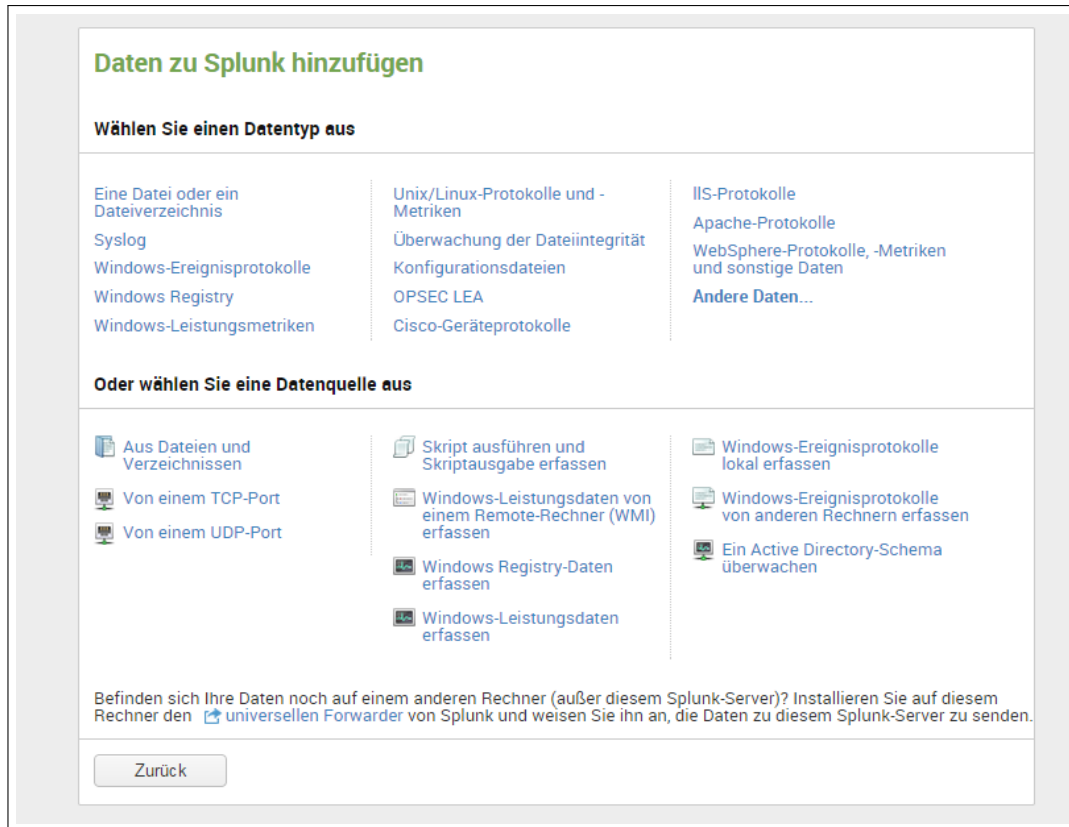


Figure 5.13. [Splunk] Importing data using the web interface

Installation and Setup

After successful registration on the website, Splunk can be downloaded as an executable file. The installation process is simple and demands the user only to click a few times “next”. Once Splunk is installed the browser will open a new window with a login screen. After login, Splunk is ready to use.

Before starting, one might want to check for useful apps to install these first. For example a DB connector such as the “Splunk DB Connect” app.

Installing apps can be done in two ways. Either by searching for apps via the web UI, or by downloading them from the Internet and importing them also by using the web UI.

5. Tool Evaluations

Data Visualization

User data is visualized in dashboards. Each dashboard is highly customizable in terms of placement. A Dashboard is made up by (optional) widgets and user charts which can be replaced by clicking on an element and dragging it to the desired position. Existing charts can be edited inside the dashboard, for example to change the date range of a query. However adding new charts must be generated manually by creating a new search as shown in figure 5.14.



Figure 5.14. [Splunk] Using the Search bar to run a query

As far as chart types go the amount of choices offered by Splunk, while reasonable, is not particularly large. The user may choose between the following chart types:

- Statistical table
- Line chart
- Area chart
- Column chart
- Bar chart
- Pie chart
- Scatter chart
- Few other which are not interesting in terms of this study such as Maps or Filling Charts

The customizability of each chart is not stunning at all. A few charts come with multiple

modes, for example the Bar Chart contains 3 modes: “no stacking”, “stacking” and “stacking to 100%”. Also a few charts automatically add additional information, e.g. a pie chart will add a percentage for each entry, compared to the total amount.

