



# Log Analysis with Splunk



Alexander Markov  
Senior Web Developer  
WCMS Competency Center

October 23, 2014

# SAP.com

- Several mln daily requests on average
- > 1 mln unique visitors per month
- Content publishing, reports



- Intensive work for DEV-OPS team

# Use Case



<http://www.phonesinmovies.com/2011/12/matrix-1999.html>

# Rescue Underway

Monitoring tool memory alert



```
17.07.2014 13:06:50.286 *ERROR*  
[GET /bin/assetreport.xls HTTP/1.1]  
    Uncaught Throwable  
java.lang.OutOfMemoryError: GC overhead  
    limit exceeded
```



Java Code Fixed – Problem Solved

# Any Data From Any Source



<http://www.splunk.com/view/splunk/SP-CAAAG57>

# Workflow

Search and time query

Indexing

MapReduce



Visualize search results

Timeline

Events

Statistics

Charts

Dashboards



Share

Alerts

Scheduled Reports

Data Export

# Event Timeline

## SAP.com PROD Exception Report

Save Save As View Close

```
index="wcms" host="sapcom-prod-publish01" sourcetype=crx_err_wcms_publish *Exception* OR OutOfMemoryError |
rex field=_raw max_match=50 "\b(?:<exception_type>[\w\.\.]+\.[\.\.]{1}[A-Z]{1}[\w\.\.]+(Exception|Error))\b\b(?:<sap_source>com\.sap[\.\.]+)" |
stats count, values(sap_source) by exception_type |
sort -count |
eval exception_type=if(mvcount(split(exception_type,".") > 3, mvjoin(mvindex(split(exception_type,"."),0,1),"."+"..." +mvindex(split(exception_type,"."),-1),
exception_type) |
rename exception_type AS "Exception Type",values(sap_source) AS "SAP.com Sources"
```

Last 24 hours

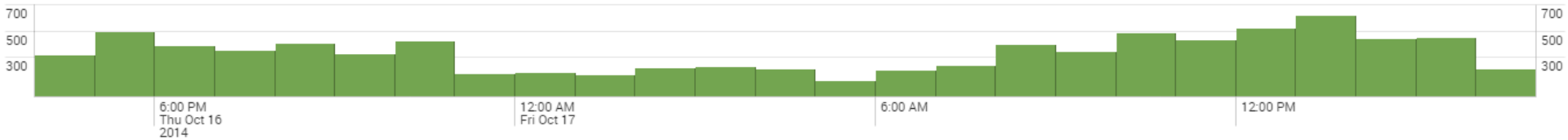
8,167 events (10/16/14 4:24:54.000 PM to 10/17/14 4:24:54.000 PM)

Job View Stop Refresh Download Print Verbose Mode

Events (8,167) Statistics (24) Visualization

Format Timeline Zoom Out Zoom to Selection Deselect

1 hour per column



List Format 20 Per Page

< Prev 1 2 3 4 5 6 7 8 9 ... Next >

All Fields		i	Time	Event
< Hide Fields		>	10/17/14 4:24:46.362 PM	17.10.2014 16:24:46.362 *ERROR* [172.16.93.131 [1413555885301] GET /content/sapcom/emea/norway/no_no/pc/tech/business-process-manag ement/resources.html HTTP/1.1] com.sap.smart.response.AssetResponse AssetResponse.fromResponse(jsonString)- exception: class org.ap ache.sling.commons.json.JSONObject\$Null:99.0 incompatible with class org.apache.sling.commons.json.JSONArray:99.0 host = sapcom-prod-publish01   source = /monsoon/opt/jcr/publish/crx-quickstart/logs/error.log   sourcetype = crx_err_wcms_publish
Selected Fields a host 1 a source 1 a sourcetype 1		>	10/17/14 4:24:38.293 PM	17.10.2014 16:24:38.293 *ERROR* [172.16.93.194 [1413555876976] GET /content/sapcom/emea/norway/no_no/services-support/svc/business- analytics/resources.html HTTP/1.1] com.sap.smart.response.AssetResponse AssetResponse.fromResponse(jsonString)- exception: class or g.apache.sling.commons.json.JSONObject\$Null:99.0 incompatible with class org.apache.sling.commons.json.JSONArray:99.0 host = sapcom-prod-publish01   source = /monsoon/opt/jcr/publish/crx-quickstart/logs/error.log   sourcetype = crx_err_wcms_publish
Interesting Fields # date_hour 24 # date_mday 2 # date_minute 60 # date_month 1		>	10/17/14 4:24:36.927 PM	17.10.2014 16:24:36.927 *ERROR* [172.16.93.131 [1413555875312] GET /content/sapcareer/global/usa/en_us/_jcr_content/home/columncont rol/par1/personalizedref.nocache.html HTTP/1.1] com.sapcareer.components.video.VideoDisplayBean Exception in AwardBean javax.jcr.Ac cessDeniedException: /content/sapcareer/global/usa/en_us/personalization-landing/video-landing/jcr:content/personalizedComponent/vi dentile 0/image/fileReference: not allowed to add or modify item

# Statistics

## SAP.com PROD Exception Report

Save Save As View Close

```
index="wcms" host="sapcom-prod-publish01" sourcetype=crx_err_wcms_publish *Exception* OR OutOfMemoryError |
rex field=_raw max_match=50 "\b(?<exception_type>[\w\.\.]+\.[\.\.]{1}[A-Z]{1}[\w]+(Exception|Error))\b|\b(?<sap_source>com\.sap[^\s]+)" |
stats count, values(sap_source) by exception_type |
sort -count |
eval exception_type=if(mvcount(split(exception_type,".") > 3, mvjoin(mvindex(split(exception_type,"."),0,1),"-")+"..." +mvindex(split(exception_type,"."),-1),
exception_type) |
rename exception_type AS "Exception Type",values(sap_source) AS "SAP.com Sources"
```

Last 24 hours

8,167 events (10/16/14 4:24:54.000 PM to 10/17/14 4:24:54.000 PM)

