Fall 2015

Big Data Insights Using Analytics

Naga Krishna Reddy Muppidi
Governors State University

Sai Kiran Merugu
Governors State University

Khambhampati Pramod
Governors State University

Follow this and additional works at: http://opus.govst.edu/capstones

Part of the Databases and Information Systems Commons

Recommended Citation
http://opus.govst.edu/capstones/150

For more information about the academic degree, extended learning, and certificate programs of Governors State University, go to http://www.govst.edu/Academics/Degree_Programs_and_Certifications/

Visit the Governors State Computer Science Department

This Project Summary is brought to you for free and open access by the Student Capstone Projects at OPUS Open Portal to University Scholarship. It has been accepted for inclusion in All Capstone Projects by an authorized administrator of OPUS Open Portal to University Scholarship. For more information, please contact opus@govst.edu.
Table of Contents

1.1 Project Abstract ................................................................................................................... 1
1.2 Competitive Information ..................................................................................................... 1
1.3 Relationship to Other Applications/Projects .................................................................... 1

2. About Splunk Enterprise ...................................................................................................... 1
  2.1 What is Splunk Enterprise ................................................................................................. 1
  2.2 Splunk Enterprise features ............................................................................................... 1
  2.3 About Splunk Enterprise users .......................................................................................... 3
  2.4 About Splunk Enterprise deployments ............................................................................. 4
      2.4.1 Splunk Enterprise and your IT infrastructure ............................................................. 4
      2.4.2 Splunk Enterprise Components ................................................................................ 5
  3. Getting started with alerts ................................................................................................ 6
      3.1 What is an alert? ............................................................................................................ 6
      3.2 Alert basics .................................................................................................................. 6
      3.3 About alert types .......................................................................................................... 7
      3.4 Choosing an alert type .................................................................................................. 8
      3.5 Managing alert frequency ............................................................................................ 8
      3.6 Using alert actions ...................................................................................................... 8
      3.7 Alert and alert action permissions ............................................................................... 9
      3.8 Scheduled reports and scheduled alerts are not the same ........................................... 9
  4. Visualization Reference ..................................................................................................... 10
  5. Accessing visualization definition features ...................................................................... 10
      Visualizations from Search ................................................................................................. 11
      5.1 Dashboard panel visualizations ..................................................................................... 11
  6. Events visualizations ......................................................................................................... 11
  7. Charts ................................................................................................................................ 12
      7.1 Column and bar charts ................................................................................................. 13
      7.2 Stacked column and bar charts ................................................................................... 14
  8. Open Issues ....................................................................................................................... 15
  9. Acknowledgements .......................................................................................................... 15
10. References ........................................................................................................................ 15
1. Project Description

1.1 Project Abstract

The main objective of this project is to find the data insights from the huge amount of data that is evolving around us day by day. In order to analyze the data we need an architecture that is suitable for all kinds of data that we see in 21st century. We are using SPLUNK architecture for analyzing the data and getting the insights that we need for taking better decisions. SPLUNK is google for datacenters. By using SPLUNK we can generate all kinds of DASHBOARDS, ALERTS, SCHEDULING, PIVOTS and a lot more important things that is very usable for managers to take a better decisions. We use SPL language for manipulating the data.

1.2 Competitive Information

Identify competitor products, applications, or services that this project will directly compete against, and whether your team has the potential to be the first to market this new project application or capability.

1.3 Relationship to Other Applications/Projects

Identify whether this project relates to other projects and/or other applications.

2. About Splunk Enterprise

2.1 What is Splunk Enterprise

Splunk Enterprise is a software platform to search, analyze, and visualize the machine-generated data gathered from the websites, applications, sensors, devices, and so on, that comprise your IT infrastructure or business.

After you define the data source, Splunk Enterprise indexes the data stream and parses it into a series of individual events that you can view and search.

You can use the search processing language or the interactive pivot feature to create reports and visualizations.