However x-values can't be edited at all. As for y-values, scaling can be changed between linear and logarithmic. Also an interval and minimal/ maximal value can be set.

Filters can be added via a dashboard widget or by editing the search query of a chart. It seems that filter of the same type which are “hard-coded” inside the query have higher priority than filters from a widget which is connected to the chart. For Example: Having a chart of some values for the years 2011 and 2012. Adding a new “time” widget and setting the time range from 1.1.2011 to 31.12.2011 will not affected the chart at all. Hovering over filter elements in the chart will enable the gray-out filter as shown in figure 5.15, while hovering over a single entry will display more informations about it.

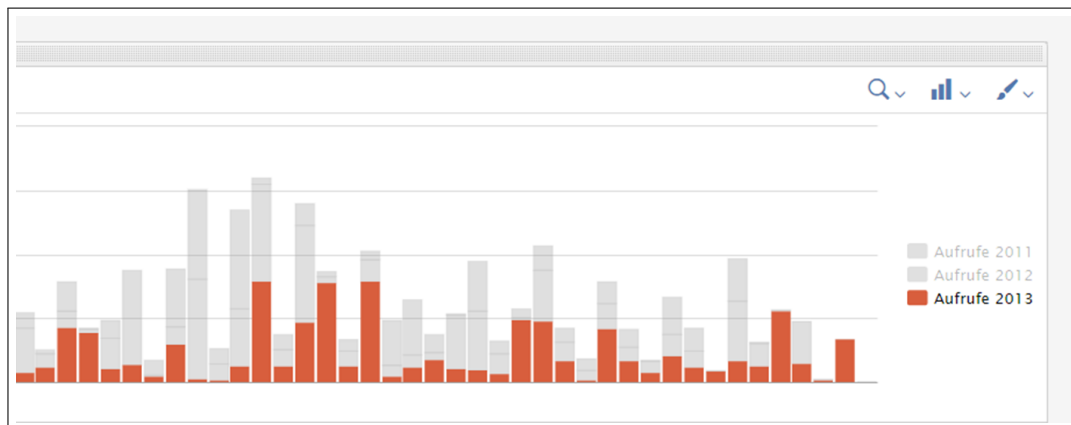


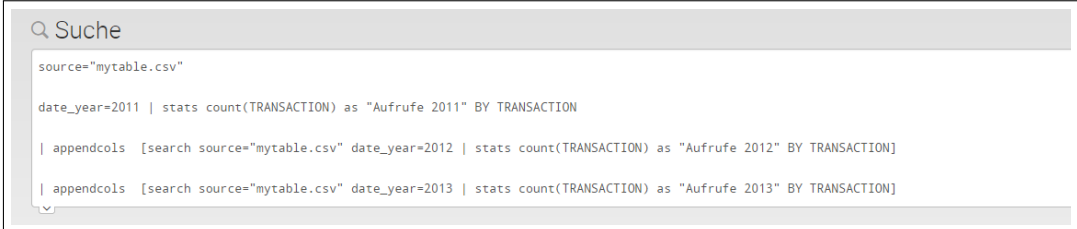
Figure 5.15. [Splunk] Hovering over "Aufrufe 2013" will gray-out all other filter elements

Data Aggregation

Unfortunately Splunk offers a little amount of predefined functions to aggregate data. The user can only choose between "average over time", "maximum value", "minimum value" and "top 10 entries" for a single field in the data set. The Splunk Search function is therefore obligatory. Using the Splunk Search for the first time, it will take quite a while of reading and searching to understand how this “language” works. Simple searches are still easy to learn. On the Splunk website, there is section for “SQL in Splunk” witch contains a table with a SQL to Splunk-Search “translation”. However setting up more difficult searches will take a lot longer to learn since the SQL pattern will not match anymore and the Splunk documentation rather explains the very basics. Using an App which allows SQL

5. Tool Evaluations

queries is therefore very helpful. Especially for queries where values from multiple sources need to be compared. Figure 5.16 and figure 5.17 show a comparison between both query structures. (Note in figure 5.16 the command *appendcols* is only one possibility. There are more ways to get the same output).



```
Suche
source="mytable.csv"
date_year=2011 | stats count(TRANSACTION) as "Aufrufe 2011" BY TRANSACTION
| appendcols [search source="mytable.csv" date_year=2012 | stats count(TRANSACTION) as "Aufrufe 2012" BY TRANSACTION]
| appendcols [search source="mytable.csv" date_year=2013 | stats count(TRANSACTION) as "Aufrufe 2013" BY TRANSACTION]
```

Figure 5.16. [Splunk] Example of Splunk Search language



```
Suche
| mysqlquery spec=MyDb query="SELECT * FROM MyTable;"
```

Figure 5.17. [Splunk] Using an SQL Connector

Data Reporting

The major benefit of Splunk is that almost everything can be done automatically. Automatic reports, as well as, searches and alerts can be set up using the search scheduler (see figure 5.18 for creating an alert). Also the interval is highly customizable. The user can chose between some basic intervals like daily or monthly up to "First Monday of each month, at 9am."