Job View Stop Refresh Download Print Verbose Mode

Events (8,167) Statistics (24) Visualization

50 Per Page Format Preview

Exception Type	count	SAP.com Sources
1 javax.jcr.AccessDeniedException	4743	com.sap.components.tags.IncludeTag.doEndTag(IncludeTag.java:79) com.sap.wcms.assetmgmt.downloadhandler.DownloadHandlerServlet com.sap.wcms.assetmgmt.downloadhandler.VideoDisplayBean com.sap.wcms.assetmgmt.downloadhandler.VideoDisplayBean.getClearOldProperties(VideoDisplayBean.java:391) com.sapcareer.components.video.VideoDisplayBean com.sapcareer.core.filter.ApplicationFilterServiceImpl.doFilter(ApplicationFilterServiceImpl.java:119)
2 java.lang.IllegalStateException	657	com.sap.wcms.profilemgmt.forms.PrivacyWTAForm.<init>(PrivacyWTAForm.java:53) com.sap.wcms.profilemgmt.servlet.LoadRegionOfCountries.doGet(LoadRegionOfCountries.java:46) com.sap.wcms.profilemgmt.servlet.ProfileMgmtServlet.doPost(ProfileMgmtServlet.java:134) com.sap.wcms.profilemgmt.servlet.ProfileMgmtServlet.getJSONResponsePrivacyWTAFromNPC(ProfileMgmtServlet.java:230) com.sap.wcms.registration.forms.CampaignForm.<init>(CampaignForm.java:36) com.sap.wcms.registration.forms.RegistrationFormFactory.getFormByType(RegistrationFormFactory.java:48) com.sap.wcms.registration.forms.UpdateProfileMangtExplicitFormComponentBean.<init>(UpdateProfileMangtExplicitFormComponentBean.java:91) com.sap.wcms.registration.servlet.RedirectServlet.doGet(RedirectServlet.java:100) com.sap.wcms.registration.servlet.RegistrationServlet.doPost(RegistrationServlet.java:115) com.sap.wcms.registration.util.EmailUtil.getConfirmationEmailJSONArray(EmailUtil.java:72) com.sap.wcms.registration.util.EmailUtil.getEmailContent(EmailUtil.java:174) com.sapcareer.core.filter.ApplicationFilterServiceImpl.doFilter(ApplicationFilterServiceImpl.java:119)
3 java.net.URISyntaxException	510	com.sapcareer.core.filter.ApplicationFilterServiceImpl.doFilter(ApplicationFilterServiceImpl.java:119)
4 org.apache...SlingException	429	com.sap.admin.components.tagging.TagInheritanceBean.getGsaTags(TagInheritanceBean.java:89) com.sap.components.AbstractComponentBean.initVariables(AbstractComponentBean.java:151)



# Visualization – Pie Chart

Search Pivot Reports Usage Reports Error Reports Alerts Dashboards Dashboards WCMS for sap.com

## SAP.com PROD Exception Report

Save Save As View Close

```
index="wcms" host="sapcom-prod-publish01" sourcetype=crx_err_wcms_publish *Exception* OR OutOfMemoryError |
rex field=_raw max_match=50 "\b(?<exception_type>[\w\.\.]+\.[A-Z]{1}[\w]+(Exception|Error))\b|\b(?<sap_source>com\.sap[^\s]+)" |
stats count, values(sap_source) by exception_type |
sort -count |
eval exception_type=if(mvcount(split(exception_type,".") > 3, mvjoin(mvindex(split(exception_type,"."),0,1),"."+"..." +mvindex(split(exception_type,"."),-1),
exception_type) |
rename exception_type AS "Exception Type",values(sap_source) AS "SAP.com Sources"
```

Last 24 hours

8,167 events (10/16/14 4:24:54.000 PM to 10/17/14 4:24:54.000 PM)

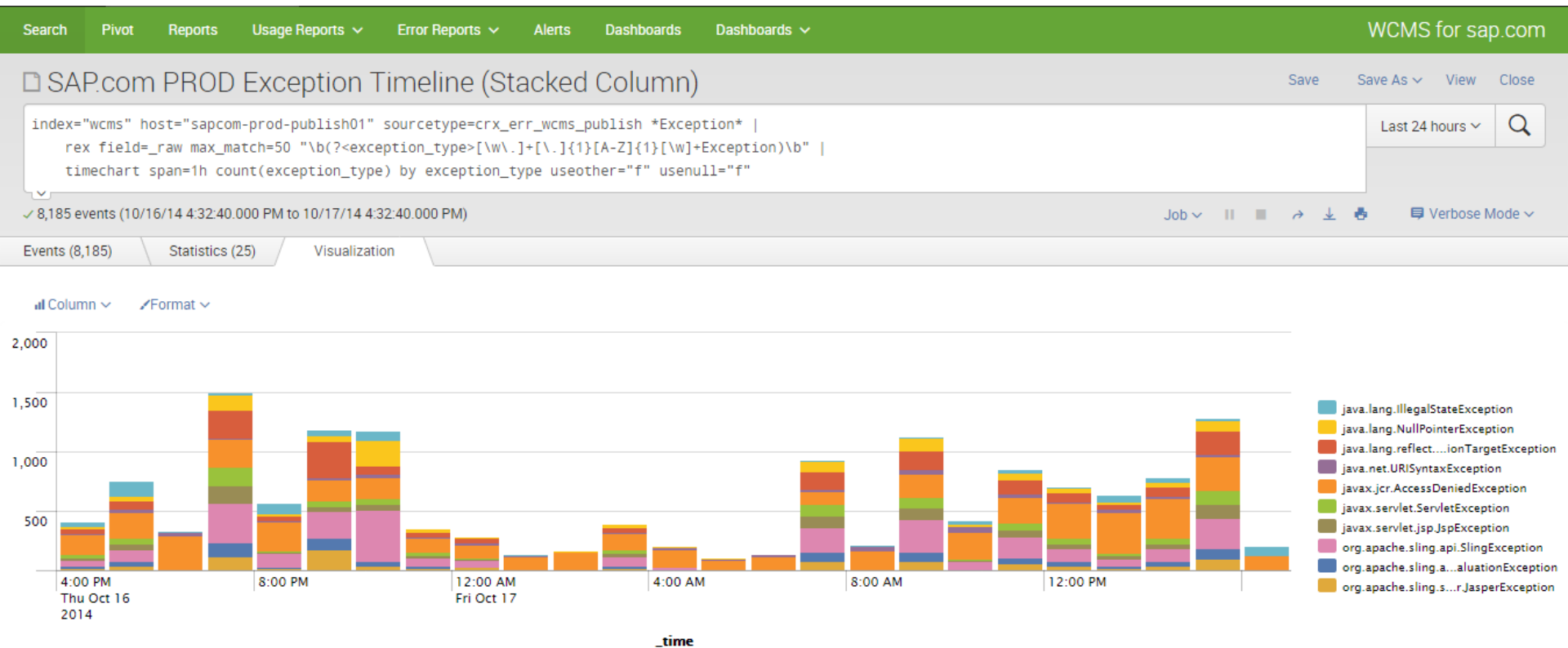
Job View Stop Refresh Download Print Verbose Mode

