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.



Figure 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’*.

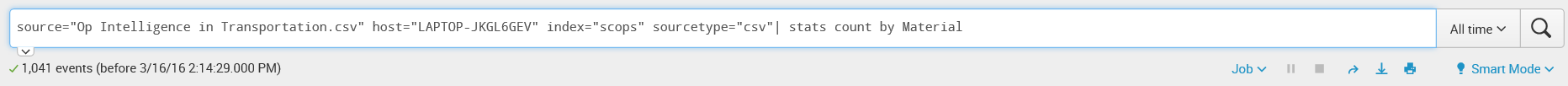
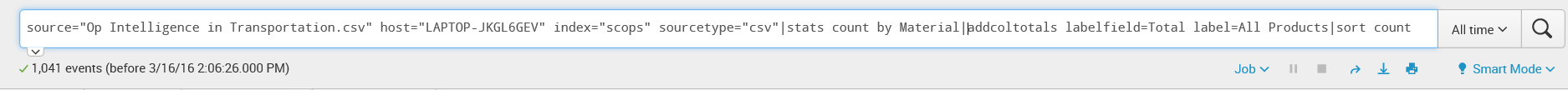
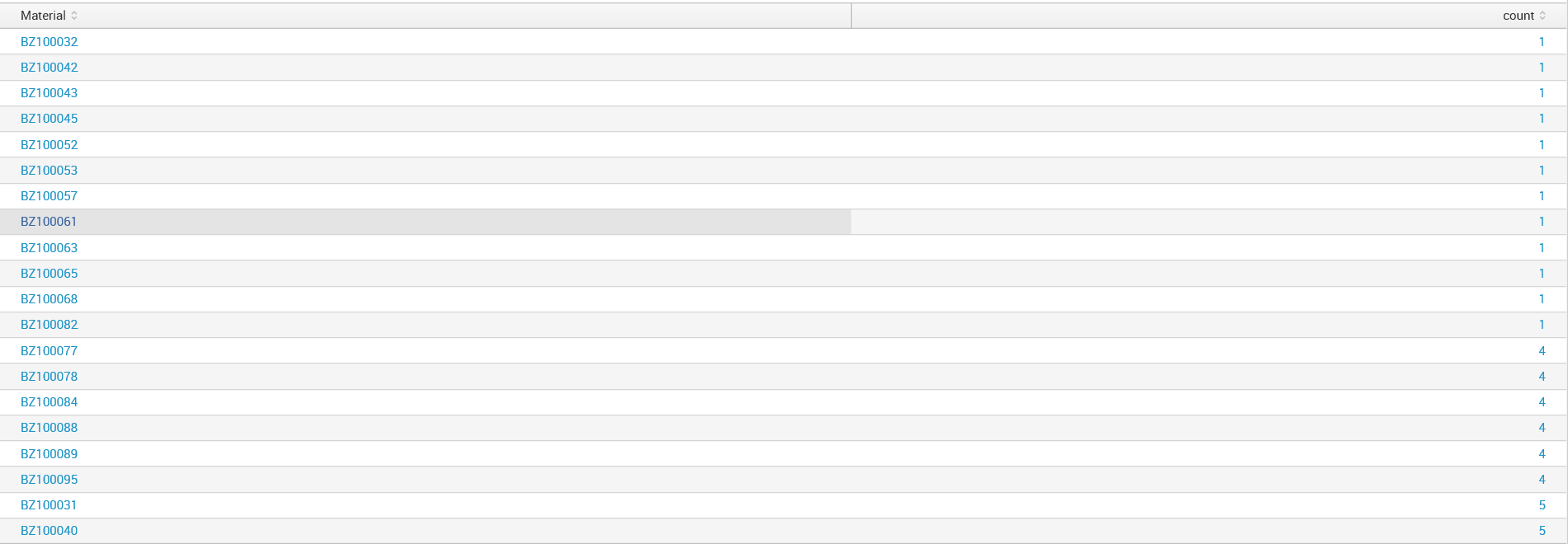




Figure Products and lines per product

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*.





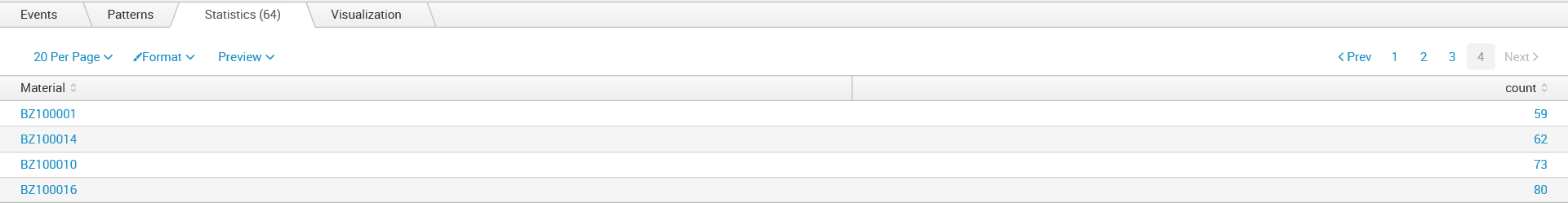
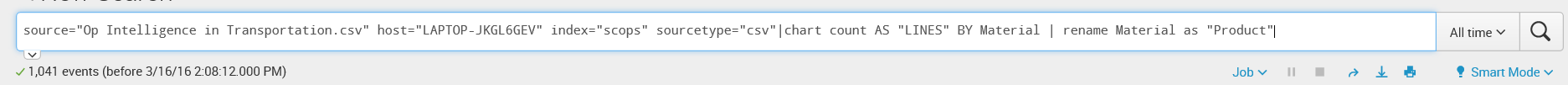