2.2 Splunk Enterprise features

The following table highlights seven Splunk Enterprise features. You can read about more features on splunk.com
<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Indexing</strong></td>
<td>Splunk indexes machine data. This includes data streaming from packaged and custom applications, application servers, web servers, databases, networks, virtual machines, telecoms equipment, operating systems, sensors, and so on, that make up your IT infrastructure. The maximum indexing volume depends on the Splunk Enterprise license.</td>
</tr>
<tr>
<td><strong>Data model</strong></td>
<td>A data model is a hierarchically-structured search-time mapping of semantic knowledge about one or more datasets. It encodes the domain knowledge necessary to build a variety of specialized searches of those datasets. These specialized searches are used by Splunk Enterprise to generate reports for Pivot users. Data model objects represent different datasets within the larger set of data indexed by Splunk Enterprise.</td>
</tr>
<tr>
<td><strong>Pivot</strong></td>
<td>Pivot refers to the table, chart, or data visualization you create using the Pivot Editor. The Pivot Editor lets users map attributes defined by data model objects to a table or chart data visualization without having to write the searches to generate them. Pivots can be saved as reports and added to dashboards.</td>
</tr>
<tr>
<td><strong>Search</strong></td>
<td>Search is the primary way users navigate data in Splunk Enterprise. You can write a search to retrieve events from an index, use statistical commands to calculate metrics and generate reports, search for specific conditions within a rolling time window, identify patterns in your data, predict future trends, and so on. Searches can be saved as reports and used to power dashboard panels.</td>
</tr>
<tr>
<td><strong>Alerts</strong></td>
<td>Alerts are triggered when conditions are met by search results for both historical and real-time searches. Alerts can be configured to trigger actions such as sending alert information to designated email addresses, post alert information to an RSS feed, and run a custom script, such as one that posts an alert event to syslog.</td>
</tr>
<tr>
<td><strong>Reports</strong></td>
<td>Reports are saved searches and pivots. You can run reports on an ad hoc basis, schedule them to run on a regular interval, set a scheduled report to generate alerts when the results of their runs meet particular conditions. Reports can be added to dashboards as dashboard panels.</td>
</tr>
<tr>
<td><strong>Dashboards</strong></td>
<td>Dashboards are made up of panels that contain modules such as search boxes, fields, charts, tables, forms, and so on. Dashboard panels are usually hooked up to saved searches or pivots. They can display the</td>
</tr>
</tbody>
</table>
results of completed searches as well as data from backgrounded real-time searches.

2.3 About Splunk Enterprise users

Splunk Enterprise serves different types of users. There are five main personas that use Splunk Enterprise:

<table>
<thead>
<tr>
<th>Persona</th>
<th>Industry Role</th>
<th>Activities</th>
</tr>
</thead>
</table>
| Administrator | network engineer, system administrator | • Configures, administers, optimizes, and secures the Splunk Enterprise deployment.  
|             |                                        | • Sets up user accounts and permissions.                                      
|             |                                        | • Gets data into Splunk Enterprise.                                           |
| Knowledge   | data analyst,                          | • Oversees knowledge object creation,                                        |
Manager

system administrator

- normalization, and usage across teams, departments, and deployments.
  - Gets the data into Splunk, or works with the administrator to do so.
  - Creates and shares data models.

Search User
data analyst, IT professional, network engineer, security analyst, system administrator

- Uses Search to investigate server problems, understand configurations, monitor user activities, and troubleshoot escalated problems.
- Builds reports and dashboards to monitor the health, performance, activity, and capacity of their IT infrastructure.
- Identifies patterns and trends that are indicators of routine problems.

Pivot User
business professional, data analyst, executive, IT professional, manager, system administrator

- Uses Pivot to build reports based on data models created by the Knowledge Manager.
- Creates reports and dashboards to monitor their businesses.
- Identifies trends in the health and performance of their businesses.

Developer
system integrator, professional developer

- Integrates data and functionality of applications with Splunk Enterprise.
- Builds Splunk Apps and add-ons with custom dashboards and data visualizations.

