Motivation: I came across Splunk recently and tried out the reporting capabilities that were available without any significant amount of coding. of the kind that is real-world and therefore requires several hundred hours to develop, test and implement. This is typically the case for the one-size-meets-all, one software solves all problems type of software that is definitely not sold at Walmart type of software. What I found was a revelation. Why would anyone have to implement reporting using complex systems whose only differentiator is that they are natively integrated into their one-solution-for the world systems.

In order to perform a quick, pilot test I pared down some data from the transportation side and cleansed the data. I can say that the data is not from a company that is peddling healthy solutions to consumers. Neither does the company sell any state of the art devices or equipment.

I generated a file with about 1040 records and some 30 columns, that is from the shipment processing side and is a look at the load sizes to customers from specific distribution sites. The data also includes ship dates, times, product information and logistical units. Some of you will recognize the provenance of the data as far as the system is concerned. I have not chosen to hide that aspect for a specific reason that will be clear by the end of this pilot.

What follows is this quick journey through Splunk is an easy exposition of its capabilities. The pilot is not intended to be exhaustive in terms of exploring the complete spectrum of possibilities. Neither did I set out to carry out any performance testing on scale. On these points, I would venture as far as to state that none of the results that I experienced would lead me to believe that there would be degradation in the realms of possibility.

Citation: The tests were run on a demo version of the Splunk Enterprise system. This system was available for download from http://www.splunk.com. I also accessed the Splunk knowledgebase at docs.splunk.com for an understanding of the commands and the syntax. The book 'Big Data Analytics Using Splunk: Deriving Operational Intelligence from Social Media, Machine Data, Existing Data Warehouses, and Other Real-Time Streaming Sources by authors Peter Zadrozny and Raghu Kodali is a tremendous asset in understanding the depth and breadth of this system, this book provides more information than I could possibly provide in a summary pilot such as this.

Process: I ingested the data into Splunk. The source was a 'csv' file which originated from a backend ERP system that is used more than 100,000 installations worldwide. I digress from my main purpose when I inform my audience that it would take atleast a 150 hours to design, develop, test and implement a report such as the one I have attempted to produce in the pages that follow.

The data that was ingested included the following columns. The file included a reduced set of shipments over a three-week period.

Shipping Point	Reason fo	Planned S	Planned S	Sales Doc	Customer	Delivery D	Sales Doc	Sold-to pa	Name	Ship To Pa	Ship To Na	Total Pall	Pallets	Total Laye	Layers	Order Qua	Total Gro	Total Vol	Net value	Net Value	Document	Material	Descriptio	Status	Status	Schedule	Document Date

Figure 1 Metadata for the shipment file

The first step was to gain an understanding of the number of the products and the number of lines per product. The basic code was *'stats count by Material'*.

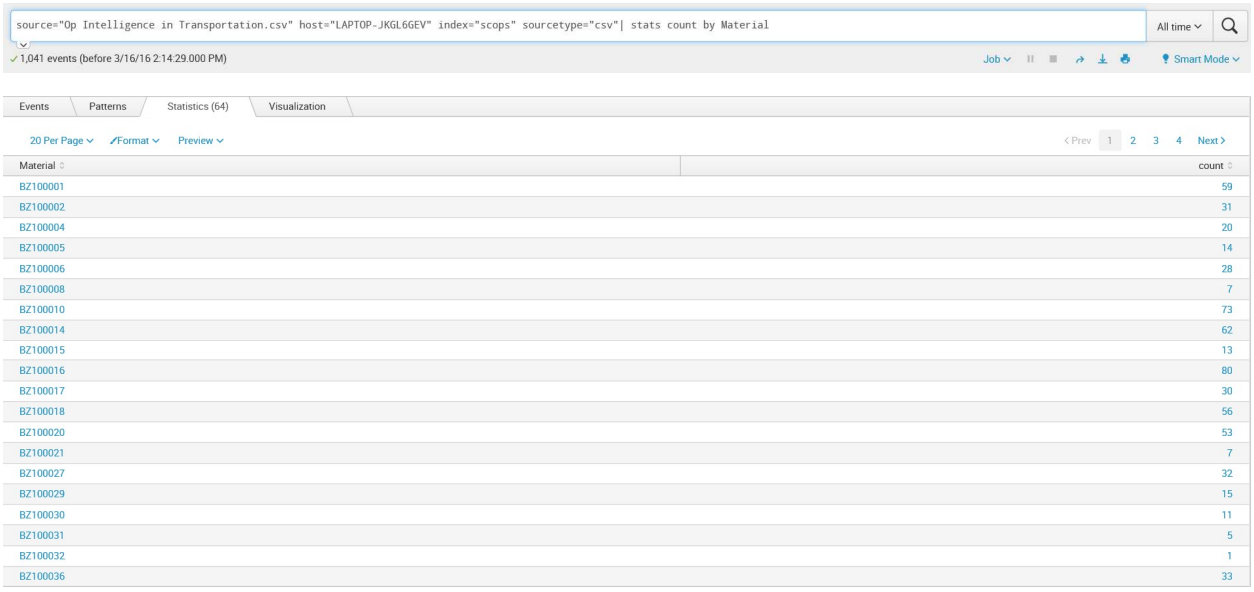


Figure 2 Products and lines per product

Following that baby step, it was time to total the number of lines and also sort the lines in an ascending order. This took an additional phrase '*addcoltotals*' followed by a *sort*.