Figure Total line count and sort by line count

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





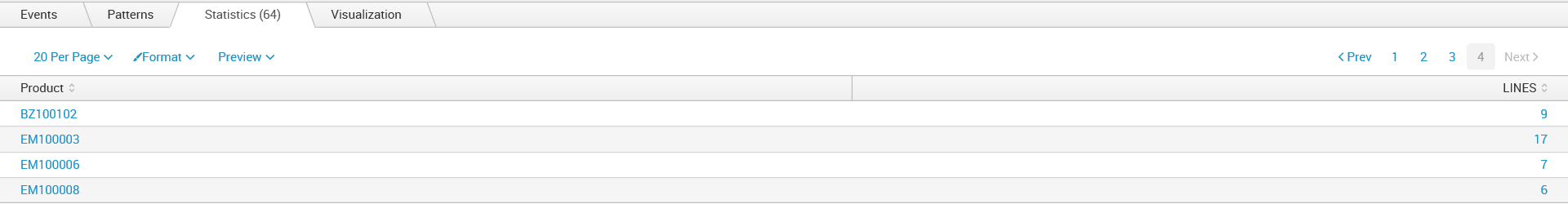


Figure Renamed column labels to make it user friendlier!

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

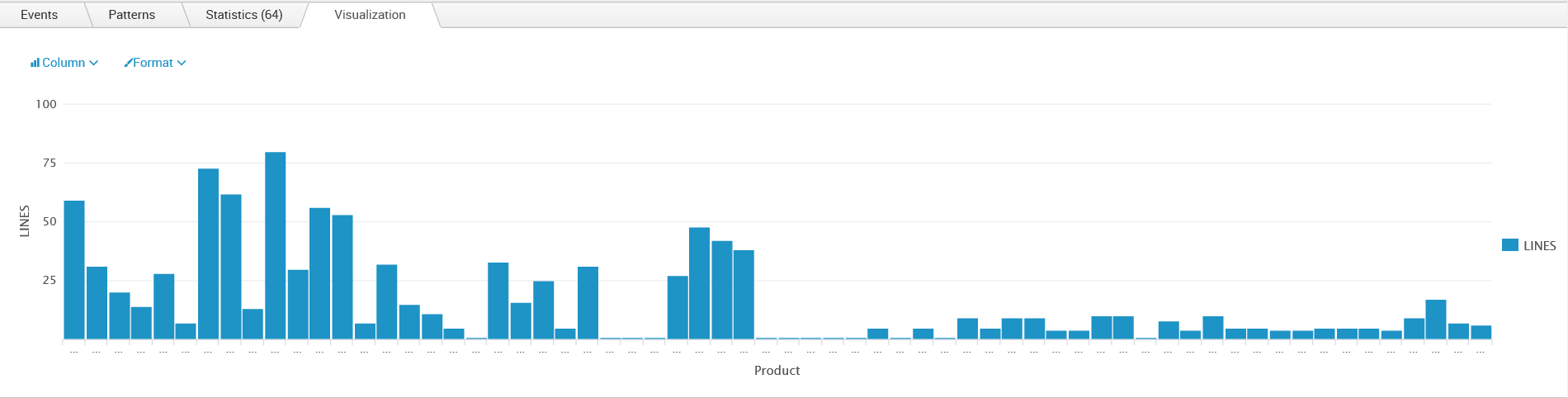


Figure Bar chart of products and lines per product

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.

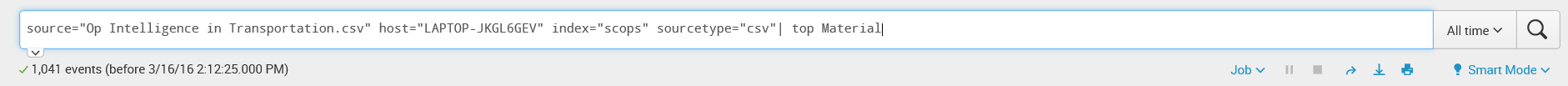




Figure Top 10 products by number of lines

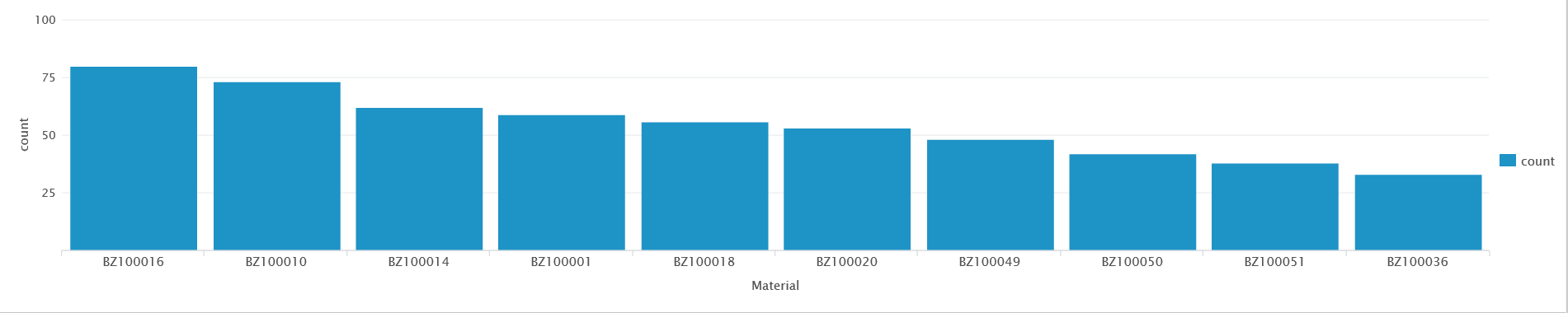
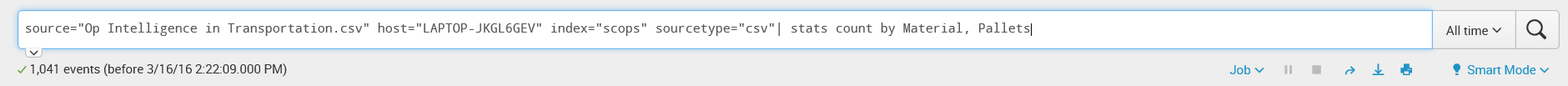


Figure Bar chart for top 10 products

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’.



