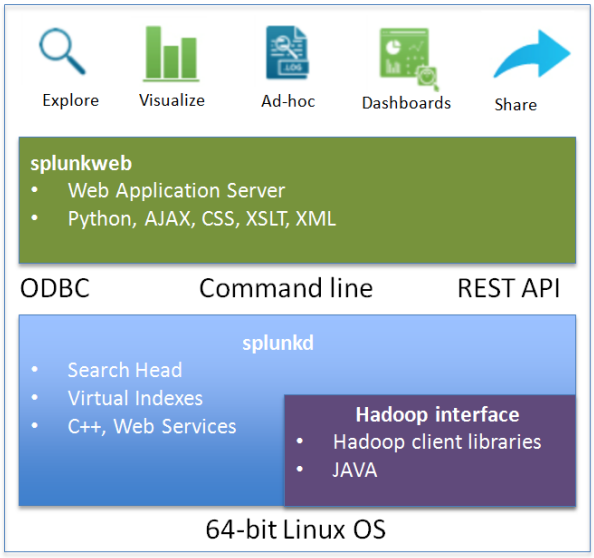
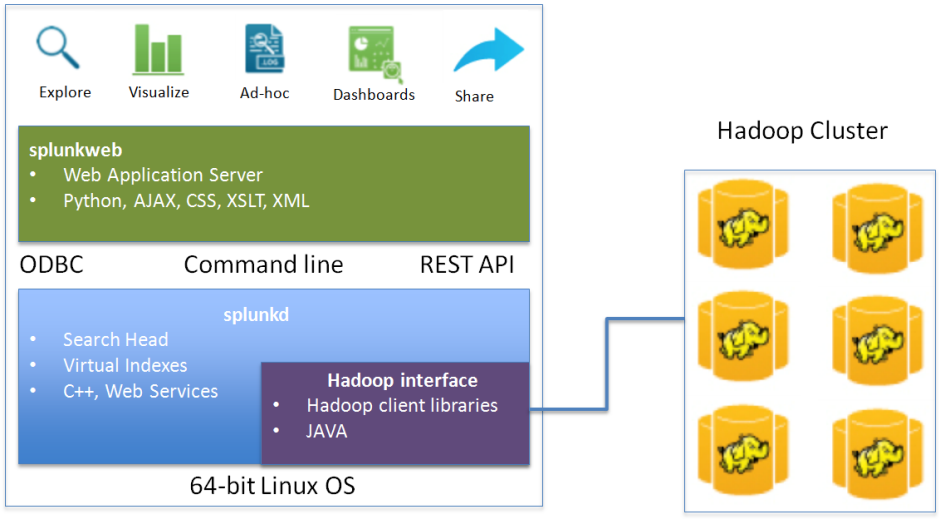
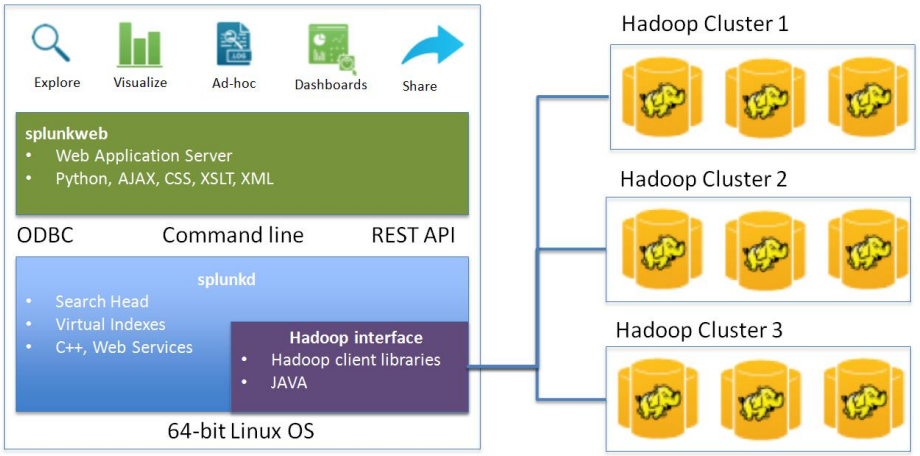
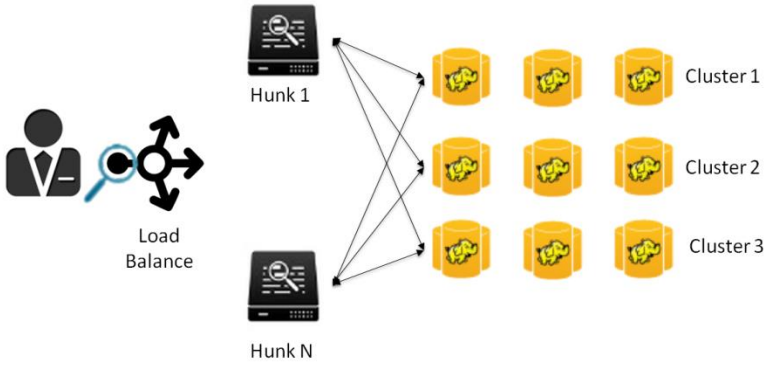
Chapter 1: Meet Hunk

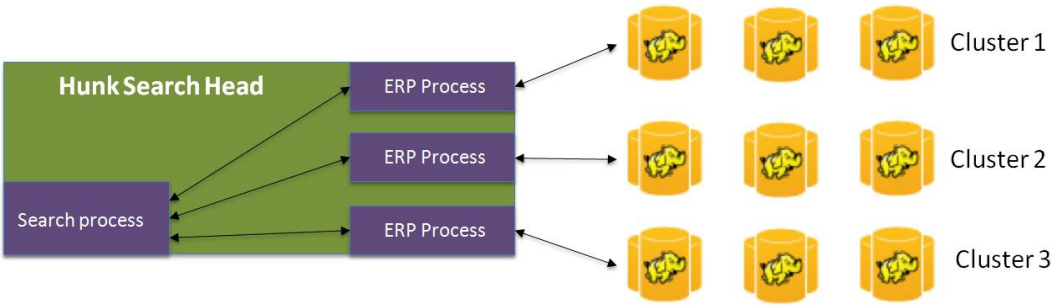
Hadoop + Splunk = Hunk

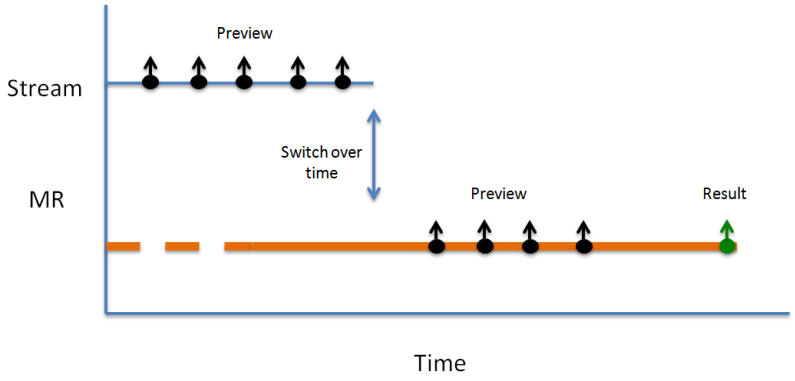

Explore
Visualize
Ad-hoc
Dashboards
Share
splunkweb
Web Application Server
Python, AJAX, CSS, XSLT, XML
ODBC
Command line
REST API
splunkd
Search Head
Virtual Indexes
C++, Web Services
Hadoop interface
Hadoop client libraries
JAVA
64-bit Linux OS

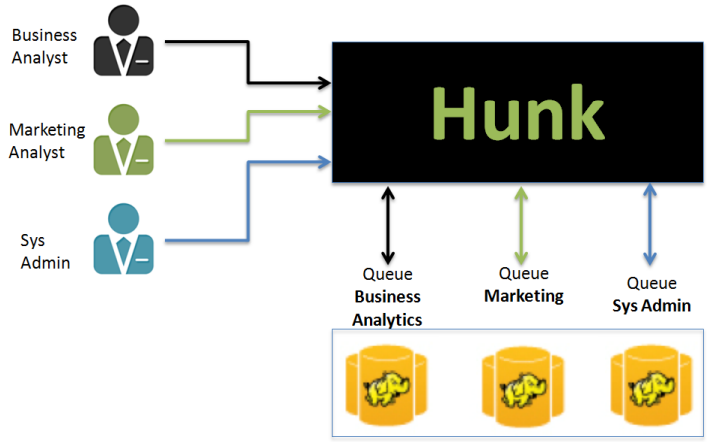

Explore
Visualize
Ad-hoc
Dashboards
Share
splunkweb
Web Application Server
Python, AJAX, CSS, XSLT, XML
ODBC
Command line
REST API
splunkd
Search Head
Virtual Indexes
C++, Web Services
Hadoop interface
Hadoop client libraries
JAVA
64-bit Linux OS
Hadoop Cluster

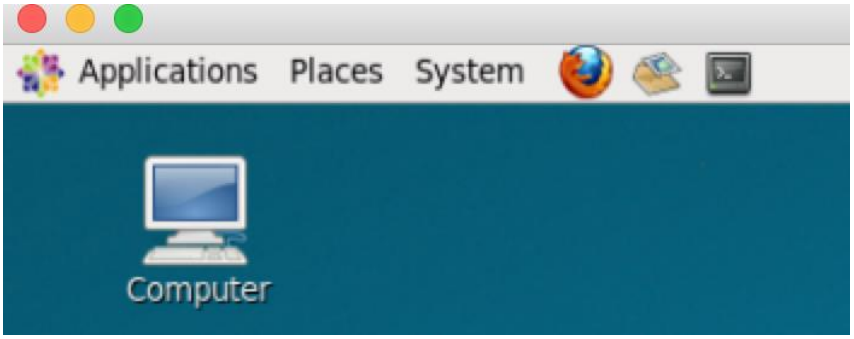

Explore
Visualize
Ad-hoc
Dashboards
Share
splunkweb
• Web Application Server
• Python, AJAX, CSS, XSLT, XML
ODBC
Command line
REST API
splunkd
• Search Head
• Virtual Indexes
• C++, Web Services
Hadoop interface
• Hadoop client libraries
• JAVA
64-bit Linux OS
Hadoop Cluster 1
Hadoop Cluster 2
Hadoop Cluster 3

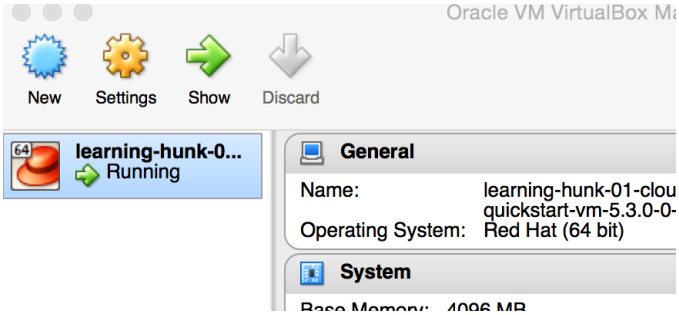

Hunk 1
Hunk N
Load
Balance
Cluster 1
Cluster 2
Cluster 3

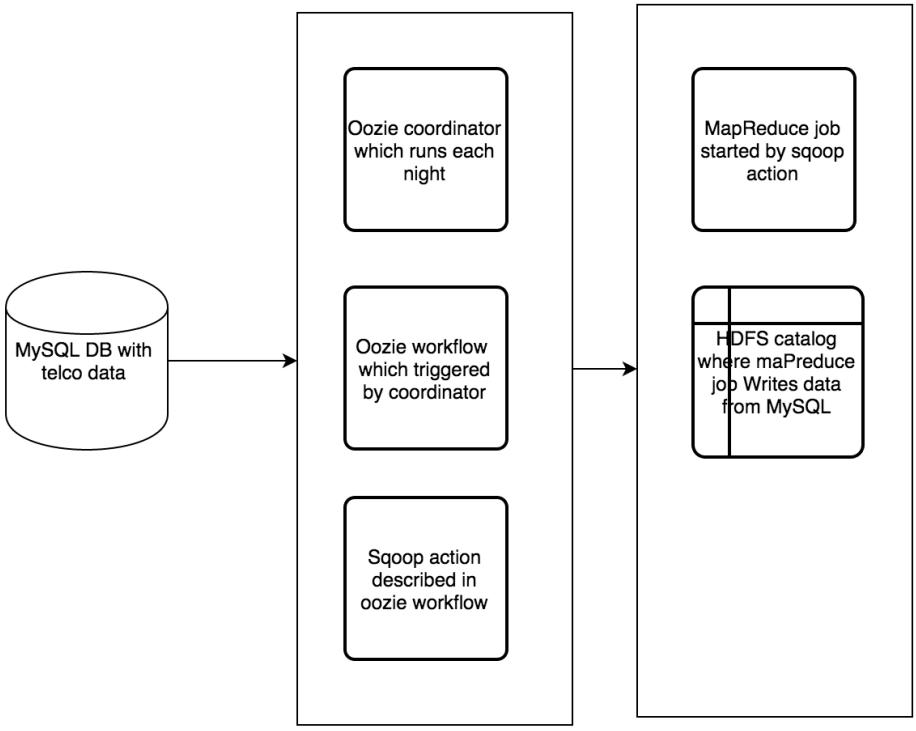

Hunk Search Head
ERP Process
ERP Process
ERP Process
Search process
Cluster 1
Cluster 2
Cluster 3

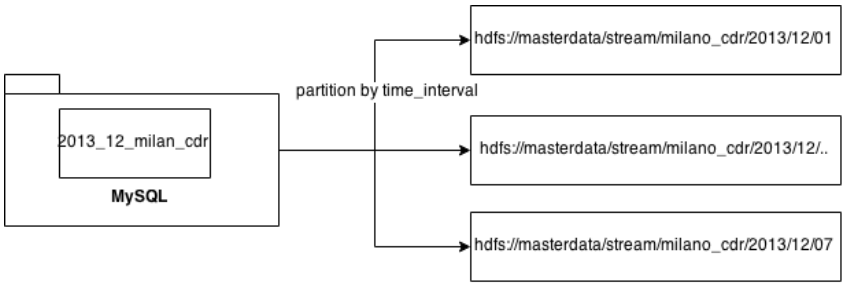

Preview
Stream
Switch over time
MR
Preview
Result
Time

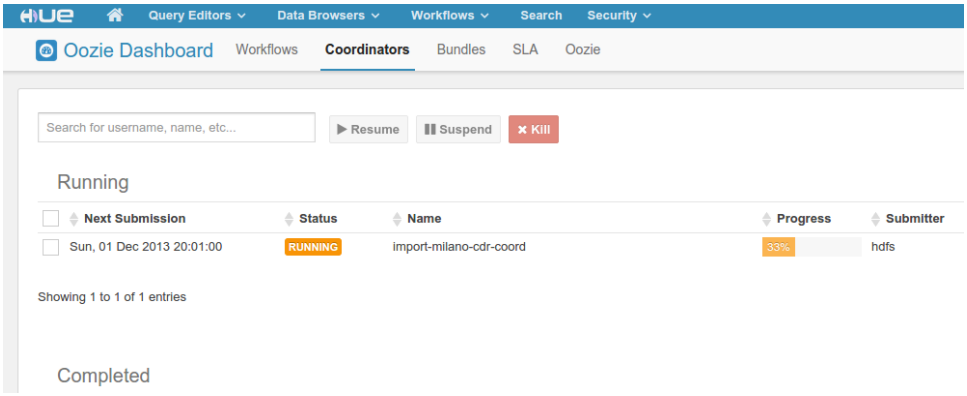

Business Analyst
Marketing Analyst
Sys Admin
Hunk
Queue Business Analytics
Queue Marketing
Queue Sys Admin

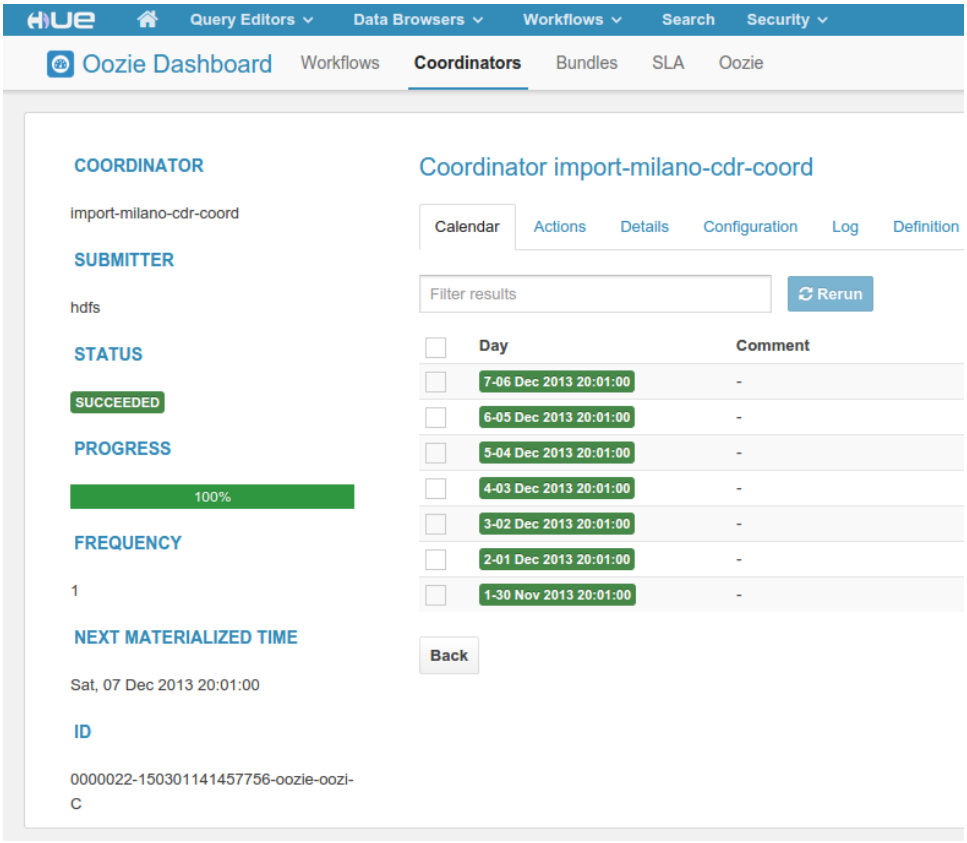

Applications
Places
System
Computer


Oracle VM VirtualBox Ma
New
Settings
Show
Discard
learning-hunk-0...
Running
General
Name:
learning-hunk-01-clou
quickstart-vm-5.3.0-0-
Operating System:
Red Hat (64 bit)
System

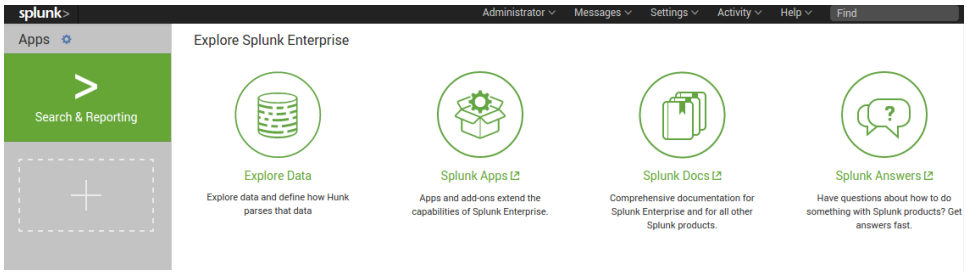

MySQL DB with telco data
Oozie coordinator which runs each night
Oozie workflow which triggered by coordinator
Sqoop action described in oozie workflow
MapReduce job started by sqoop action
HDFS catalog where maPreduce job Writes data from MySQL

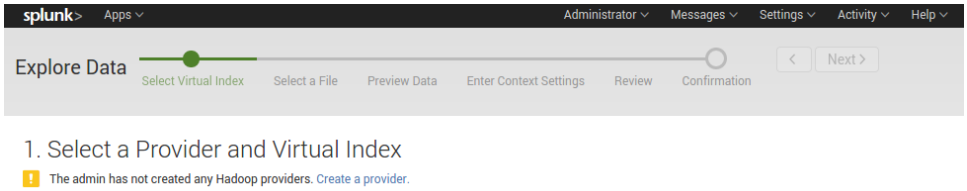

hdfs://masterdata/stream/milano_cdr/2013/12/01
partition by time_interval
2013_12_milan_cdr
MySQL
hdfs://masterdata/stream/milano_cdr/2013/12/..
hdfs://masterdata/stream/milano_cdr/2013/12/07

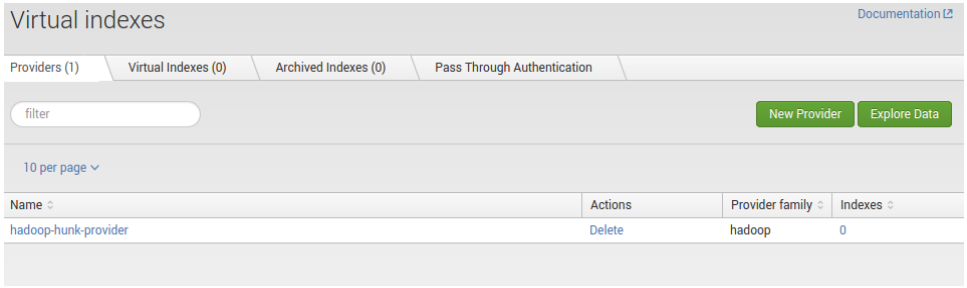

HUE
Query Editors
Data Browsers
Workflows
Search
Security
Oozie Dashboard
Workflows
Coordinators
Bundles
SLA
Oozie
Search for username, name, etc...
Resume
Suspend
Kill
Running
Next Submission
Status
Name
Progress
Submitter
Sun, 01 Dec 2013 20:01:00
RUNNING
import-milano-cdr-coord
33%
hdfs
Showing 1 to 1 of 1 entries
Completed

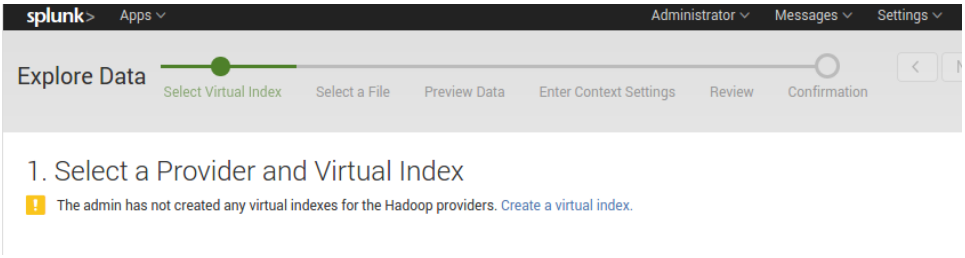

Query Editors
Data Browsers
Workflows
Search
Security
Oozie Dashboard
Workflows
Coordinators
Bundles
SLA
Oozie
COORDINATOR
import-milano-cdr-coord
SUBMITTER
hdfs
STATUS
SUCCEEDED
PROGRESS
100%
FREQUENCY
1
NEXT MATERIALIZED TIME
Sat, 07 Dec 2013 20:01:00
ID
0000022-150301141457756-oozie-oozi-C
Coordinator import-milano-cdr-coord
Calendar
Actions
Details
Configuration
Log
Definition
Filter results
Rerun
Day
Comment
7-06 Dec 2013 20:01:00
6-05 Dec 2013 20:01:00
5-04 Dec 2013 20:01:00
4-03 Dec 2013 20:01:00
3-02 Dec 2013 20:01:00
2-01 Dec 2013 20:01:00
1-30 Nov 2013 20:01:00
Back

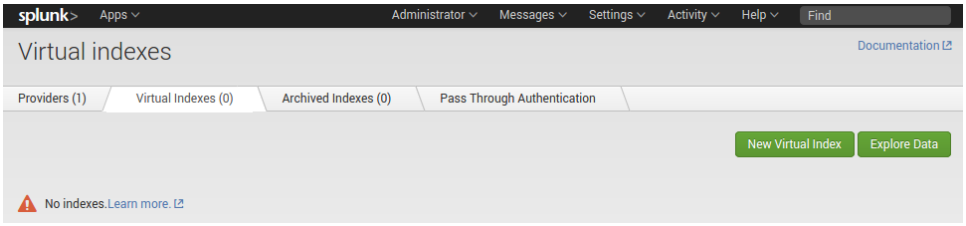
Chapter 2: Explore Hadoop Data with Hunk

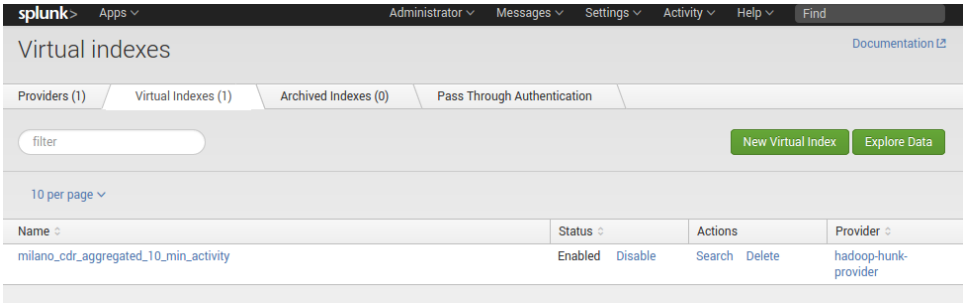


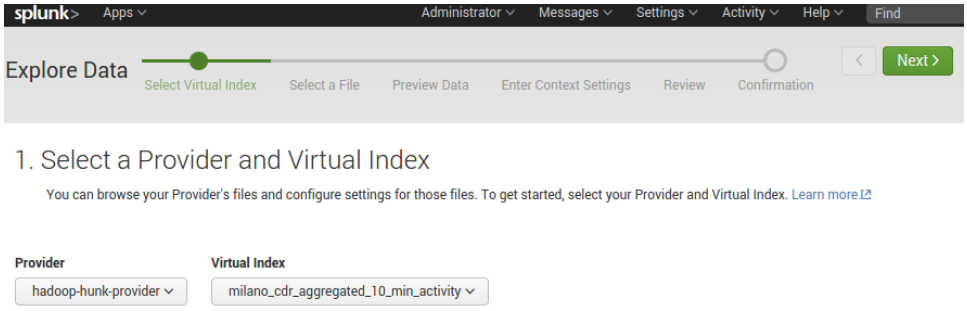


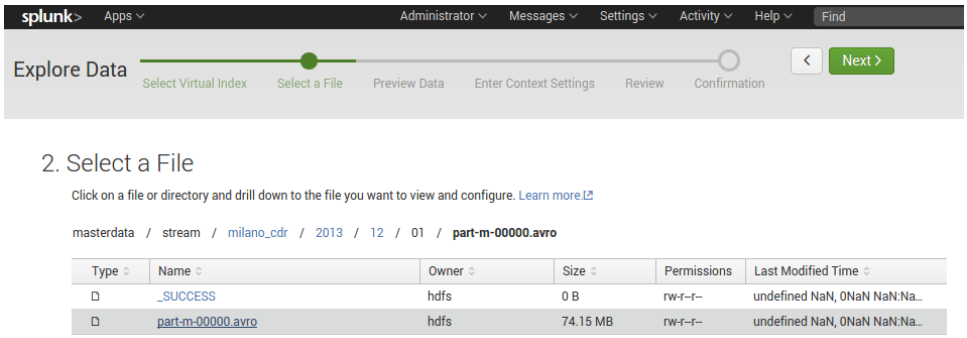


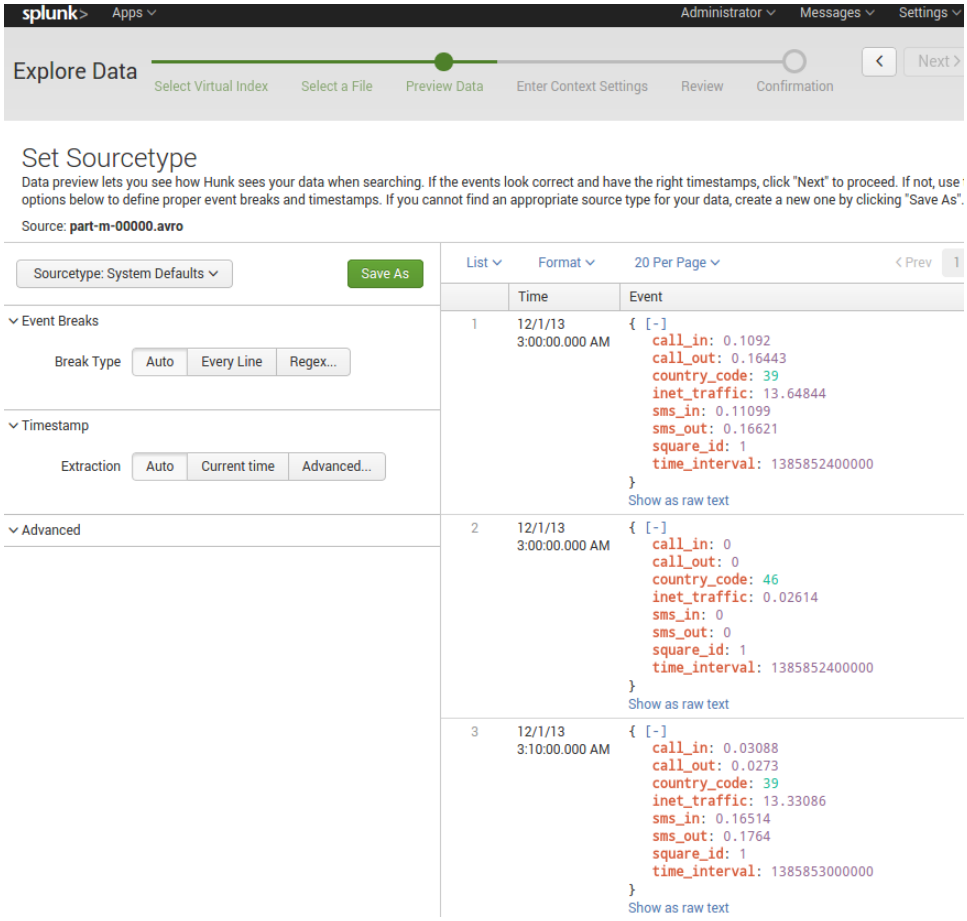

splunk> Apps
Administrator
Messages
Settings
Explore Data
Select Virtual Index
Select a File
Preview Data
Enter Context Settings
Review
Confirmation
1. Select a Provider and Virtual Index
The admin has not created any virtual indexes for the Hadoop providers. Create a virtual index.

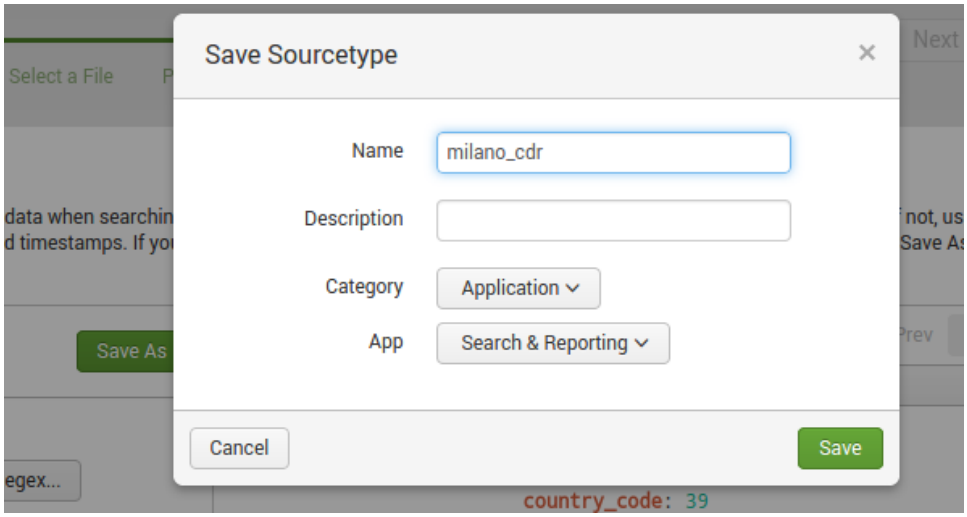

splunk> Apps
Administrator
Messages
Settings
Activity
Help
Find
Virtual indexes
Documentation
Providers (1)
Virtual Indexes (0)
Archived Indexes (0)
Pass Through Authentication
New Virtual Index
Explore Data
No indexes. Learn more.