2.4 About Splunk Enterprise deployments

2.4.1 Splunk Enterprise and your IT infrastructure

Splunk Enterprise indexes data from the servers, applications, databases, network devices, virtual machines, and so on, that make up your IT infrastructure. As long as the machine that generates the data is a part of your network, Splunk Enterprise can collect the data from machines located anywhere, whether it is local (on-the-premises in a server room), remote...
(off-the-premises in a datacenter), entirely in the cloud, or a hybrid (such as on-premise and in the cloud).

Most users connect to Splunk Enterprise with a web browser and use splunk web to administer their deployment, manage and create knowledge objects, run searches, create pivots and reports, and so on. You can also use the command-line interface to administer your Splunk Enterprise deployment.

Splunk Enterprise supports a multi-user and distributed product architecture. This means that you can search and report on data spanning multiple Splunk Enterprise deployments within a single datacenter or globally across multiple datacenters and cloud infrastructures.

### 2.4.2 Splunk Enterprise Components

<table>
<thead>
<tr>
<th>Component</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apps</td>
<td>Apps are a collection of configurations, knowledge objects, and customer designed views and dashboards that extend the Splunk Enterprise environment to fit the specific needs of organizational teams such as Unix or Windows system administrators, network security specialists, website managers, business analysts, and so on. A single Splunk Enterprise installation can run multiple apps simultaneously.</td>
</tr>
<tr>
<td>Forwarder</td>
<td>A forwarder is a Splunk Enterprise instance that forwards data to another Splunk Enterprise instance (an indexer or another forwarder) or to a third-party system. Most forwarders are lightweight instances, with minimal resource utilization, allowing them to reside easily on the machine generating the data.</td>
</tr>
<tr>
<td>Indexer</td>
<td>An indexer is the Splunk Enterprise instance that indexes data. It typically receives data from a group of forwarders. The indexer transforms the data into events and stores the events into an index. The indexer also searches the indexed data in response to search requests. In a distributed search deployment, you might have multiple indexers, also known as search peers.</td>
</tr>
</tbody>
</table>

To ensure high data availability and protect against data loss, or just to
simplify the management of multiple indexers, you can deploy multiple indexers in indexer clusters.

**Search head**

In a distributed search deployment, the search head is the Splunk Enterprise instance that handles search management functions, directing search requests to a set of indexers and then merging the results back to the user. In a single-instance deployment, the one instance serves as both search head and indexer.

To ensure high availability and simplify horizontal scaling, you can deploy multiple search heads in search head clusters.

---

3. Getting started with alerts

3.1 What is an alert?

If you want to receive notifications about certain events, you can use alerts. When you set up an alert, search results trigger an alert action if they match the alert's conditions.

3.2 Alert basics
To get started with an alert, there are a few things to consider.

- **Conditions**: What do you want to know about?
  You can start with a search for the events you want to track. As an example, if you have an online store you can track when customers purchase your newest product. You can use an alert whose conditions are website purchase events that also involve this product.

- **Type and Frequency**: How often do you want to know about the event?
  You can receive a notification about every customer purchase of a new product as it occurs. Or, you can get a notification on a weekly basis. You can choose continuous per-result, rolling, or scheduled alerts, and adjust their frequency.

- **Alert Action**: What should happen when an alert is triggered?
  Once you set up an alert, when customer purchases of the new product show up in search results, they match the alert's conditions. Matching results trigger an alert action according to the frequency you choose. There are several options for alert actions. For example, you can receive an email or update a web resource in response to the triggered alert.

### 3.3 About alert types

There are a few alert types that you can use. Each type works differently with a search to trigger alert actions. You can choose an alert type depending on what event you are tracking and when you want to know about it.

Here is a quick reference guide to alert types and behavior:

<table>
<thead>
<tr>
<th>Alert type</th>
<th>How it works with searches</th>
<th>Triggering this alert</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Per-result alert</strong></td>
<td>Based on a continuous real time search</td>
<td>This basic alert triggers any time its search returns a result.</td>
</tr>
<tr>
<td><strong>Scheduled alert</strong></td>
<td>Runs a search according to a schedule that you specify when creating the alert.</td>
<td>You can specify which search results trigger the alert.</td>
</tr>
<tr>
<td><strong>Rolling-window alert</strong></td>
<td>Based on a continuous real time search</td>
<td>You can specify the time window and the conditions</td>
</tr>
</tbody>
</table>
that, together, trigger the alert.