Events (8,167) Statistics (24) Visualization

Pie Format

Exception Type	count	SAP.com Sources
1 javax.jcr.AccessDeniedException	4743	com.sap.components.tags.IncludeTag.doEndTag(IncludeTag.java:79) com.sap.wcms.assetmgmt.downloadhandler.DownloadHandlerServlet com.sap.wcms.assetmgmt.downloadhandler.VideoDisplayBean

# Visualization – Timechart

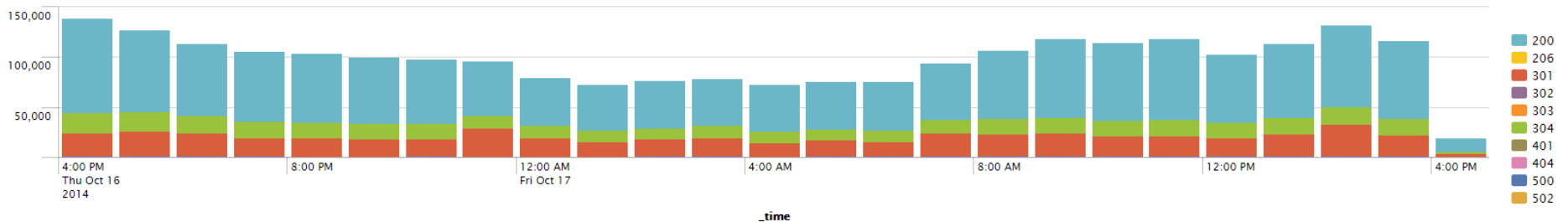


# Real-Time Dashboards

## SAP.com PROD Dashboard

Edit More Info

### User Activity with Response Codes



### Top Pages 9m ago

WCMS_URL	count
41 /services-support.html	376
42 /brazil/index.html	359
43 /solutions/tech/cloud.html	354
44 /japan/index.html	351
45 /solutions/analytics/business-intelligence.html	349
46 /cis/index.html	346
47 /solutions/sme.html	344
48 /uk/index.html	339
49 /software-free-trials/index.html	322
50 /france/index.html	308

### Top 404 URLs 9m ago

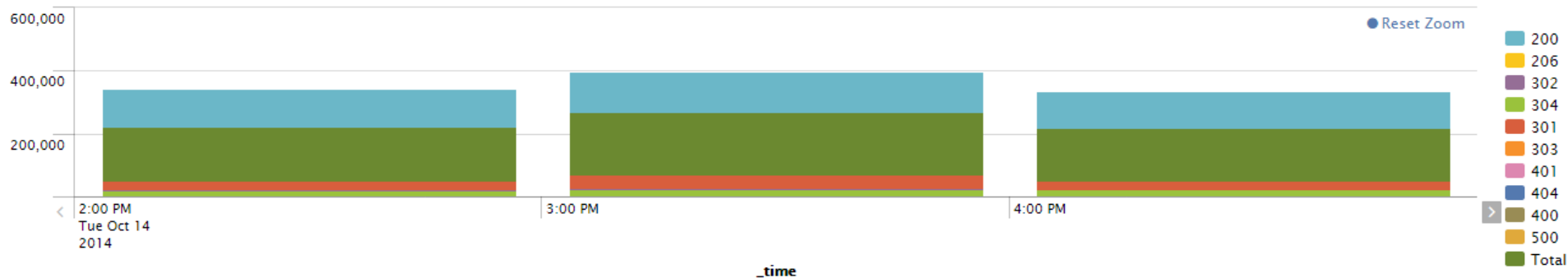
WCMS_URL	count
21 /etc/designs/sapcom/resources/www_sap_com/global/ui/images/backgrounds/bkg_wh	21
22 /etc/designs/sapcom/resources/www_sap_com/global/ui/images/icons/ico_facebook_t	22
23 /etc/designs/sapcom/resources/global/ui/css/images/background.png	23
24 /etc/designs/sapcom/resources/global/ui/ux/icons/icon-whitearrowright.gif	24
25 /etc/designs/sapcom/resources/www_sap_com/global/ui/images/icons/icon-social-net	25
26 /etc/designs/sapcom/resources/global/ui/images/icons/icon-internal-search-results-pre	27
27 /etc/designs/sapcom/resources/global/ui/images/icons/icon-internal-search-results-pre	27
28 /etc/designs/sapcom/resources/global/ui/images/icons/socialmedia/icon-social-networ	28
29 /etc/designs/sapcom/resources/global/ui/images/icons/socialmedia/icon-social-networ	29
30 /etc/designs/sapcom/resources/global/ui/images/icons/socialmedia/icon-social-xing-w	30