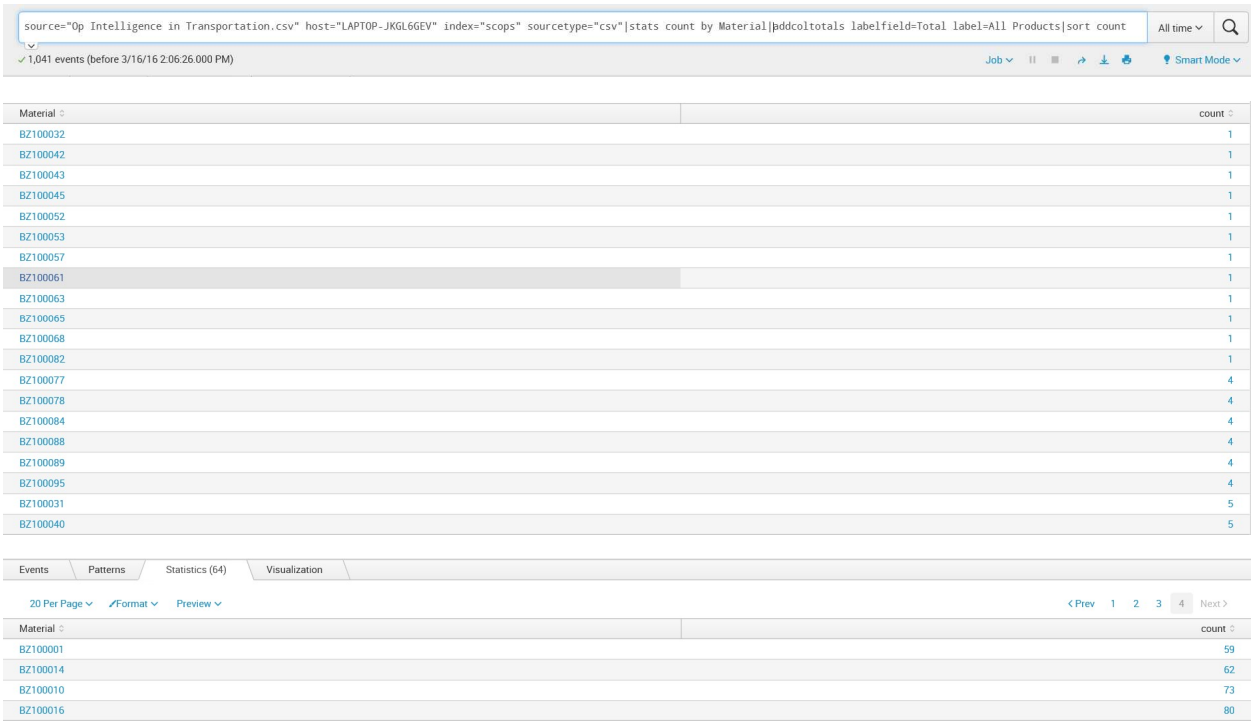


Figure 3 Total line count and sort by line count

At this point, the veteran in the transportation department who was used to ‘Products’ and could care less about ‘Materials’ got his wish.