To learn about choosing an alert type for different scenarios, see alert types and scenarios.

For more information on setting up specific alerts, check out resources on alerts, scheduled alerts, and rolling windows-alerts in this manual.

You can also check out alert examples to get an idea of how each alert type can work.

3.4 Choosing an alert type

You can consider using different alerts for different scenarios. Depending on how you want to search for results and set up an alert, you can opt for a per result, scheduled, or rolling windows alert.

To see some example scenarios and learn about choosing an alert type, see alert types and scenarios.

3.5 Managing alert frequency

You can throttle an alert if you want to change how often it runs an alert action. Throttling an alert does not change how often search results meet the alert conditions. Instead, it changes how often search results matching the alert conditions trigger an alert action.

To learn about changing alert frequency, look at throttle alerts in this manual.

3.6 Using alert actions
When search results match an alert's conditions, they trigger the alert action. What happens next?

There are many options for configuring alert actions. For example, you can opt for an email based on the search results. If you want to see updates in a chat room, blog, or other web resource, you can use a webhook alert action.

3.7 Alert and alert action permissions

Alerts and alert actions are knowledge objects with defined permissions. User roles and capabilities determine alert and alert action permissions.

By default, only users with the Admin or Power roles can:

- Create alerts.
- Run real-time searches.
- Schedule searches.
- Save searches.
- Share alerts.

3.8 Scheduled reports and scheduled alerts are not the same

A scheduled report is similar to a scheduled or rolling-window alert in some ways. You can schedule a report and set up an action to run each time the scheduled report runs.

Scheduled reports are different from alerts, however, because a scheduled report's action will run every time the report is run. The report action does not depend on trigger conditions like an alert action does.

As an example, you can monitor guest check-ins at a hotel using an hourly search. Here are the differences between a scheduled report and an alert with email notification actions.

- **Scheduled report**: runs its action and sends an email every time the report completes, even if there are no search results showing check-ins. In this case, you get an email notification every hour.

- **Alert**: only runs alert action when it is triggered by search results showing one or more check-in events. In this case, you only get an email notification if results trigger the alert action.
4. Visualization Reference

Splunk software has many options for search result visualizations. Beyond a straightforward events list, you can visualize data in tables and charts. When a search returns a single discrete number, you can use single value visualizations or gauges for the result.

Visualizations have specific data structure requirements. You can create a search that returns data in different structures depending on how you want to see your results. For example, use a transforming command (such as `stats`, `timechart`, or `top`) to return search results in a data structure supporting tables and charts.

5. Accessing visualization definition features

Splunk software provides user interface tools to create and modify visualizations. You can access these tools from various places in Splunk Web.

- Search
- Dashboards
- Dashboard Editor
- Pivot
- Reports

You can also create and modify visualizations directly in simple XML code.
**Visualizations from Search**

You can modify how search results display in the Search page. After running a search, select the **Visualization** tab, then select the type of visualization to display. You can specify formatting options for the selected visualization. The search must be a reporting search that returns results that can be formatted as a visualization.

**Edit visualizations** provides details for editing visualizations in the Dashboard Editor. The instructions for the Dashboard Editor also apply to visualizations from the Search page.

**5.1 Dashboard panel visualizations**

When you base a new dashboard panel on search results you can choose the visualization that best represents the data returned by the search. You can then use the Visualization Editor to fine-tune the way the panel visualization displays.

To create a dashboard panel from search results, after you run the search click **Save As > Dashboard Panel**. For more information about creating and editing dashboards, see the **About the Dashboard Editor** and **Edit visualizations**.

**6. Events visualizations**

Events visualizations are essentially raw lists of events.

You get events visualizations from any search that does not include a transform operation, such as a search that uses reporting commands.
If you add a transforming command to the search, you get statistical results that you can present either as a table or a chart:

With event listing visualizations, you can:

- Determine the number of events listed.
- Determine whether numbers appear to the left of each event.
- Have event text wrap to fit within the dashboard panel.