### Top User Agent Strings 8m ago

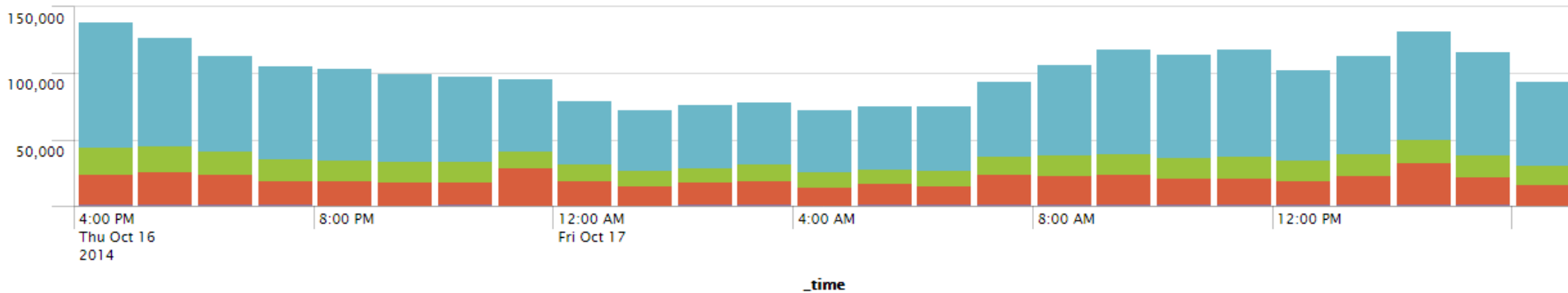
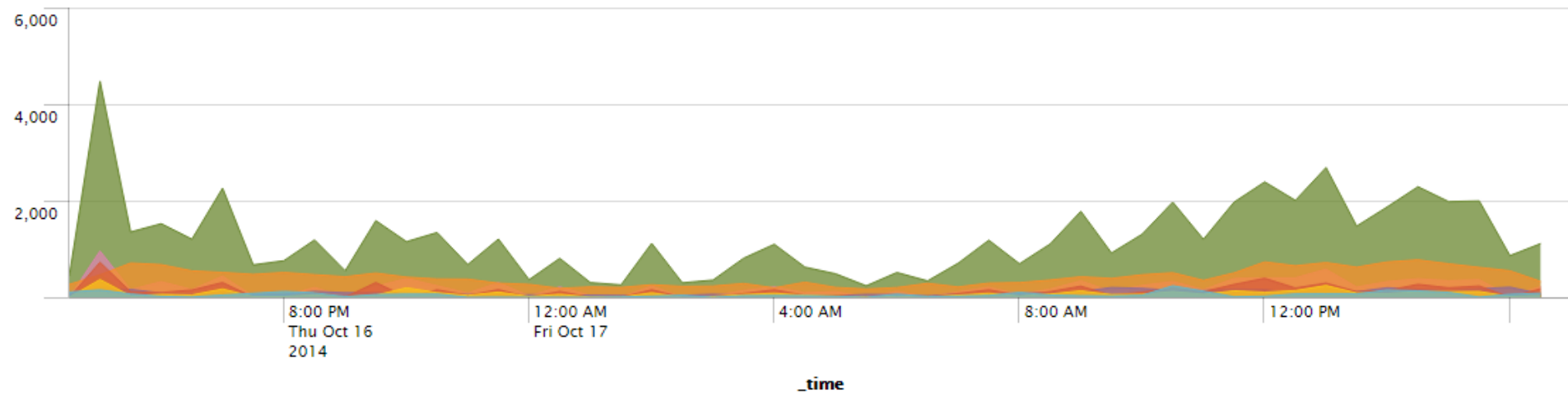
WCMS_UserAgentString	count
21 Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 6.1; WOW64; Trident/4.0; SLCC2; .NET CLR 2.0.50727; .NET CLR 3.5.30729; .NET CLR 3.0.30729; Media Center PC 6.0; infoPath.3)	15009
22 Mozilla/5.0 (iPhone; CPU iPhone OS 8_0_2 like Mac OS X) AppleWebKit/600.1.4 (KHTML, like Gecko) Version/8.0 Mobile/12A405 Safari/600.1.4	14786
23 Mozilla/5.0 (Windows NT 6.3; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/38.0.2125.104 Safari/537.36	14730
24 Mozilla/5.0 (Macintosh; Intel Mac OS X 10_9_5) AppleWebKit/600.1.17 (KHTML, like Gecko) Version/7.1 Safari/537.85.10	14523
25 Mozilla/5.0 (Windows NT 5.1) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/37.0.2062.124 Safari/537.36	14163

# Pan and Zoom

Narrow your search by click on time intervals/table rows/histogram columns



# Number of Exceptions vs User Activity



# Alerts and Scheduled Reports

## Search query

- `index="wcms" source="/error.log" OutOfMemoryError`

## Cron schedule/real time

- `00 0,12 * * * | 30 7 * * 1,4`

## Time range

- `-12h@h - now`

## Trigger condition

- `number of results > 0`

## Email recipients, subject, CSV/Table/PDF results

- `Operations team distribution list`

# Search Processing Language

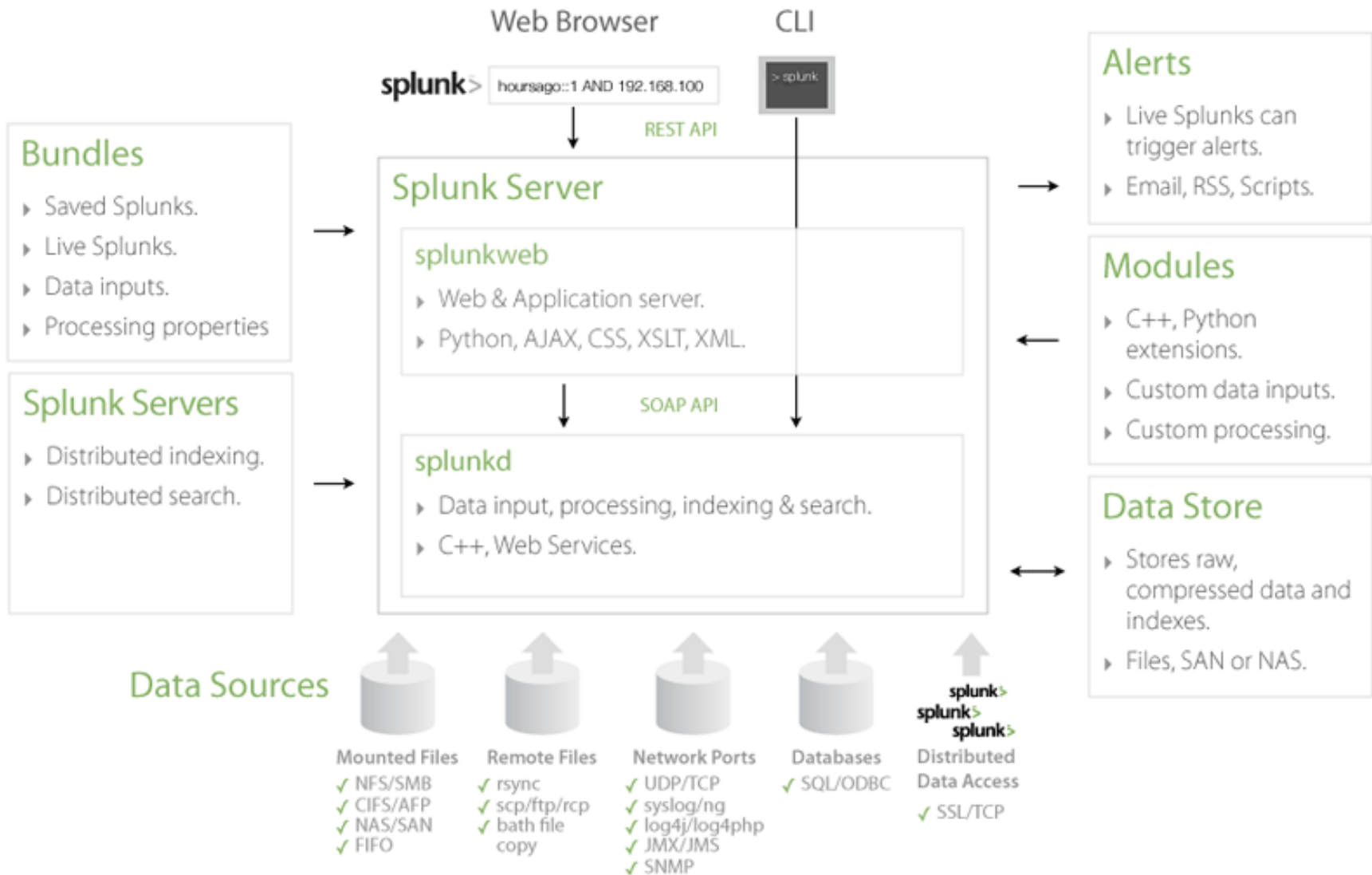
```
index="wcms" host="sapcom*" source="/error.log"
*Exception* OR OutOfMemoryError |
rex field=_raw max_match=50
"\b(?<exc_type>[\w\.]++[\.]{1}[A-Z]{1}[\w]+(Exception|Error))\b|\b(?<sap_source>com
\sap[^\s]+)" |
stats count, values(sap_source) by exc_type |
sort -count |
eval exc_type=if(mvcount(split(exc_type, ".")) > 3,
mvjoin(mvindex(split(exc_type, "."), 0, 1), ".")+"..."
+mvindex(split(exc_type, "."), -1), exc_type) |
rename exc_type AS "Exception
Type", values(sap_source) AS "SAP.com Sources"
```