Alerts can observe specific values and execute a defined behaviour, if a condition is violated.

Create Alert

— 1 Save Search — 2 Set Up Alert — 3 Define Actions —

Send email Enable

Splunk Alert: \$name\$

support@splunk.com

To send email you must set a valid MTA in [Email alert settings](#).

Include search results as CSV inline as PDF

To send PDF's, [learn more](#) as PDF server.

Add to RSS Enable

RSS link displays after alert is created.

Run a script Enable

\$SPLUNK_HOME/bin/scripts/

Tracking Show triggered alerts in [Alert manager](#)

Cancel « Back Finish »

Figure 5.18. [Splunk] Setting up an alert

Final Thoughts

"People + Product + Passion = Splunk", is written on the *aboutus* page of Splunk. And it really doesn't feel exaggerated. Splunk is a great and powerful tool. It feels well thought through and offers the user a lot of useful and highly customizable functions. There are only 3 things one could criticize about: The price, the chart visualization and the 1-lined search bar. There might be an alternative search app which allows the user to write queries

5. Tool Evaluations

like you would expect in a compiler with highlighted keywords, however we haven't one.

5.6 Tabular Synopsis

	SiSense	QlikView	Birst	Splunk
Development Status	+	+	+	+
License Cost	?	-	?	-
Product Support	+	+	+	+
Available Documentation	+	~	-	+
Trial version features	+	+	-	+
Supported Operating Systems	~	~	+	+
Supported Import Data Formats	+	+	+	+
Chart Type Diversity	+	+	+	~
Chart Data Selection	~	+	+	+
Chart Malleability	~	+	+	~
Ease Of Data Selection	~	+	+	~
Scope Of Offered Functions	-	+	+	~
Versatility of Formula Application	-	+	+	+
Supported Data Export Formats	~	~	+	+
Reporting Automation	+	-	+	+
Report Sharing	+	-	+	+

Final Recommendation

For this final recommendation let us slim down our criteria catalogue such that only two broad metrics remain: On the one hand there is *power* - a tool's ability to apply functions to and analyse and aggregate interconnected data sets and to visualize the results in meaningful ways. On the other hand there is *usability* - defining how easily the former may be accomplished, how difficult it would be to adjust a new dataset or a new use case and how quickly a tool's behaviour can be understood and how much sense it makes. One usually comes at the expense of the other, hence we are looking for a tool that maximizes both, while being well balanced. Thus equipped it becomes much easier to give each tool a final assessment.

Graylog and Logstash score a zero on the usability scale. We were not even able to get them to function in any meaningful way.

SiSense offers good, on the import side even great, usability, but its power is unable to keep with the demands imposed by the reports in our criteria catalogue.

Next up is Splunk, a well rounded candidate, capable of executing the tasks we threw at it. However it is neither as powerful nor as simple as Birst or QlikView. Its visualisation capability is somewhat meagre compared to these two, while aggregation oftentimes has to take place by writing out scripts by hand.

The final decision is now to be made between QlikView and Birst and, to some degree, between power and usability. QlikView offers the former in spades, such that we feel justified in calling it the Microsoft Excel of analysis tools. But this power comes at the price of a steep learning curve. Anything scenario beyond elementary use cases will require the consultation of either QlikView's documentation or its community. The latter is of course a feature in its own right that only Splunk is capable of keeping up with.

Birst on the other hand offers a good visualisation capability alongside a guided function and filter creation that allows for a nearly non-existent learning curve. While Birst is still not quite as powerful as QlikView, the difference is unlikely to matter in but a very few edge cases, if at all. Mind that Birst's ease of use might after all be a necessity as there exists hardly any public documentation and no online community to turn to for help.

Our final recommendation is thus a draw between Birst and QlikView as we deem both tools to be capable of dealing with the kind of problems we defined in our criteria catalogue.

Declaration

I declare that this thesis is the solely effort of the author. I did not use any other sources and references than the listed ones. I have marked all contained direct or indirect statements from other sources as such. Neither this work nor significant parts of it were part of another review process. I did not publish this work partially or completely yet. The electronic copy is consistent with all submitted copies.

place, date, signature