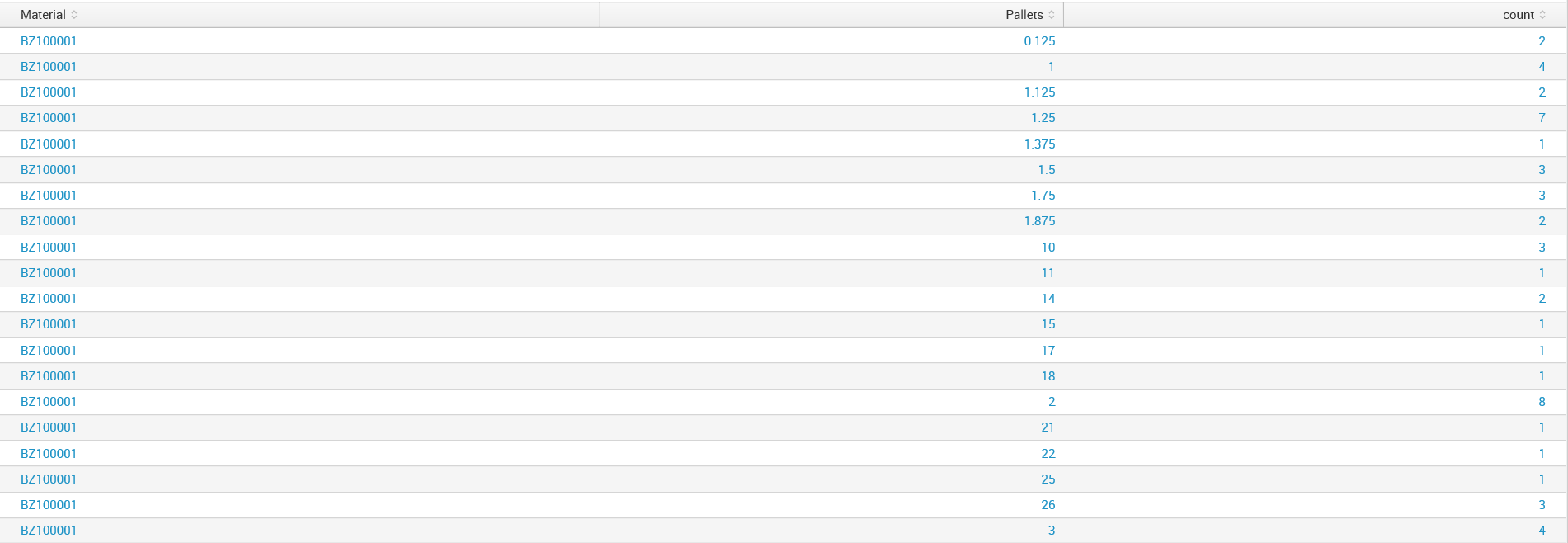


Figure Products by number of pallets

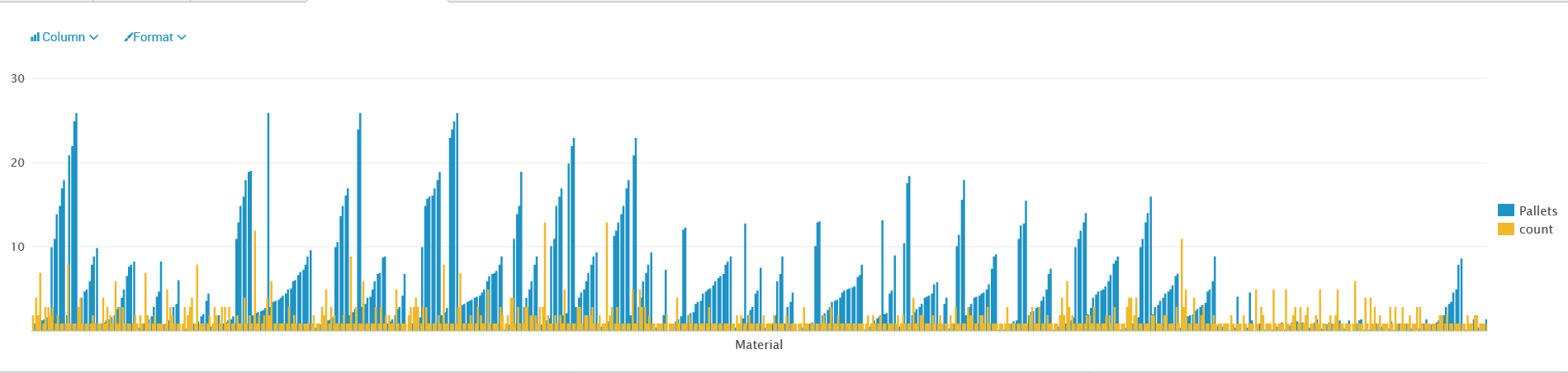


Figure Column chart of pallets by product

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.