# Timechart Query

```
index="wcms" host="sapcom*"
source="/error.log" *Exception* |
rex field=_raw max_match=50
"\b(?<exc_type>[\w\.]++[\.]{1}[A-Z]{1}[\w]+(Exception|Error))\b" |
timechart span=1h count(exc_type) by
exc_type useother="f" usenull="f"
```

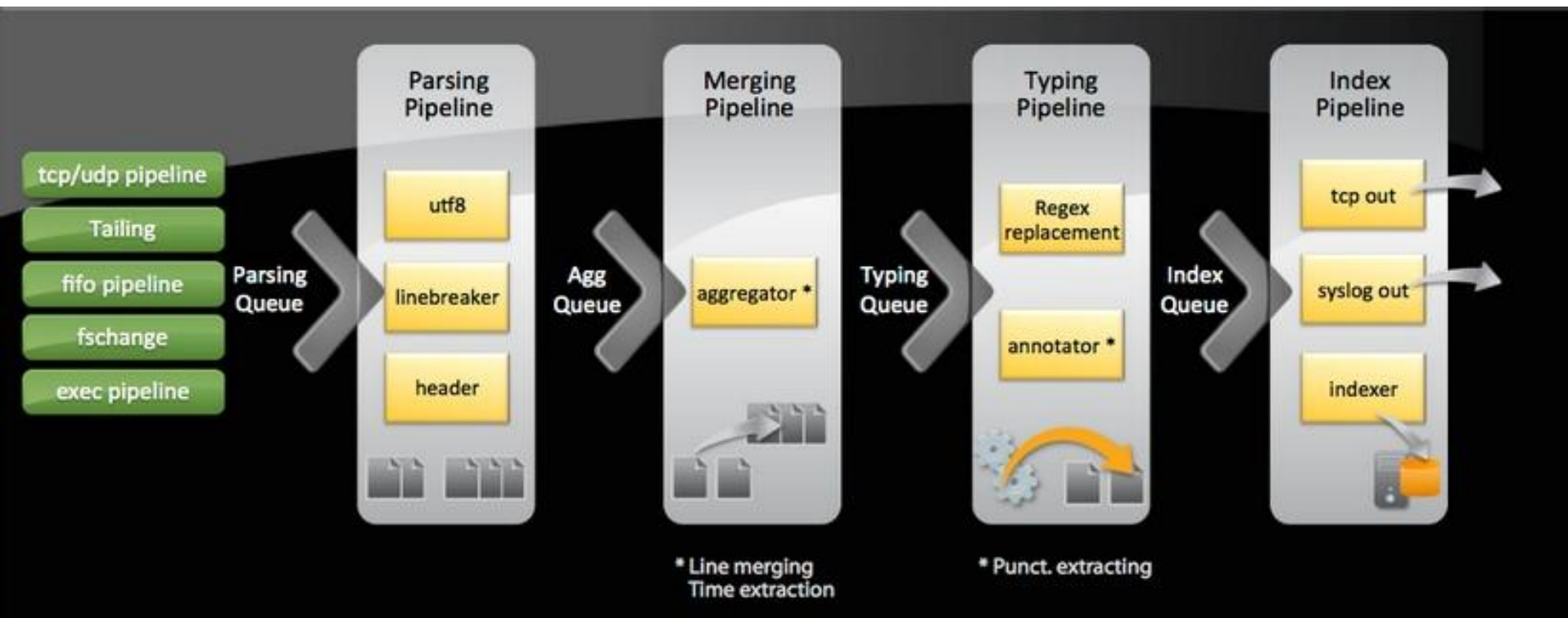


# Architecture



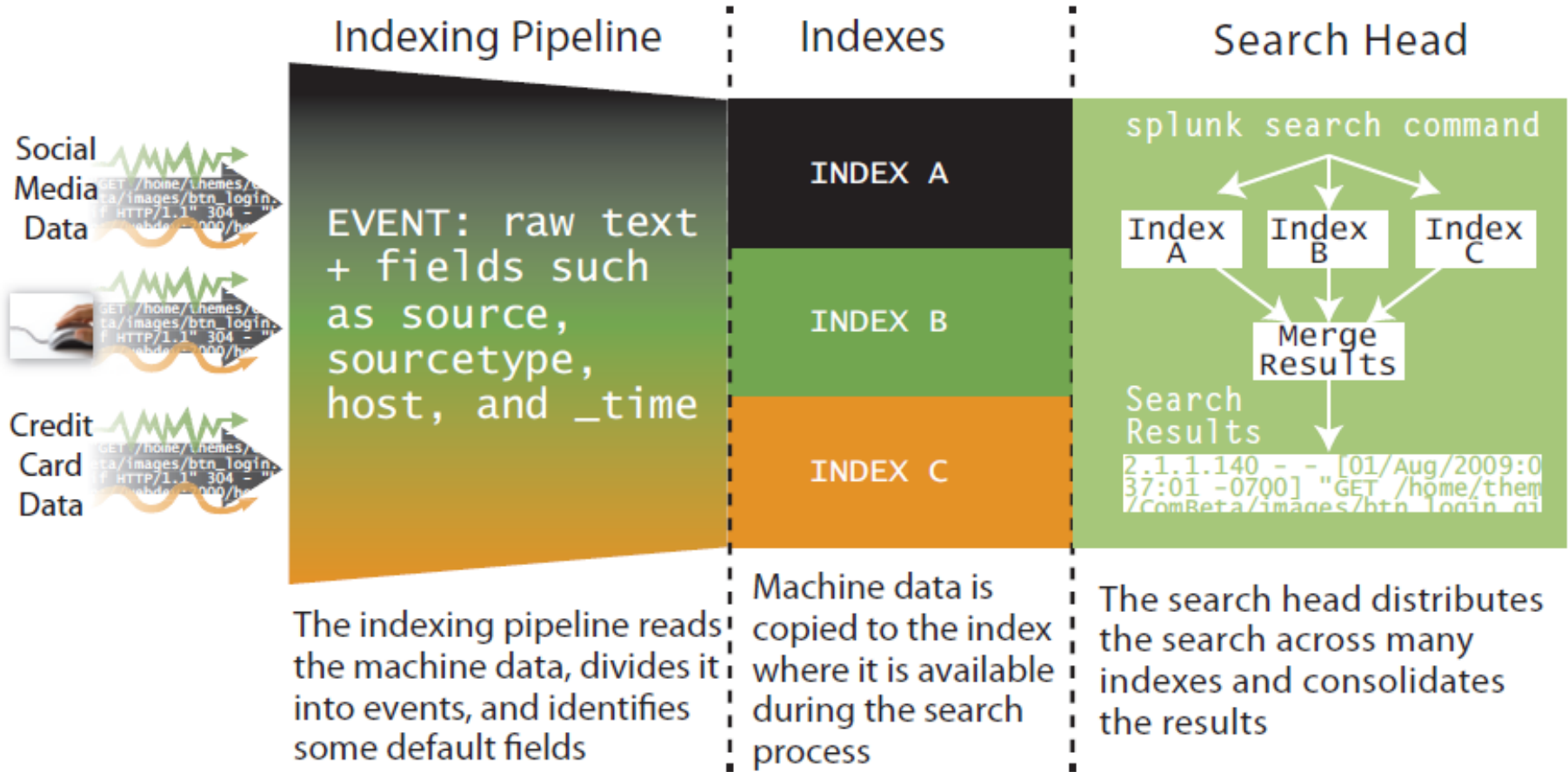
[http://www.splunk.com/themes/splunk\\_com/img/assets/images/developers/splunkarchitecture.gif](http://www.splunk.com/themes/splunk_com/img/assets/images/developers/splunkarchitecture.gif)