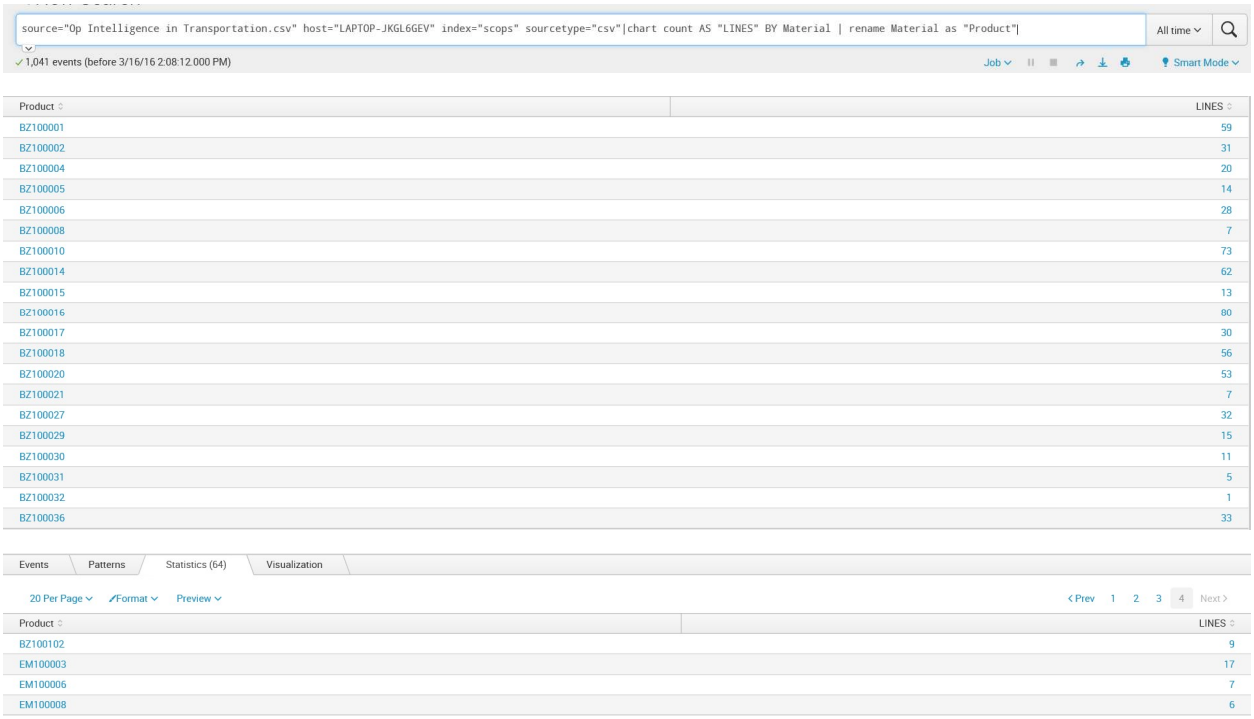


Figure 4 Renamed column labels to make it user friendlier!

As an added bonus, Splunk also delivered a bar chart, this was done without any additional scripting.

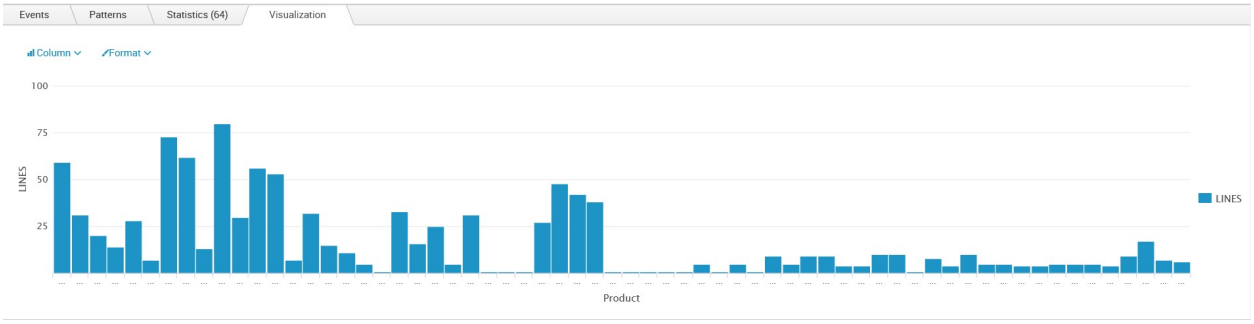
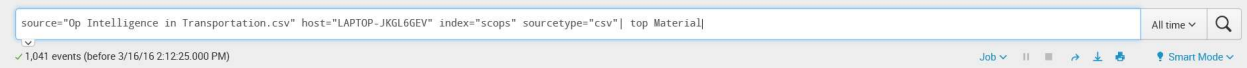


Figure 5 Bar chart of products and lines per product

The traffic supervisor was eager to know the top 10 products being shipped for the next 3 weeks, and this could have been because some products might have required special transport equipment. With that did I give away the industry? Let me know.

Before you say something like, 'But I would like to know my top 10 based on expiry dates or values or any other criteria'. My response is those criteria you should bring those over and Splunk can help you find your top ten charts.



Material	count	percent
BZ100016	80	7.684918
BZ100010	73	7.012488
BZ100014	62	5.955812
BZ100001	59	5.667627
BZ100018	56	5.379443
BZ100020	53	5.091258
BZ100049	48	4.610951
BZ100050	42	4.034582
BZ100051	38	3.650336
BZ100036	33	3.170029

Figure 6 Top 10 products by number of lines

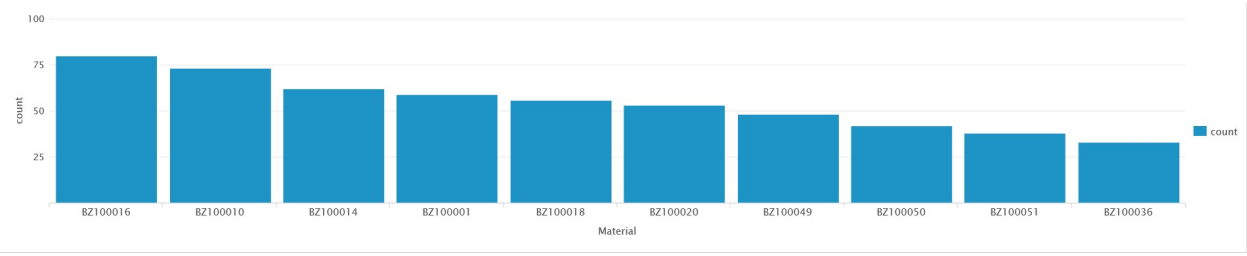


Figure 7 Bar chart for top 10 products

Ok, says the warehouse guy, 'I need to know the number of pallets to prepare, stage and load, lines do not mean much to me'.