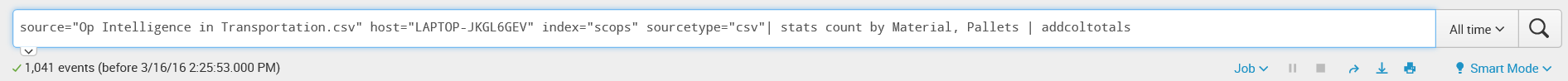
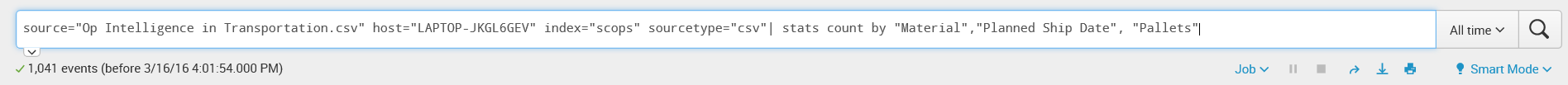




Figure Totals by pallets and lines

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



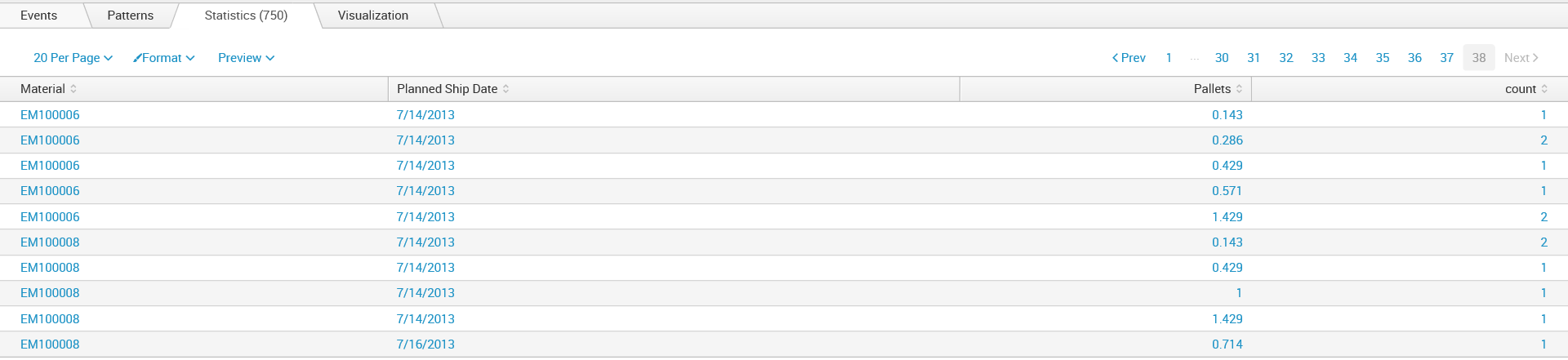
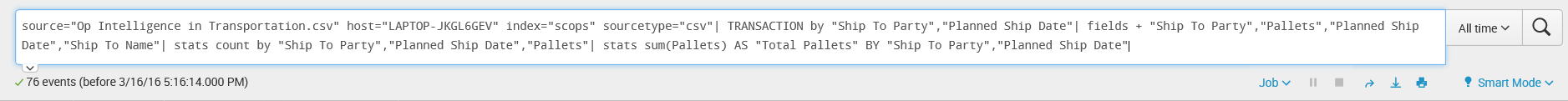


Figure Pallets by Ship Date

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.