# Data Pipeline



<http://wiki.splunk.com/Community:HowIndexingWorks>

# Indexing



[http://www.splunk.com/web\\_assets/v5/book/Exploring\\_Splunk.pdf](http://www.splunk.com/web_assets/v5/book/Exploring_Splunk.pdf)

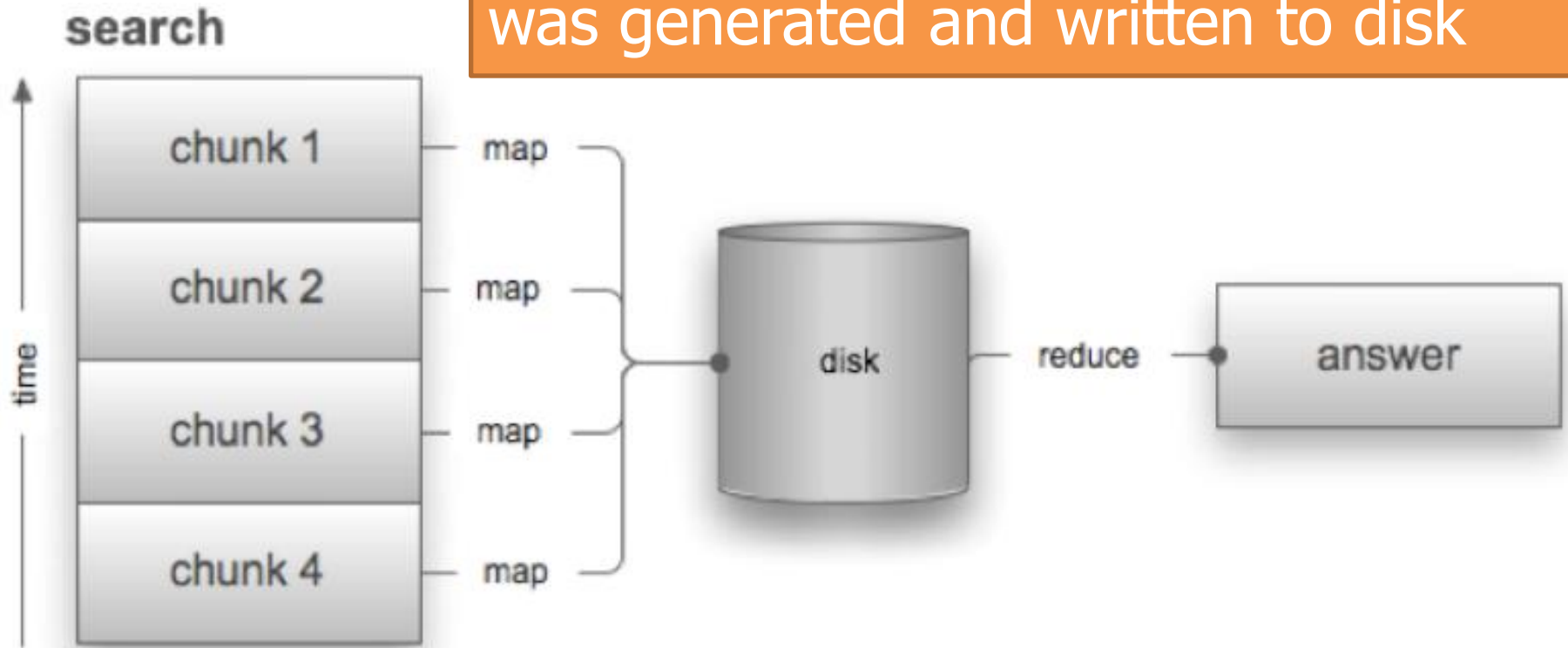
# MapReduce (1)

Introduced by Google, 2004

- concurrent map functions that process source data into sufficient statistics
- reduce function that merges statistics into final answer – see stats/timechart

# MapReduce (2)

Data is almost **always available** for search and reporting seconds after it was generated and written to disk



[http://www.splunk.com/web\\_assets/pdfs/secure/Splunk\\_and\\_MapReduce.pdf](http://www.splunk.com/web_assets/pdfs/secure/Splunk_and_MapReduce.pdf)

# Conclusion

1. Fast problem/cause detection and fix
2. Data centralization from multiple sources
3. Real-time web traffic analysis
4. Data visualization in timelines, tables, charts
5. User stats for deeper marketing research
6. Trends, forecasts ...