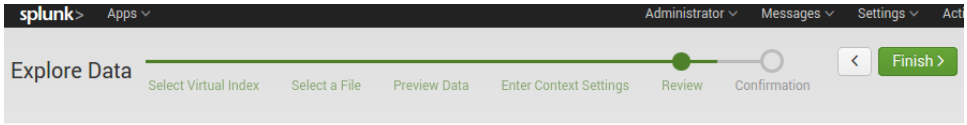

splunk> Apps
Administrator
Messages
Settings
Activity
Help
Find
Virtual indexes
Documentation
Providers (1)
Virtual Indexes (1)
Archived Indexes (0)
Pass Through Authentication
filter
New Virtual Index
Explore Data
10 per page
Name
Status
Actions
Provider
milano_cdr_aggregated_10_min_activity
Enabled
Disable
Search
Delete
hadoop-hunk-provider


splunk>
Apps
Administrator
Messages
Settings
Activity
Help
Find
Explore Data
Select Virtual Index
Select a File
Preview Data
Enter Context Settings
Review
Confirmation
Next
1. Select a Provider and Virtual Index
You can browse your Provider's files and configure settings for those files. To get started, select your Provider and Virtual Index. Learn more
Provider
hadoop-hunk-provider
Virtual Index
milano_cdr_aggregated_10_min_activity


splunk>
Apps
Administrator
Messages
Settings
Activity
Help
Find
Explore Data
Select Virtual Index
Select a File
Preview Data
Enter Context Settings
Review
Confirmation
Next
2. Select a File
Click on a file or directory and drill down to the file you want to view and configure. Learn more
masterdata / stream / milano_cdr / 2013 / 12 / 01 / part-m-00000.avro
Type
Name
Owner
Size
Permissions
Last Modified Time
_SUCCESS
hdfs
0 B
rw-r--r--
undefined NaN, 0NaN NaN:Na...
part-m-00000.avro
hdfs
74.15 MB
rw-r--r--
undefined NaN, 0NaN NaN:Na...


splunk> Apps
Administrator
Messages
Settings
Explore Data
Select Virtual Index
Select a File
Preview Data
Enter Context Settings
Review
Confirmation
Next
Set Sourcetype
Data preview lets you see how Hunk sees your data when searching. If the events look correct and have the right timestamps, click "Next" to proceed. If not, use
options below to define proper event breaks and timestamps. If you cannot find an appropriate source type for your data, create a new one by clicking "Save As".
Source: part-m-00000.avro
Sourcetype: System Defaults
Save As
Event Breaks
Break Type
Auto
Every Line
Regex...
Timestamp
Extraction
Auto
Current time
Advanced...
Advanced
List
Format
20 Per Page
Prev
1
Time
Event
1
12/1/13
3:00:00.000 AM
{ [-]
call_in: 0.1092
call_out: 0.16443
country_code: 39
inet_traffic: 13.64844
sms_in: 0.11099
sms_out: 0.16621
square_id: 1
time_interval: 1385852400000
}
Show as raw text
2
12/1/13
3:00:00.000 AM
{ [-]
call_in: 0
call_out: 0
country_code: 46
inet_traffic: 0.02614
sms_in: 0
sms_out: 0
square_id: 1
time_interval: 1385852400000
}
Show as raw text
3
12/1/13
3:10:00.000 AM
{ [-]
call_in: 0.03088
call_out: 0.0273
country_code: 39
inet_traffic: 13.33086
sms_in: 0.16514
sms_out: 0.1764
square_id: 1
time_interval: 1385853000000
}
Show as raw text


Save Sourcetype
Name
milano_cdr
Description
Category
Application
App
Search & Reporting
Cancel
Save
Select a File
Next
Save As
country_code: 39



5. Review

Review your saved configuration details. Learn more

Provider	hadoop-hunk-provider
Virtual Index	milano_cdr_aggregated_10_min_activity
Source	/masterdata/stream/milano_cdr/2013/12/01/part-m-00000.avro Edit
Sourcetype	milano_cdr
App context	search
Sharing Preference	global

The following configuration has been added to your props.conf file. You can copy these to the clipboard for reuse.

```
[source::/masterdata/stream/milano_cdr/2013/12/01/part-m-00000.avro]
sourcetype = milano_cdr
```

Name *

digital_analytics

Description

Provider

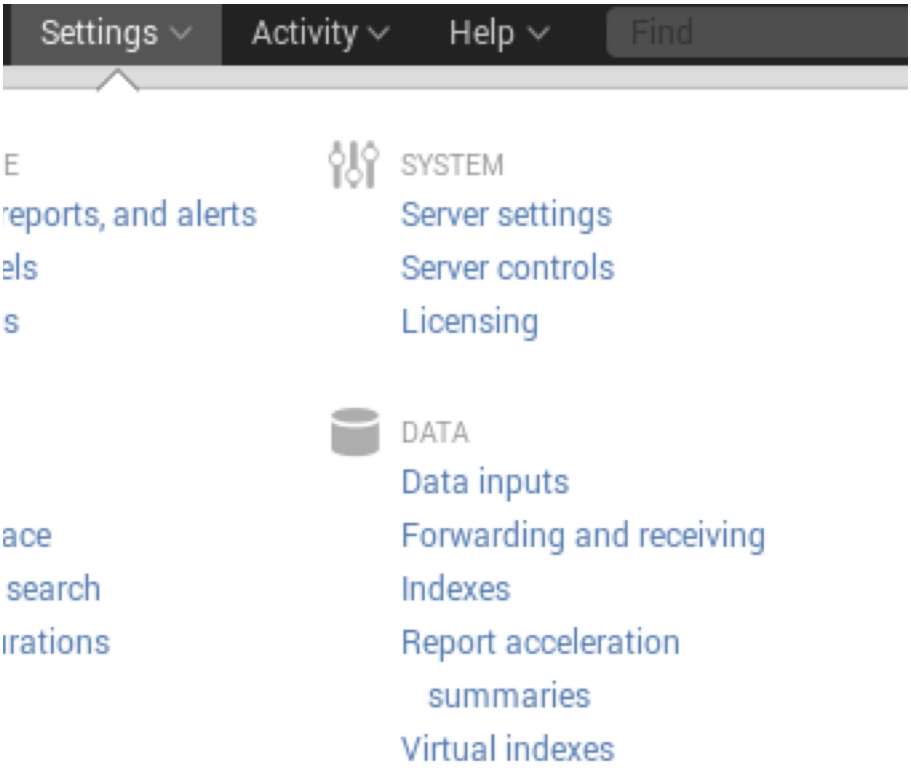
hadoop-hunk-provider

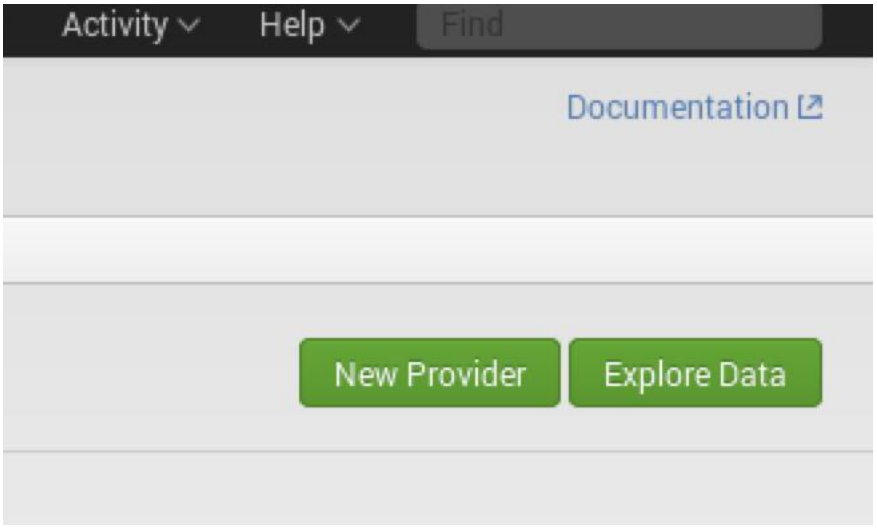
Paths

Path to data in HDFS ? *

/staging/web_logs

Example: /home/data/apache/logs/


Settings
Activity
Help
Find
E
eports, and alerts
els
s
ace
search
irations
SYSTEM
Server settings
Server controls
Licensing
DATA
Data inputs
Forwarding and receiving
Indexes
Report acceleration
summaries
Virtual indexes



1. Select a Provider and Virtual Index

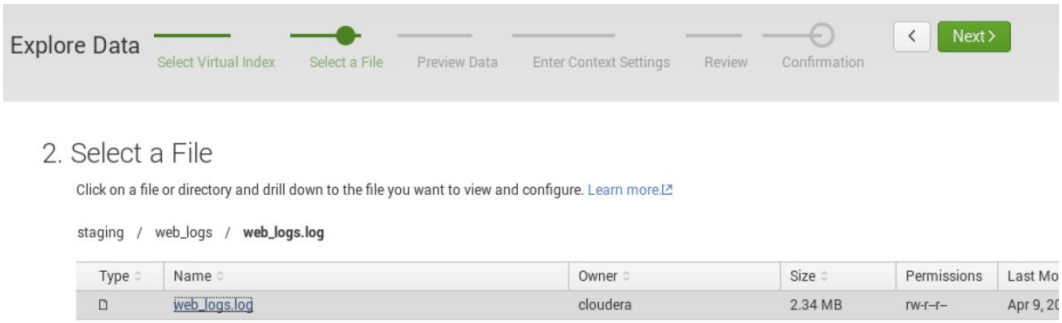
You can browse your Provider's files and configure settings for those f

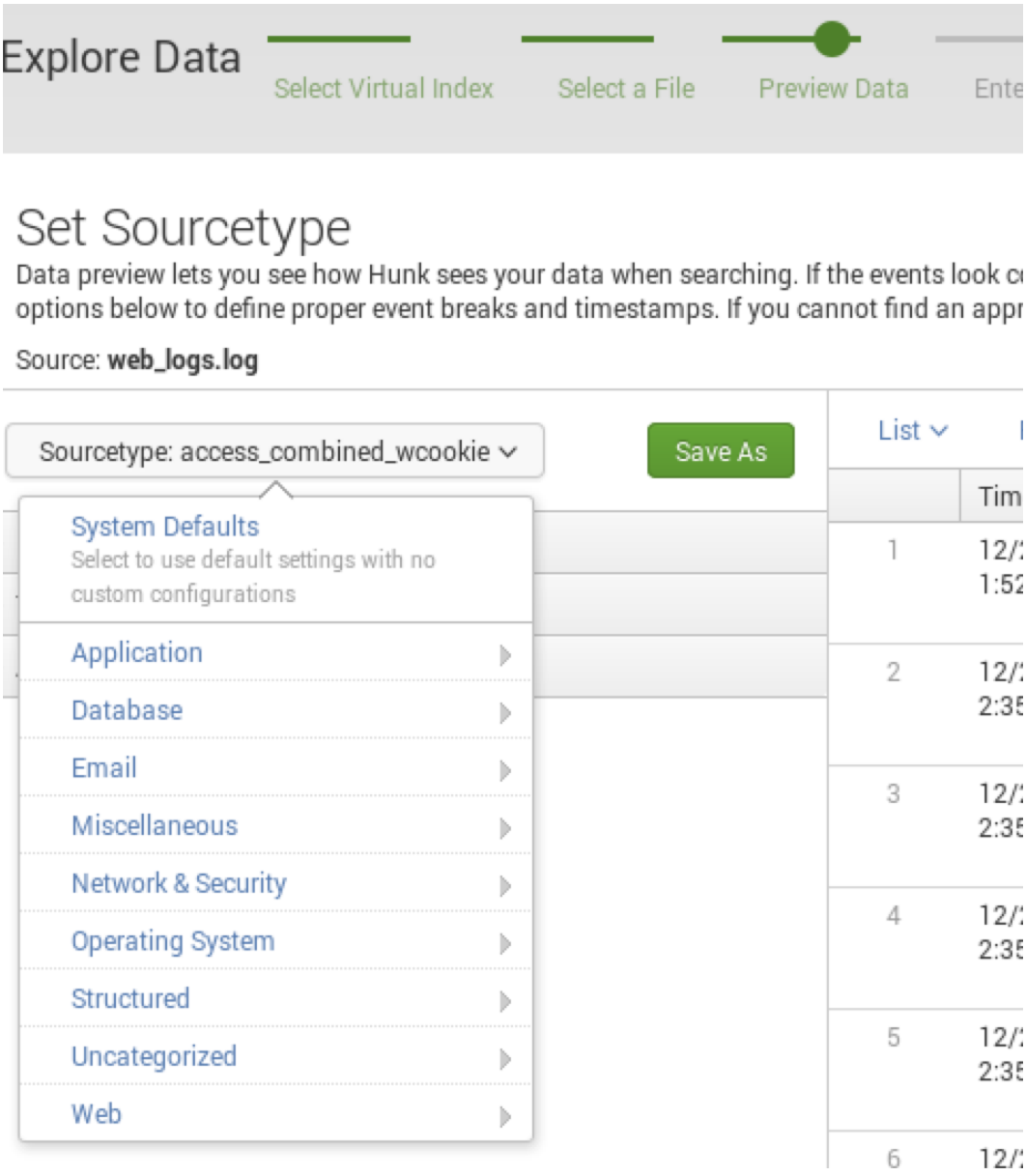
Provider

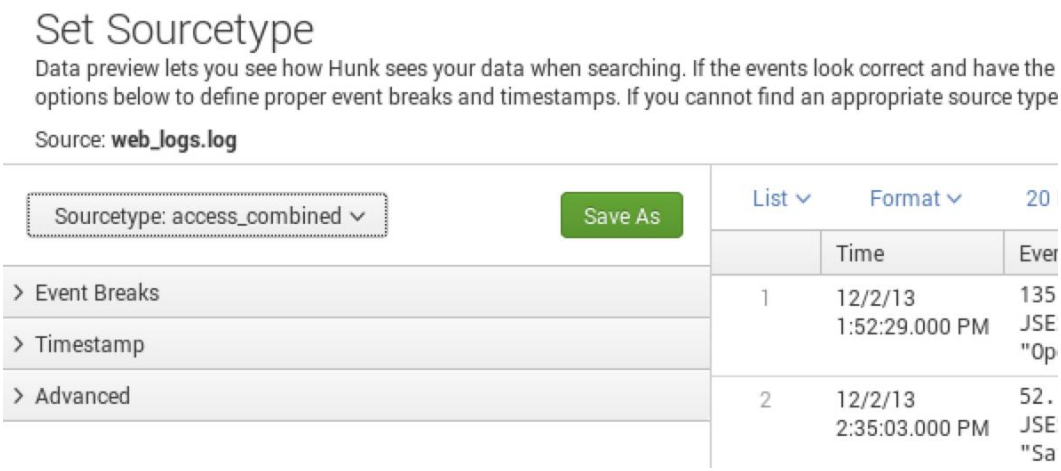
hadoop-hunk-provider

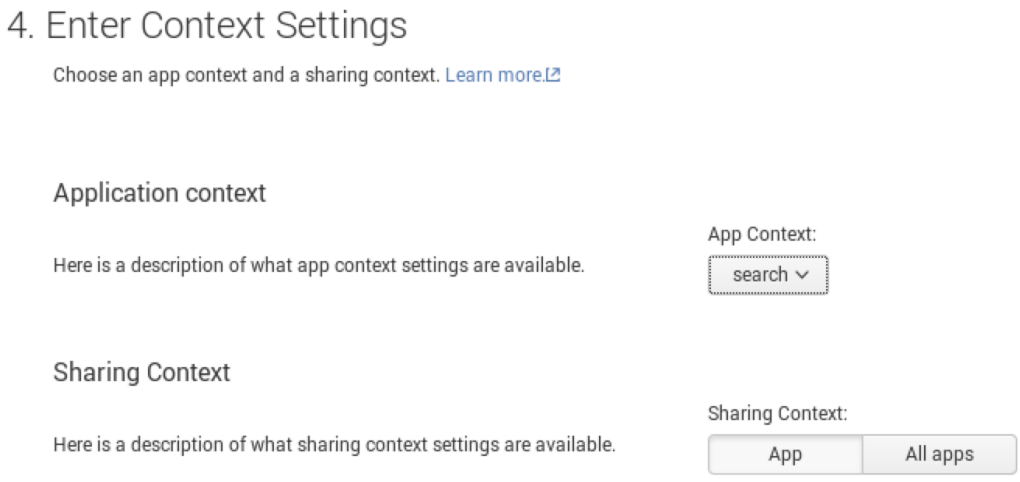
Virtual Index

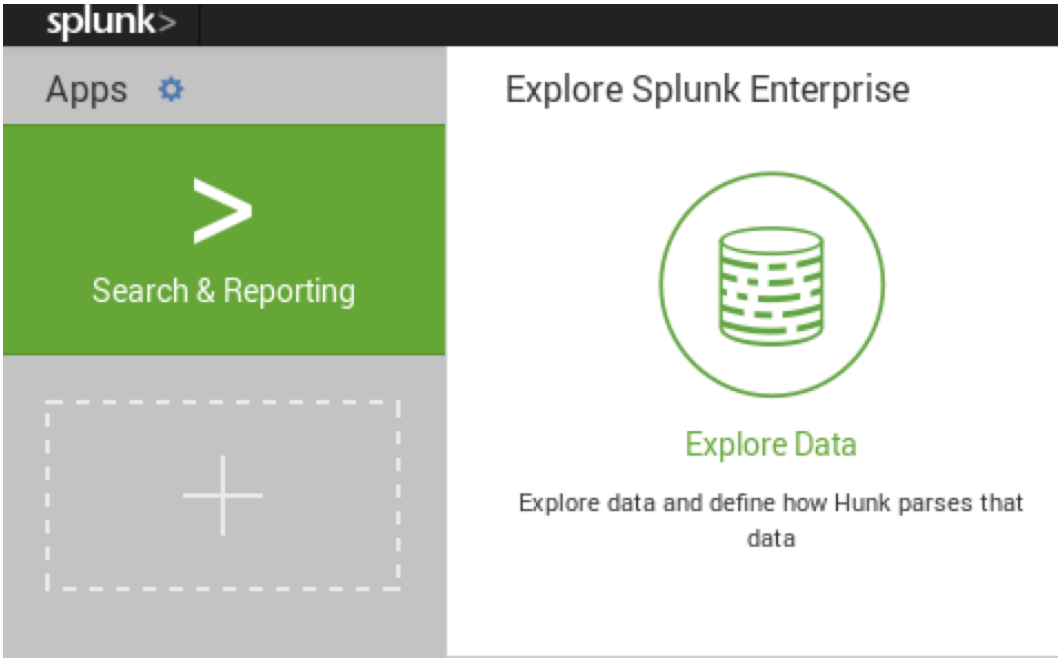
digital_analytics

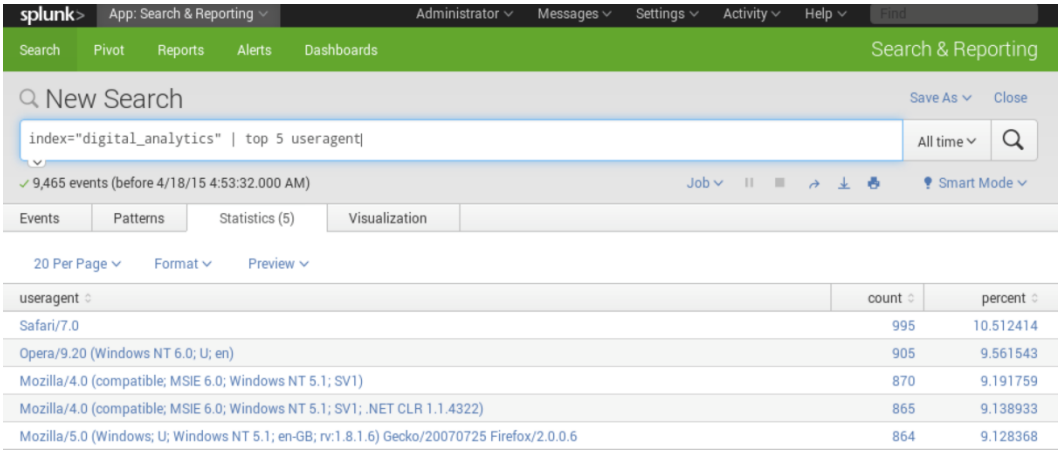

Explore Data
Select Virtual Index
Select a File
Preview Data
Enter Context Settings
Review
Confirmation
Next
2. Select a File
Click on a file or directory and drill down to the file you want to view and configure. Learn more.
staging / web_logs / web_logs.log
Type
Name
Owner
Size
Permissions
Last Mo
web_logs.log
cloudera
2.34 MB
rw-r--r--
Apr 9, 20

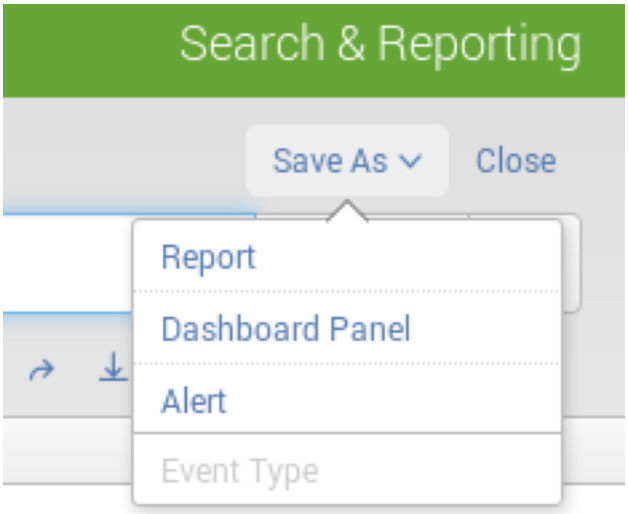

Explore Data
Select Virtual Index
Select a File
Preview Data
Ente
Set Sourcetype
Data preview lets you see how Hunk sees your data when searching. If the events look c
options below to define proper event breaks and timestamps. If you cannot find an appr
Source: web_logs.log
Sourcetype: access_combined_wcookie
Save As
List
Tim
System Defaults
Select to use default settings with no custom configurations
Application
Database
Email
Miscellaneous
Network & Security
Operating System
Structured
Uncategorized
Web
1
12/
1:52
2
12/
2:35
3
12/
2:35
4
12/
2:35
5
12/
2:35
6
12/

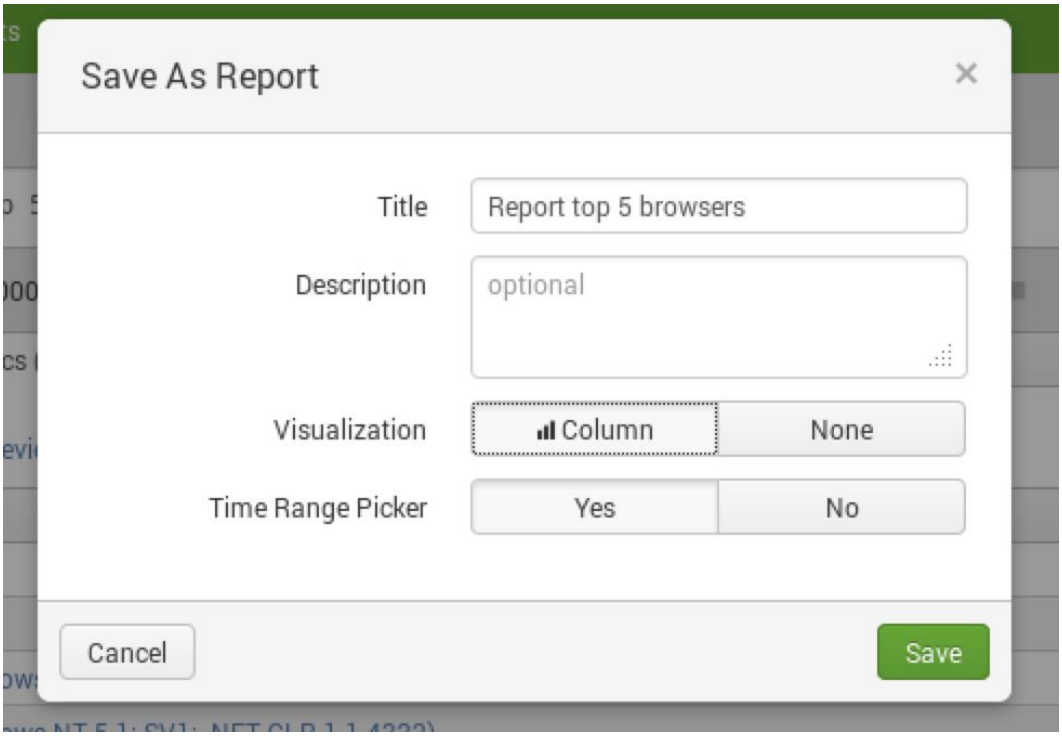

Set Sourcetype
Data preview lets you see how Hunk sees your data when searching. If the events look correct and have the options below to define proper event breaks and timestamps. If you cannot find an appropriate source type
Source: web_logs.log
Sourcetype: access_combined
Save As
List
Format
20
Time
Event Breaks
Timestamp
Advanced
1
12/2/13 1:52:29.000 PM
2
12/2/13 2:35:03.000 PM

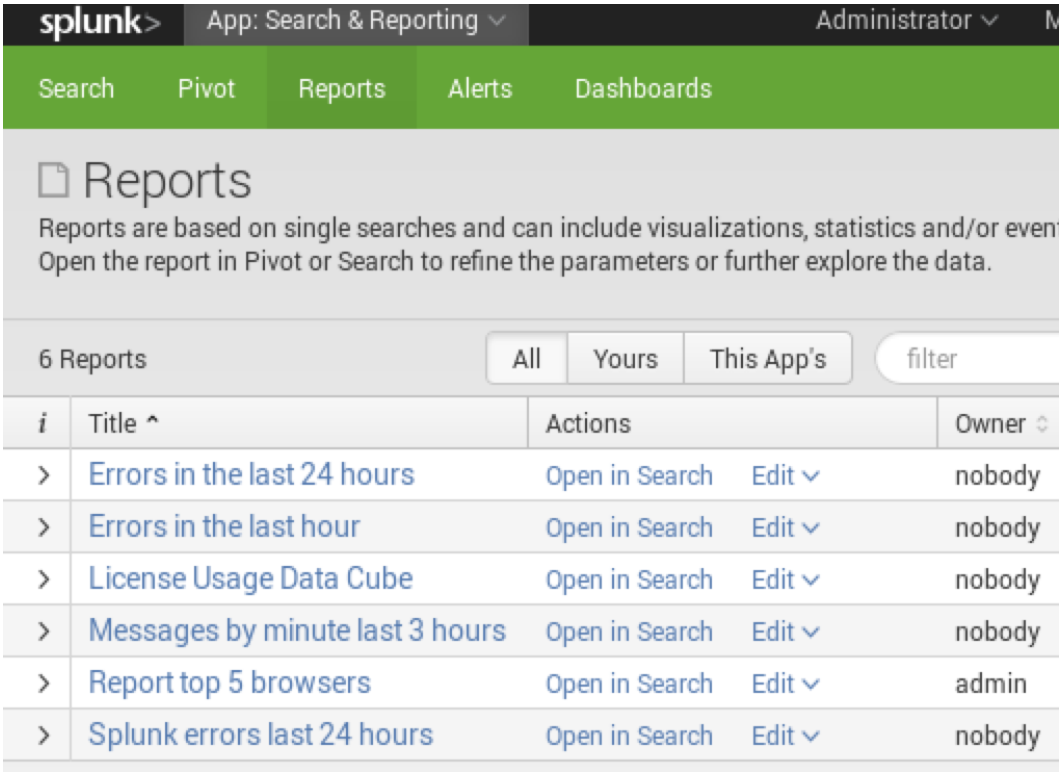

4. Enter Context Settings
Choose an app context and a sharing context. Learn more
Application context
Here is a description of what app context settings are available.
App Context:
search
Sharing Context
Here is a description of what sharing context settings are available.
Sharing Context:
App
All apps

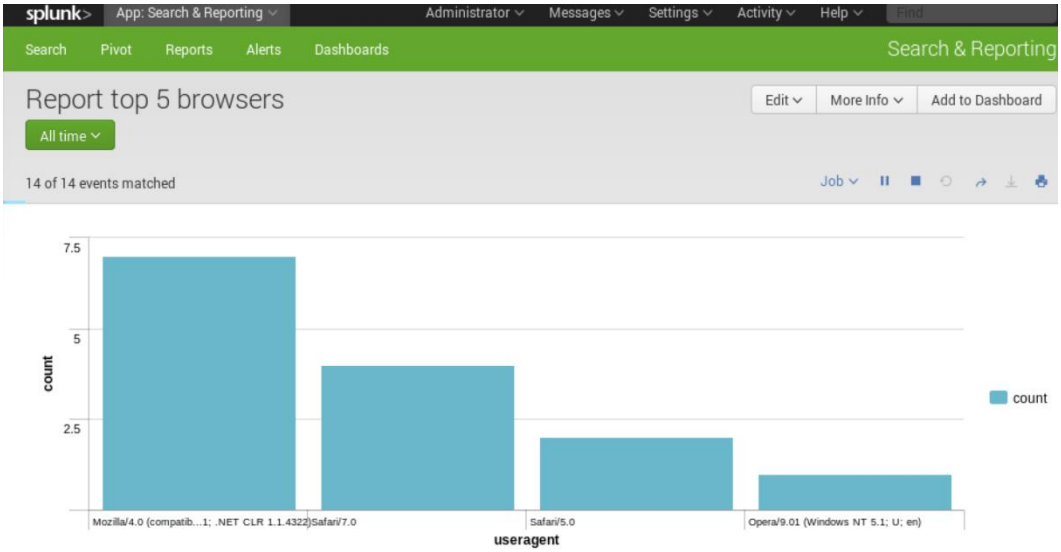

splunk>
Apps
Search & Reporting
Explore Splunk Enterprise
Explore Data
Explore data and define how Hunk parses that data


splunk>
App: Search & Reporting
Administrator
Messages
Settings
Activity
Help
Find
Search
Pivot
Reports
Alerts
Dashboards
Search & Reporting
New Search
Save As
Close
index="digital_analytics" | top 5 useragent
All time
9,465 events (before 4/18/15 4:53:32.000 AM)
Job
Smart Mode
Events
Patterns
Statistics (5)
Visualization
20 Per Page
Format
Preview
useragent
count
percent
Safari/7.0
995
10.512414
Opera/9.20 (Windows NT 6.0; U; en)
905
9.561543
Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1)
870
9.191759
Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4322)
865
9.138933
Mozilla/5.0 (Windows; U; Windows NT 5.1; en-GB; rv:1.8.1.6) Gecko/20070725 Firefox/2.0.0.6
864
9.128368