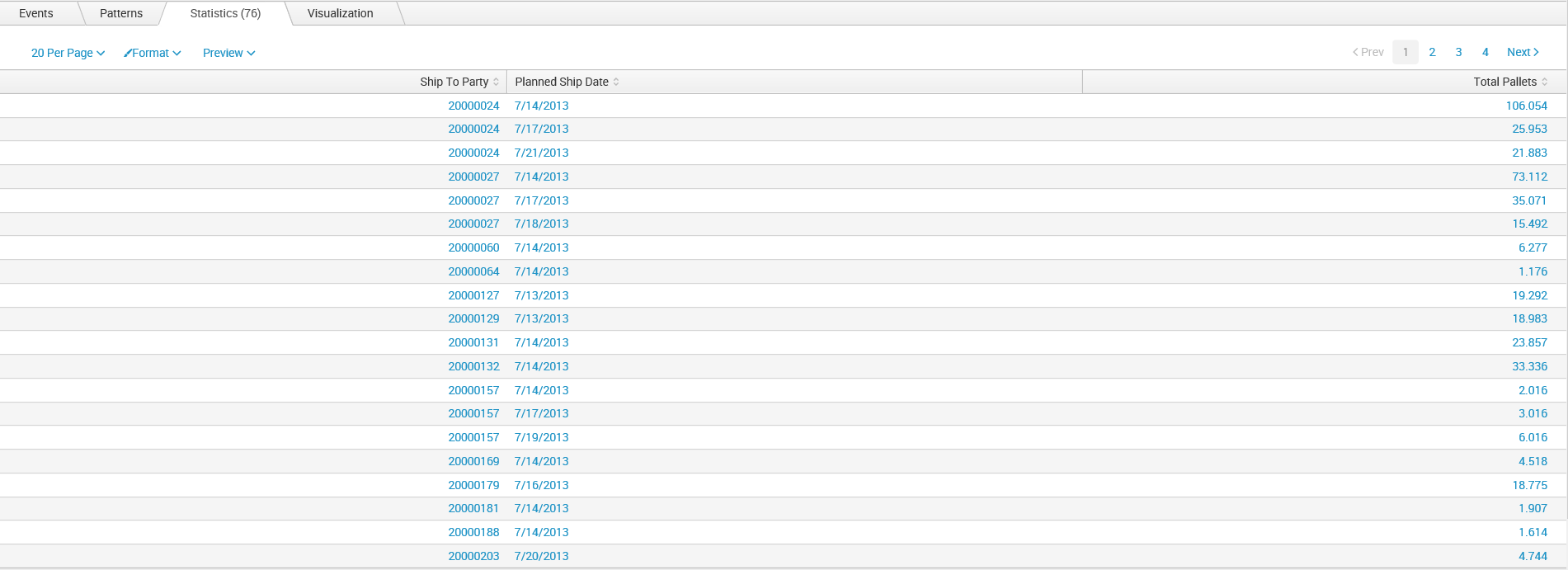


Figure Ship to, Ship date and Pallets

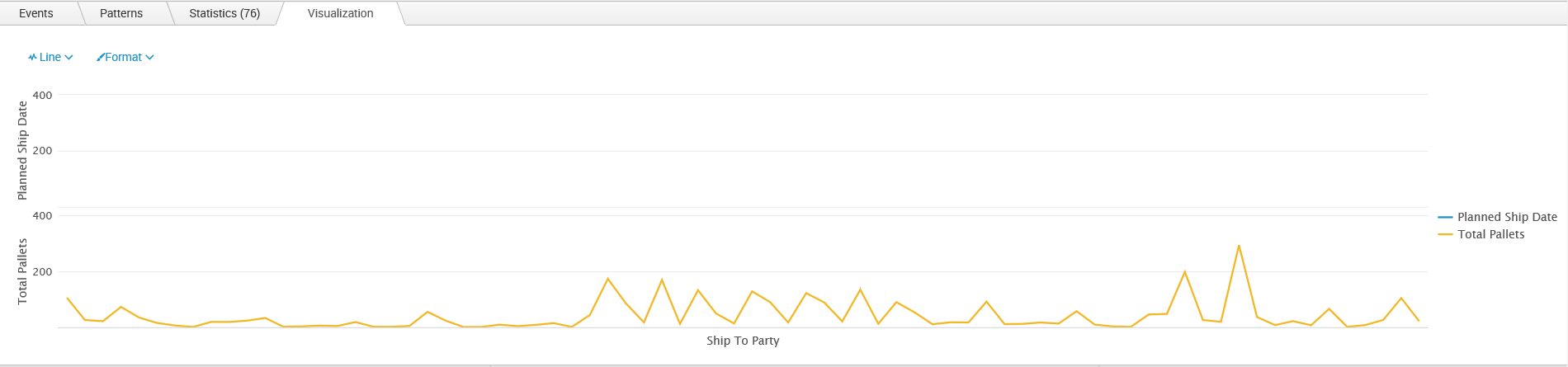
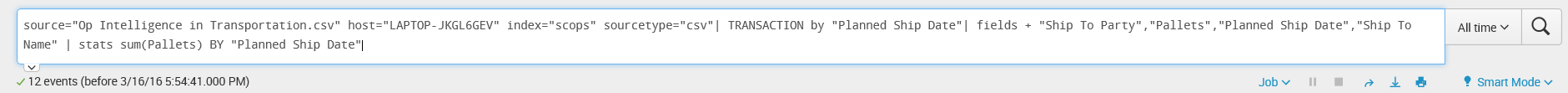


Figure Line chart by ship to and load size

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.