# P.S. Exception Handling

1. Log exceptions, use loggers and facades
2. Create readable messages with useful info about the error and use proper log level
3. Handle exceptions on the upper levels  
(throw from DAO – catch and log in web)
4. Throw app-specific exceptions with info for the client side

**THANK YOU  
FOR YOUR ATTENTION!**

[alexander.markov@sap.com](mailto:alexander.markov@sap.com)



# Q & A

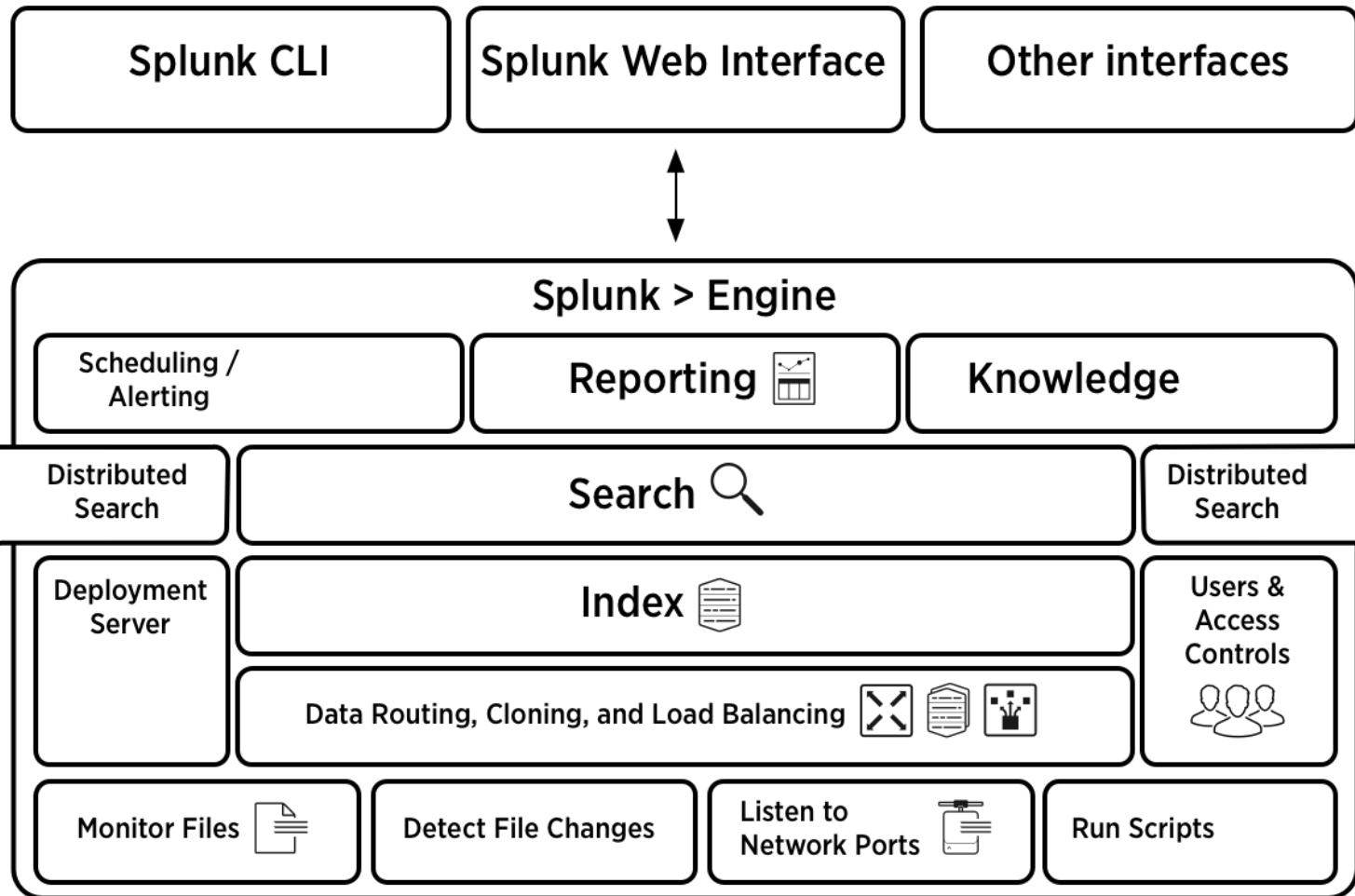
# Use Cases

- A. Boss alert: website is down
- B. Monitoring tool memory alert
- C. Website shows error page
- D. Extraordinary user activity spike

# History

- v1-3: like Google web search with 100 quickly displayed results
- v4: asynchronous multi-index search with MapReduce on an indexed datastore
- v5: report acceleration, PDF delivery, REST API, multisearch & predict
- v6: UI enhancements like dashboard editor, Pan and Zoom chart controls etc

# Architecture (1)



<http://docs.splunk.com/Documentation/Splunk/6.1.1/Installation/Splunksarchitectureandwhatgetsinstalled>

# Processes

**splunkd** distributed C/C++ server that accesses, processes and indexes data by streaming it through a series of pipelines, each made up of a series of processors; supports cmd interface for searching and viewing results