Search & Reporting
Save As
Close
Report
Dashboard Panel
Alert
Event Type

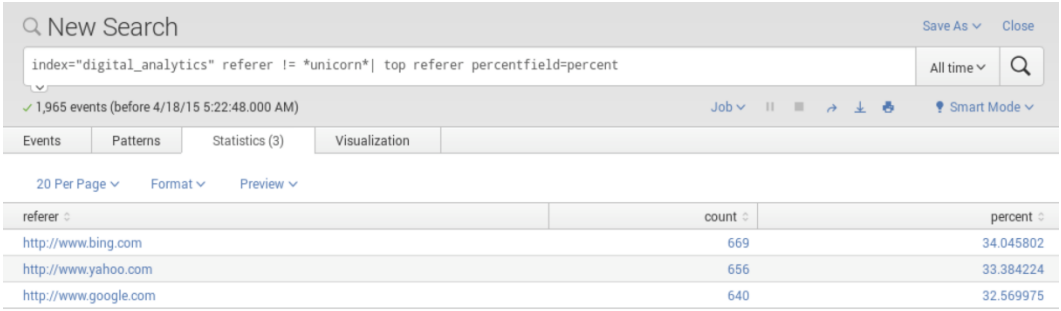

Save As Report
Title
Report top 5 browsers
Description
optional
Visualization
Column
None
Time Range Picker
Yes
No
Cancel
Save


splunk>
App: Search & Reporting
Administrator
Search
Pivot
Reports
Alerts
Dashboards
Reports
Reports are based on single searches and can include visualizations, statistics and/or even
Open the report in Pivot or Search to refine the parameters or further explore the data.
6 Reports
All
Yours
This App's
filter
i
Title
Actions
Owner
Errors in the last 24 hours
Open in Search
Edit
nobody
Errors in the last hour
Open in Search
Edit
nobody
License Usage Data Cube
Open in Search
Edit
nobody
Messages by minute last 3 hours
Open in Search
Edit
nobody
Report top 5 browsers
Open in Search
Edit
admin
Splunk errors last 24 hours
Open in Search
Edit
nobody

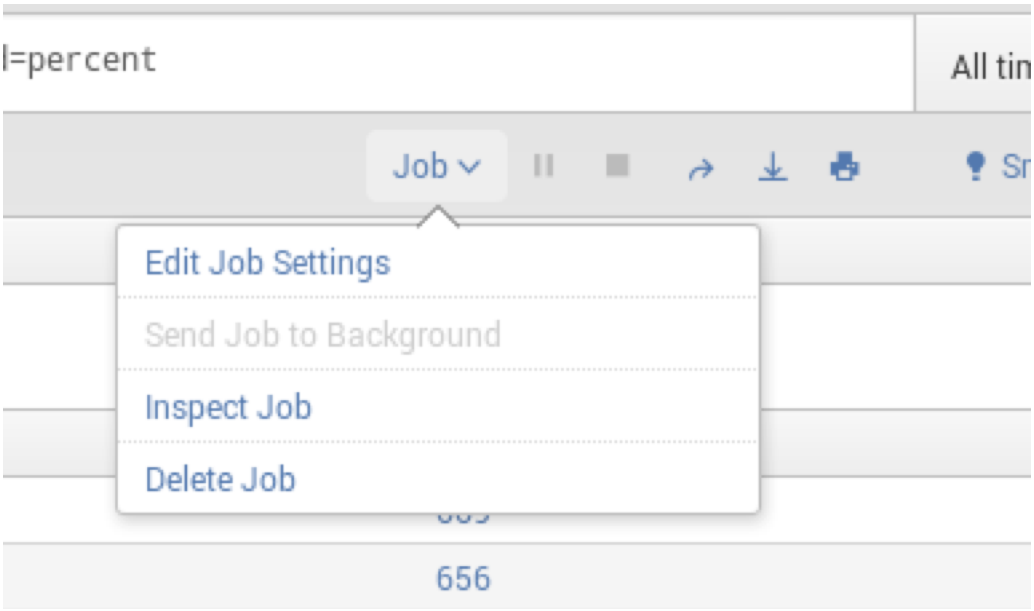


4 results 20 per page

useragent	count	percent
Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4322)	7	50.000000
Safari/7.0	4	28.571429
Safari/5.0	2	14.285714



referer	count	percent
http://www.bing.com	669	34.045802
http://www.yahoo.com	656	33.384224
http://www.google.com	640	32.569975



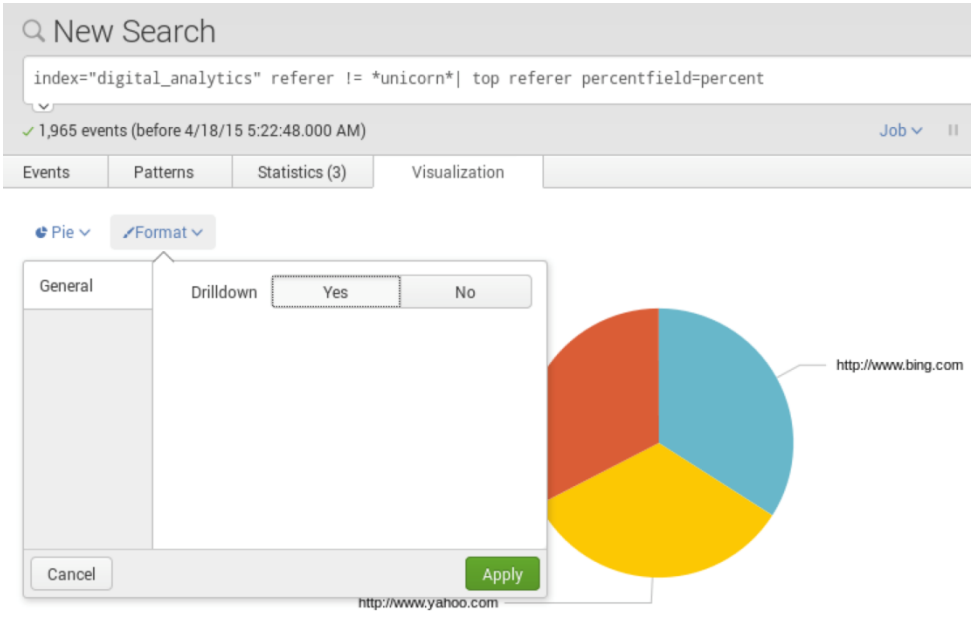
Search job inspector

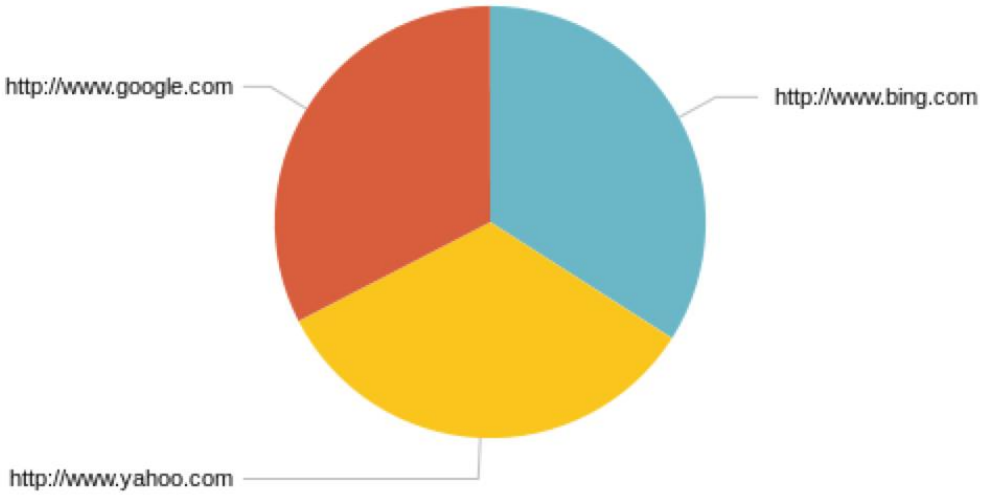
This search has completed and has returned **3** results by scanning **9,465** events in **27.765** seconds.

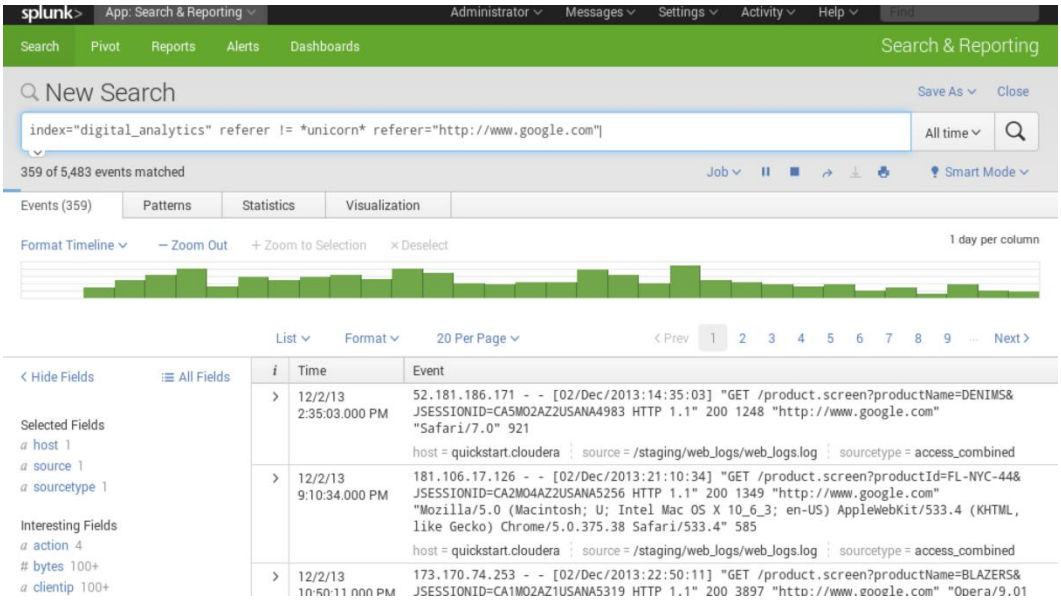
(SID: 1429359768.55) search.log

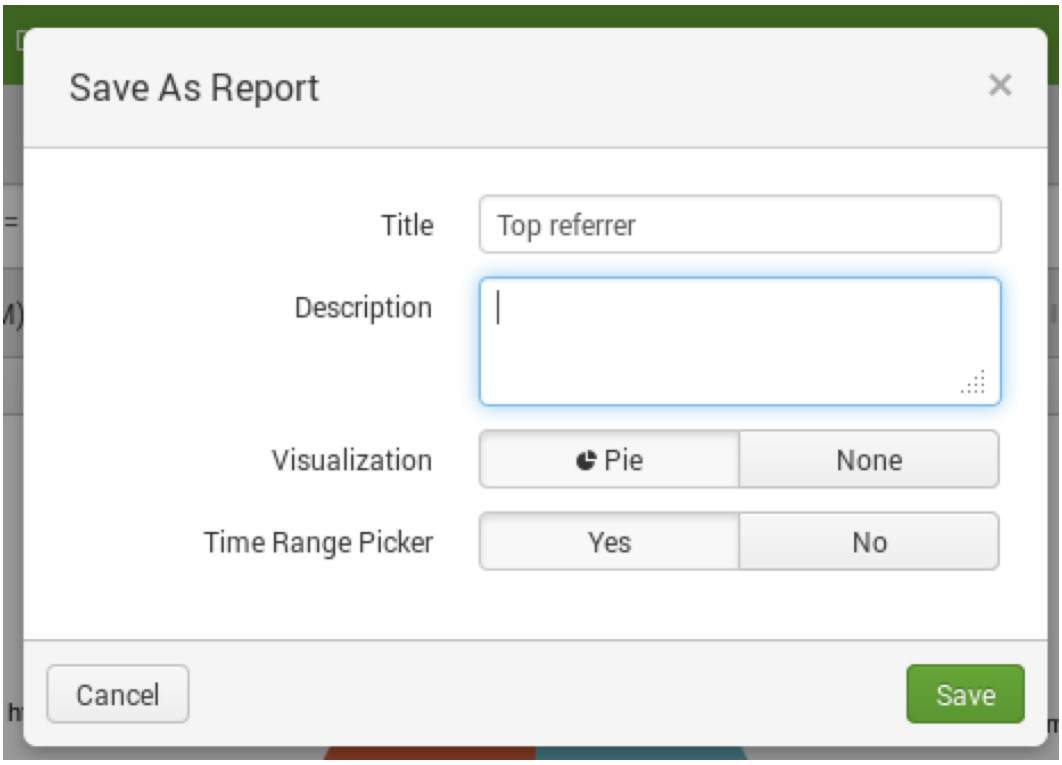
Execution costs

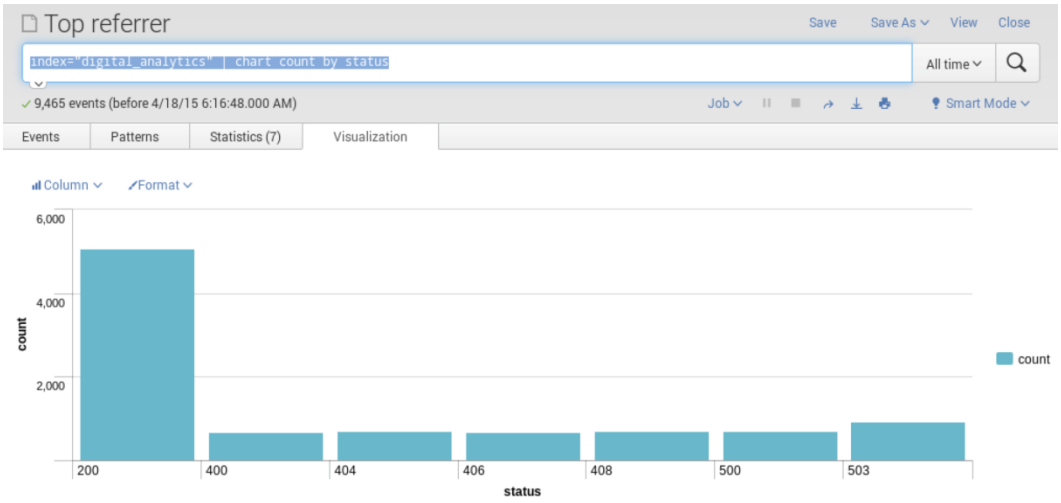
Duration (seconds)		Component	Invocations	Input count	Output count
	0.01	command.addinfo	6	1,965	1,965
	0.00	command.fields	6	1,965	1,965
	0.02	command.pretop	6	1,965	16
	0.05	command.search	6	9,465	1,965
	0.05	command.search.filter	6	-	-
	21.93	command.stdin	5	-	9,465
	0.07	command.stdin.cpd2sr	4	9,465	9,465
	0.00	command.stdin.calcfields	4	9,465	9,465

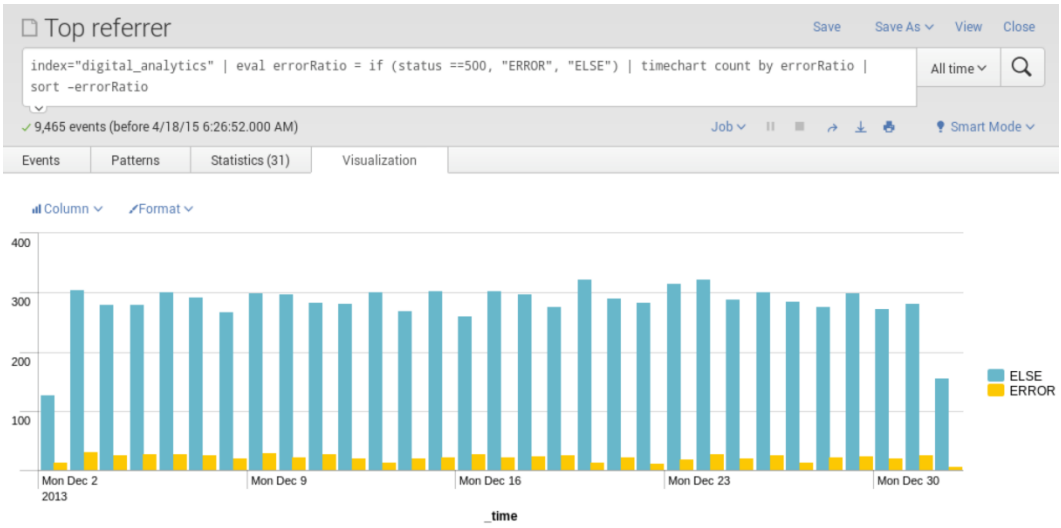

New Search
index="digital_analytics" referer != *unicorn*| top referer percentfield=percent
1,965 events (before 4/18/15 5:22:48.000 AM)
Job
Events
Patterns
Statistics (3)
Visualization
Pie
Format
General
Drilldown
Yes
No
Cancel
Apply
http://www.bing.com
http://www.yahoo.com

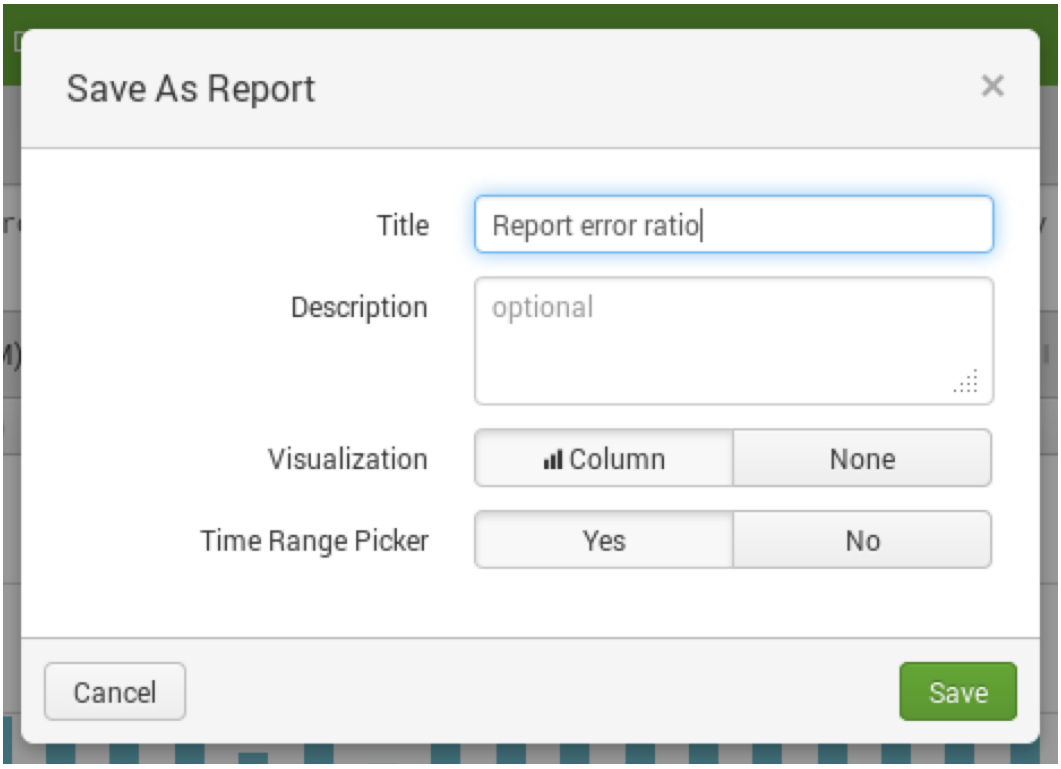

http://www.google.com
http://www.bing.com
http://www.yahoo.com

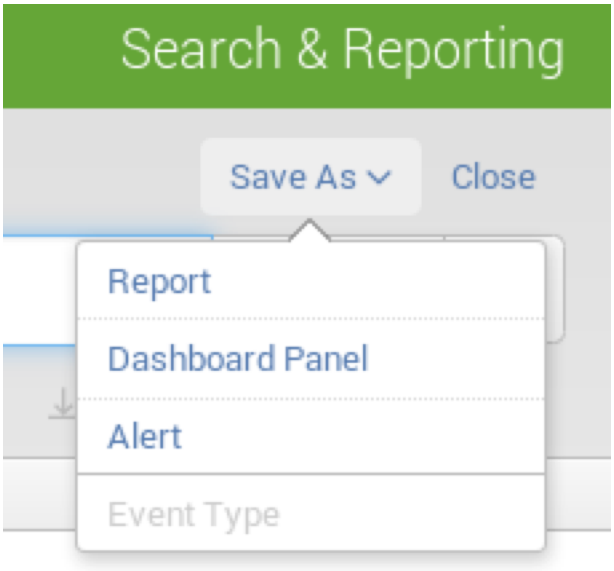

splunk> App: Search & Reporting
Administrator Messages Settings Activity Help
Search Pivot Reports Alerts Dashboards
Search & Reporting
New Search
Save As Close
index="digital_analytics" referer != *unicorn* referer="http://www.google.com"
All time
359 of 5,483 events matched
Job Smart Mode
Events (359) Patterns Statistics Visualization
Format Timeline Zoom Out Zoom to Selection Deselect
1 day per column
List Format 20 Per Page
Prev 1 2 3 4 5 6 7 8 9 Next
Hide Fields All Fields
Selected Fields
a host 1
a source 1
a sourcetype 1
Interesting Fields
a action 4
bytes 100+
a clientip 100+
i Time Event
12/2/13 2:35:03.000 PM
52.181.186.171 - - [02/Dec/2013:14:35:03] "GET /product.screen?productName=DENIMS& JSESSIONID=CA5MO2AZ2USANA4983 HTTP 1.1" 200 1248 "http://www.google.com" "Safari/7.0" 921
host = quickstart.cloudera source = /staging/web_logs/web_logs.log sourcetype = access_combined
12/2/13 9:10:34.000 PM
181.106.17.126 - - [02/Dec/2013:21:10:34] "GET /product.screen?productId=FL-NYC-44& JSESSIONID=CA2MO4AZ2USANA5256 HTTP 1.1" 200 1349 "http://www.google.com" "Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10_6_3; en-US) AppleWebKit/533.4 (KHTML, like Gecko) Chrome/5.0.375.38 Safari/533.4" 585
host = quickstart.cloudera source = /staging/web_logs/web_logs.log sourcetype = access_combined
12/2/13
173.170.74.253 - - [02/Dec/2013:22:50:11] "GET /product.screen?productName=BLAZERS& JSESSIONID=CA1MO2AZ1USANA5319 HTTP 1.1" 200 3897 "http://www.google.com" "Opera/9.01

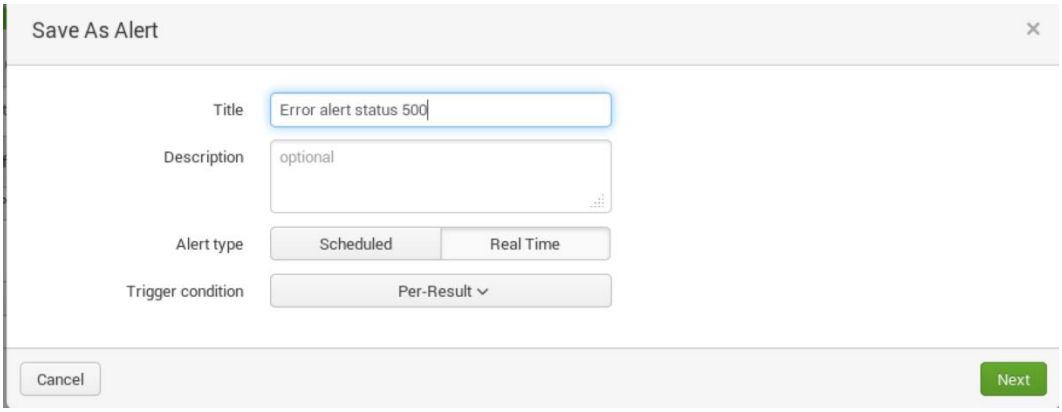

Save As Report
Title
Top referrer
Description
Visualization
Pie
None
Time Range Picker
Yes
No
Cancel
Save

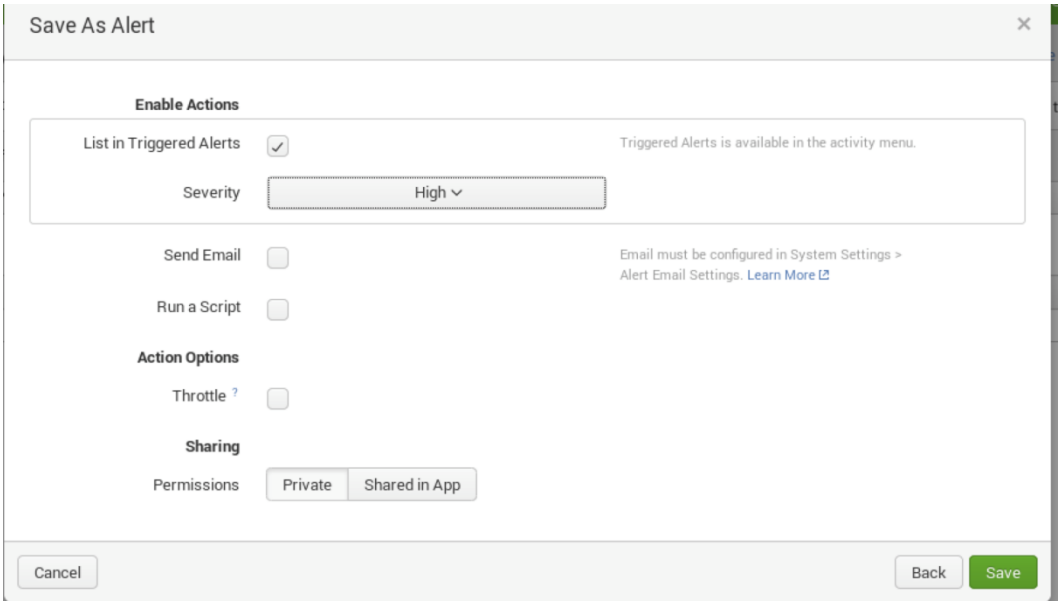

Top referrer
Save
Save As
View
Close
index="digital_analytics" | chart count by status
All time
9,465 events (before 4/18/15 6:16:48.000 AM)
Job
Smart Mode
Events
Patterns
Statistics (7)
Visualization
Column
Format
6,000
4,000
2,000
count
200
400
404
406
408
500
503
status
count

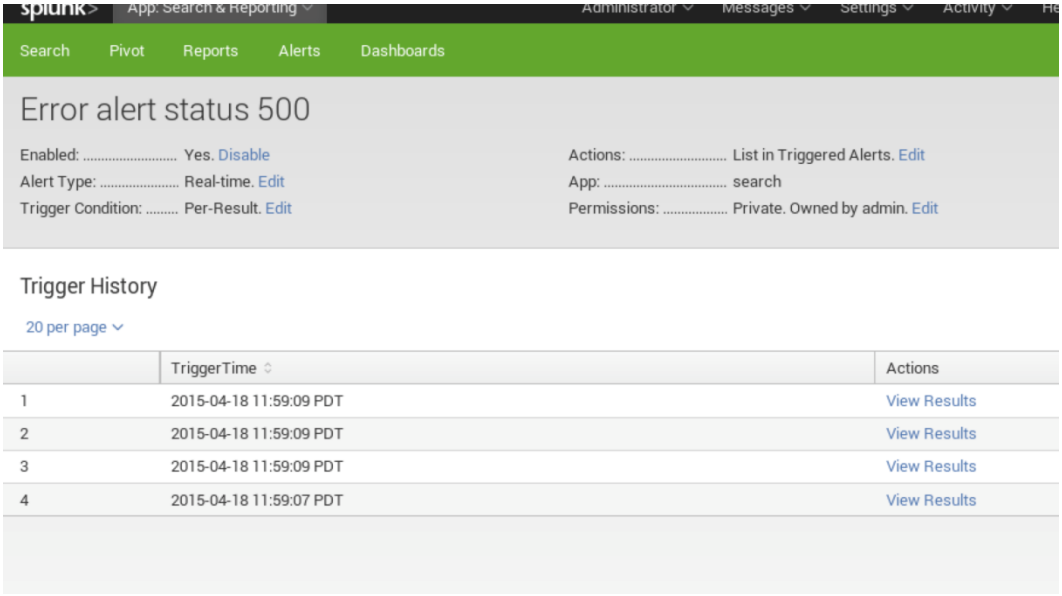

Top referrer
Save
Save As
View
Close
index="digital_analytics" | eval errorRatio = if (status ==500, "ERROR", "ELSE") | timechart count by errorRatio | sort -errorRatio
All time
9,465 events (before 4/18/15 6:26:52.000 AM)
Job
Smart Mode
Events
Patterns
Statistics (31)
Visualization
Column
Format
400
300
200
100
Mon Dec 2
2013
Mon Dec 9
Mon Dec 16
Mon Dec 23
Mon Dec 30
_time
ELSE
ERROR

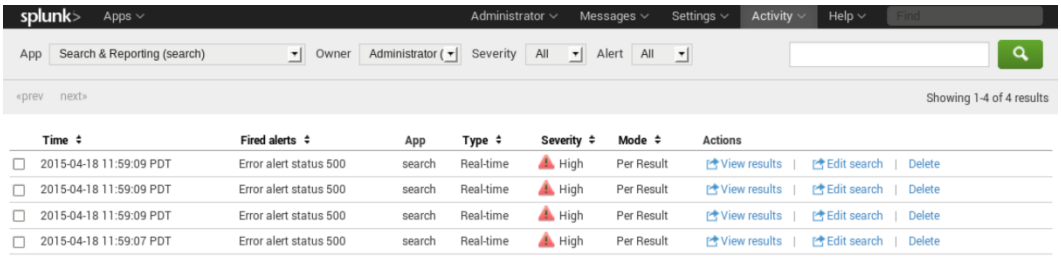

Save As Report
Title
Report error ratio
Description
optional
Visualization
Column
None
Time Range Picker
Yes
No
Cancel
Save

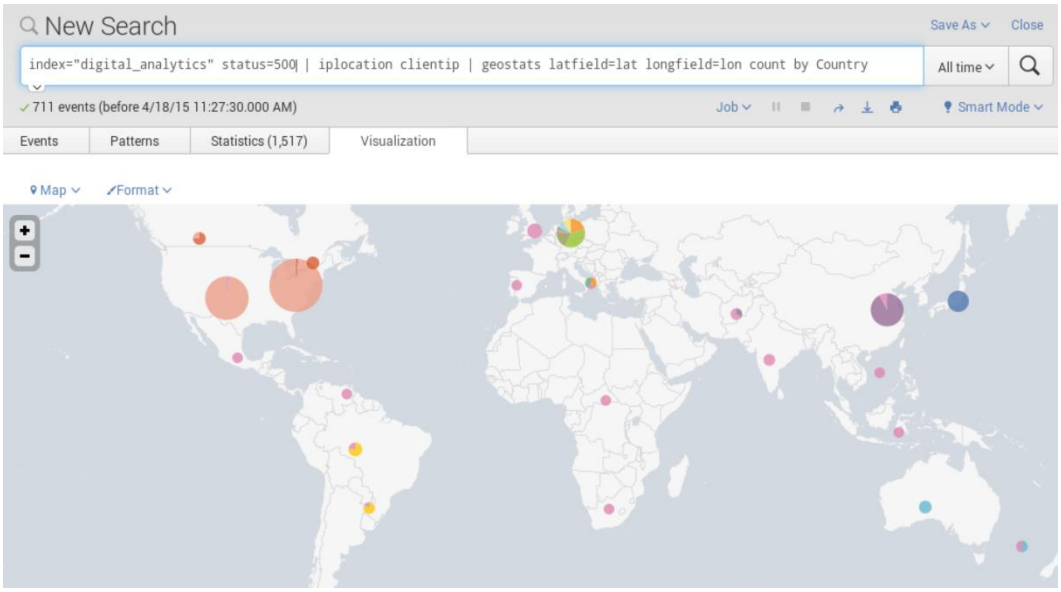

Search & Reporting
Save As
Close
Report
Dashboard Panel
Alert
Event Type

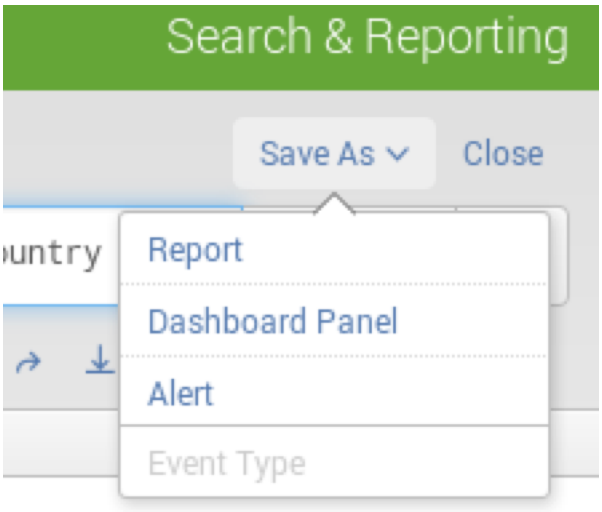

Save As Alert
Title
Error alert status 500
Description
optional
Alert type
Scheduled
Real Time
Trigger condition
Per-Result
Cancel
Next