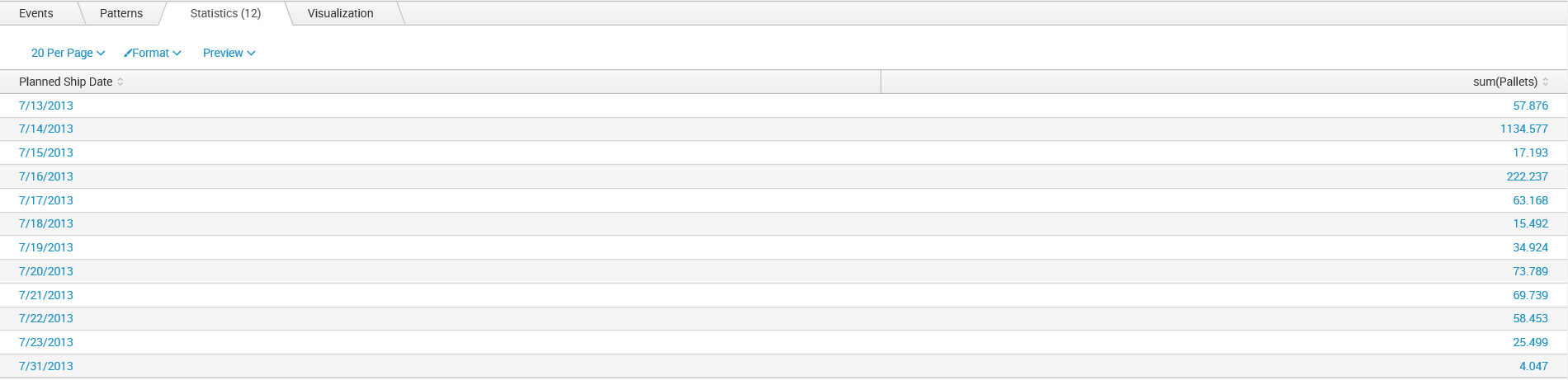


Figure Loads by ship dates

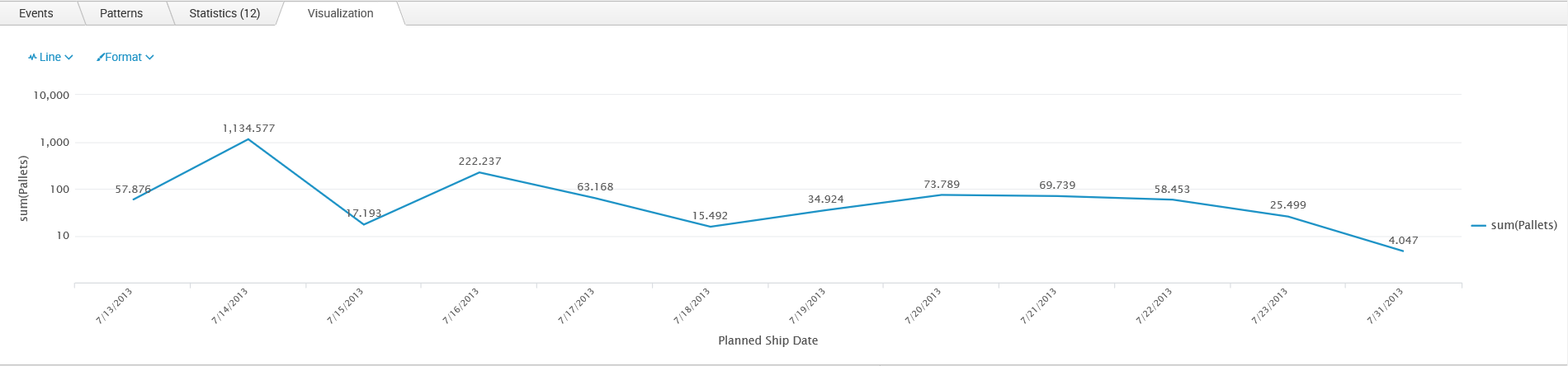
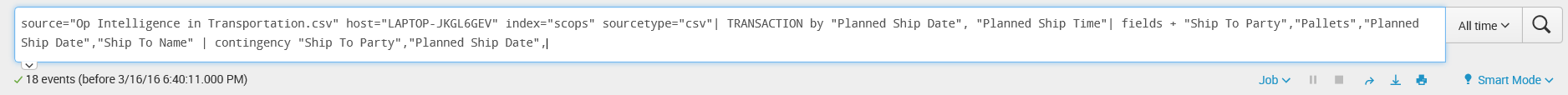


Figure Line chart of loads by ship date

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.



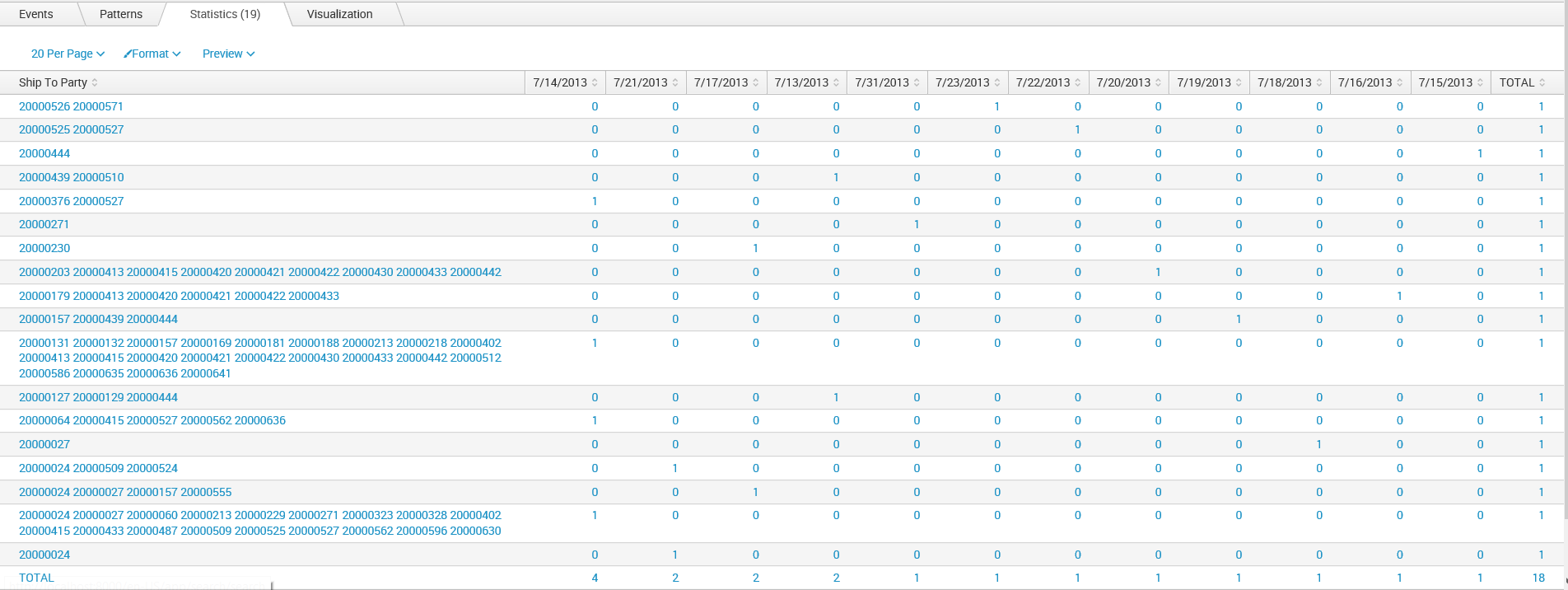
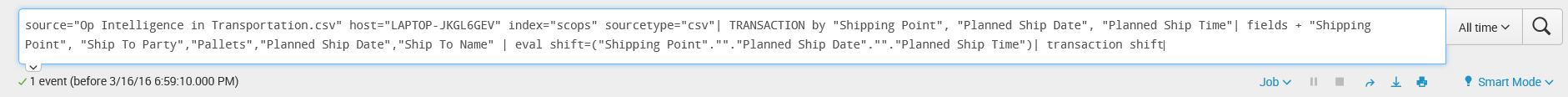


Figure Contingency table - Ship to by Ship date

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.