Material	Pallets	count
BZ100001	0.125	2
BZ100001	1	4
BZ100001	1.125	2
BZ100001	1.25	7
BZ100001	1.375	1
BZ100001	1.5	3
BZ100001	1.75	3
BZ100001	1.875	2
BZ100001	10	3
BZ100001	11	1
BZ100001	14	2
BZ100001	15	1
BZ100001	17	1
BZ100001	18	1
BZ100001	2	8
BZ100001	21	1
BZ100001	22	1
BZ100001	25	1
BZ100001	26	3
BZ100001	3	4

Figure 8 Products by number of pallets

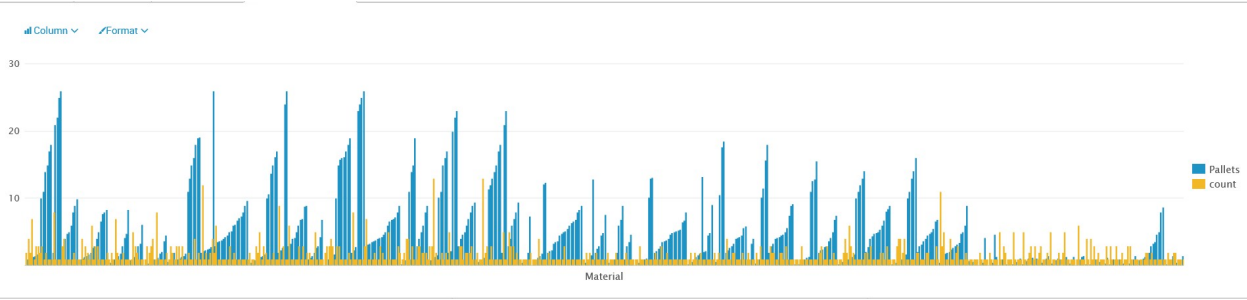
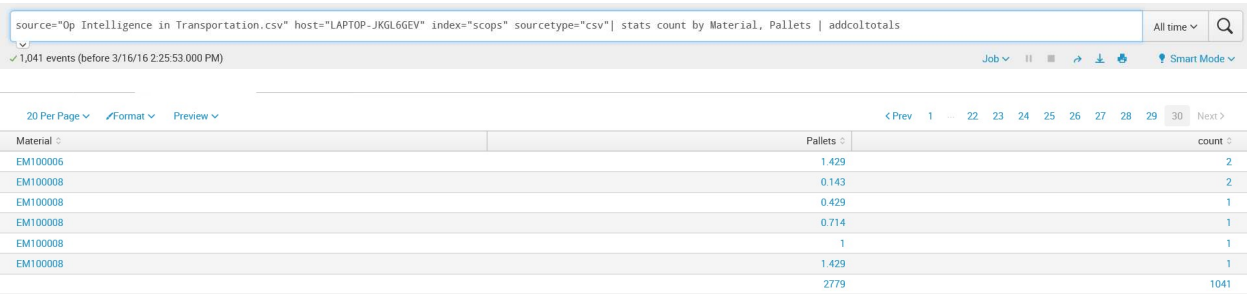


Figure 9 Column chart of pallets by product

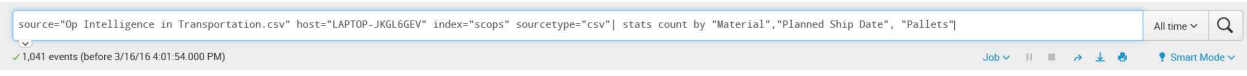
Splunk helps you do the quick arithmetic on the number of pallets or number of lines or number of anything you would like to get a picture of! All of the reporting so far has involved a basic phrase and nothing more.



Material	Pallets	count
EM100006	1.429	2
EM100008	0.143	2
EM100008	0.429	1
EM100008	0.714	1
EM100008	1	1
EM100008	1.429	1
	2779	1041

Figure 10 Totals by pallets and lines

All right, let us bring in the ship date and research pallets by Ship Date.



Material	Planned Ship Date	Pallets	count
EM100006	7/14/2013	0.143	1
EM100006	7/14/2013	0.286	2
EM100006	7/14/2013	0.429	1
EM100006	7/14/2013	0.571	1
EM100006	7/14/2013	1.429	2
EM100008	7/14/2013	0.143	2
EM100008	7/14/2013	0.429	1
EM100008	7/14/2013	1	1
EM100008	7/14/2013	1.429	1
EM100008	7/16/2013	0.714	1

Figure 11 Pallets by Ship Date

How do I know the customer, the day on which it needs to ship and the size of the load? That was easy, with the equivalent of a simple phrase, Splunk is also provided with the list of fields and we have our result.