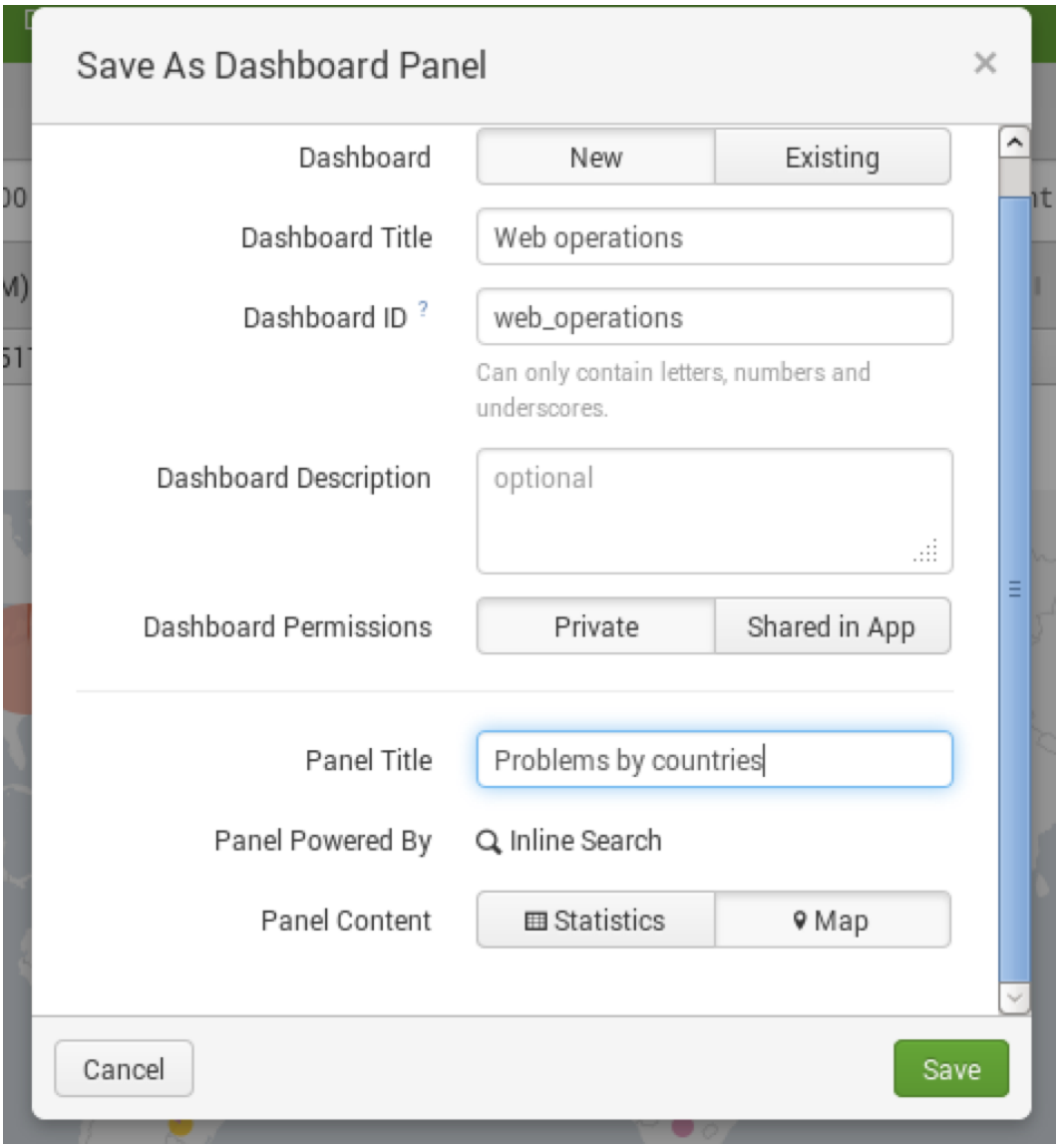

Save As Alert
Enable Actions
List in Triggered Alerts
Triggered Alerts is available in the activity menu.
Severity
High
Send Email
Email must be configured in System Settings > Alert Email Settings. Learn More
Run a Script
Action Options
Throttle
Sharing
Permissions
Private
Shared in App
Cancel
Back
Save

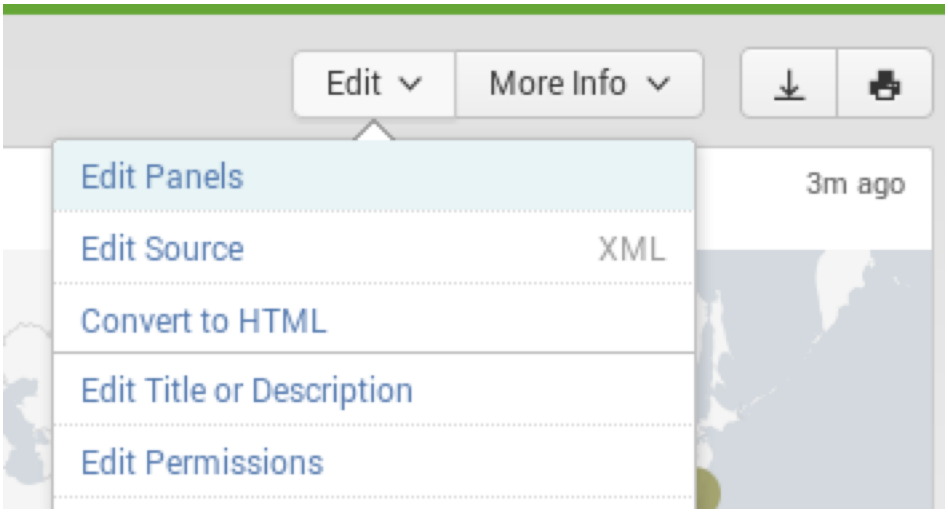

splunk> App: Search & Reporting
Administrator
Messages
Settings
Activity
Search
Pivot
Reports
Alerts
Dashboards
Error alert status 500
Enabled: ..... Yes. Disable
Alert Type: ..... Real-time. Edit
Trigger Condition: ..... Per-Result. Edit
Actions: ..... List in Triggered Alerts. Edit
App: ..... search
Permissions: ..... Private. Owned by admin. Edit
Trigger History
20 per page
TriggerTime
Actions
1 2015-04-18 11:59:09 PDT View Results
2 2015-04-18 11:59:09 PDT View Results
3 2015-04-18 11:59:09 PDT View Results
4 2015-04-18 11:59:07 PDT View Results

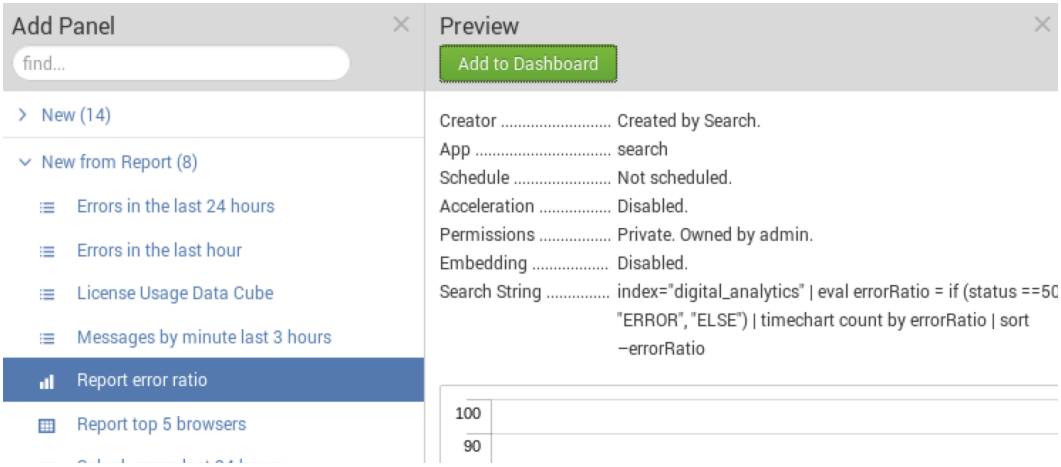

splunk> Apps
Administrator
Messages
Settings
Activity
Help
Find
App Search & Reporting (search)
Owner Administrator
Severity All
Alert All
«prev next»
Showing 1-4 of 4 results
Time
Fired alerts
App
Type
Severity
Mode
Actions
2015-04-18 11:59:09 PDT Error alert status 500 search Real-time High Per Result View results | Edit search | Delete
2015-04-18 11:59:09 PDT Error alert status 500 search Real-time High Per Result View results | Edit search | Delete
2015-04-18 11:59:09 PDT Error alert status 500 search Real-time High Per Result View results | Edit search | Delete
2015-04-18 11:59:07 PDT Error alert status 500 search Real-time High Per Result View results | Edit search | Delete

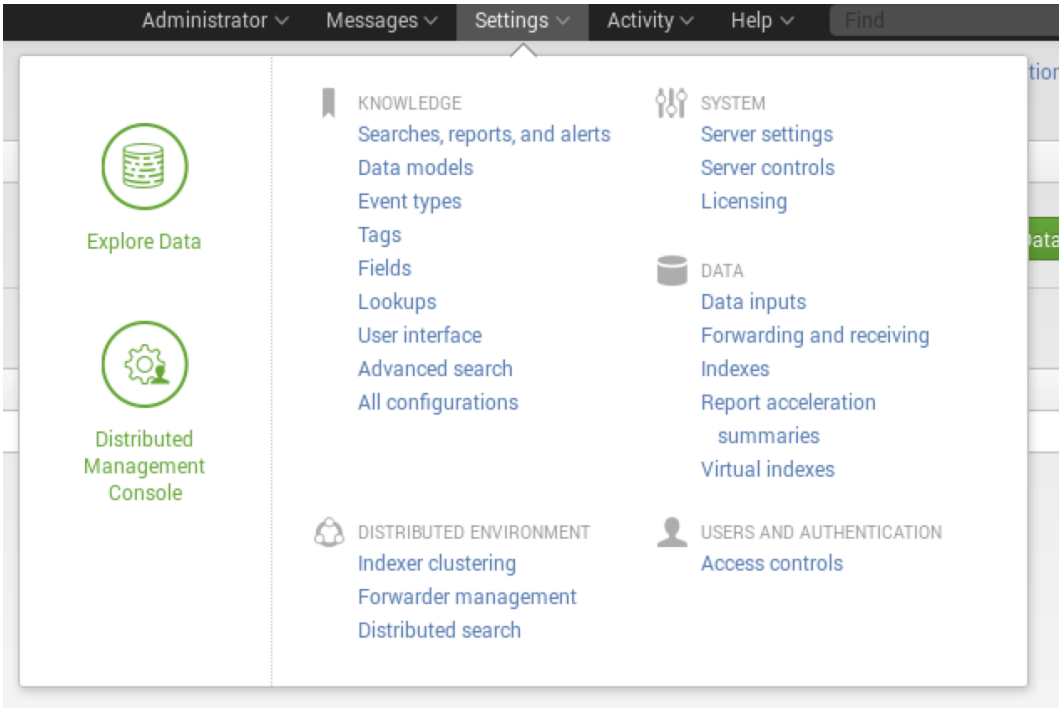

New Search
Save As
Close
index="digital_analytics" status=500 | iplocation clientip | geostats latfield=lat longfield=lon count by Country
All time
711 events (before 4/18/15 11:27:30.000 AM)
Job
Smart Mode
Events
Patterns
Statistics (1,517)
Visualization
Map
Format

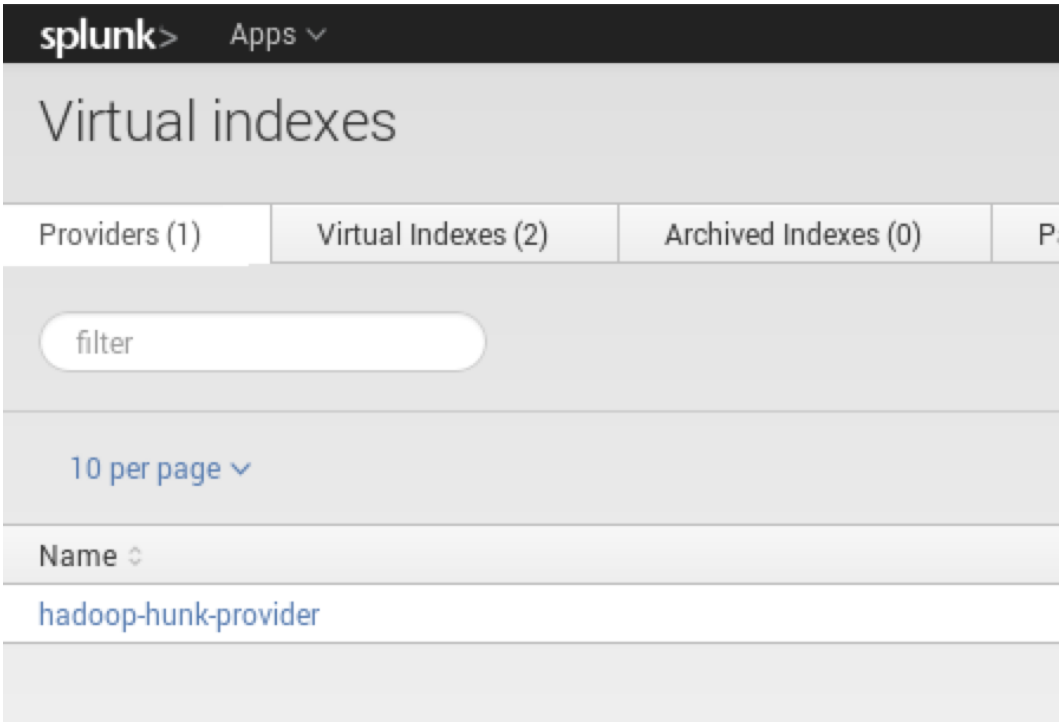

Search & Reporting
Save As
Close
Report
Dashboard Panel
Alert
Event Type


Save As Dashboard Panel
Dashboard
New
Existing
Dashboard Title
Web operations
Dashboard ID
web_operations
Can only contain letters, numbers and underscores.
Dashboard Description
optional
Dashboard Permissions
Private
Shared in App
Panel Title
Problems by countries
Panel Powered By
Inline Search
Panel Content
Statistics
Map
Cancel
Save

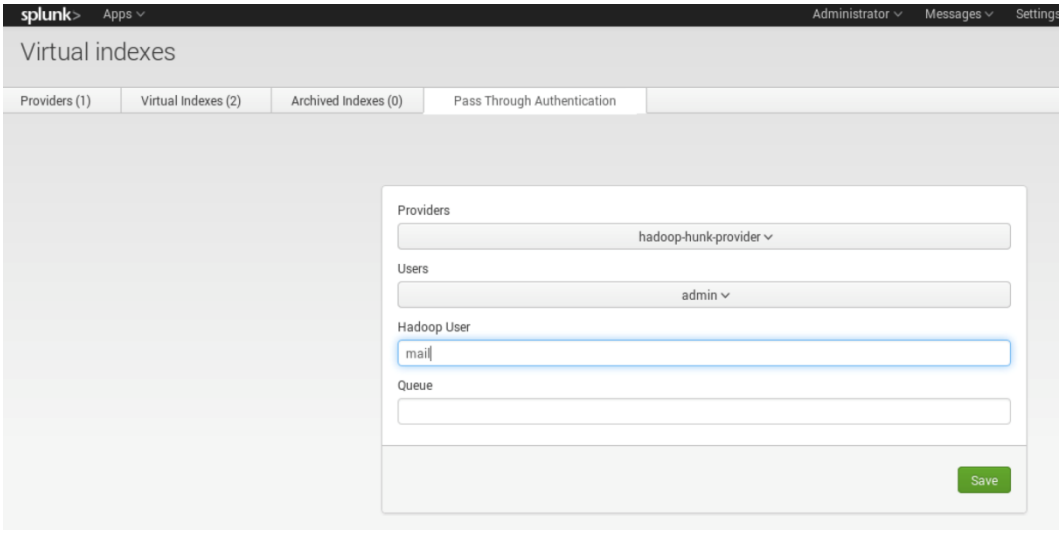

Edit
More Info
Edit Panels
Edit Source
XML
Convert to HTML
Edit Title or Description
Edit Permissions
3m ago

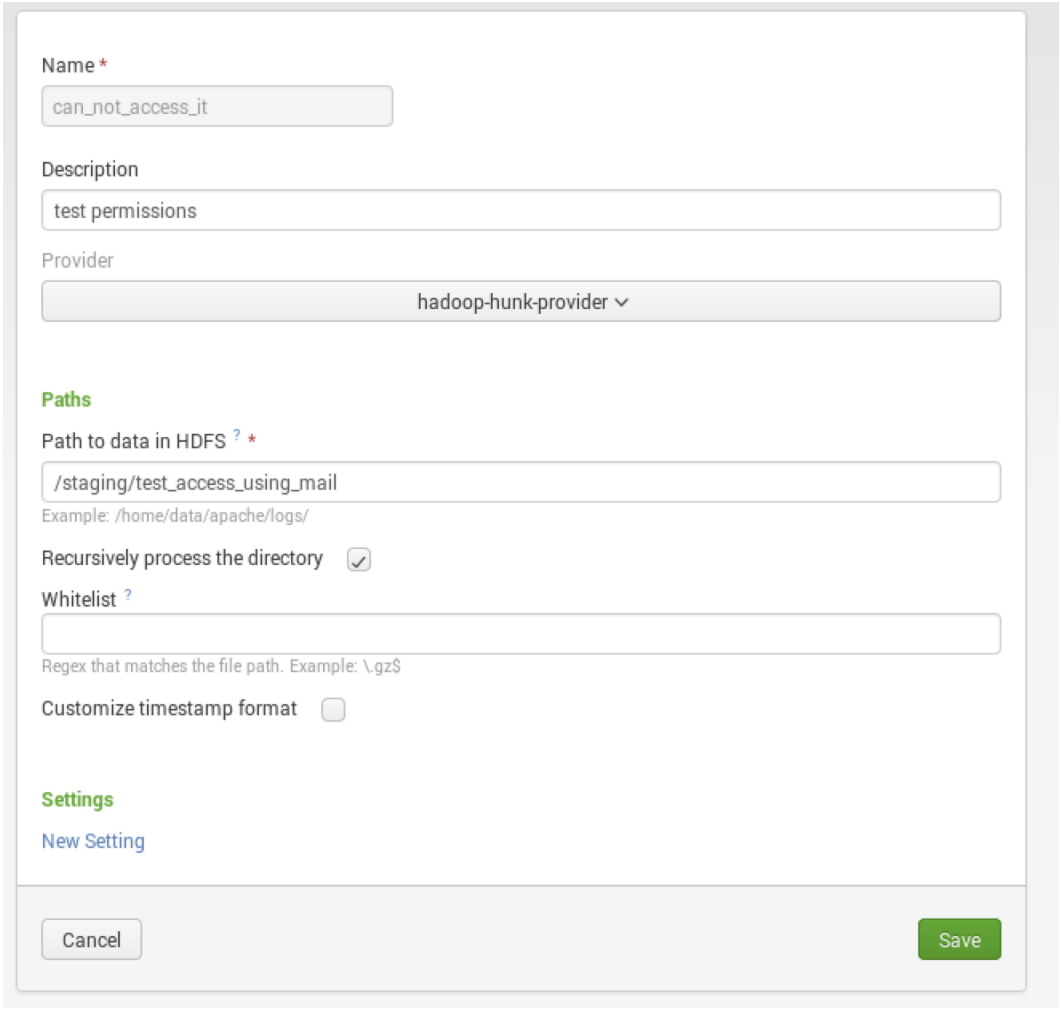

Add Panel
find...
New (14)
New from Report (8)
Errors in the last 24 hours
Errors in the last hour
License Usage Data Cube
Messages by minute last 3 hours
Report error ratio
Report top 5 browsers
Preview
Add to Dashboard
Creator ..... Created by Search.
App ..... search
Schedule ..... Not scheduled.
Acceleration ..... Disabled.
Permissions ..... Private. Owned by admin.
Embedding ..... Disabled.
Search String ..... index="digital_analytics" | eval errorRatio = if (status ==50
"ERROR", "ELSE") | timechart count by errorRatio | sort
–errorRatio
100
90

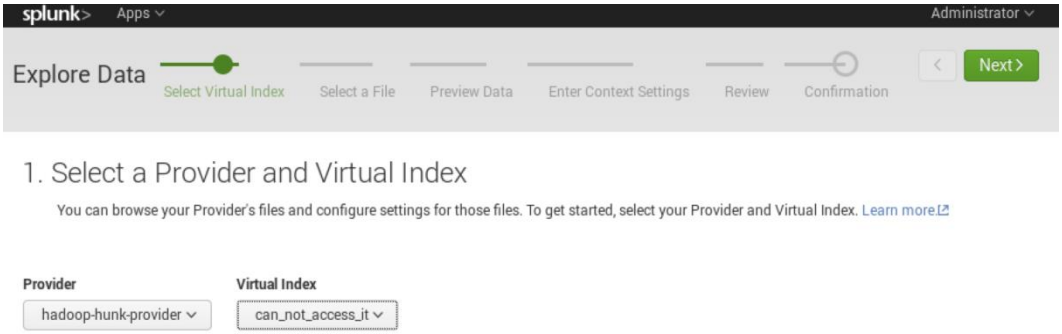

Administrator
Messages
Settings
Activity
Help
Find
Explore Data
Distributed Management Console
KNOWLEDGE
Searches, reports, and alerts
Data models
Event types
Tags
Fields
Lookups
User interface
Advanced search
All configurations
SYSTEM
Server settings
Server controls
Licensing
DATA
Data inputs
Forwarding and receiving
Indexes
Report acceleration summaries
Virtual indexes
DISTRIBUTED ENVIRONMENT
Indexer clustering
Forwarder management
Distributed search
USERS AND AUTHENTICATION
Access controls

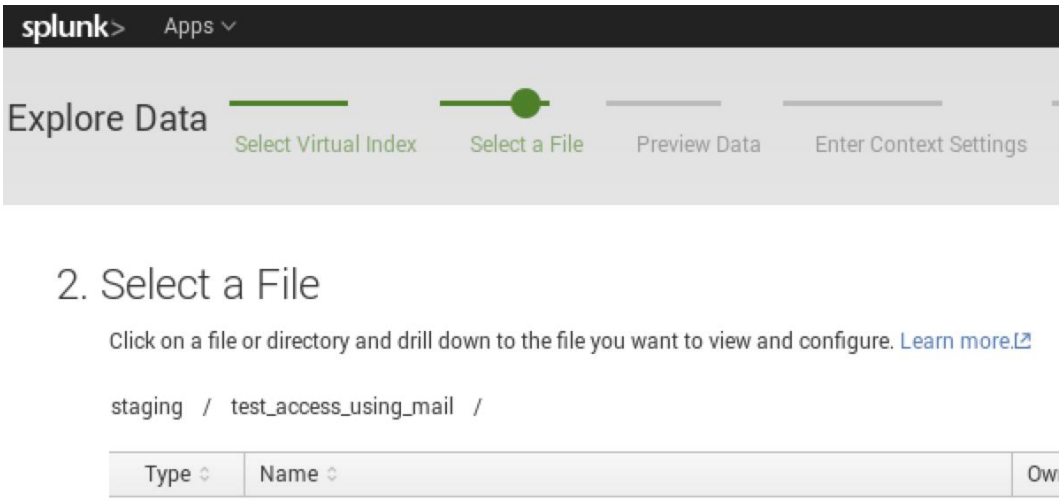

splunk> Apps
Virtual indexes
Providers (1)
Virtual Indexes (2)
Archived Indexes (0)
filter
10 per page
Name
hadoop-hunk-provider

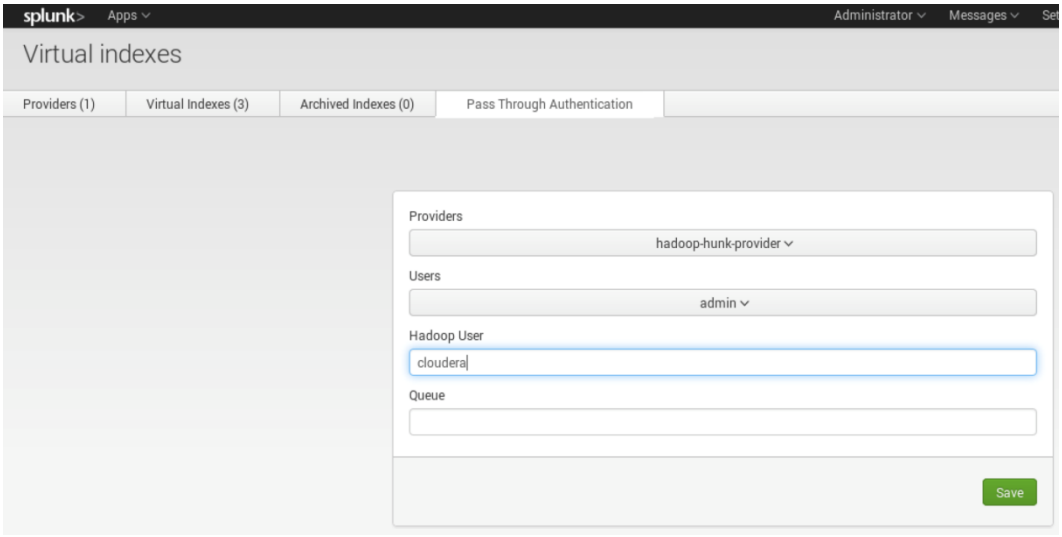

Hadoop Cluster Information
Hadoop Version
Hadoop 2.x, (Yarn)
File System *
hdfs://quickstart.cloudera:8020
Example: hdfs://namenode.example.com:8020
Enable Pass Through Authentication

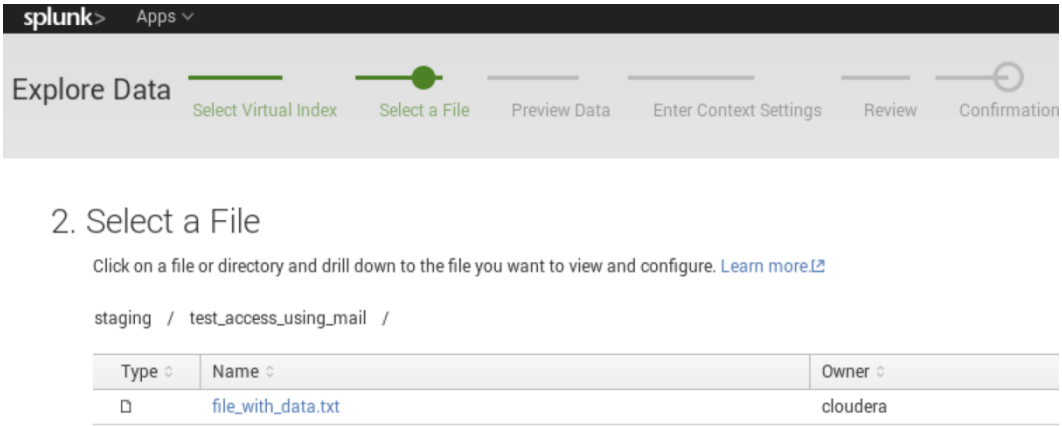

splunk>
Apps
Administrator
Messages
Settings
Virtual indexes
Providers (1)
Virtual Indexes (2)
Archived Indexes (0)
Pass Through Authentication
Providers
hadoop-hunk-provider
Users
admin
Hadoop User
mail
Queue
Save


Name *
can_not_access_it
Description
test permissions
Provider
hadoop-hunk-provider
Paths
Path to data in HDFS *
/staging/test_access_using_mail
Example: /home/data/apache/logs/
Recursively process the directory
Whitelist
Regex that matches the file path. Example: \.gz$
Customize timestamp format
Settings
New Setting
Cancel
Save

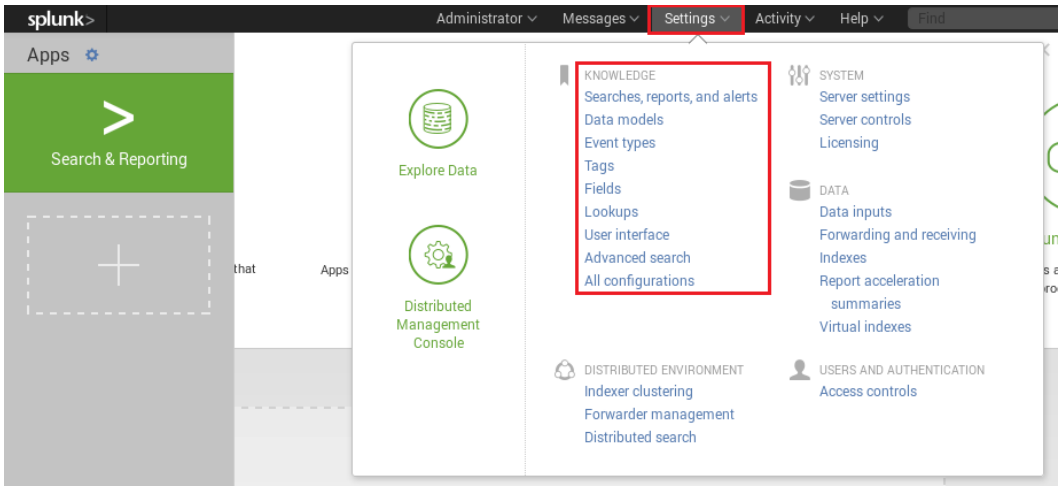

splunk> Apps
Administrator
Explore Data
Select Virtual Index
Select a File
Preview Data
Enter Context Settings
Review
Confirmation
Next
1. Select a Provider and Virtual Index
You can browse your Provider's files and configure settings for those files. To get started, select your Provider and Virtual Index. Learn more
Provider
hadoop-hunk-provider
Virtual Index
can_not_access_it

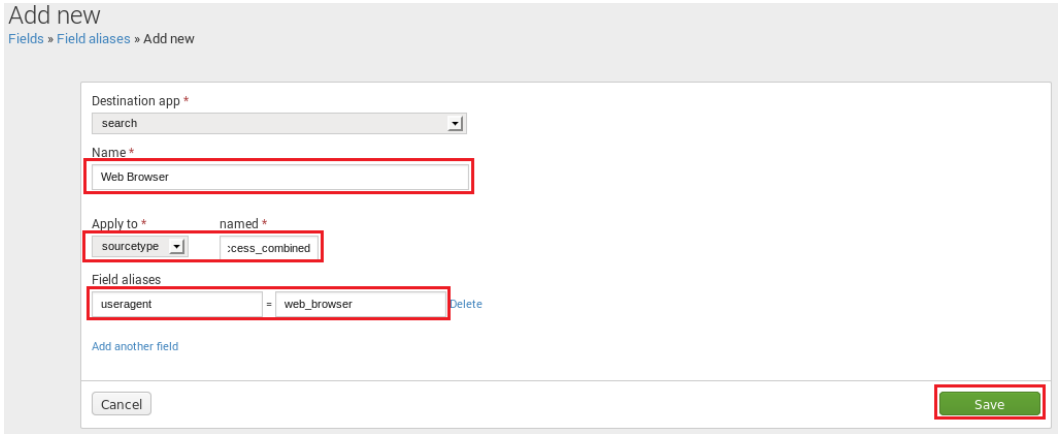

splunk> Apps
Explore Data
Select Virtual Index
Select a File
Preview Data
Enter Context Settings
2. Select a File
Click on a file or directory and drill down to the file you want to view and configure. Learn more
staging / test_access_using_mail /
Type
Name
Ow