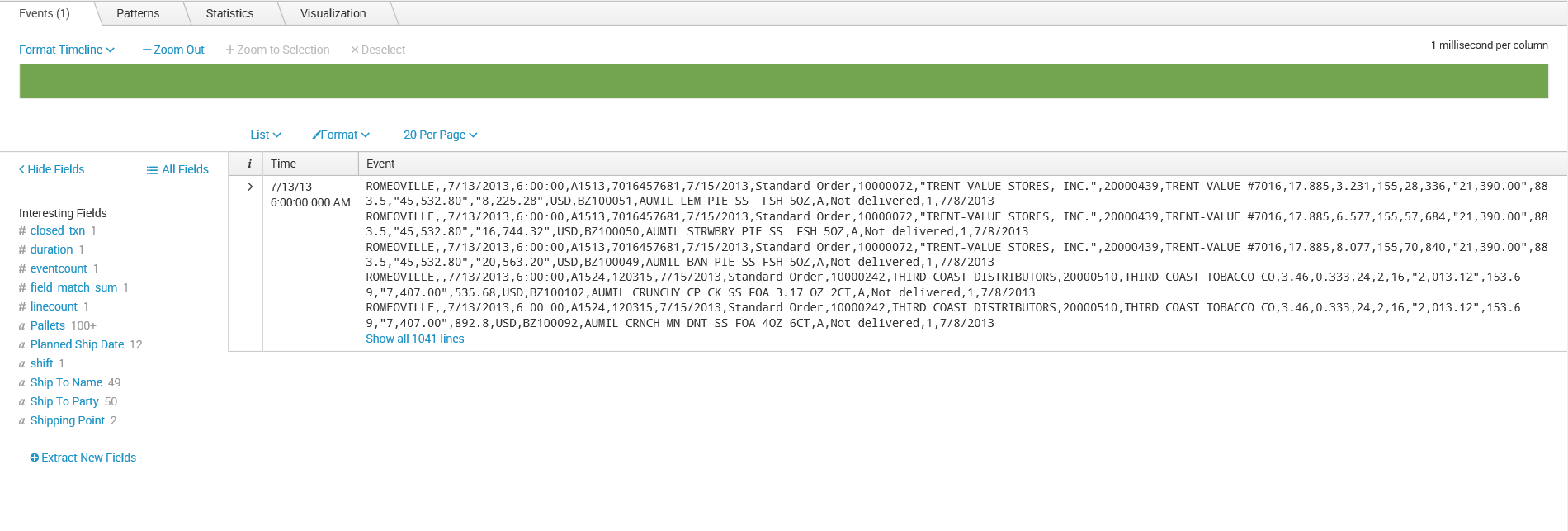
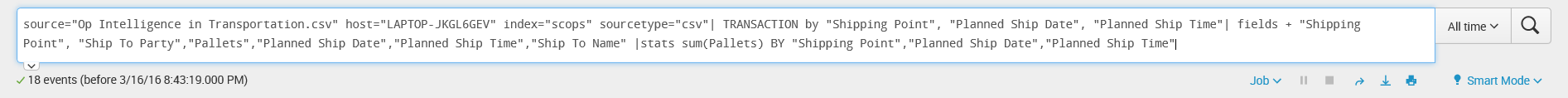
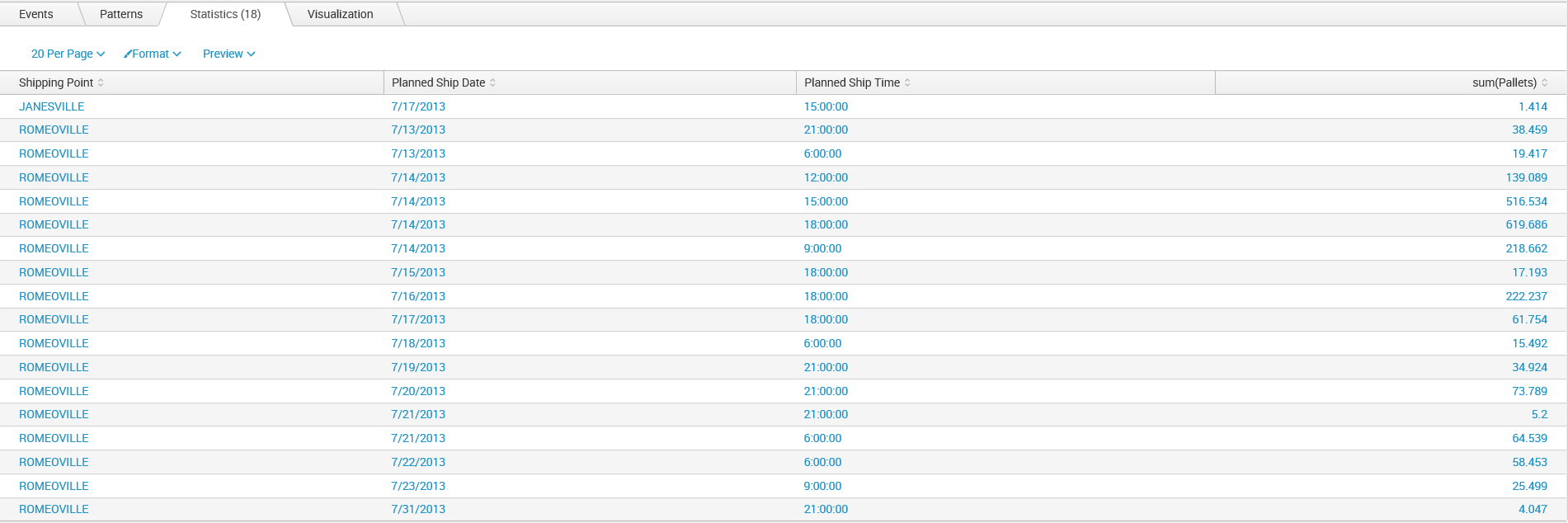