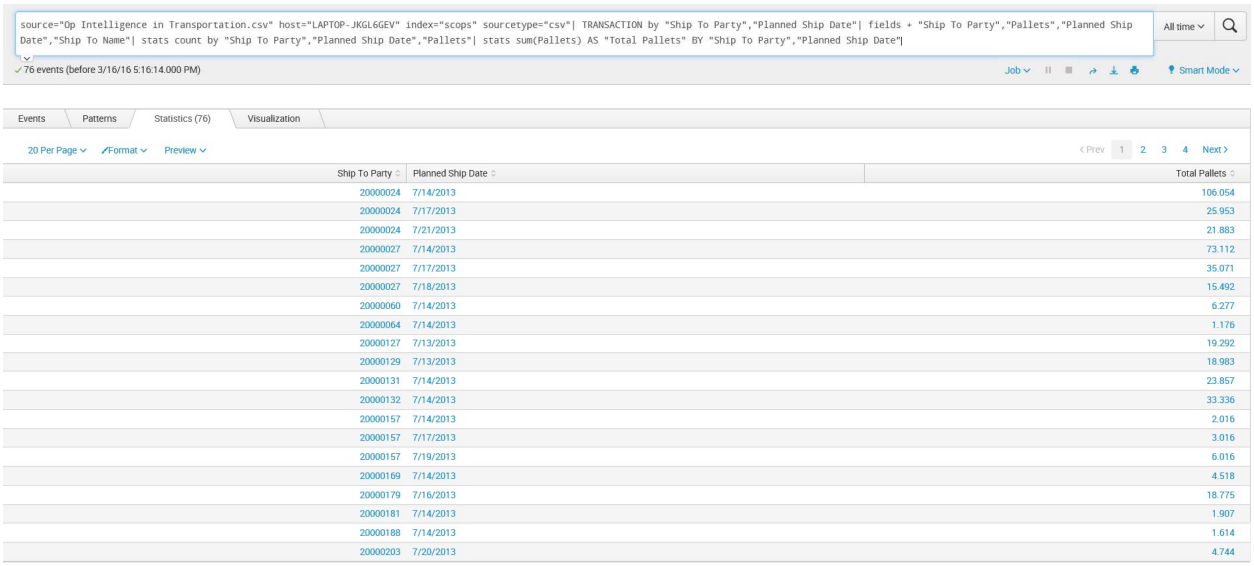


Figure 12 Ship to, Ship date and Pallets

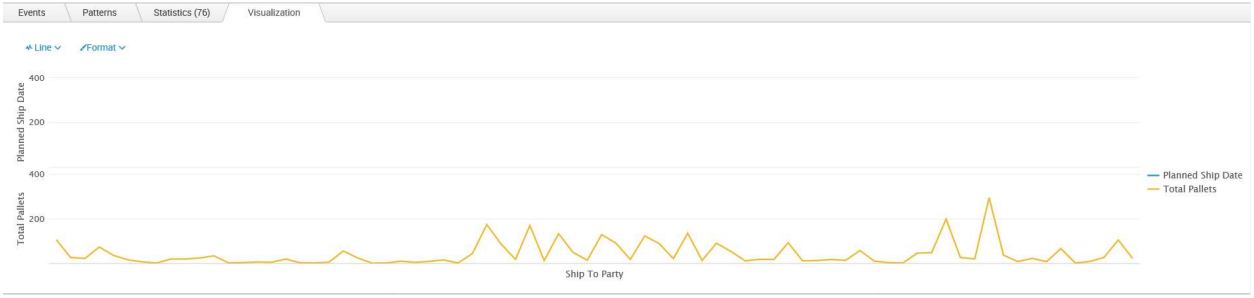


Figure 13 Line chart by ship to and load size

Another slice at the data, this time trying to find loads by ship dates, this helps me take a quick look at capacity by day.