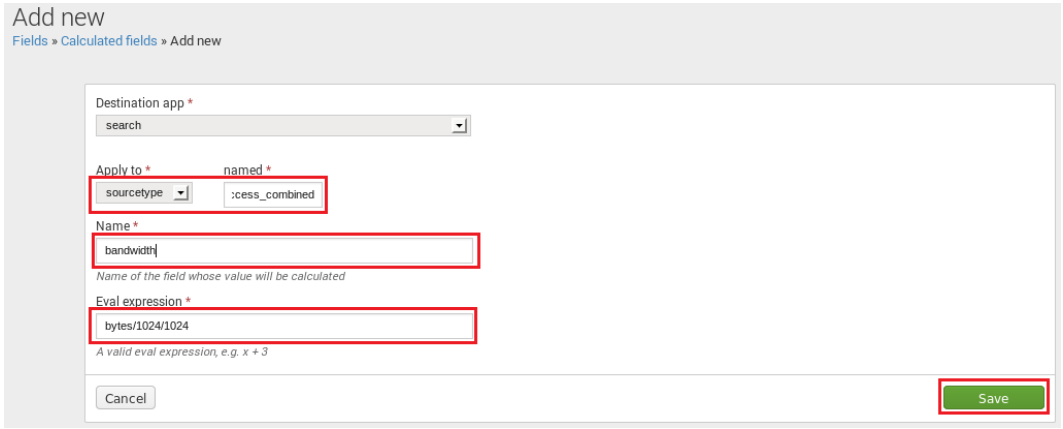

splunk> Apps
Administrator
Messages
Virtual indexes
Providers (1)
Virtual Indexes (3)
Archived Indexes (0)
Pass Through Authentication
Providers
hadoop-hunk-provider
Users
admin
Hadoop User
cloudera
Queue
Save


splunk> Apps
Explore Data
Select Virtual Index
Select a File
Preview Data
Enter Context Settings
Review
Confirmation
2. Select a File
Click on a file or directory and drill down to the file you want to view and configure. Learn more
staging / test_access_using_mail /
Type
Name
Owner
file_with_data.txt
cloudera

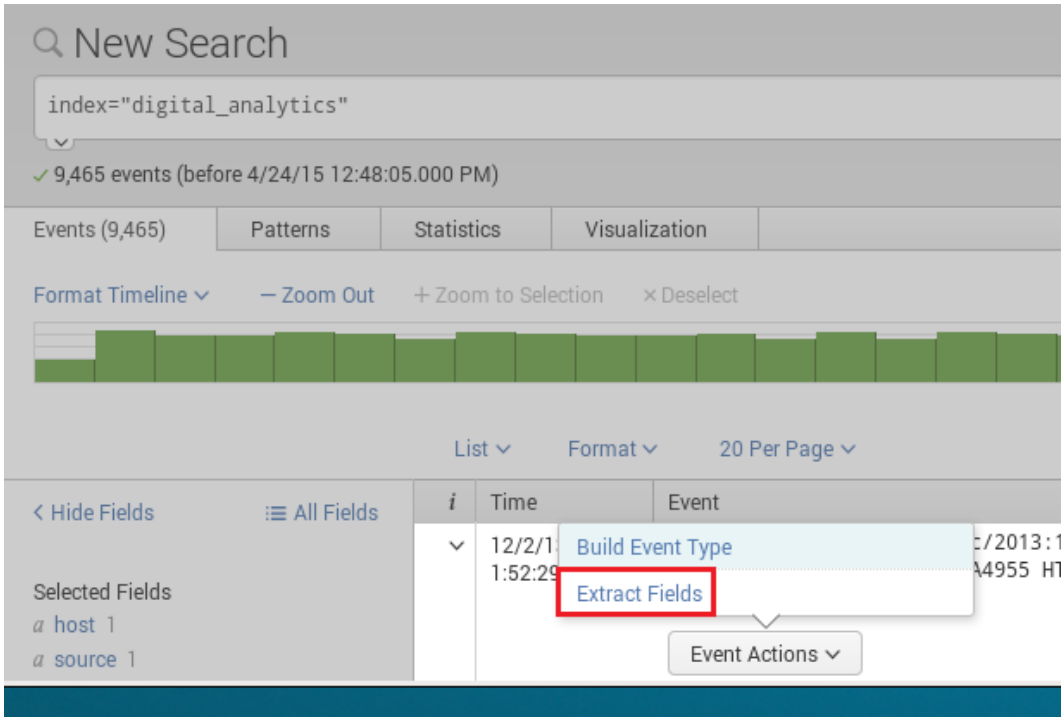
Chapter 3: Meeting Hunk Features

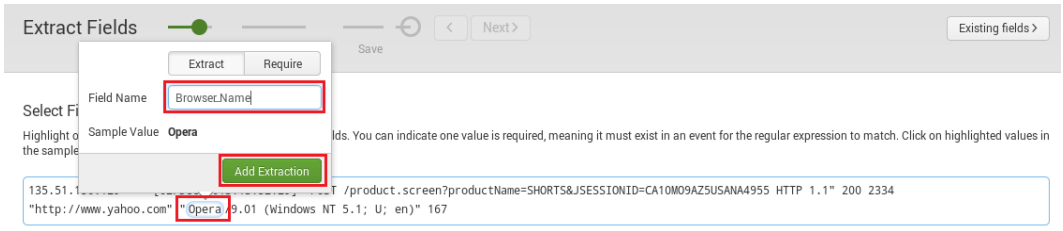




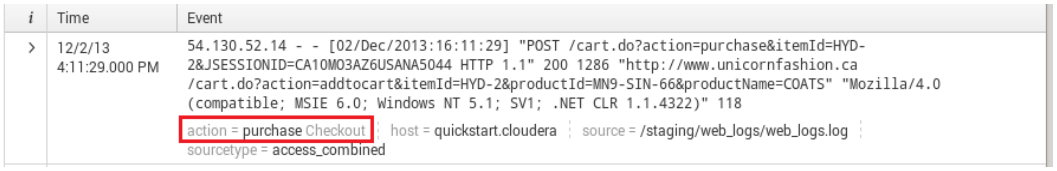


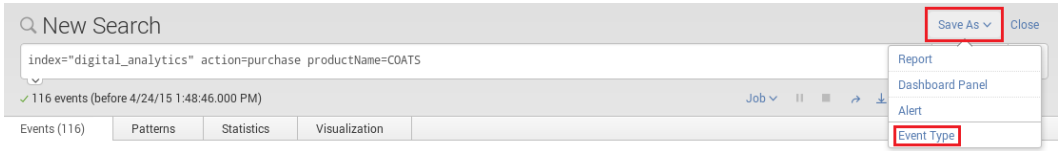
Country	sum(bandwidth)
United States	6.786556
China	1.636466
Japan	0.971759
United Kingdom	0.627292
Germany	0.570653
Canada	0.395550
Brazil	0.379452

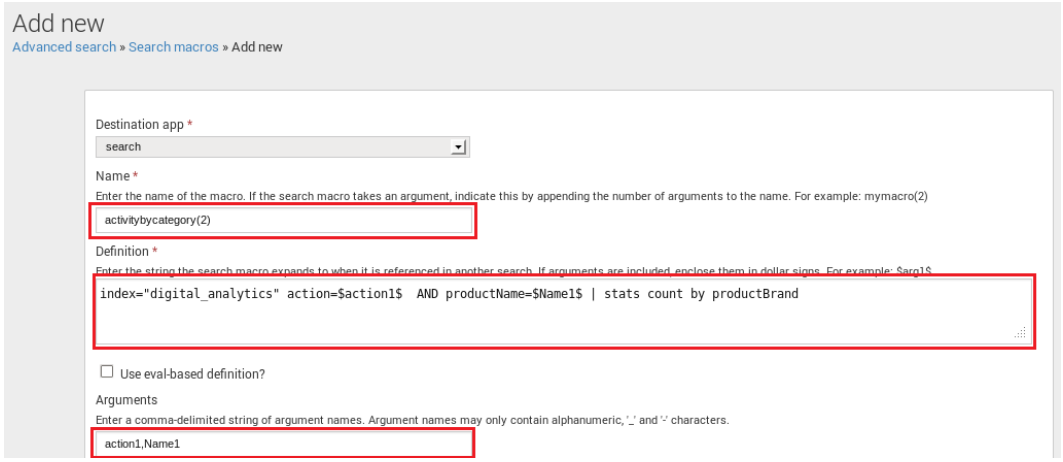

New Search
index="digital_analytics"
9,465 events (before 4/24/15 12:48:05.000 PM)
Events (9,465)
Patterns
Statistics
Visualization
Format Timeline
Zoom Out
Zoom to Selection
Deselect
List
Format
20 Per Page
Hide Fields
All Fields
Selected Fields
a host 1
a source 1
i
Time
Event
Build Event Type
Extract Fields
Event Actions

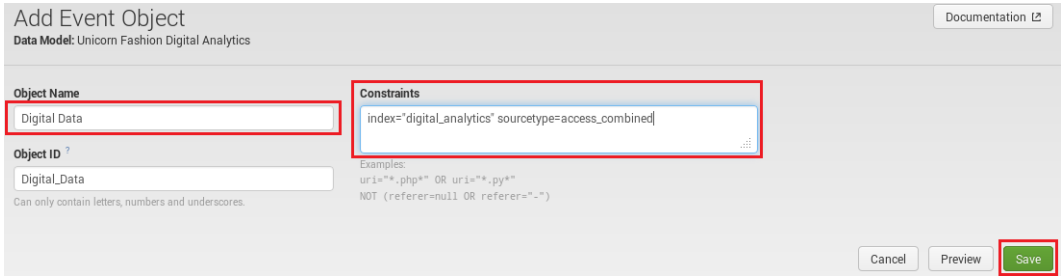

Extract Fields
Next
Save
Existing fields
Extract
Require
Field Name
Browser.Name
Sample Value
Opera
Add Extraction
/product.screen?productName=SHORTS&JSESSIONID=CA10MO9AZ5USANA4955 HTTP 1.1" 200 2334
"http://www.yahoo.com"
9.01 (Windows NT 5.1; U; en)" 167

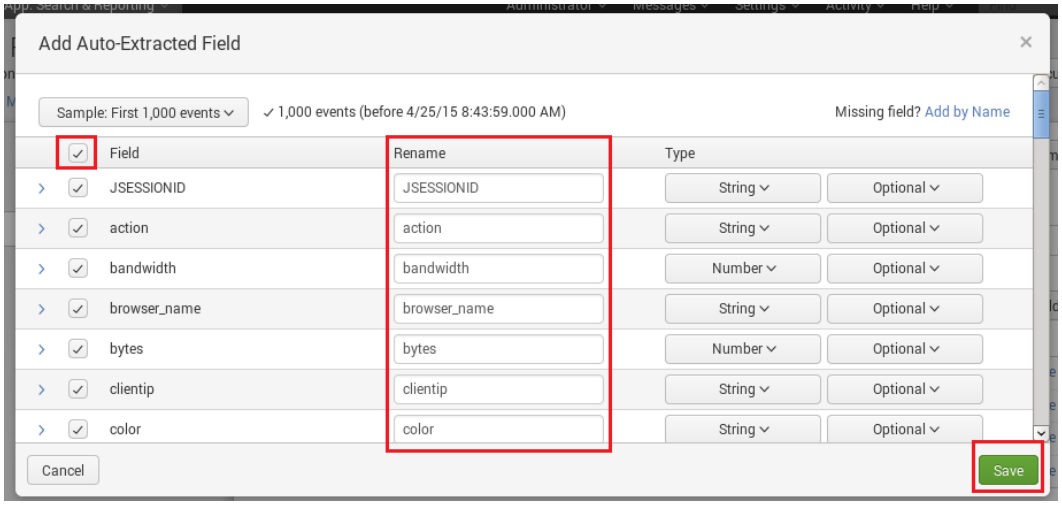

12/2/13
4:11:29.000 PM
54.130.52.14 - - [02/Dec/2013:16:11:29] "POST /cart.do?action=purchase&itemId=HYD-2&JSESSIONID=CA10MO3AZ6USANA5044 HTTP 1.1" 200 1286 "http://www.unicornfashion.ca /cart.do?action=addtocart&itemId=HYD-2&productId=MN9-SIN-66&productName=COATS" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4322)" 118
Event Actions
Type
Field
Value
Actions
Selected
host
quickstart.cloudera
source
/staging/web_logs/web_logs.log
sourcetype
access_combined
Event
JSESSIONID
CA10MO3AZ6USANA5044
action
purchase
bandwidth
0.001226
browser_name
Mozilla
bytes
1286
clientip
54.130.52.14
file
cart.do
Edit Tags

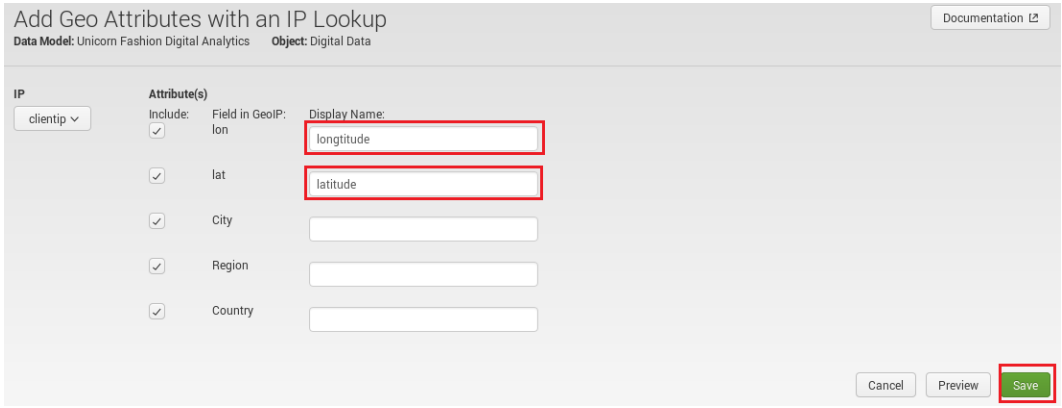

i
Time
Event
12/2/13
4:11:29.000 PM
54.130.52.14 - - [02/Dec/2013:16:11:29] "POST /cart.do?action=purchase&itemId=HYD-2&JSESSIONID=CA10MO3AZ6USANA5044 HTTP 1.1" 200 1286 "http://www.unicornfashion.ca /cart.do?action=addtocart&itemId=HYD-2&productId=MN9-SIN-66&productName=COATS" "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4322)" 118
action = purchase Checkout
host = quickstart.cloudera
source = /staging/web_logs/web_logs.log
sourcetype = access_combined

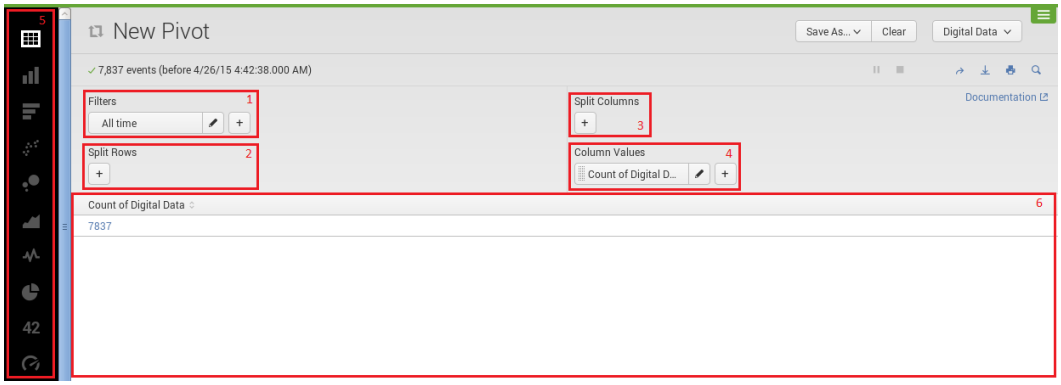

New Search
Save As
Close
index="digital_analytics" action=purchase productName=COATS
116 events (before 4/24/15 1:48:46.000 PM)
Job
Events (116)
Patterns
Statistics
Visualization
Report
Dashboard Panel
Alert
Event Type

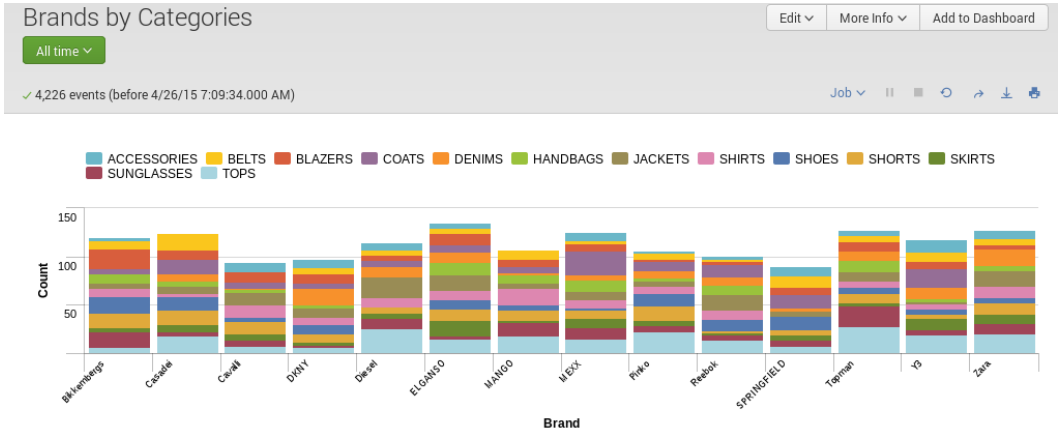

Add new
Advanced search » Search macros » Add new
Destination app *
search
Name *
Enter the name of the macro. If the search macro takes an argument, indicate this by appending the number of arguments to the name. For example: mymacro(2)
activitybycategory(2)
Definition *
index="digital_analytics" action=$action1$ AND productName=$Name1$ | stats count by productBrand
Use eval-based definition?
Arguments
Enter a comma-delimited string of argument names. Argument names may only contain alphanumeric, '_' and '-' characters.
action1,Name1


Add Event Object
Data Model: Unicorn Fashion Digital Analytics
Documentation
Object Name
Digital Data
Object ID
Digital_Data
Can only contain letters, numbers and underscores.
Constraints
index="digital_analytics" sourcetype=access_combined
Examples:
uri="*.php*" OR uri="*.py*"
NOT (referer=null OR referer="-")
Cancel
Preview
Save

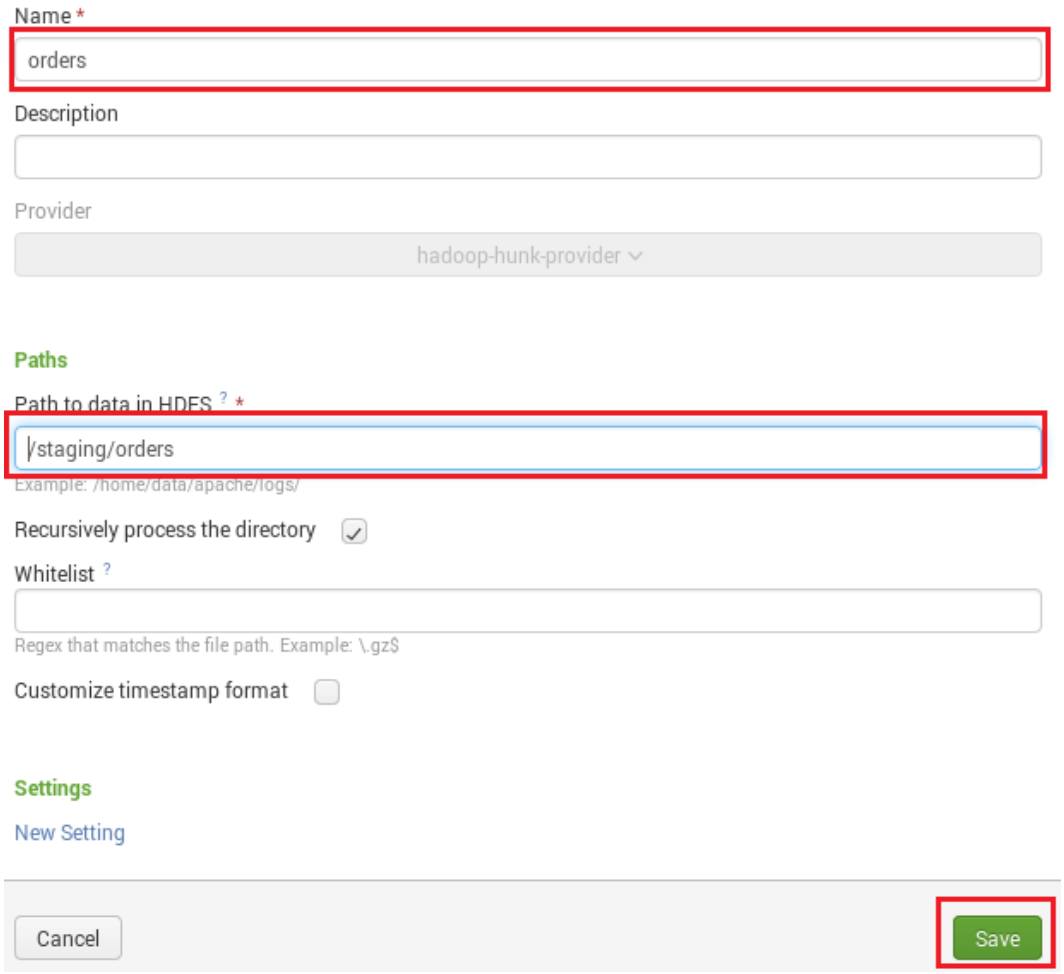

Add Auto-Extracted Field
Sample: First 1,000 events
✓ 1,000 events (before 4/25/15 8:43:59.000 AM)
Missing field? Add by Name
Field
Rename
Type
JSESSIONID
action
bandwidth
browser_name
bytes
clientip
color
String
Number
Optional
Cancel
Save

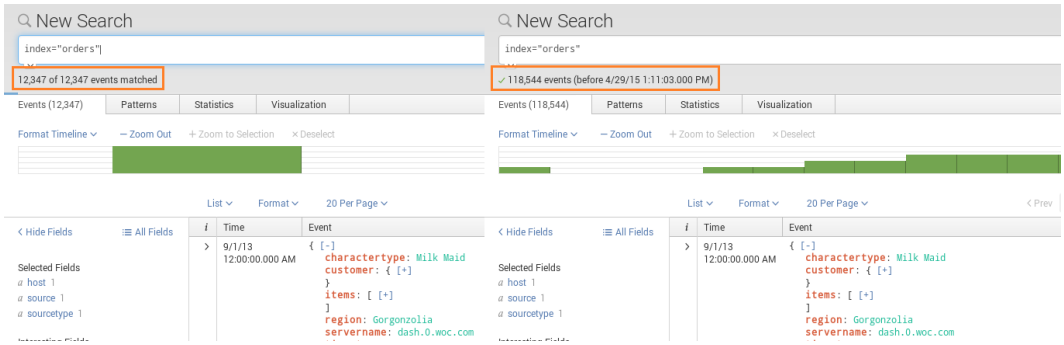

Add Geo Attributes with an IP Lookup
Data Model: Unicorn Fashion Digital Analytics
Object: Digital Data
Documentation
IP
clientip
Attribute(s)
Include:
Field in GeoIP:
Display Name:
lon
longtitude
lat
latitude
City
Region
Country
Cancel
Preview
Save

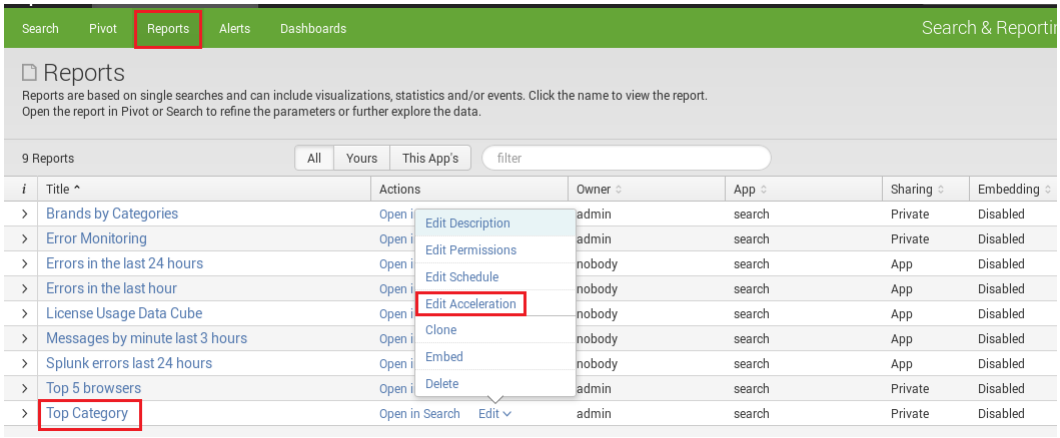

New Pivot
Save As...
Clear
Digital Data
7,837 events (before 4/26/15 4:42:38.000 AM)
Documentation
Filters
1
All time
Split Columns
3
Split Rows
2
Column Values
4
Count of Digital D...
Count of Digital Data
6
7837
42

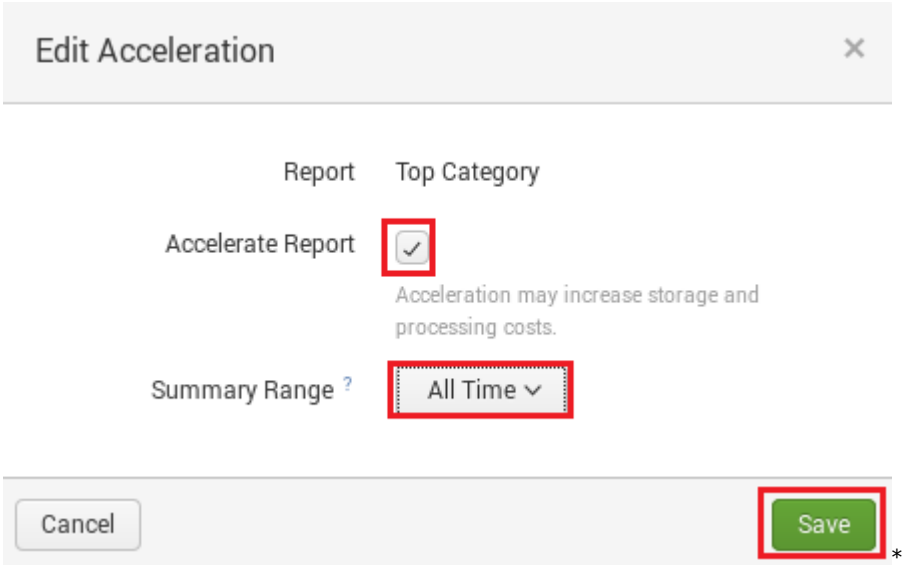

Brands by Categories
Edit
More Info
Add to Dashboard
All time
4,226 events (before 4/26/15 7:09:34.000 AM)
Job
ACCESSORIES
BELTS
BLAZERS
COATS
DENIMS
HANDBAGS
JACKETS
SHIRTS
SHOES
SHORTS
SKIRTS
SUNGLASSES
TOPS
150
100
50
Count
Bikkembergs
Casadei
Cavalli
DKNY
Diesel
ELGANSO
MANGO
MEXX
Pinko
Reebok
SPRINGFIELD
Topman
Y3
Zara
Brand

Chapter 4: Adding Speed to Reports


Name *
orders
Description
Provider
hadoop-hunk-provider
Paths
Path to data in HDFS *
/staging/orders
Example: /home/data/apache/logs/
Recursively process the directory
Whitelist
Regex that matches the file path. Example: \.gz$
Customize timestamp format
Settings
New Setting
Cancel
Save


New Search
index="orders"
12,347 of 12,347 events matched
Events (12,347)
Patterns
Statistics
Visualization
Format Timeline
Zoom Out
Zoom to Selection
Deselect
List
Format
20 Per Page
Hide Fields
All Fields
Selected Fields
host 1
source 1
sourcetype 1
Time
Event
9/1/13
12:00:00.000 AM
{ [-]
charactertype: Milk Maid
customer: { [+]
}
items: [[+]
]
region: Gorgonzolia
servername: dash.0.woc.com
New Search
index="orders"
118,544 events (before 4/29/15 1:11:03.000 PM)
Events (118,544)
Patterns
Statistics
Visualization
Format Timeline
Zoom Out
Zoom to Selection
Deselect
List
Format
20 Per Page
Prev
Hide Fields
All Fields
Selected Fields
host 1
source 1
sourcetype 1
Time
Event
9/1/13
12:00:00.000 AM
{ [-]
charactertype: Milk Maid
customer: { [+]
}
items: [[+]
]
region: Gorgonzolia
servername: dash.0.woc.com


Search
Pivot
Reports
Alerts
Dashboards
Search & Reporti
Reports
Reports are based on single searches and can include visualizations, statistics and/or events. Click the name to view the report.
Open the report in Pivot or Search to refine the parameters or further explore the data.
9 Reports
All
Yours
This App's
filter
Title
Actions
Owner
App
Sharing
Embedding
Brands by Categories
Error Monitoring
Errors in the last 24 hours
Errors in the last hour
License Usage Data Cube
Messages by minute last 3 hours
Splunk errors last 24 hours
Top 5 browsers
Top Category
Open in Search
Edit
Edit Description
Edit Permissions
Edit Schedule
Edit Acceleration
Clone
Embed
Delete
admin
nobody
search
App
Private
Disabled



Dispatched at	Owner	Application	Size	Events	Run time	Expires	Status	Actions
4/29/15 2:06:13 PM	admin	search	0.08MB	118,544	00:00:07	Apr 29, 2015 2:17:52 PM	Done	Inspect | Save | Delete
index="orders" | top "items{}.category" | rename "items{}.category" as Category | sort -Category | eval message = "Hello World" *[earliest time, latest time]*								
4/29/15 2:03:25 PM	admin	search	0.10MB	118,544	00:01:27	Apr 29, 2015 2:15:23 PM	Done	Inspect | Save | Delete
index="orders" | top "items{}.category" | rename "items{}.category" as Category | sort -Category | eval message = "Hello World" *[earliest time, latest time]*								

Showing 1-2 of 2 items

Results per page 25

Summary ID	Normalized Summary ID	Reports Using Summary	Summarization Load	Access Count	Summary Status
d1f911276107e5a8	NS6c17c90570bf72ce	Top Category	0.1012	0 Last Access: Never	Complete Updated: 3m ago
918af19ead627c15	NS06aa76b5b48c65f2	1	0.0982	0 Last Access: Never	Complete Updated: 1m ago