Figure Grouping loads

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





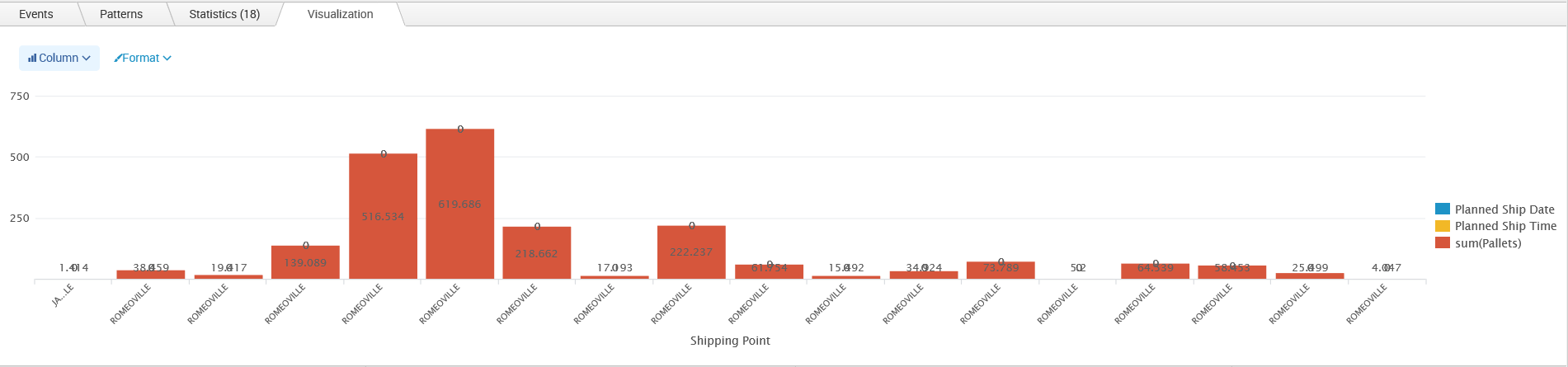


Figure Column chart for loads from a DC

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

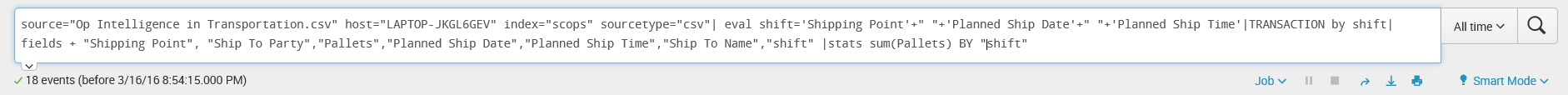




Figure Loads by shift, by location

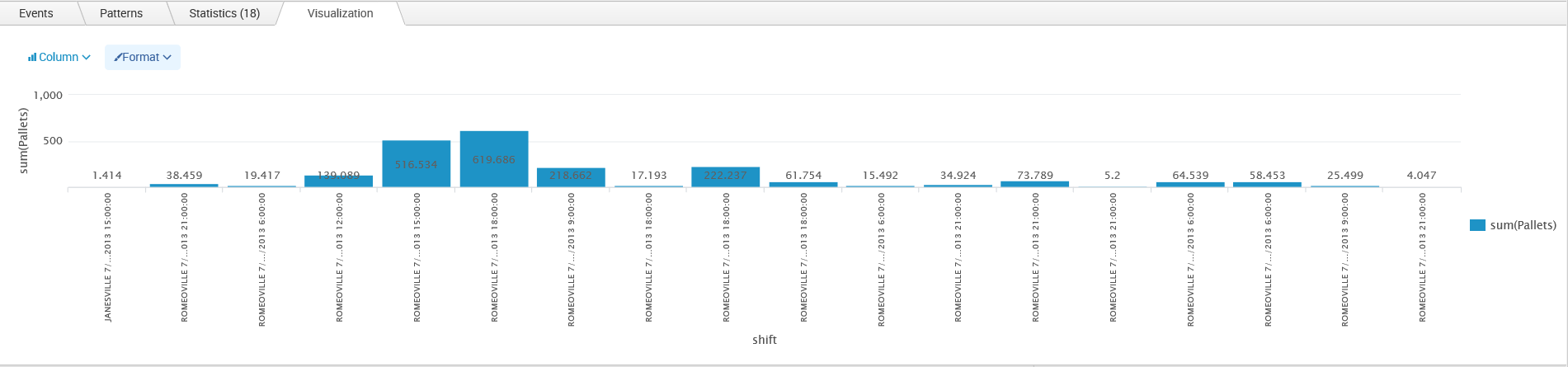


Figure Column chart of loads by shift and location

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.