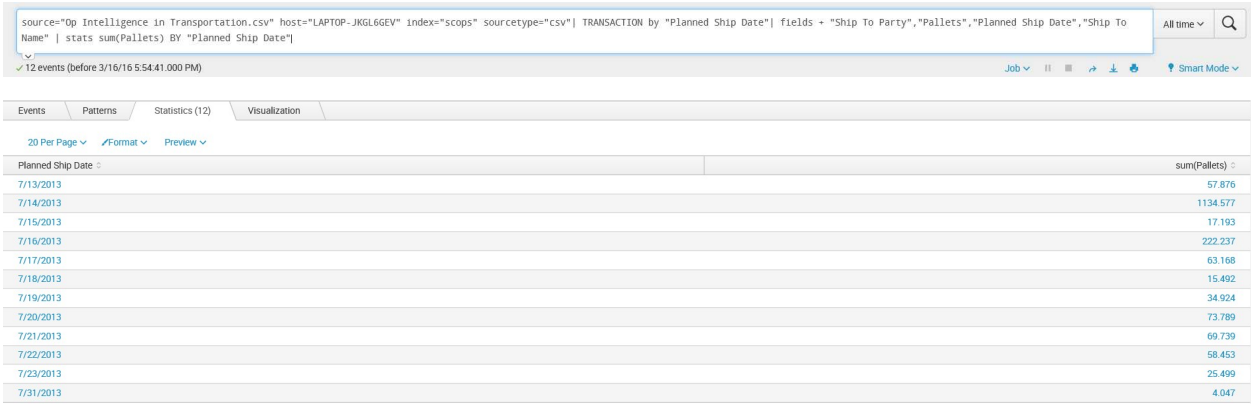


Figure 14 Loads by ship dates

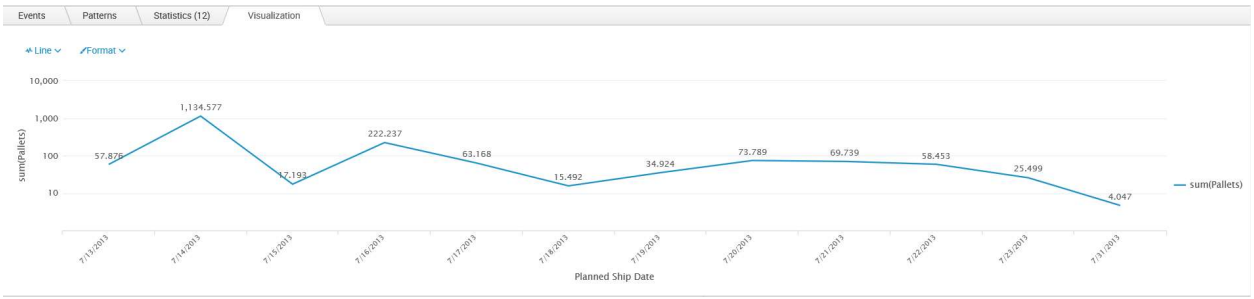
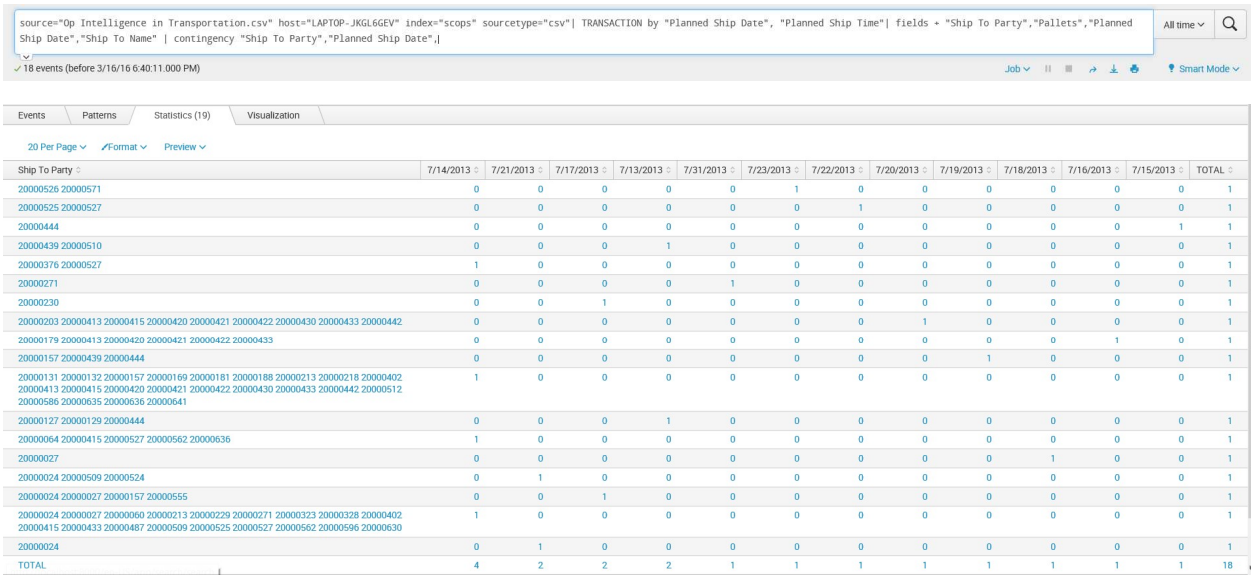


Figure 15 Line chart of loads by ship date

Did someone ask for a quick table of shipments by dates, we have that as well. Again this did not take more than a simple phrase.



Ship To Party	7/14/2013	7/21/2013	7/17/2013	7/13/2013	7/31/2013	7/23/2013	7/22/2013	7/20/2013	7/19/2013	7/18/2013	7/16/2013	7/15/2013	TOTAL
20000526 20000571	0	0	0	0	0	1	0	0	0	0	0	0	1
20000525 20000527	0	0	0	0	0	0	1	0	0	0	0	0	1
20000444	0	0	0	0	0	0	0	0	0	0	0	1	1
20000439 20000510	0	0	0	1	0	0	0	0	0	0	0	0	1
20000376 20000527	1	0	0	0	0	0	0	0	0	0	0	0	1
20000271	0	0	0	0	1	0	0	0	0	0	0	0	1
20000230	0	0	1	0	0	0	0	0	0	0	0	0	1
20000203 20000413 20000415 20000420 20000421 20000422 20000430 20000433 20000442	0	0	0	0	0	0	0	1	0	0	0	0	1
20000179 20000413 20000420 20000421 20000422 20000433	0	0	0	0	0	0	0	0	0	0	1	0	1
20000157 20000439 20000444	0	0	0	0	0	0	0	0	1	0	0	0	1
20000131 20000132 20000157 20000169 20000181 20000188 20000213 20000218 20000402 20000413 20000415 20000420 20000421 20000422 20000430 20000433 20000442 20000512 20000586 20000635 20000636 20000641	1	0	0	0	0	0	0	0	0	0	0	0	1
20000127 20000129 20000444	0	0	0	1	0	0	0	0	0	0	0	0	1
20000064 20000415 20000527 20000562 20000636	1	0	0	0	0	0	0	0	0	0	0	0	1
20000027	0	0	0	0	0	0	0	0	0	1	0	0	1
20000024 20000509 20000524	0	1	0	0	0	0	0	0	0	0	0	0	1
20000024 20000027 20000157 20000555	0	0	1	0	0	0	0	0	0	0	0	0	1
20000024 20000027 20000060 20000213 20000229 20000271 20000323 20000328 20000402 20000415 20000433 20000487 20000509 20000525 20000527 20000562 20000596 20000630	1	0	0	0	0	0	0	0	0	0	0	0	1
20000024	0	1	0	0	0	0	0	0	0	0	0	0	1
TOTAL	4	2	2	2	1	1	1	1	1	1	1	1	18

Figure 16 Contingency table - Ship to by Ship date

Things get interesting, I do need to plan centrally and locally for each of my distribution centers. I can combine the shipments in such a way as to create self-contained groupings. My grouping is based on DC, Ship date and Ship time. This I have called my shift. Needless to say, any type of grouping is possible.