Showing 1-2 of 2 items

Summary: NS6c17c90570bf72ce

Summary Status

Pending Updated: 17m ago

Actions

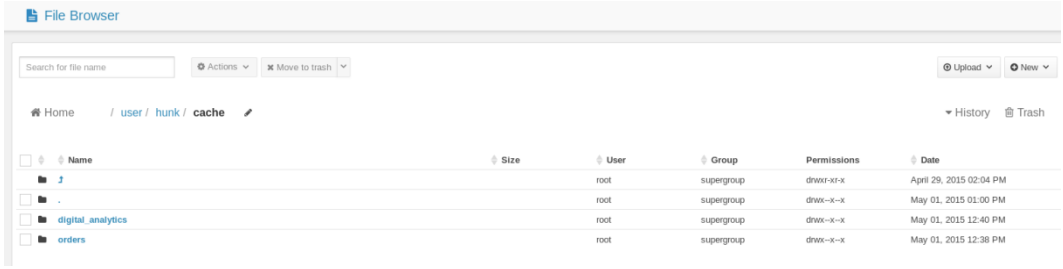
Verify | Update | Rebuild | Delete

Reports Using This Summary

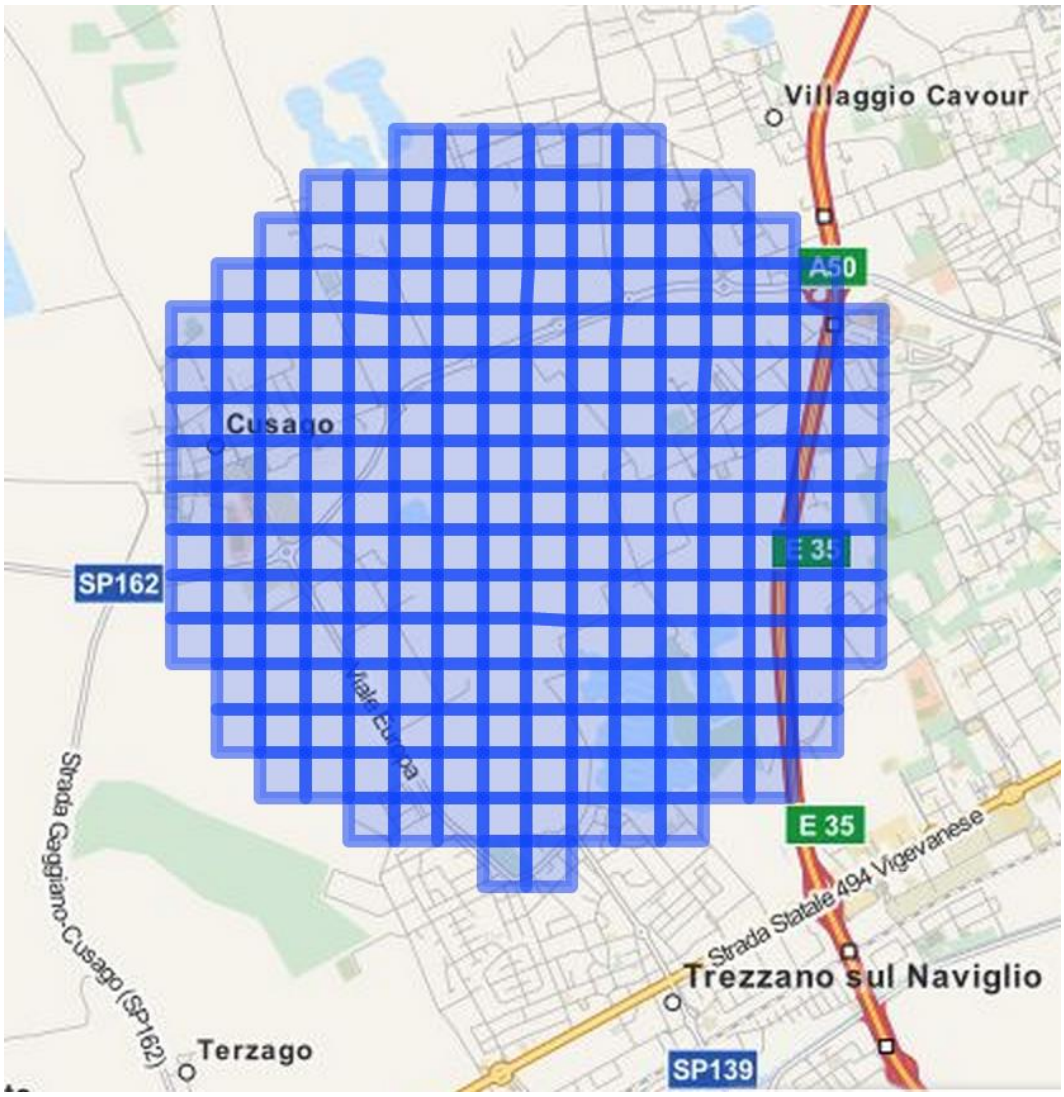
Search name	Owner	App
Top Category	admin	search

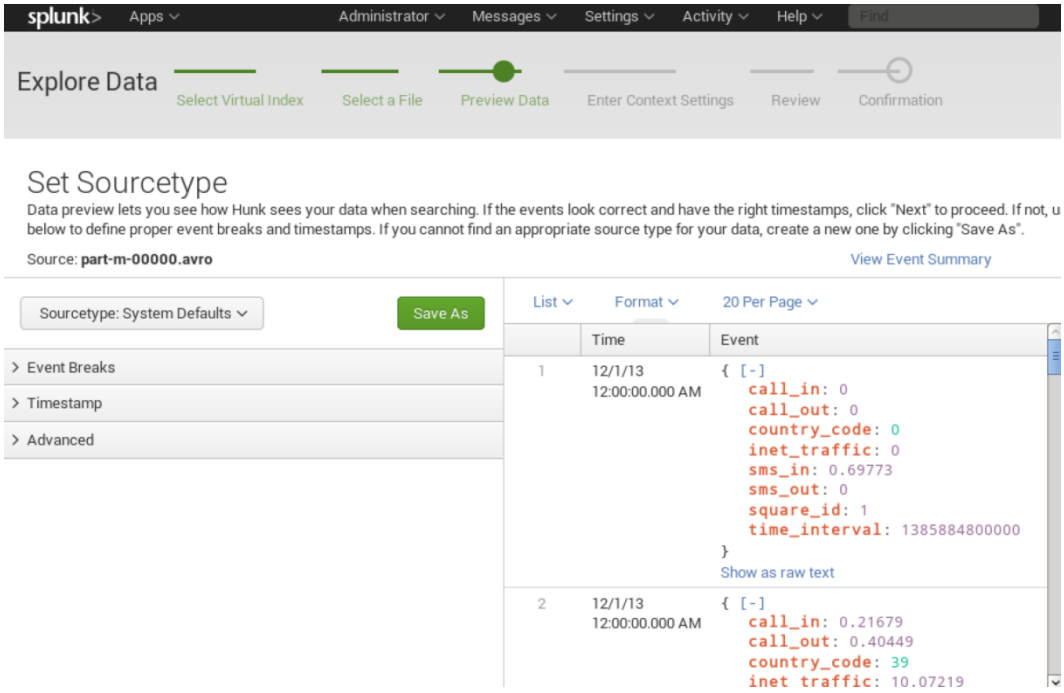
Details Learn more.

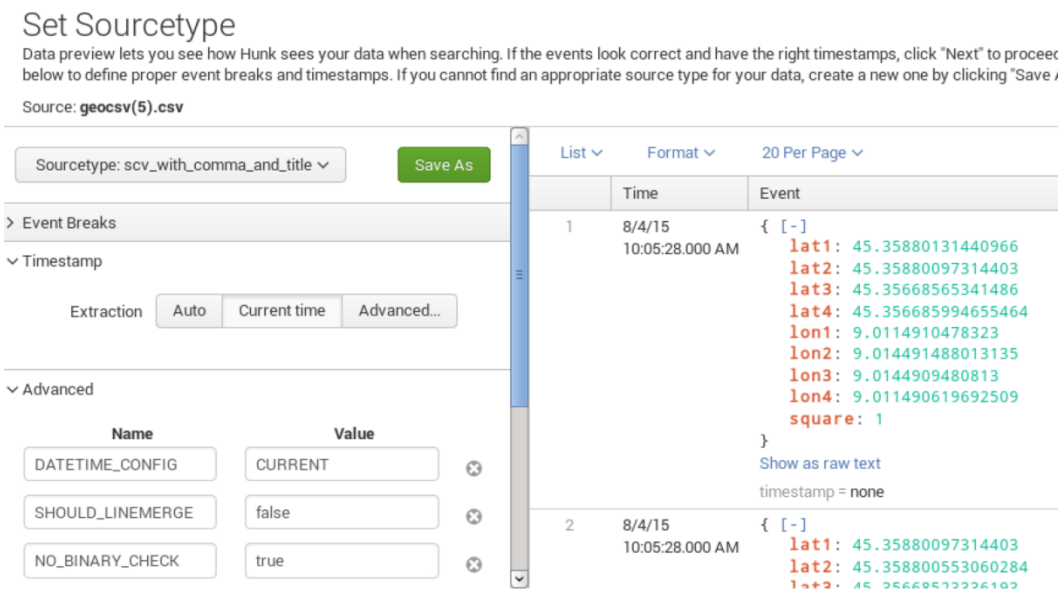
Summarization Load	0.0688
Access Count	1 Last Access: < 1 min ago
Size on Disk	0.00MB
Summary Range	All Time
Timespans	
Buckets	1
Chunks	1

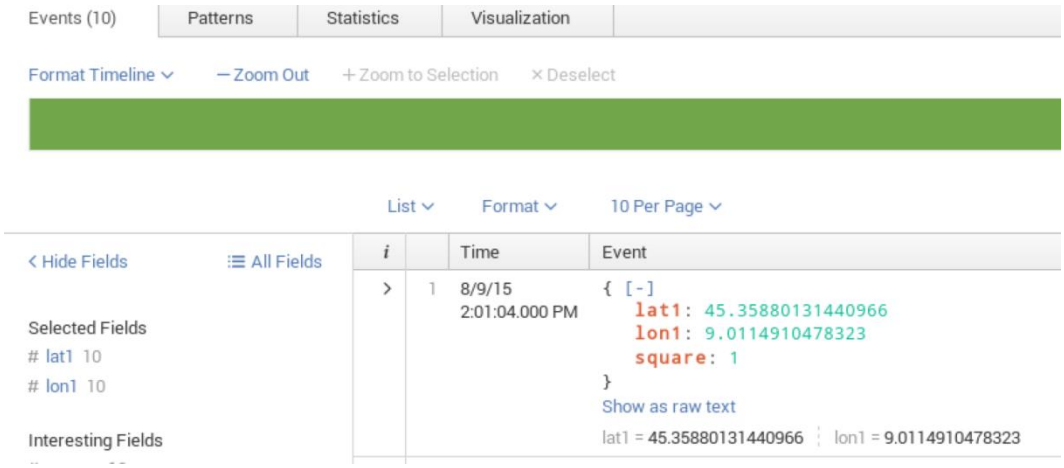


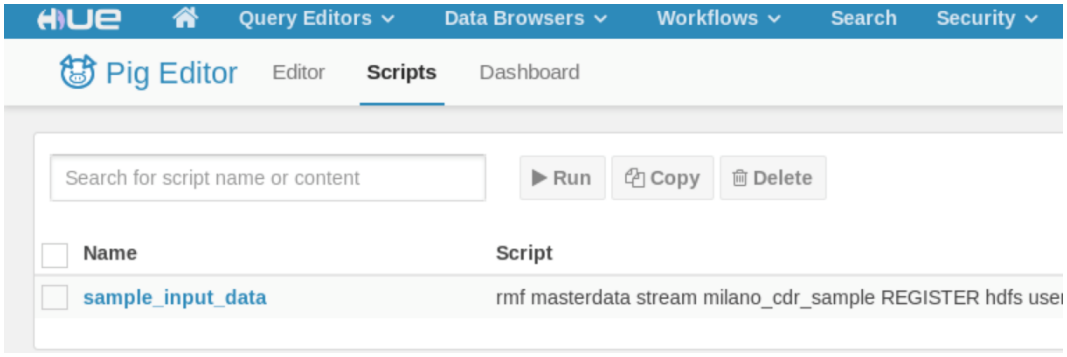
Chapter 5: Customizing Hunk

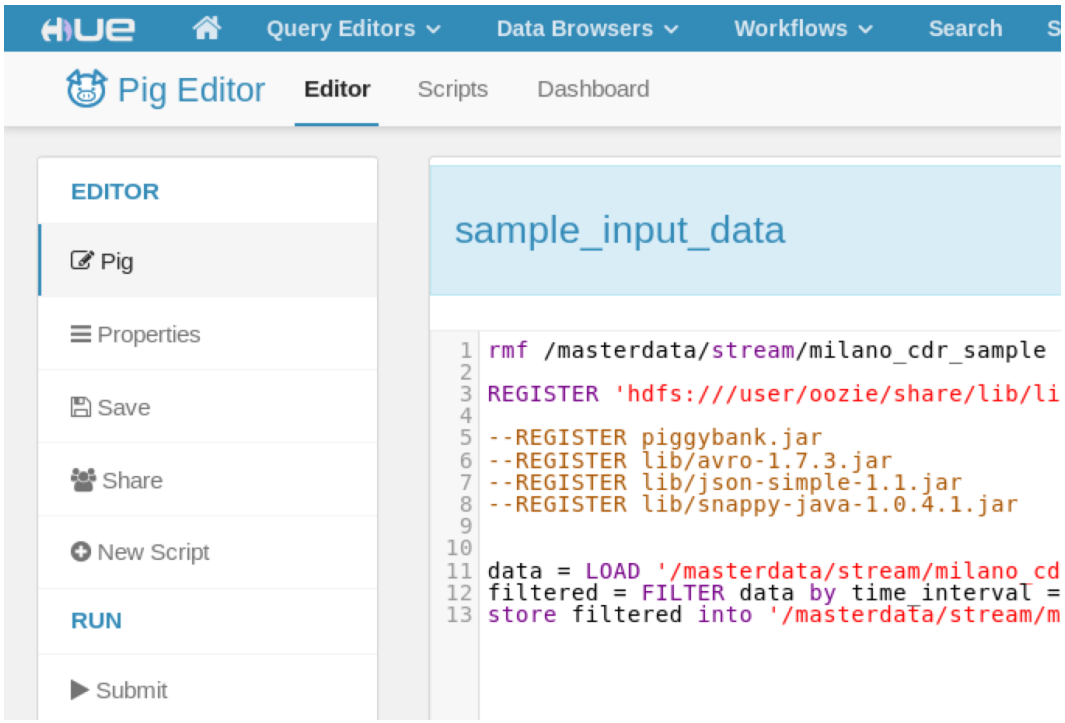


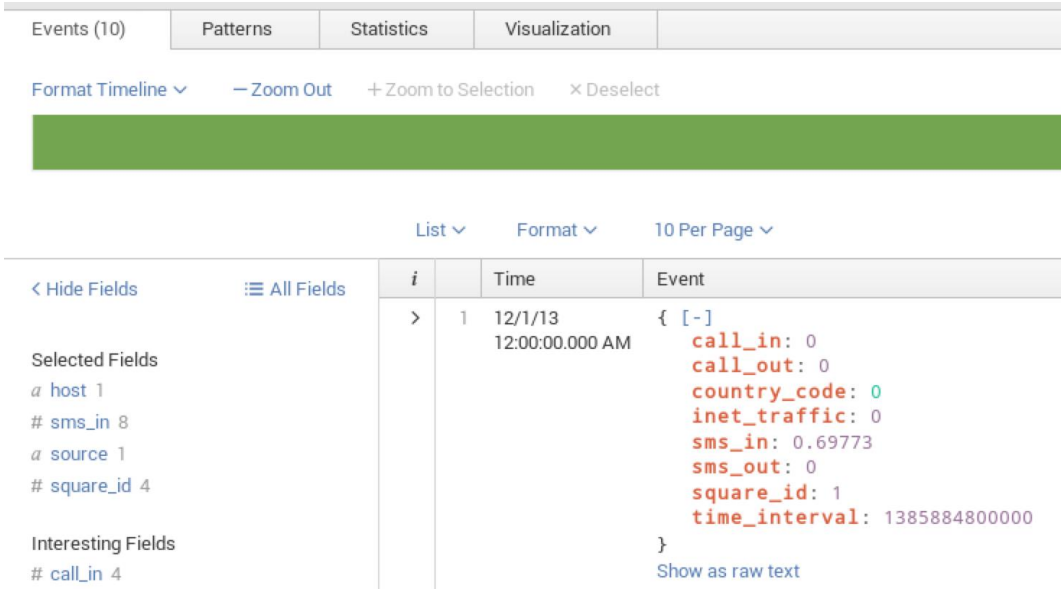

splunk>
Apps
Administrator
Messages
Settings
Activity
Help
Find
Explore Data
Select Virtual Index
Select a File
Preview Data
Enter Context Settings
Review
Confirmation
Set Sourcetype
Data preview lets you see how Hunk sees your data when searching. If the events look correct and have the right timestamps, click "Next" to proceed. If not, u
below to define proper event breaks and timestamps. If you cannot find an appropriate source type for your data, create a new one by clicking "Save As".
Source: part-m-00000.avro
View Event Summary
Sourcetype: System Defaults
Save As
Event Breaks
Timestamp
Advanced
List
Format
20 Per Page
Time
Event
1
12/1/13
12:00:00.000 AM
{ [-]
call_in: 0
call_out: 0
country_code: 0
inet_traffic: 0
sms_in: 0.69773
sms_out: 0
square_id: 1
time_interval: 1385884800000
}
Show as raw text
2
12/1/13
12:00:00.000 AM
{ [-]
call_in: 0.21679
call_out: 0.40449
country_code: 39
inet_traffic: 10.07219

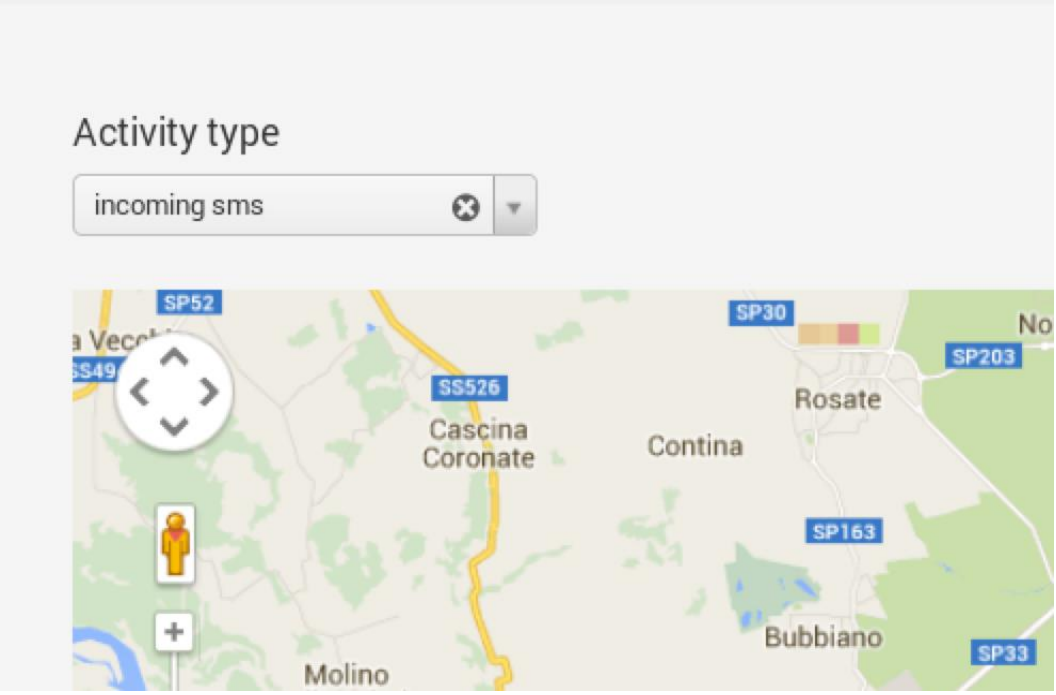

Set Sourcetype
Data preview lets you see how Hunk sees your data when searching. If the events look correct and have the right timestamps, click "Next" to proceed
below to define proper event breaks and timestamps. If you cannot find an appropriate source type for your data, create a new one by clicking "Save A
Source: geocsv(5).csv
Sourcetype: scv_with_comma_and_title
Save As
Event Breaks
Timestamp
Extraction
Auto
Current time
Advanced...
Advanced
Name
Value
DATETIME_CONFIG
CURRENT
SHOULD_LINEMERGE
false
NO_BINARY_CHECK
true
List
Format
20 Per Page
Time
Event
1
8/4/15
10:05:28.000 AM
{ [-]
lat1: 45.35880131440966
lat2: 45.35880097314403
lat3: 45.35668565341486
lat4: 45.356685994655464
lon1: 9.0114910478323
lon2: 9.014491488013135
lon3: 9.0144909480813
lon4: 9.011490619692509
square: 1
}
Show as raw text
timestamp = none
2
8/4/15
10:05:28.000 AM
{ [-]
lat1: 45.35880097314403
lat2: 45.358800553060284

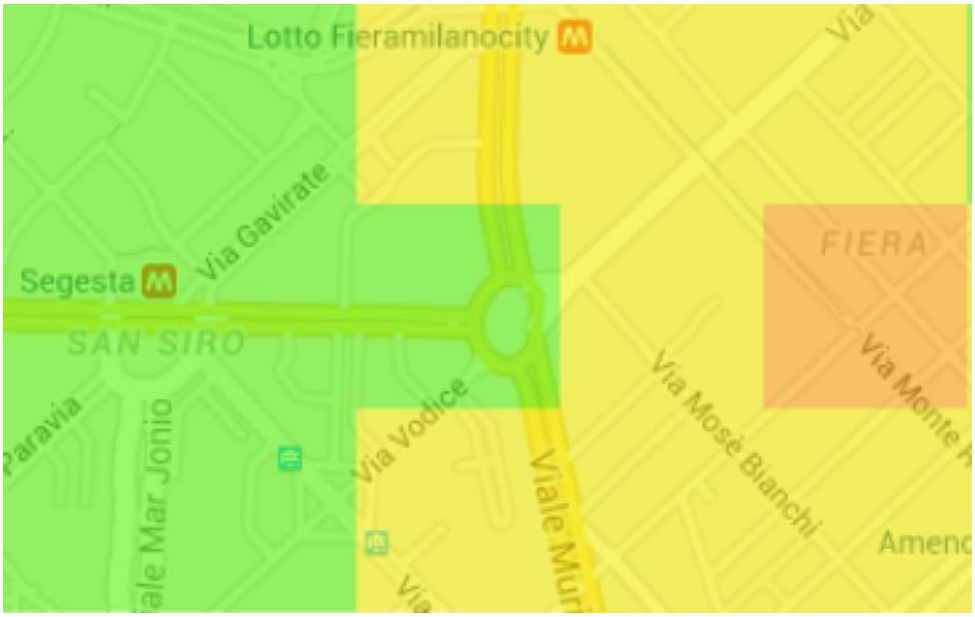

Events (10)
Patterns
Statistics
Visualization
Format Timeline
– Zoom Out
+ Zoom to Selection
× Deselect
List
Format
10 Per Page
Hide Fields
All Fields
Selected Fields
lat1 10
lon1 10
Interesting Fields
i
Time
Event
1
8/9/15
2:01:04.000 PM
{ [-]
lat1: 45.35880131440966
lon1: 9.0114910478323
square: 1
}
Show as raw text
lat1 = 45.35880131440966
lon1 = 9.0114910478323

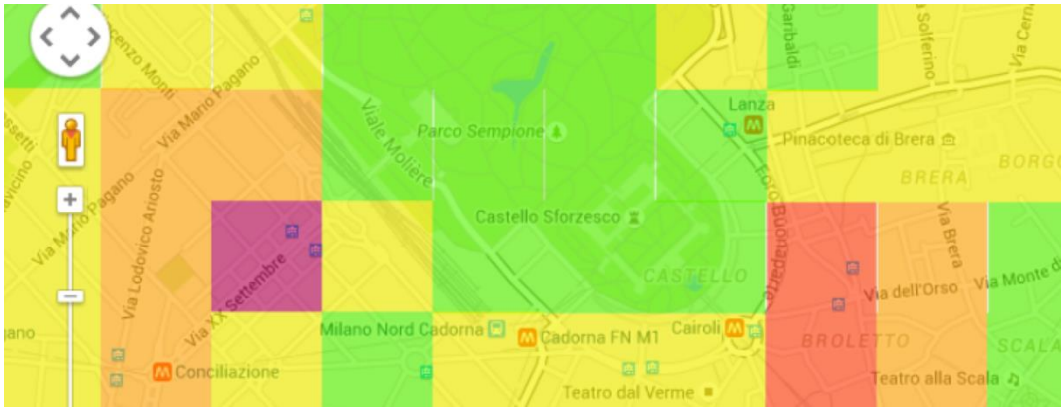

Query Editors
Data Browsers
Workflows
Search
Security
Pig Editor
Editor
Scripts
Dashboard
Search for script name or content
Run
Copy
Delete
Name
Script
sample_input_data
rmf masterdata stream milano_cdr_sample REGISTER hdfs use


Query Editors
Data Browsers
Workflows
Search
Pig Editor
Editor
Scripts
Dashboard
EDITOR
Pig
Properties
Save
Share
New Script
RUN
Submit
sample_input_data
rmf /masterdata/stream/milano_cdr_sample
REGISTER 'hdfs:///user/oozie/share/lib/li
--REGISTER piggybank.jar
--REGISTER lib/avro-1.7.3.jar
--REGISTER lib/json-simple-1.1.jar
--REGISTER lib/snappy-java-1.0.4.1.jar
data = LOAD '/masterdata/stream/milano_cd
filtered = FILTER data by time_interval =
store filtered into '/masterdata/stream/m

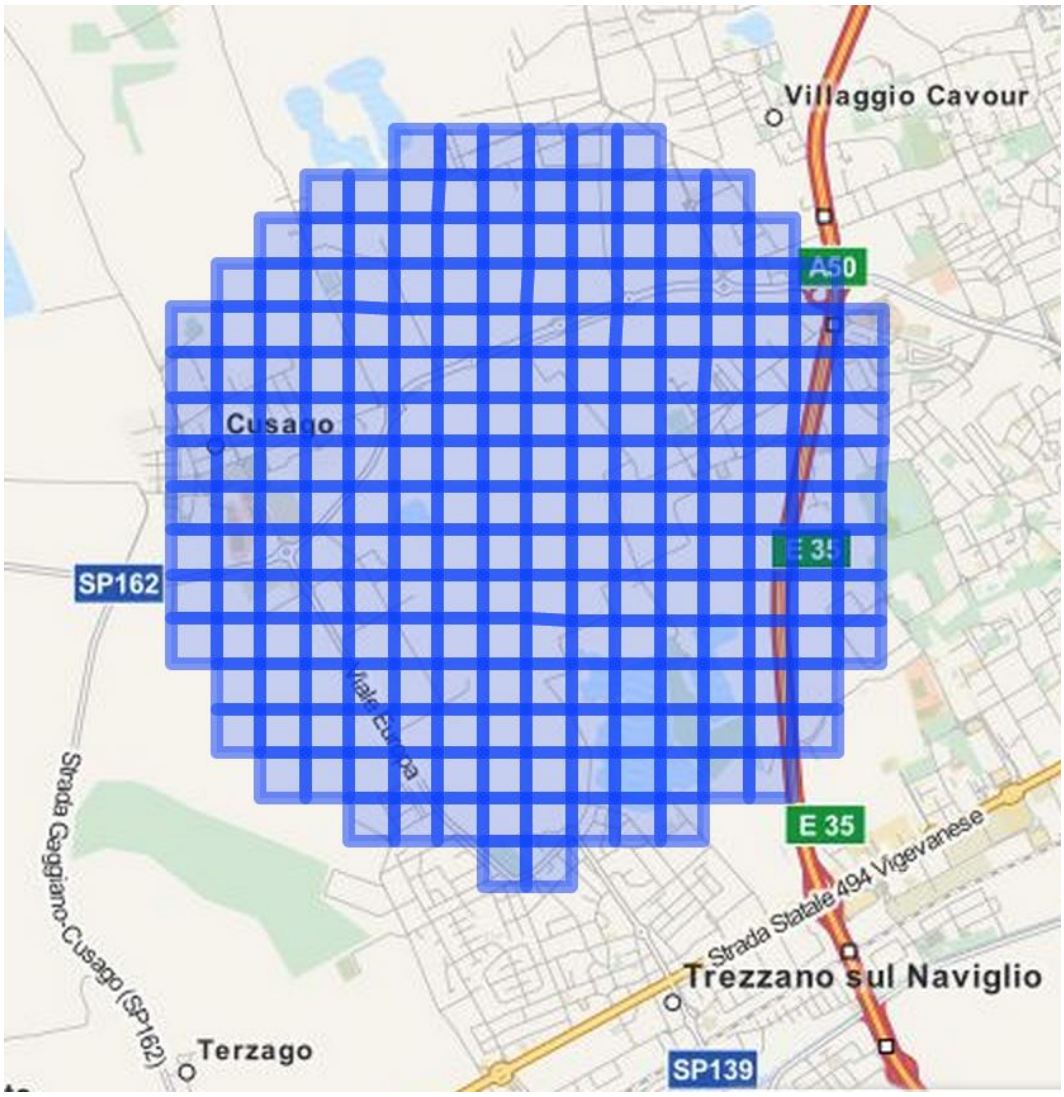

Events (10)
Patterns
Statistics
Visualization
Format Timeline
Zoom Out
Zoom to Selection
Deselect
List
Format
10 Per Page
Hide Fields
All Fields
Selected Fields
a host 1
sms_in 8
a source 1
square_id 4
Interesting Fields
call_in 4
i
Time
Event
1
12/1/13
12:00:00.000 AM
{ [-]
call_in: 0
call_out: 0
country_code: 0
inet_traffic: 0
sms_in: 0.69773
sms_out: 0
square_id: 1
time_interval: 1385884800000
}
Show as raw text

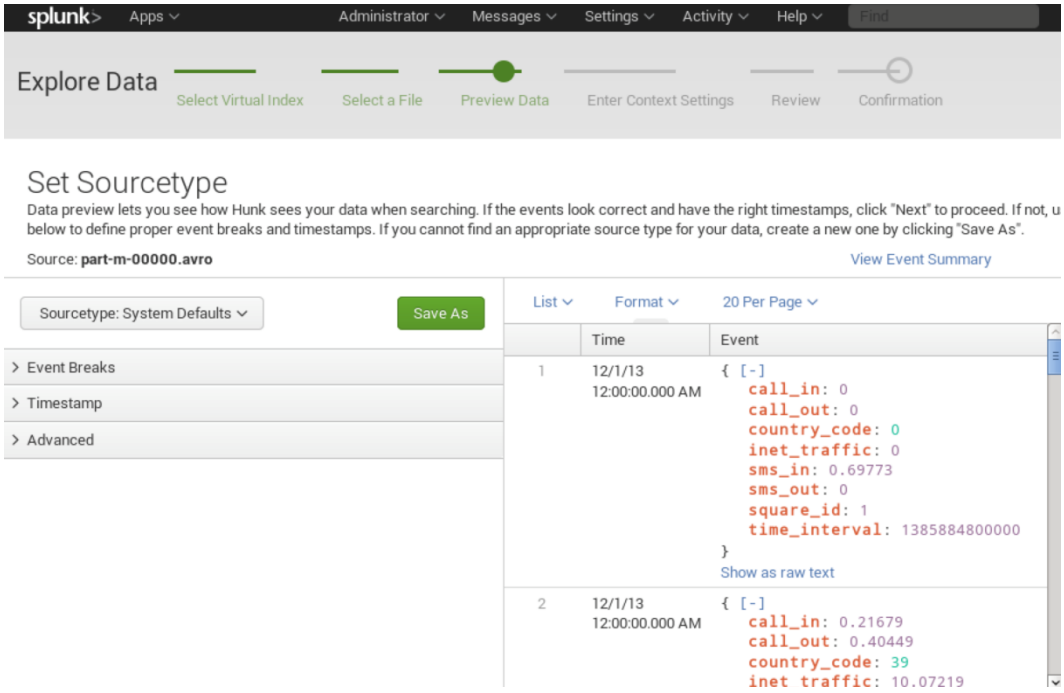

Activity type
incoming sms
SP52
SS49
SS526
Cascina
Coronate
Contina
SP30
SP203
Rosate
SP163
Bubbiano
SP33
Molino