7. Charts

Splunk software provides a variety of chart visualizations, such as column, line, area, scatter, and pie charts. These visualizations require transforming searches whose results involve one or more series.

A series is a sequence of related data points that can be plotted on a chart. For example, each line plotted on a line chart represents an individual series. You can design transforming searches that produce a single series, or you can set them up so the results provide data for multiple series.
Consider a table that a transforming search generates. Each column in the table after the first column represents a different series. A "single series" search produces a table with only two columns, while a "multiple series" search produces a table with three or more columns.

If you enable multi-series mode for line, bar, area, and column charts, you can also select independent Y-axis ranges for each series. This option can help you compare spikes or trends across multiple series. Once you select a line, bar, column, or area chart, select Format > General to enable Multi-series Mode. Then, select the Y-Axis formatting panel to enable the Independent axis range.

- Note: Enabling independent y-axis ranges is only available in multi-series mode.

All chart visualizations can display single-series searches. However the bar, column, line, and pie chart visualizations usually display the data best. Pie charts can only display data from single series searches.

If a search produces multiple series, bar, column, line, area, and scatter chart visualizations display the data best.

### 7.1 Column and bar charts

Use a column chart or bar chart to compare the frequency of values of fields in your data. In a column chart, the x-axis values are typically field values. If the search uses the timechart transforming command, the x-axis represents time. The y-axis can be any other field value, count of values, or statistical calculation of a field value. Column charts and bar charts represent data similarly, except that the x-axis and y-axis values are reversed. For more information, see the Data structure requirements for visualizations in this manual.

The following bar chart presents the results of a search that uses internal metrics. It calculates the sum of CPU seconds by processor in the last 15 minutes. It then arranges the processors with the top ten sums in descending order. This example also shows how you can mouse over a single bar or column to get detailed information.

The following search drives the bar chart visualization.

For column and bar chart visualizations, you can do the following:

- Set the chart titles, as well as the titles of the x-axis and y-axis.
- Set the minimum y-axis values.
• Set the unit scale to logarithmic values

• Configure charts as stacked, 100% stacked, and unstacked. Bar and column charts are unstacked by default. See the following subsection for details on stacking bar and column charts.

• Set the major unit for the y-axis. For example, configure tick marks in units that work best for your data.

• Determine the position of the chart legend and the manner in which the legend labels are truncated.

7.2 Stacked column and bar charts

When a base search involves more than one data series, you can use stacked column charts and stacked bar charts to compare the frequency of field values in your data.

Unstacked charts

In an unstacked column chart, the columns for different series appear alongside each other. An unstacked column chart is useful for relatively simple search results. But when the series count increases an unstacked column chart can appear cluttered and confusing.

Stacked charts

A stacked column chart displays all the series columns for a single data point as segments of a single column. The total value of the column is the sum of the segments. You typically use a stacked column or bar chart to highlight the relative weight, or importance, of the different types of data that make up a specific data set.

With line and area charts, you can do the following:

• Set the chart titles, as well as the titles of the x-axis and y-axis.

• Determine how to display null y-axis values. You can leave gaps for null datap points, connect to zero data points, or just connect to the next positive data point. If you choose to leave gaps, the chart displays markers for data points that are disconnected. In this case, they are not adjacent to other positive data points.

• Set the minimum y-axis value.

• Set the unit scale to logarithmic values Logarithmic values are useful with a mix of very small and very large y-axis values.
• Set the major unit for the y-axis
  For example, configure tick marks in units that work best for your data.

• Determine the position of the chart legend and the manner in which the legend labels are truncated.

• Enable or disable **drilldown** functionality.
  For more information about drilldown, see Understand basic table and chart drilldown actions in this manual.

8. Open Issues
As of now it's very clear. In future enhancements there may be chances of getting some issues while developing the project.

9. Acknowledgements
Any update regarding the book from the library is to be recorded to have update & correct values.

10. References