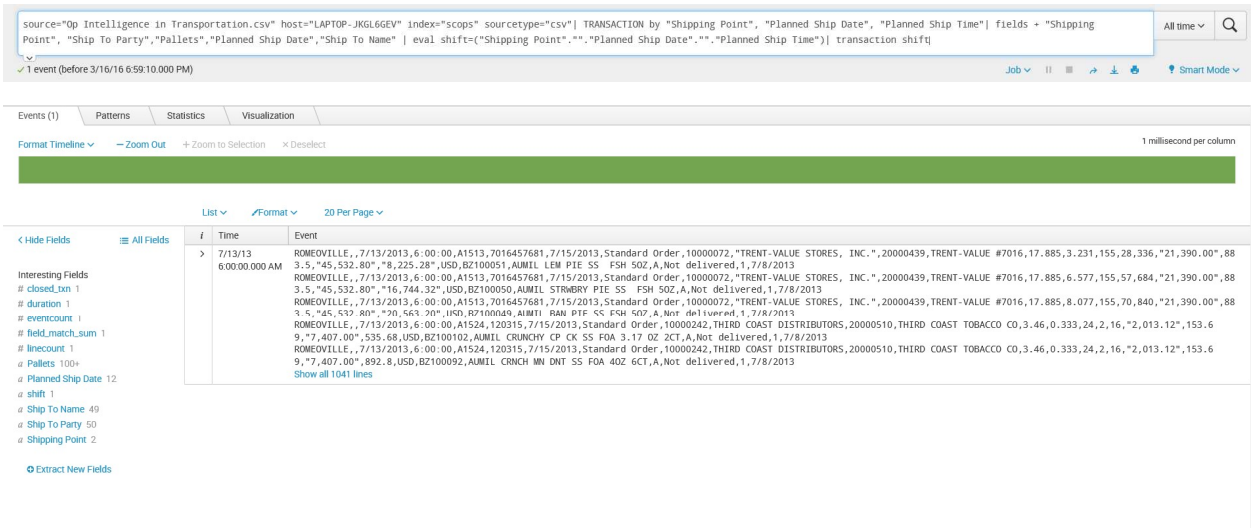


Figure 17 Grouping loads

Here is another slice, this time the loads from a specific DC.

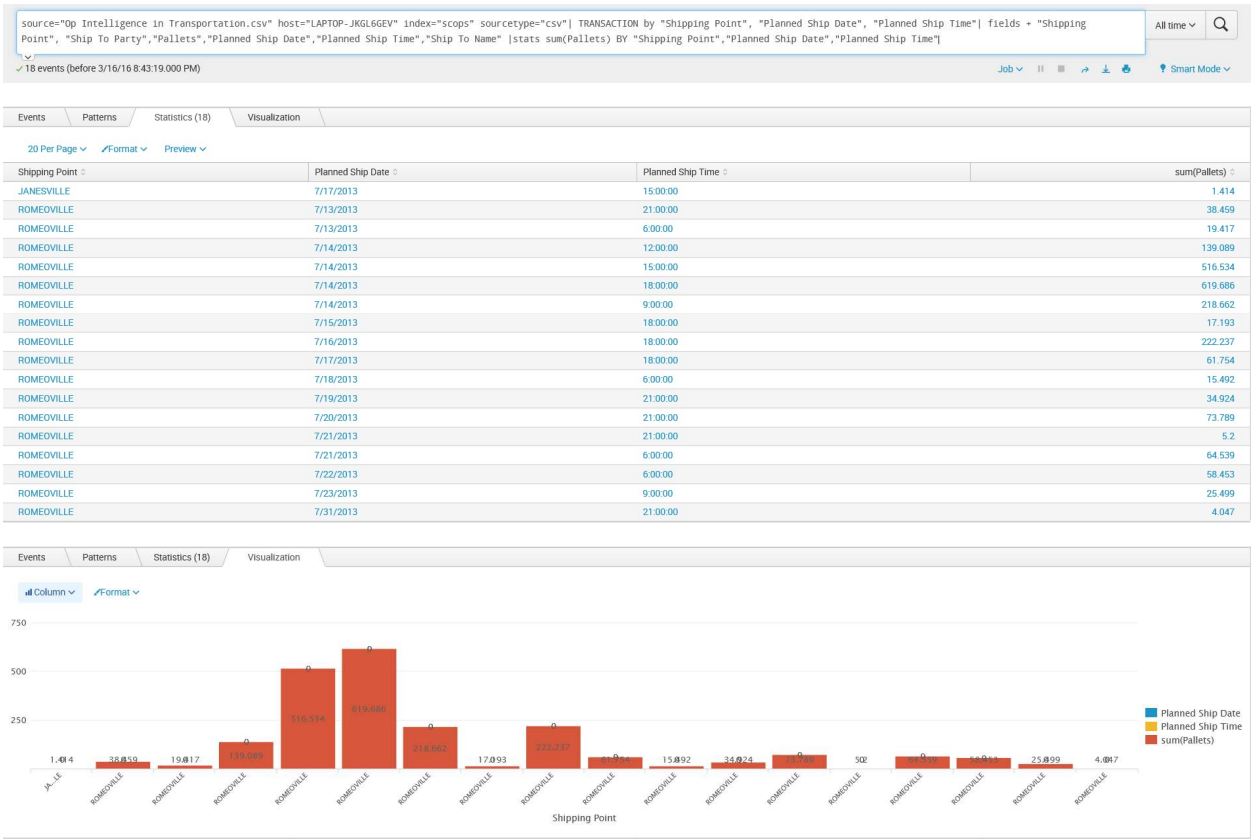
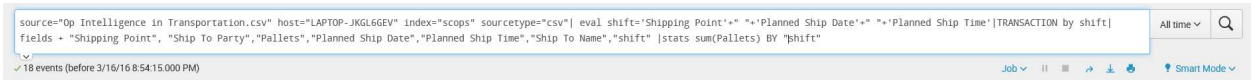