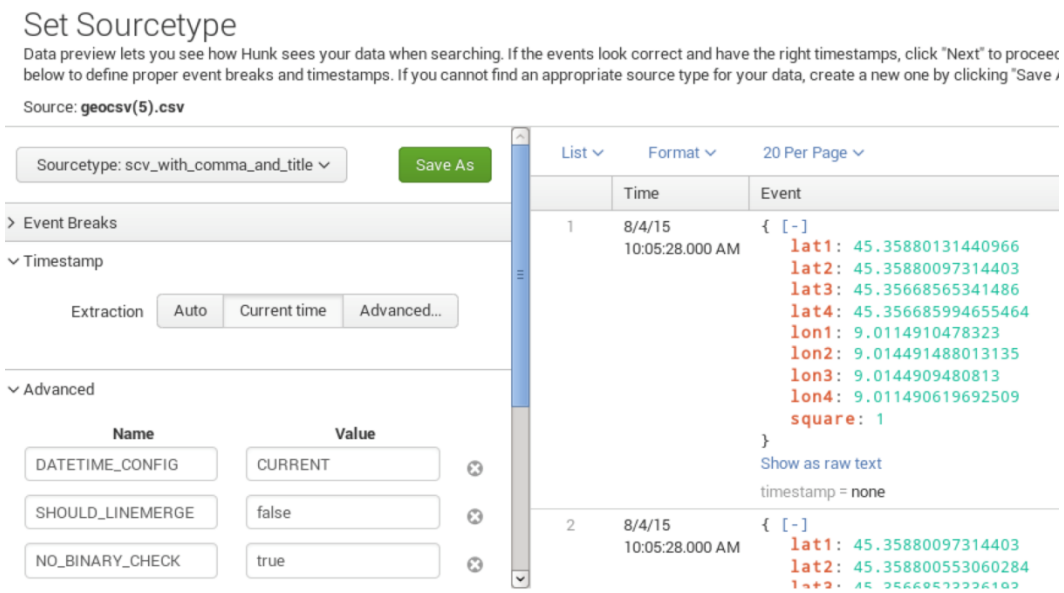

Lotto Fieramilanocity
Segesta
Via Gavirate
SAN SIRO
FIERA
Via Vodice
Viale Mar Jonio
Via Mosè Bianchi

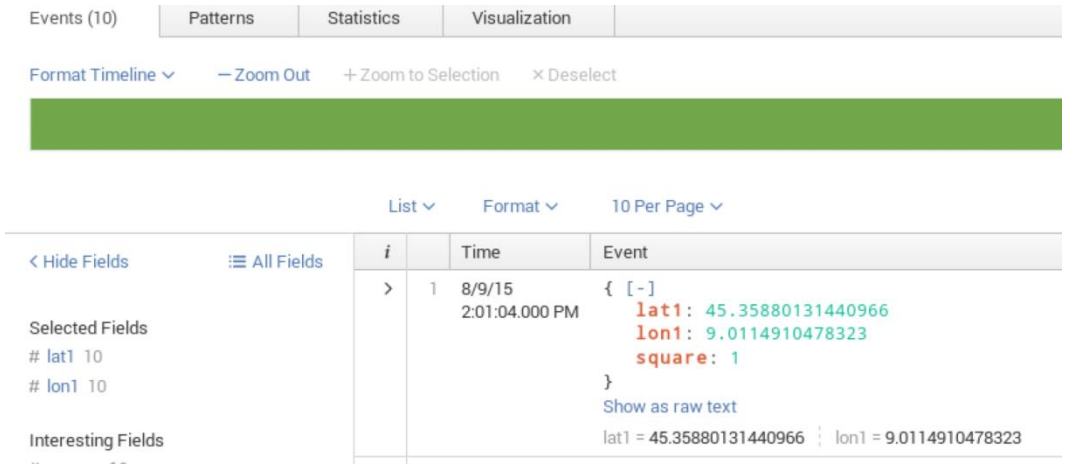

Parco Sempione
Lanza
Pinacoteca di Brera
BRERA
Castello Sforzesco
CASTELLO
Via Mario Pagano
Via Lodovico Ariosto
Via XX Settembre
Milano Nord Cadorna
Cadorna FN M1
Cairoli
Conciliazione
Teatro dal Verme
BROLETTO
Via dell'Orso
Teatro alla Scala

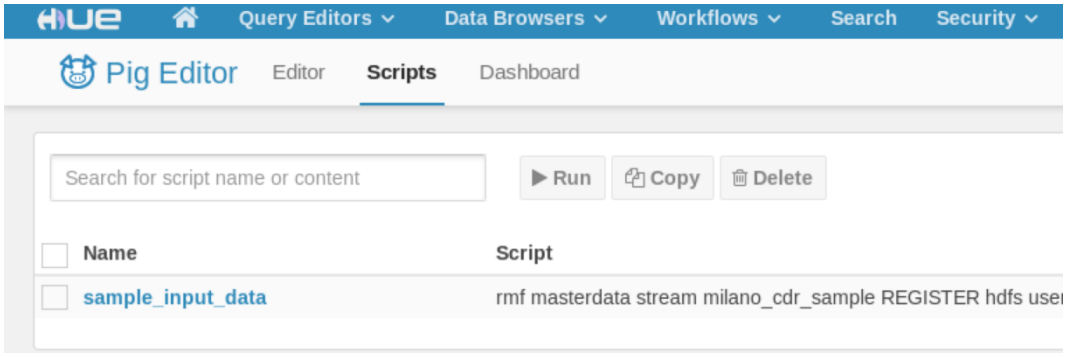
Chapter 5: Customizing Hunk

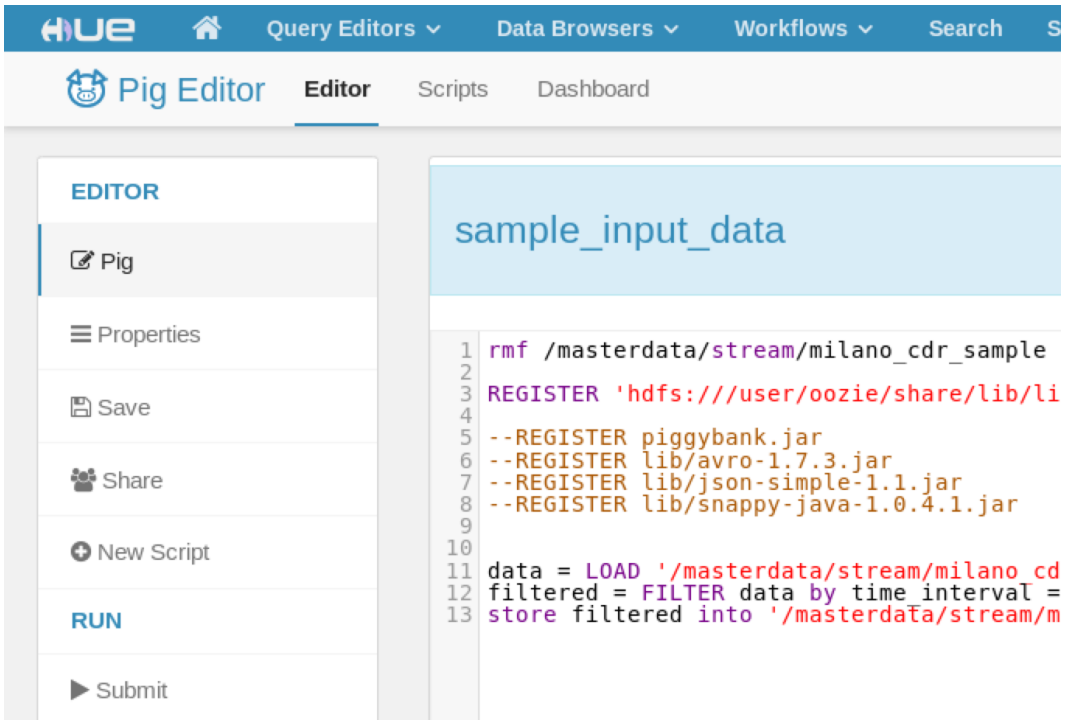


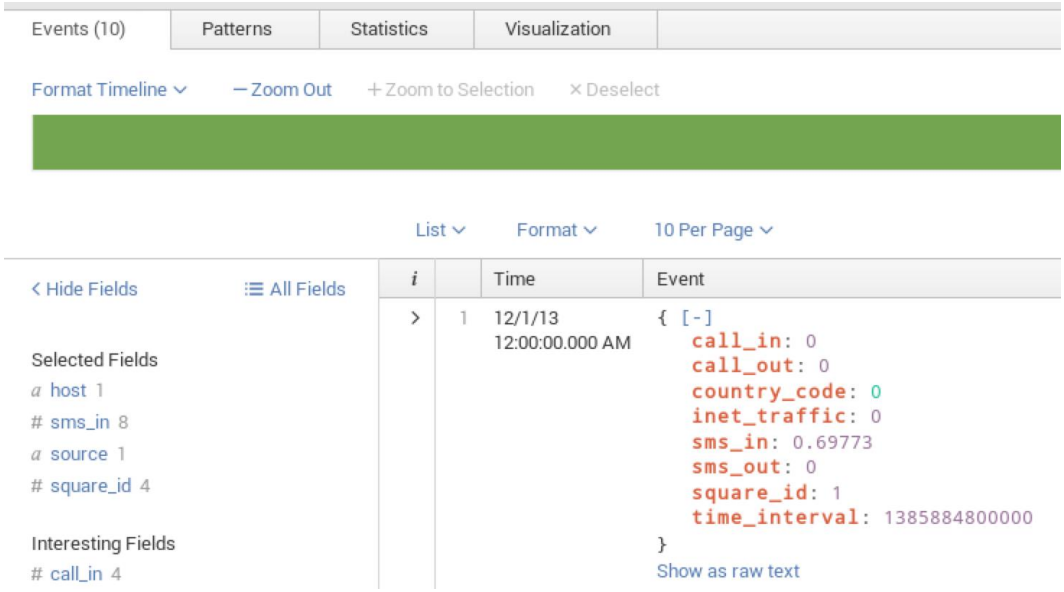

splunk>
Apps
Administrator
Messages
Settings
Activity
Help
Find
Explore Data
Select Virtual Index
Select a File
Preview Data
Enter Context Settings
Review
Confirmation
Set Sourcetype
Data preview lets you see how Hunk sees your data when searching. If the events look correct and have the right timestamps, click "Next" to proceed. If not, u
below to define proper event breaks and timestamps. If you cannot find an appropriate source type for your data, create a new one by clicking "Save As".
Source: part-m-00000.avro
View Event Summary
Sourcetype: System Defaults
Save As
Event Breaks
Timestamp
Advanced
List
Format
20 Per Page
Time
Event
1
12/1/13
12:00:00.000 AM
{ [-]
call_in: 0
call_out: 0
country_code: 0
inet_traffic: 0
sms_in: 0.69773
sms_out: 0
square_id: 1
time_interval: 1385884800000
}
Show as raw text
2
12/1/13
12:00:00.000 AM
{ [-]
call_in: 0.21679
call_out: 0.40449
country_code: 39
inet_traffic: 10.07219

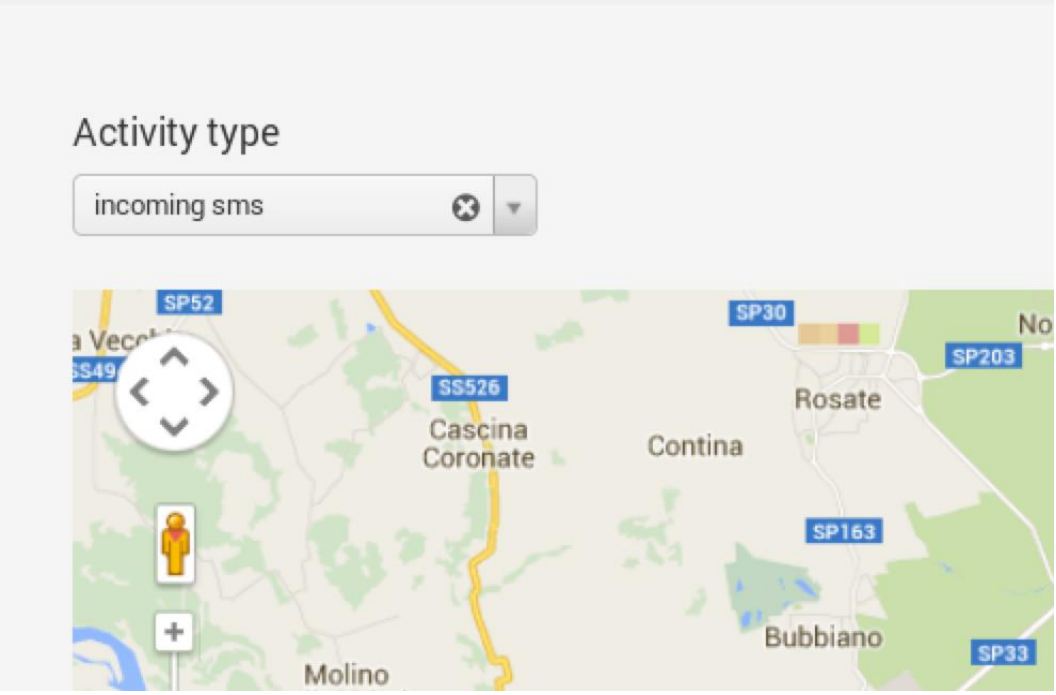

Set Sourcetype
Data preview lets you see how Hunk sees your data when searching. If the events look correct and have the right timestamps, click "Next" to proceed
below to define proper event breaks and timestamps. If you cannot find an appropriate source type for your data, create a new one by clicking "Save
Source: geocsv(5).csv
Sourcetype: scv_with_comma_and_title
Save As
Event Breaks
Timestamp
Extraction
Auto
Current time
Advanced...
Advanced
Name
Value
DATETIME_CONFIG
CURRENT
SHOULD_LINEMERGE
false
NO_BINARY_CHECK
true
List
Format
20 Per Page
Time
Event
1
8/4/15
10:05:28.000 AM
{ [-]
lat1: 45.35880131440966
lat2: 45.35880097314403
lat3: 45.35668565341486
lat4: 45.356685994655464
lon1: 9.0114910478323
lon2: 9.014491488013135
lon3: 9.0144909480813
lon4: 9.011490619692509
square: 1
}
Show as raw text
timestamp = none
2
8/4/15
10:05:28.000 AM
{ [-]
lat1: 45.35880097314403
lat2: 45.358800553060284

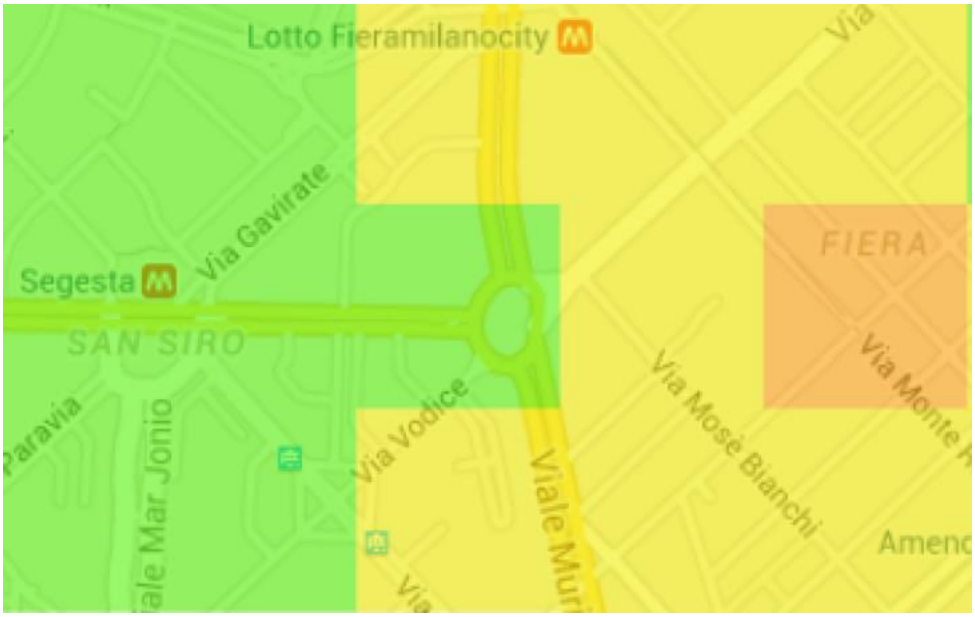

Events (10)
Patterns
Statistics
Visualization
Format Timeline
– Zoom Out
+ Zoom to Selection
× Deselect
List
Format
10 Per Page
Hide Fields
All Fields
Selected Fields
lat1 10
lon1 10
Interesting Fields
i
Time
Event
1
8/9/15
2:01:04.000 PM
{ [-]
lat1: 45.35880131440966
lon1: 9.0114910478323
square: 1
}
Show as raw text
lat1 = 45.35880131440966
lon1 = 9.0114910478323

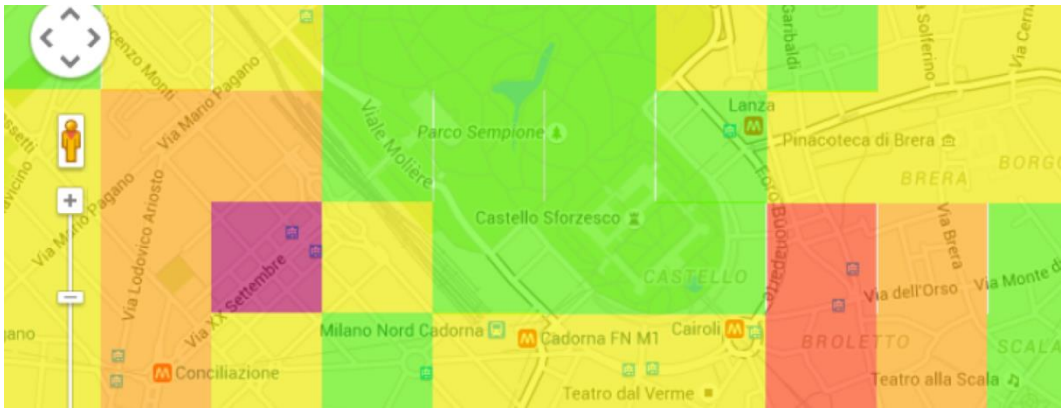

Query Editors
Data Browsers
Workflows
Search
Security
Pig Editor
Editor
Scripts
Dashboard
Search for script name or content
Run
Copy
Delete
Name
Script
sample_input_data
rmf masterdata stream milano_cdr_sample REGISTER hdfs use


Query Editors
Data Browsers
Workflows
Search
Pig Editor
Editor
Scripts
Dashboard
EDITOR
Pig
Properties
Save
Share
New Script
RUN
Submit
sample_input_data
1 rmf /masterdata/stream/milano_cdr_sample
2
3 REGISTER 'hdfs:///user/oozie/share/lib/li
4
5 --REGISTER piggybank.jar
6 --REGISTER lib/avro-1.7.3.jar
7 --REGISTER lib/json-simple-1.1.jar
8 --REGISTER lib/snappy-java-1.0.4.1.jar
9
10
11 data = LOAD '/masterdata/stream/milano_cd
12 filtered = FILTER data by time_interval =
13 store filtered into '/masterdata/stream/m

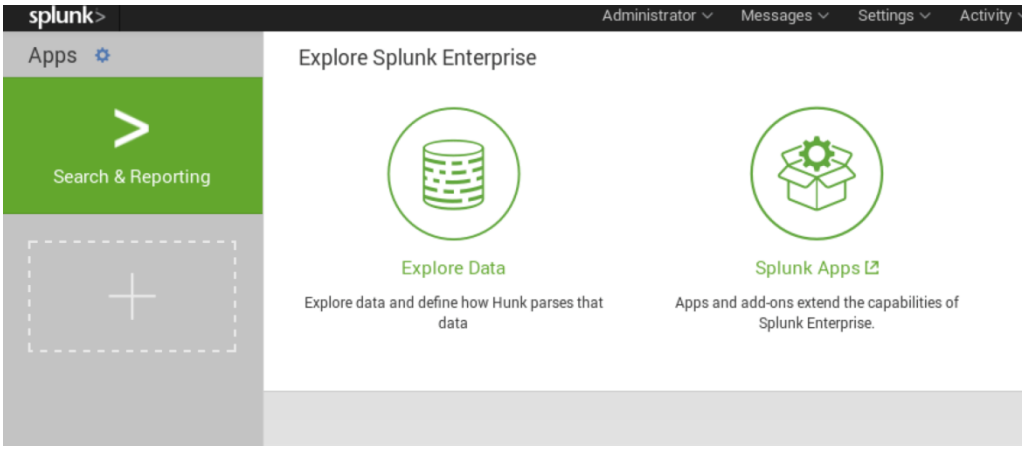

Events (10)
Patterns
Statistics
Visualization
Format Timeline
Zoom Out
Zoom to Selection
Deselect
List
Format
10 Per Page
Hide Fields
All Fields
Selected Fields
a host 1
sms_in 8
a source 1
square_id 4
Interesting Fields
call_in 4
i
Time
Event
1
12/1/13
12:00:00.000 AM
{ [-]
call_in: 0
call_out: 0
country_code: 0
inet_traffic: 0
sms_in: 0.69773
sms_out: 0
square_id: 1
time_interval: 1385884800000
}
Show as raw text

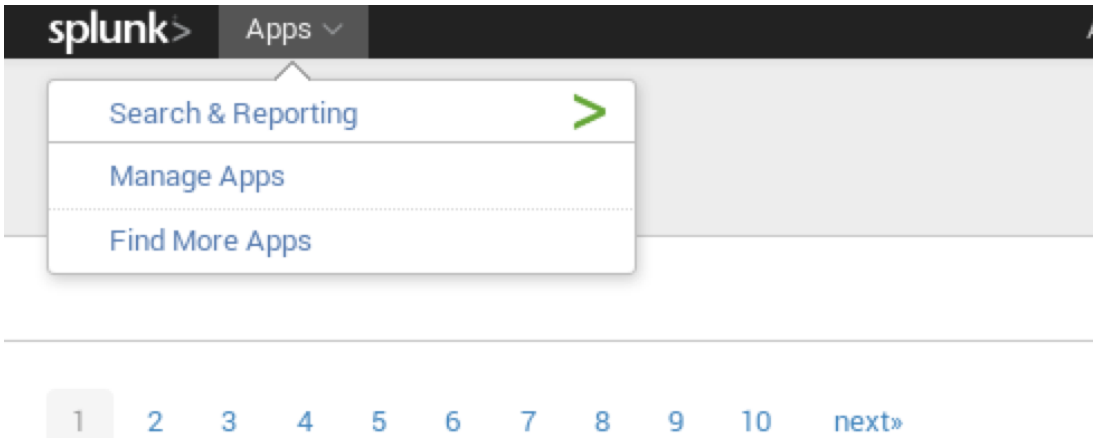

Activity type
incoming sms
SP52
SS49
SS526
Cascina
Coronate
Contina
SP30
SP203
Rosate
SP163
Bubbiano
SP33
Molino

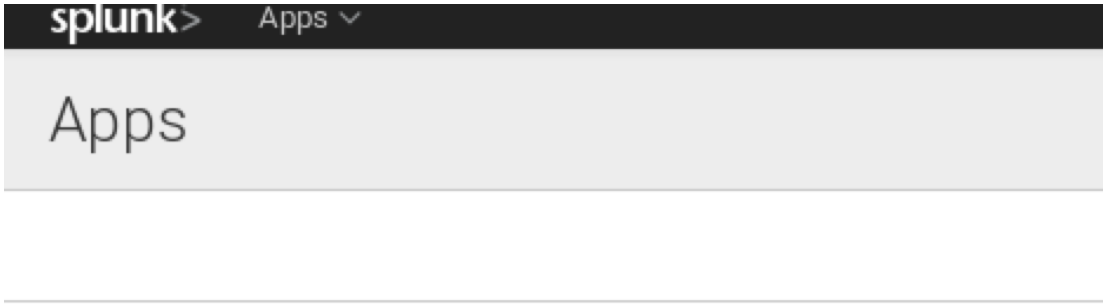

Lotto Fieramilanocity
Segesta
Via Gavirate
SAN SIRO
FIERA
Via Vodice
Via Mosè Bianchi
Viale Mar Jonio

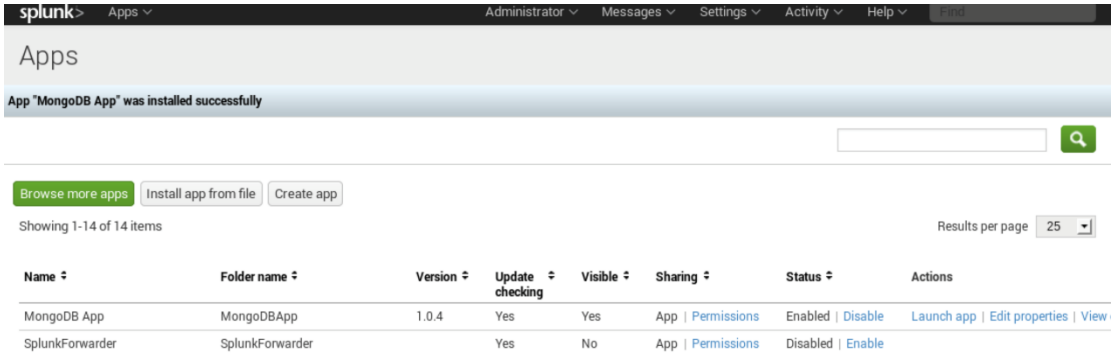

Parco Sempione
Lanza
Pinacoteca di Brera
BRERA
Castello Sforzesco
CASTELLO
Via Mario Pagano
Via Lodovico Ariosto
Via XX Settembre
Milano Nord Cadorna
Cadorna FN M1
Cairoli
Conciliazione
Teatro dal Verme
BROLETTO
Via dell'Orso
Teatro alla Scala

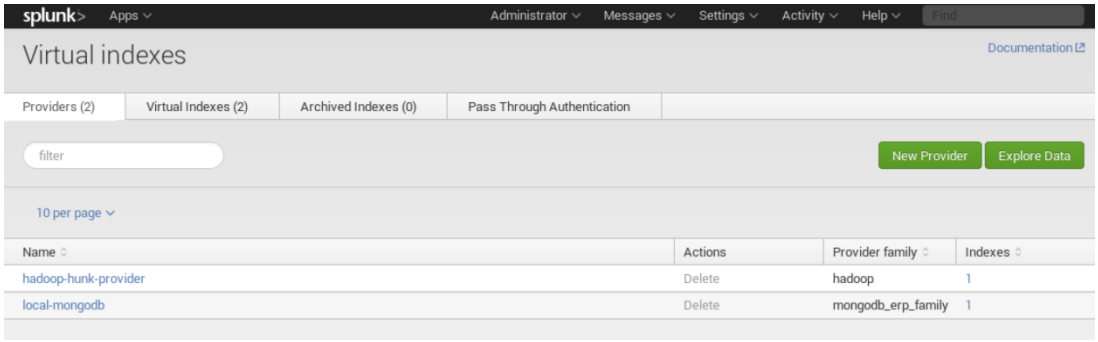
Chapter 6: Discovering Hunk Integration Apps

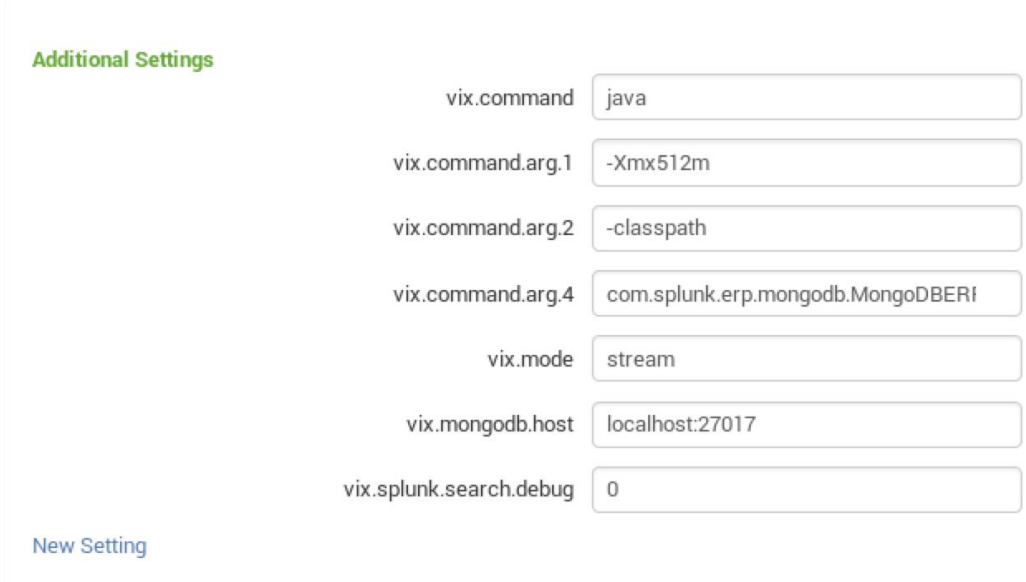


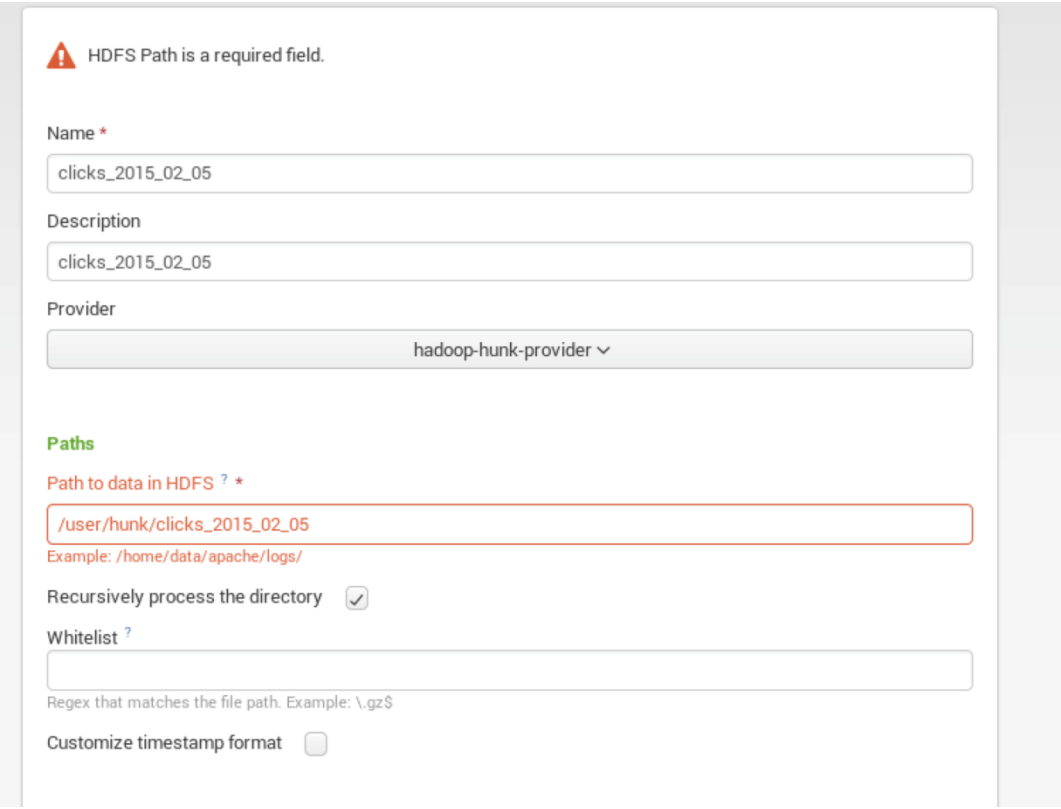


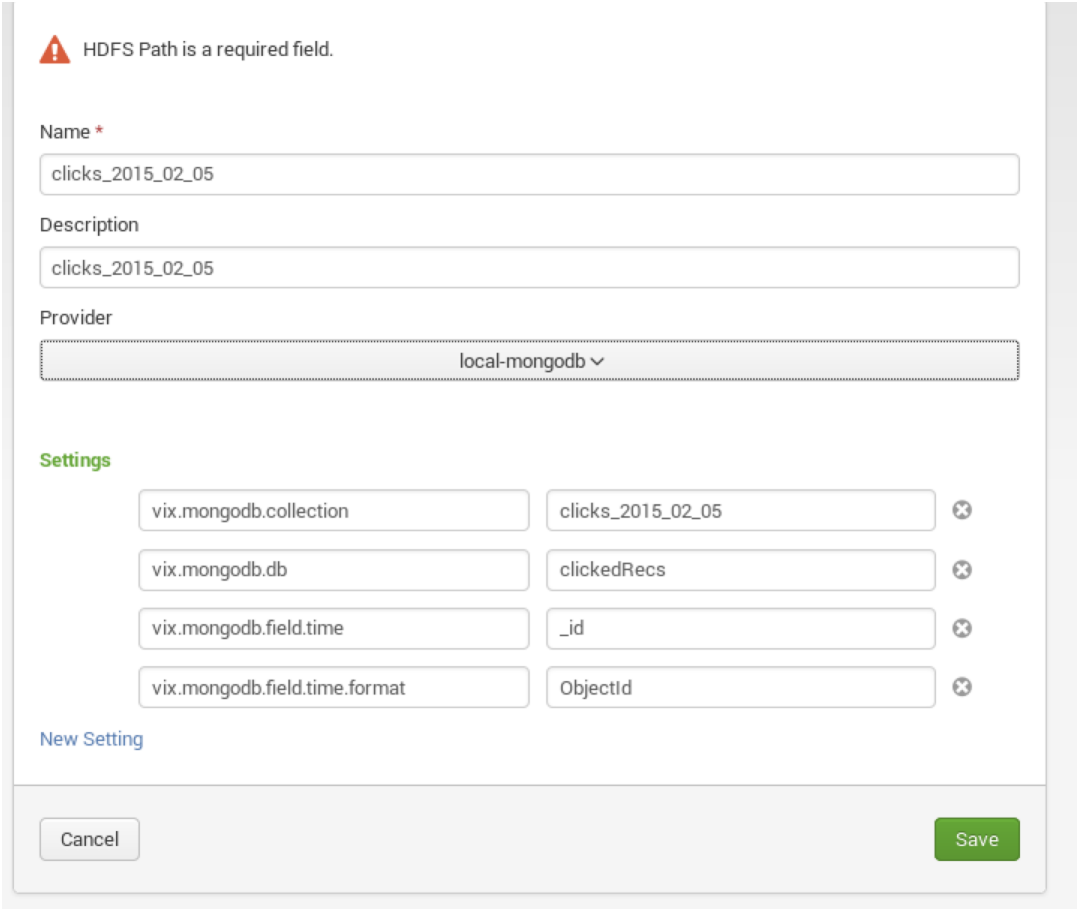

splunk> Apps
Apps
Browse more apps
Install app from file
Create app
Showing 1-13 of 13 items


splunk> Apps
Administrator
Messages
Settings
Activity
Help
Apps
App "MongoDB App" was installed successfully
Browse more apps
Install app from file
Create app
Showing 1-14 of 14 items
Results per page 25
Name
Folder name
Version
Update checking
Visible
Sharing
Status
Actions
MongoDB App
MongoDBApp
1.0.4
Yes
Yes
App | Permissions
Enabled | Disable
Launch app | Edit properties | View
SplunkForwarder
SplunkForwarder
Yes
No
App | Permissions
Disabled | Enable


splunk> Apps
Administrator
Messages
Settings
Activity
Help
Find
Virtual indexes
Documentation
Providers (2)
Virtual Indexes (2)
Archived Indexes (0)
Pass Through Authentication
filter
New Provider
Explore Data
10 per page
Name
Actions
Provider family
Indexes
hadoop-hunk-provider
Delete
hadoop
1
local-mongodb
Delete
mongodb_erp_family
1