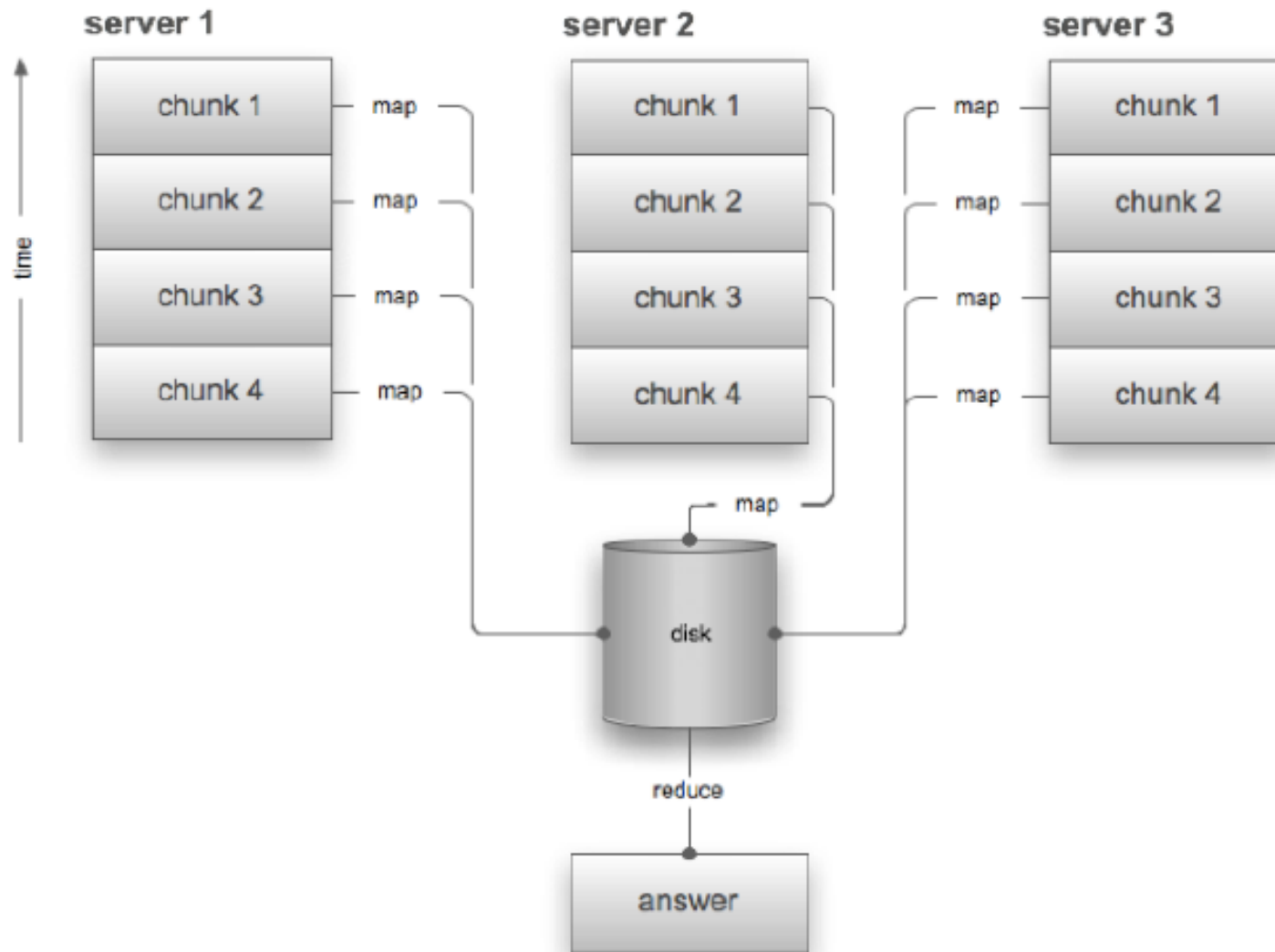
- **Pipelines** – single threads in the splunkd process with XML config, can pass data to one another via **queues**
- **Processors** – individual, reusable C/C++ functions that act on the stream of IT data passing through a pipeline

**splunkweb** Python-based application server based on CherryPy that provides the Splunk Web UI: allows users to search and navigate data stored by Splunk servers, to manage your Splunk deployment with a Web interface

splunkweb and splunkd can communicate with Web browser via REST:

- splunkd runs Web server on port 8089 with SSL/HTTPS by default
- splunkweb runs Web server on port 8000 without SSL/HTTPS by default

# MapReduce (3)



[http://www.splunk.com/web\\_assets/pdfs/secure/Splunk\\_and\\_MapReduce.pdf](http://www.splunk.com/web_assets/pdfs/secure/Splunk_and_MapReduce.pdf)

# MapReduce (4)

## **Distributed Search on a cluster**

1. Search formulated into the map and reduce functions
2. Network connections are established to each Splunk Indexer in the search cluster
3. The map function is sent to each of these Splunk instances and each begins processing data with MapReduce
4. As data streams back to the instance that initiated the search, it is persisted to disk for the reduce function

For pure reporting searches with a map function that compresses data for network transport, reporting speed is linear with the number of index servers in the cluster

# System Requirements

## OS

- Linux, FreeBSD, Mac OS X, Solaris etc
- Windows Server/7/8

## Browsers

- Firefox/Chrome/Safari
- IE 9+

Platform	Recommended hardware capacity/configuration	Minimum supported hardware capacity
Non-Windows platforms	2x six-core, 2+ GHz CPU, 12 GB RAM, Redundant Array of Independent Disks (RAID) 0 or 1+0, with a 64 bit OS installed.	1x1.4 GHz CPU, 1 GB RAM
Windows platforms	2x six-core, 2+ GHz CPU, 12 GB RAM, RAID 0 or 1+0, with a 64 bit OS installed.	Intel Nehalem CPU or equivalent at 2 GHz, 2 GB RAM

<http://docs.splunk.com/Documentation/Splunk/6.1.4/Installation/Systemrequirements>

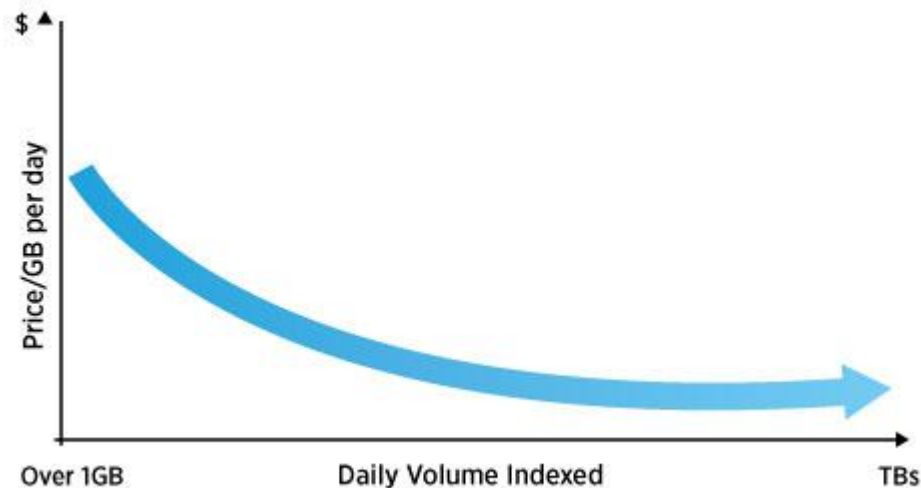


# Free vs Enterprise

Feature	Free	Enterprise	Cloud
<b>Volume</b>	500 MB/day	Unlimited	From 5 GB/day
<b>Distributed search across multiple Splunk deployments</b>	No	Yes	No
<b>Cluster management and reporting</b>	No	Yes	No
<b>High performance analytics, reports, PDFs</b>	No	Yes	Yes
<b>Real-Time Alerts</b>	No	Yes	Yes
<b>Premium apps and customer support</b>	No	Yes	Yes

<https://www.splunk.com/view/free-vs-enterprise/SP-CAAAE8W>

# Enterprise Pricing



- **Perpetual license:** one-time fee starts as low as \$4,500 for 1 GB/day not including annual support fees
- **Term license:** starts at \$1,800 per year including annual support fees
- **Splunk Cloud** is priced by subscription plans that start at \$675 per month for data volumes up to 5GB/day and scales to 5TB/day  
<https://www.splunk.com/view/pricing/SP-CAAADFV>

# Interesting points to investigate

1. Data from streams – how is it processed/saved?
2. Persistence of already aggregated data?
3. Splunk performance, benchmarks
4. Compare with Google Analytics and Yandex.Metrica