Figure 18 Column chart for loads from a DC

I decided to combine three columns to form my own grouping variable. This helps me look at loads by shift.



shift	sum(Pallets)
JANESVILLE 7/17/2013 15:00:00	1.414
ROMEOVILLE 7/13/2013 21:00:00	38.459
ROMEOVILLE 7/13/2013 6:00:00	19.417
ROMEOVILLE 7/14/2013 12:00:00	139.089
ROMEOVILLE 7/14/2013 15:00:00	516.534
ROMEOVILLE 7/14/2013 18:00:00	619.686
ROMEOVILLE 7/14/2013 9:00:00	218.662
ROMEOVILLE 7/15/2013 18:00:00	17.193
ROMEOVILLE 7/16/2013 18:00:00	222.237
ROMEOVILLE 7/17/2013 18:00:00	61.754
ROMEOVILLE 7/18/2013 6:00:00	15.492
ROMEOVILLE 7/19/2013 21:00:00	34.924
ROMEOVILLE 7/20/2013 21:00:00	73.789
ROMEOVILLE 7/21/2013 21:00:00	5.2
ROMEOVILLE 7/21/2013 6:00:00	64.539
ROMEOVILLE 7/22/2013 6:00:00	58.453
ROMEOVILLE 7/23/2013 9:00:00	25.499
ROMEOVILLE 7/31/2013 21:00:00	4.047

Figure 19 Loads by shift, by location

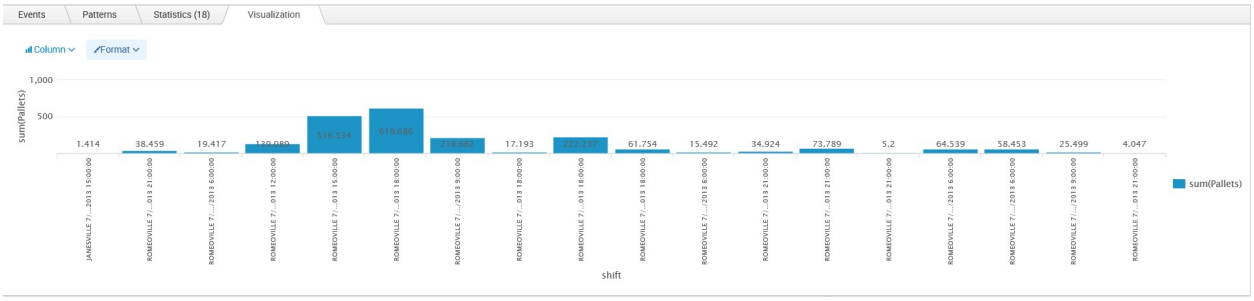


Figure 20 Column chart of loads by shift and location

Summary: I was duly impressed with the system. I had spent less than an hour and I was able to generate the equivalent of some twenty reports from a backend system with the corresponding data visualization. Even a BI system with dashboards would match this at best. The concept of operational intelligence takes on a new interpretation with a savvy user being able to generate reports on the fly. I did not show alerts in my pilot. Splunk does provide for alerts based on historical information from the recent past as well as the older information. In addition to this it is possible to generate real time alerts. For example, an alert can be configured to notify the traffic supervisor if the number of full loads on a given day exceeds a certain limit as this means that additional trucks will be required. The number of pallets can be used to inform shift supervisors so that they can plan for the right resources. If there are shipments with a ship date in the past an alert can be issued to determine the causes and to take appropriate action. Splunk is definitely a candidate in the space of operational intelligence. Splunk can ingest files in an automated fashion from specific locations. The reports that we have looked at can be saved and applied periodically to the newly ingested files, which in turn can provide alerts that are useful in making timely decisions that matter for the business.

Footnotes¶

1) **Splunk** Inc. provides the leading platform for Operational Intelligence. Customers use **Splunk** to search, monitor, analyze and visualize machine data.