Additional Settings
vix.command
java
vix.command.arg.1
-Xmx512m
vix.command.arg.2
-classpath
vix.command.arg.4
com.splunk.erp.mongodb.MongoDBERI
vix.mode
stream
vix.mongodb.host
localhost:27017
vix.splunk.search.debug
0
New Setting


HDFS Path is a required field.
Name *
clicks_2015_02_05
Description
clicks_2015_02_05
Provider
hadoop-hunk-provider
Paths
Path to data in HDFS * *
/user/hunk/clicks_2015_02_05
Example: /home/data/apache/logs/
Recursively process the directory
Whitelist
Regex that matches the file path. Example: \.gz$
Customize timestamp format


HDFS Path is a required field.
Name *
clicks_2015_02_05
Description
clicks_2015_02_05
Provider
local-mongodb
Settings
vix.mongodb.collection
clicks_2015_02_05
vix.mongodb.db
clickedRecs
vix.mongodb.field.time
_id
vix.mongodb.field.time.format
ObjectId
New Setting
Cancel
Save

10 per page

Name	Status	Actions	Provider
clicks_2015_02_01	Enabled Disable	Search Delete	hadoop-hunk-provider
clicks_2015_02_02	Enabled Disable	Search Delete	hadoop-hunk-provider
clicks_2015_02_03	Enabled Disable	Search Delete	hadoop-hunk-provider
clicks_2015_02_04	Enabled Disable	Search Delete	hadoop-hunk-provider
clicks_2015_02_05	Enabled Disable	Search Delete	local-mongodb
digital_analytics	Enabled Disable	Search Delete	hadoop-hunk-provider
mongodb_vix	Enabled Disable	Search Delete	local-mongodb

New Search

Save As Close

```
index="clicks_2015_02_01" | stats count by shop_id
```

All time

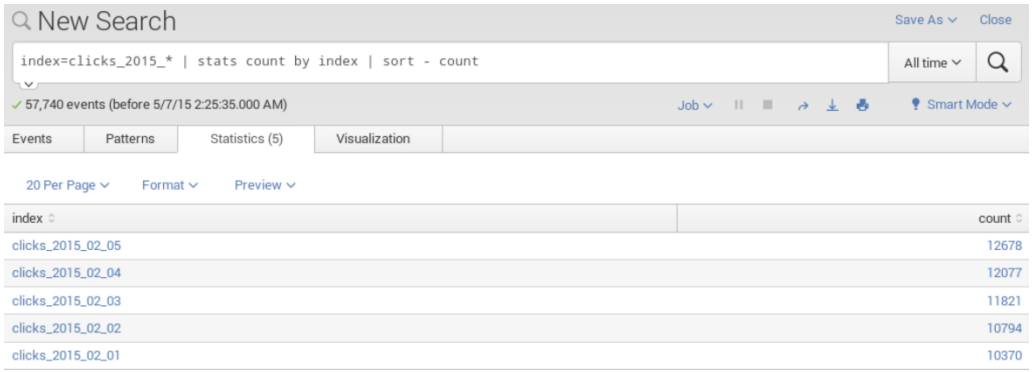
✓ 10,370 events (before 5/6/15 2:52:45.000 PM)

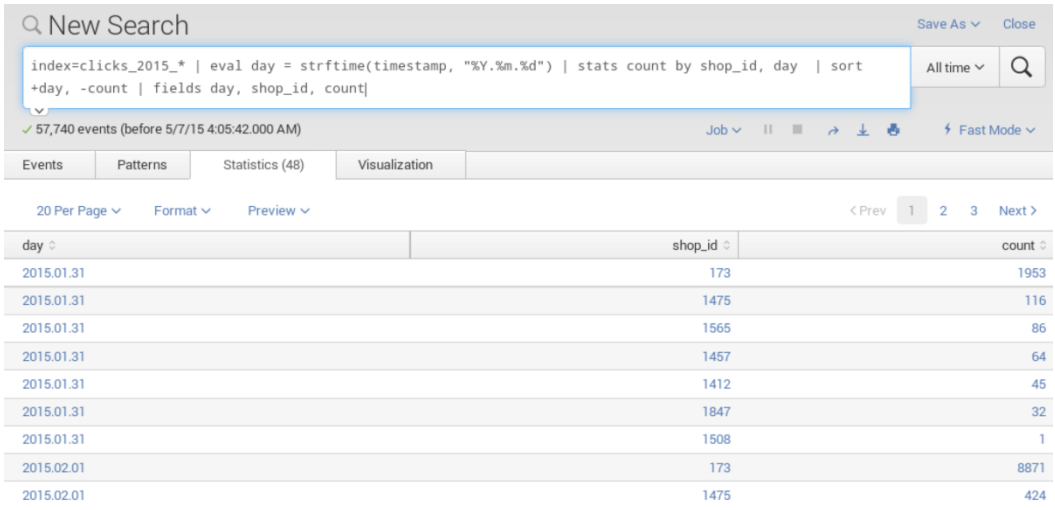
Job Smart Mode

Events Patterns Statistics (7) Visualization

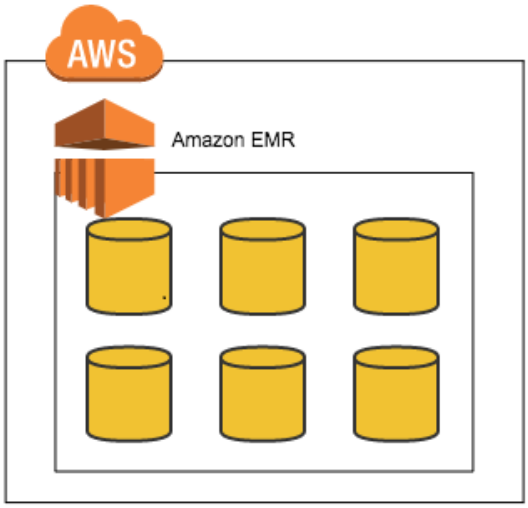
20 Per Page Format Preview

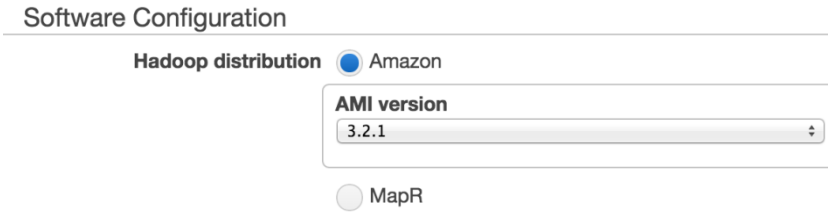
shop_id	count
1412	240
1457	300
1475	430
1508	40
1565	406
173	8763
1847	191


New Search
Save As
Close
index=clicks_2015_* | stats count by index | sort - count
All time
57,740 events (before 5/7/15 2:25:35.000 AM)
Job
Smart Mode
Events
Patterns
Statistics (5)
Visualization
20 Per Page
Format
Preview
index
count
clicks_2015_02_05 12678
clicks_2015_02_04 12077
clicks_2015_02_03 11821
clicks_2015_02_02 10794
clicks_2015_02_01 10370


New Search
Save As
Close
index=clicks_2015_* | eval day = strftime(timestamp, "%Y.%m.%d") | stats count by shop_id, day | sort +day, -count | fields day, shop_id, count
All time
57,740 events (before 5/7/15 4:05:42.000 AM)
Job
Fast Mode
Events
Patterns
Statistics (48)
Visualization
20 Per Page
Format
Preview
Prev
1
2
3
Next
day
shop_id
count
2015.01.31 173 1953
2015.01.31 1475 116
2015.01.31 1565 86
2015.01.31 1457 64
2015.01.31 1412 45
2015.01.31 1847 32
2015.01.31 1508 1
2015.02.01 173 8871
2015.02.01 1475 424

Chapter 7: Exploring Data in the Cloud





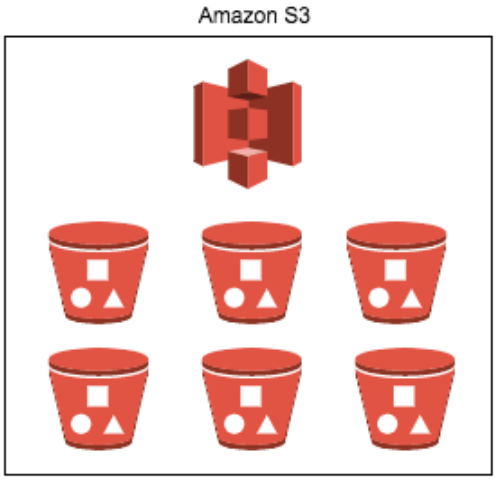
Type	Name	EC2 instance type	Count
Master	Master instance group – 1	m1.medium	1
Core	Core instance group – 2	m1.medium	2
Task	Task instance group – 3	m1.medium	0

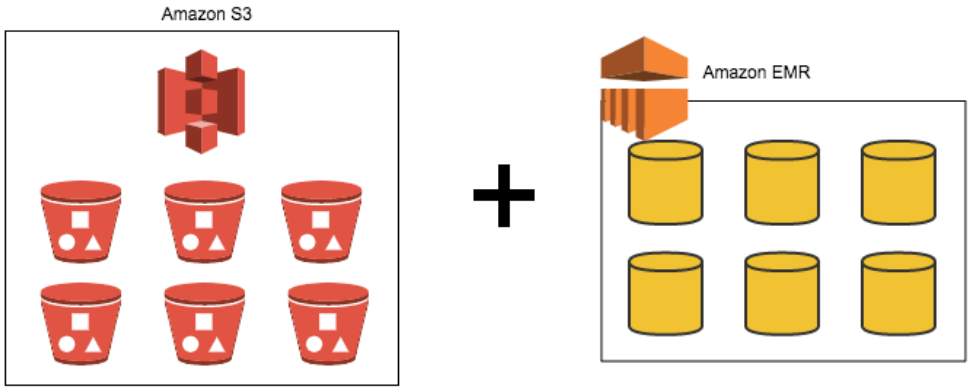
Applications to be installed	Version			
Hive	0.13.1			
Pig	0.12.0			
Hunk				

Hue is unsupported on selected AMI.

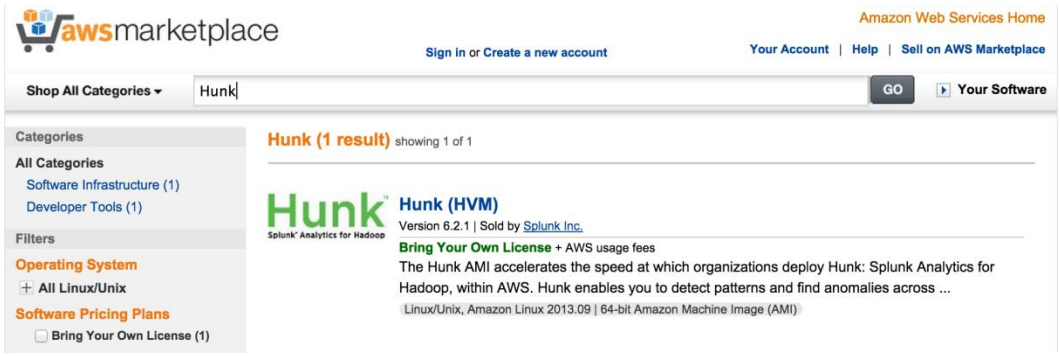
Additional applications: Select an application

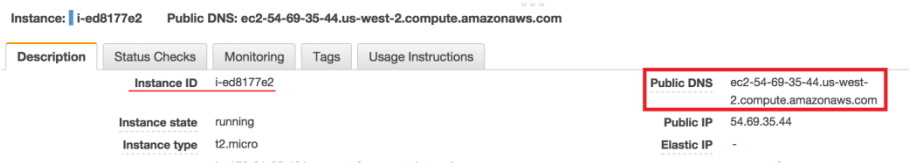
Configure and add

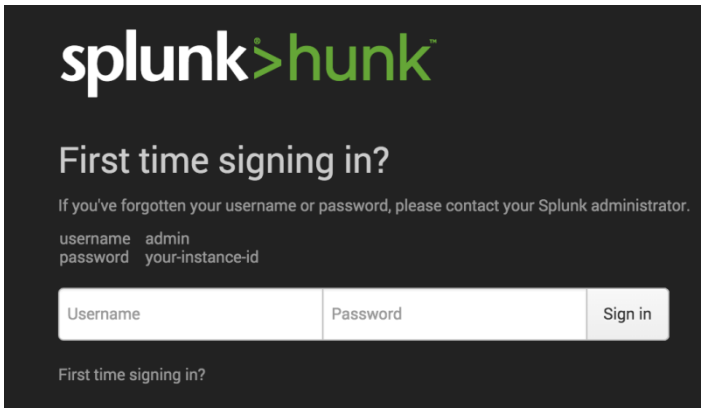

Amazon S3

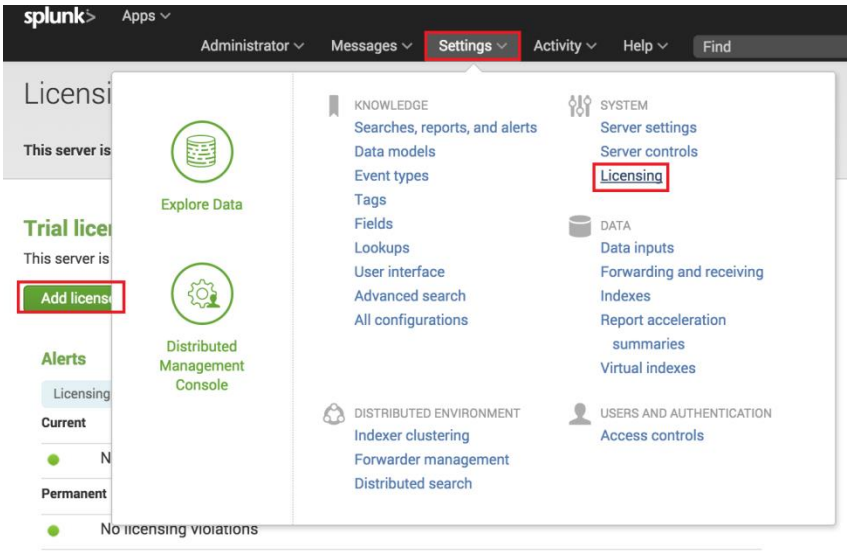

Amazon S3
Amazon EMR

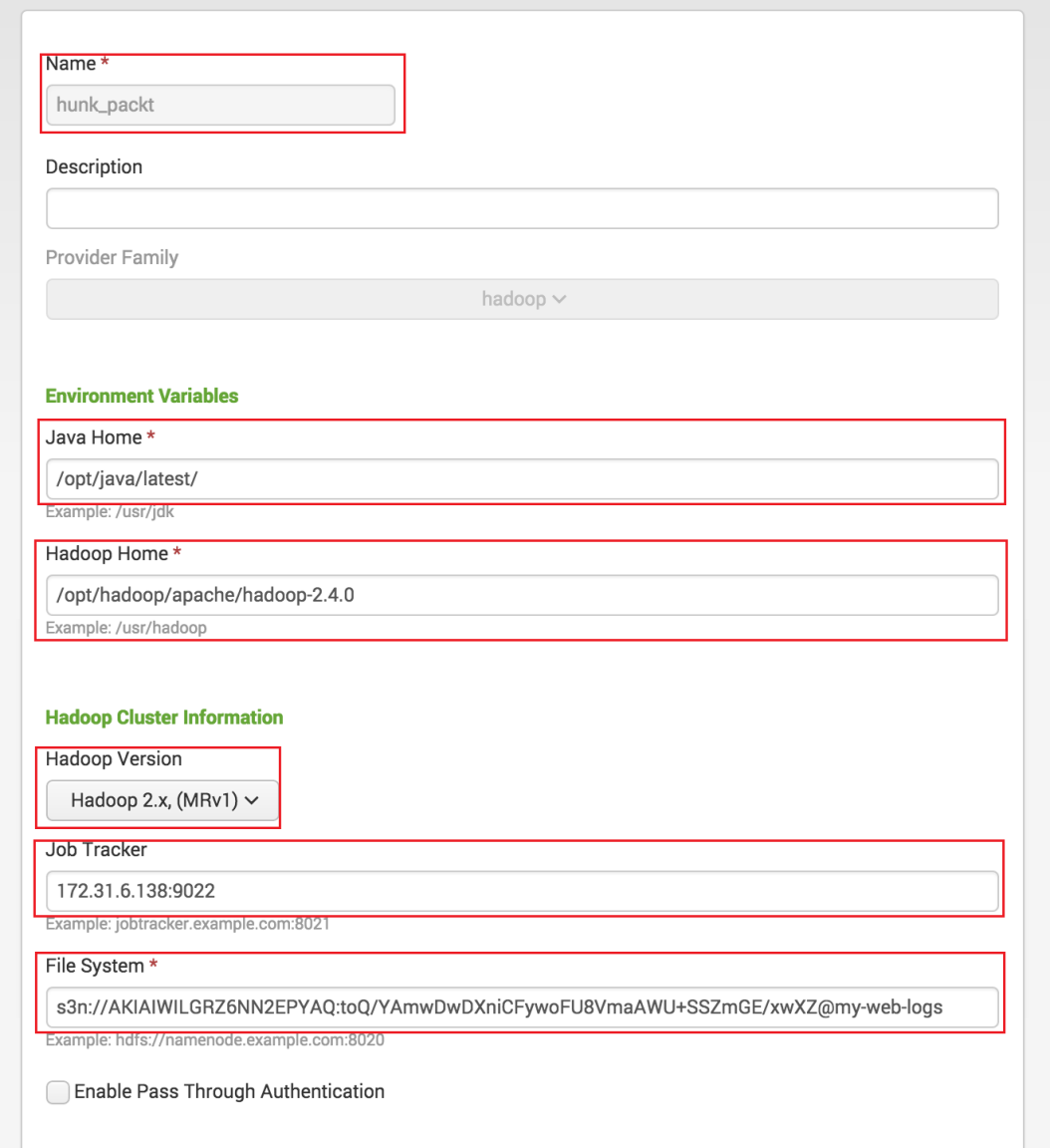

Explore
Visualize
Ad-hoc
Dashboards
Share
Hunk
Amazon EMR
Amazon S3 or HDFS

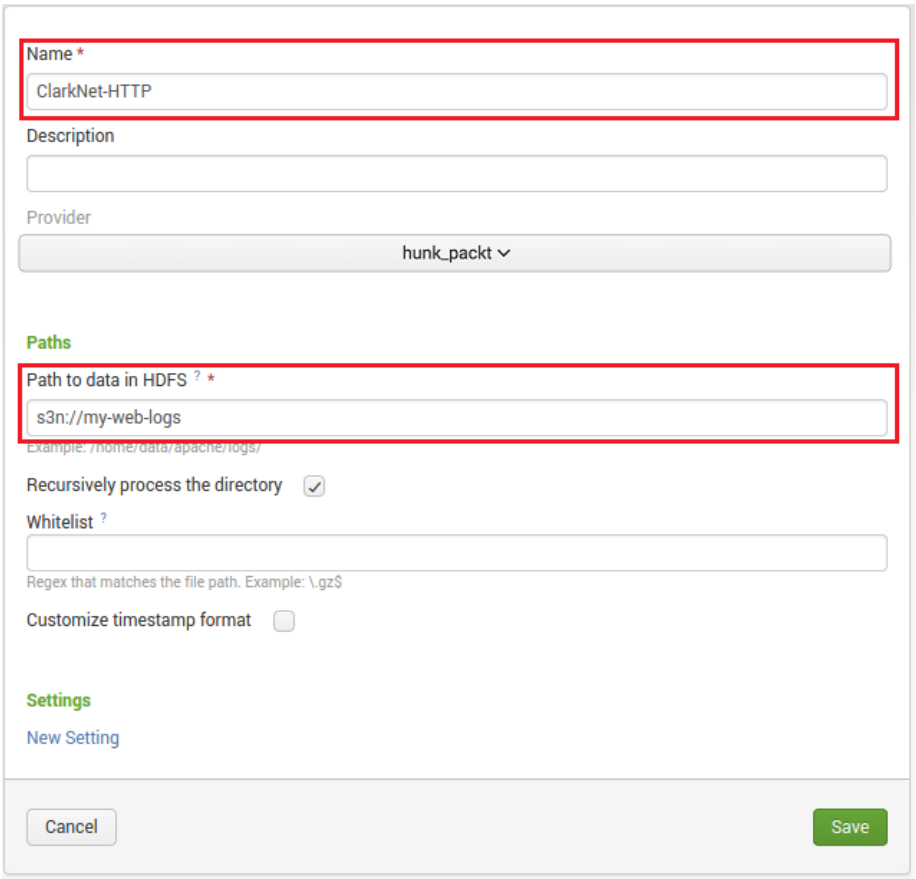

awsmarketplace
Amazon Web Services Home
Sign in or Create a new account
Your Account | Help | Sell on AWS Marketplace
Shop All Categories
Hunk
GO
Your Software
Categories
All Categories
Software Infrastructure (1)
Developer Tools (1)
Filters
Operating System
All Linux/Unix
Software Pricing Plans
Bring Your Own License (1)
Hunk (1 result) showing 1 of 1
Hunk
Splunk® Analytics for Hadoop
Hunk (HVM)
Version 6.2.1 | Sold by Splunk Inc.
Bring Your Own License + AWS usage fees
The Hunk AMI accelerates the speed at which organizations deploy Hunk: Splunk Analytics for Hadoop, within AWS. Hunk enables you to detect patterns and find anomalies across ...
Linux/Unix, Amazon Linux 2013.09 | 64-bit Amazon Machine Image (AMI)

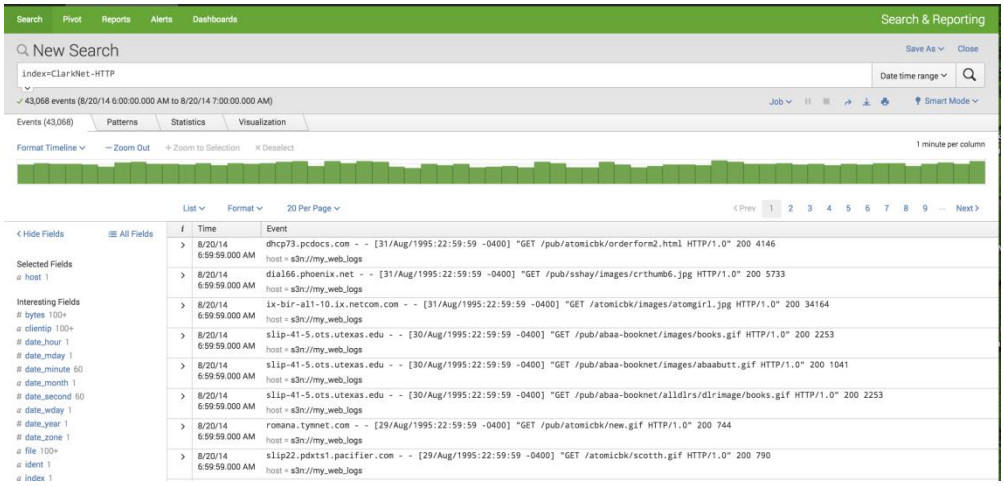

Instance: i-ed8177e2
Public DNS: ec2-54-69-35-44.us-west-2.compute.amazonaws.com
Description
Status Checks
Monitoring
Tags
Usage Instructions
Instance ID i-ed8177e2
Instance state running
Instance type t2.micro
Public DNS ec2-54-69-35-44.us-west-2.compute.amazonaws.com
Public IP 54.69.35.44
Elastic IP -


splunk>hunk
First time signing in?
If you've forgotten your username or password, please contact your Splunk administrator.
username admin
password your-instance-id
Username
Password
Sign in
First time signing in?


splunk>
Apps
Administrator
Messages
Settings
Activity
Help
Find
Licensi
This server is
Trial licer
This server is
Add licens
Alerts
Licensing
Current
Permanent
No licensing violations
Explore Data
Distributed Management Console
KNOWLEDGE
Searches, reports, and alerts
Data models
Event types
Tags
Fields
Lookups
User interface
Advanced search
All configurations
SYSTEM
Server settings
Server controls
Licensing
DATA
Data inputs
Forwarding and receiving
Indexes
Report acceleration summaries
Virtual indexes
DISTRIBUTED ENVIRONMENT
Indexer clustering
Forwarder management
Distributed search
USERS AND AUTHENTICATION
Access controls


Name *
hunk_packt
Description
Provider Family
hadoop
Environment Variables
Java Home *
/opt/java/latest/
Example: /usr/jdk
Hadoop Home *
/opt/hadoop/apache/hadoop-2.4.0
Example: /usr/hadoop
Hadoop Cluster Information
Hadoop Version
Hadoop 2.x, (MRv1)
Job Tracker
172.31.6.138:9022
Example: jobtracker.example.com:8021
File System *
s3n://AKIAIWILGRZ6NN2EPYAQ:toQ/YAmwDwDXniCFywoFU8VmaAWU+SSZmGE/xwXZ@my-web-logs
Example: hdfs://namenode.example.com:8020
Enable Pass Through Authentication


Name *
ClarkNet-HTTP
Description
Provider
hunk_packt
Paths
Path to data in HDFS *
s3n://my-web-logs
Example: /home/data/apache/logs/
Recursively process the directory
Whitelist
Regex that matches the file path. Example: \.gz$
Customize timestamp format
Settings
New Setting
Cancel
Save


Search
Pivot
Reports
Alerts
Dashboards
Search & Reporting
New Search
Save As
Close
index=ClarkNet-HTTP
Date time range
43,068 events (8/20/14 6:00:00.000 AM to 8/20/14 7:00:00.000 AM)
Job
Smart Mode
Events (43,068)
Patterns
Statistics
Visualization
Format Timeline
Zoom Out
Zoom to Selection
Deselect
1 minute per column
List
Format
20 Per Page
Prev
1
2
3
4
5
6
7
8
9
Next
Hide Fields
All Fields
Selected Fields
host 1
Interesting Fields
bytes 100+
clientip 100+
date_hour 1
date_mday 1
date_minute 60
date_month 1
date_second 60
date_wday 1
date_year 1
date_zone 1
file 100+
ident 1
index 1
Time
Event
8/20/14 6:59:59.000 AM
dhcp73.pcdocs.com - - [31/Aug/1995:22:59:59 -0400] "GET /pub/atomicbk/orderform2.html HTTP/1.0" 200 4146
host = s3n://my_web_logs
8/20/14 6:59:59.000 AM
dial66.phoenix.net - - [31/Aug/1995:22:59:59 -0400] "GET /pub/sshay/images/crthumb6.jpg HTTP/1.0" 200 5733
host = s3n://my_web_logs
8/20/14 6:59:59.000 AM
ix-bir-al1-10.ix.netcom.com - - [31/Aug/1995:22:59:59 -0400] "GET /atomicbk/images/atomgirl.jpg HTTP/1.0" 200 34164
host = s3n://my_web_logs
8/20/14 6:59:59.000 AM
slip-41-5.ots.utexas.edu - - [30/Aug/1995:22:59:59 -0400] "GET /pub/abaa-booknet/images/books.gif HTTP/1.0" 200 2253
host = s3n://my_web_logs
8/20/14 6:59:59.000 AM
slip-41-5.ots.utexas.edu - - [30/Aug/1995:22:59:59 -0400] "GET /pub/abaa-booknet/images/abaabutt.gif HTTP/1.0" 200 1041
host = s3n://my_web_logs
8/20/14 6:59:59.000 AM
slip-41-5.ots.utexas.edu - - [30/Aug/1995:22:59:59 -0400] "GET /pub/abaa-booknet/alldlrs/dlrimage/books.gif HTTP/1.0" 200 2253
host = s3n://my_web_logs
8/20/14 6:59:59.000 AM
romana.tymnet.com - - [29/Aug/1995:22:59:59 -0400] "GET /pub/atomicbk/new.gif HTTP/1.0" 200 744
host = s3n://my_web_logs
8/20/14 6:59:59.000 AM
slip22.pdxts1.pacifier.com - - [29/Aug/1995:22:59:59 -0400] "GET /atomicbk/scotth.gif HTTP/1.0" 200 790
host = s3n://my_web_logs