Log Analysis with Splunk

Alexander Markov
Senior Web Developer
WCMS Competency Center

SAP

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

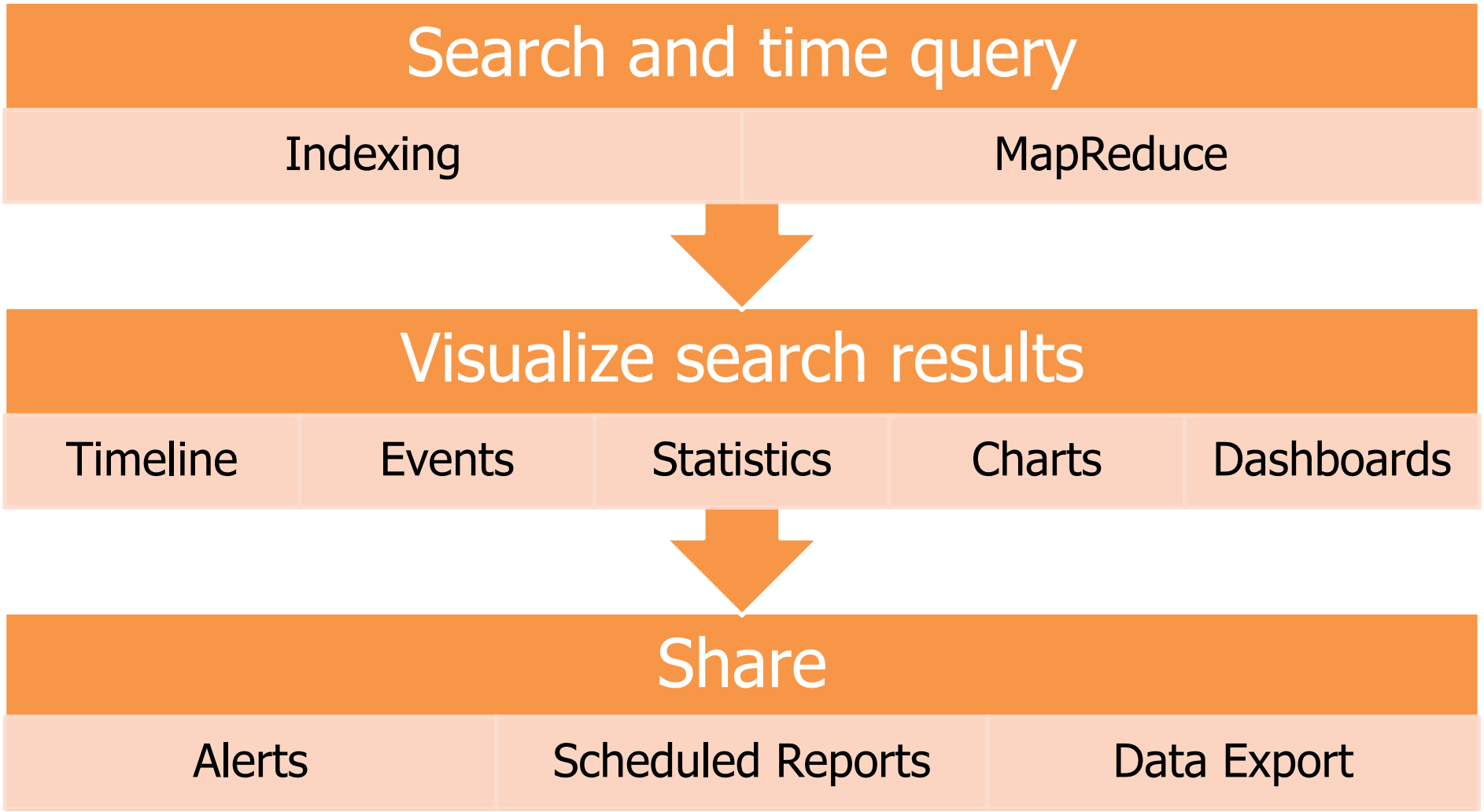
```
17.07.2014 13:06:50.286 *ERROR*
[GET /bin/assetreport.xls HTTP/1.1]
Uncaught Throwable
java.lang.OutOfMemoryError: GC overhead
limit exceeded
```

Java Code Fixed – Problem Solved

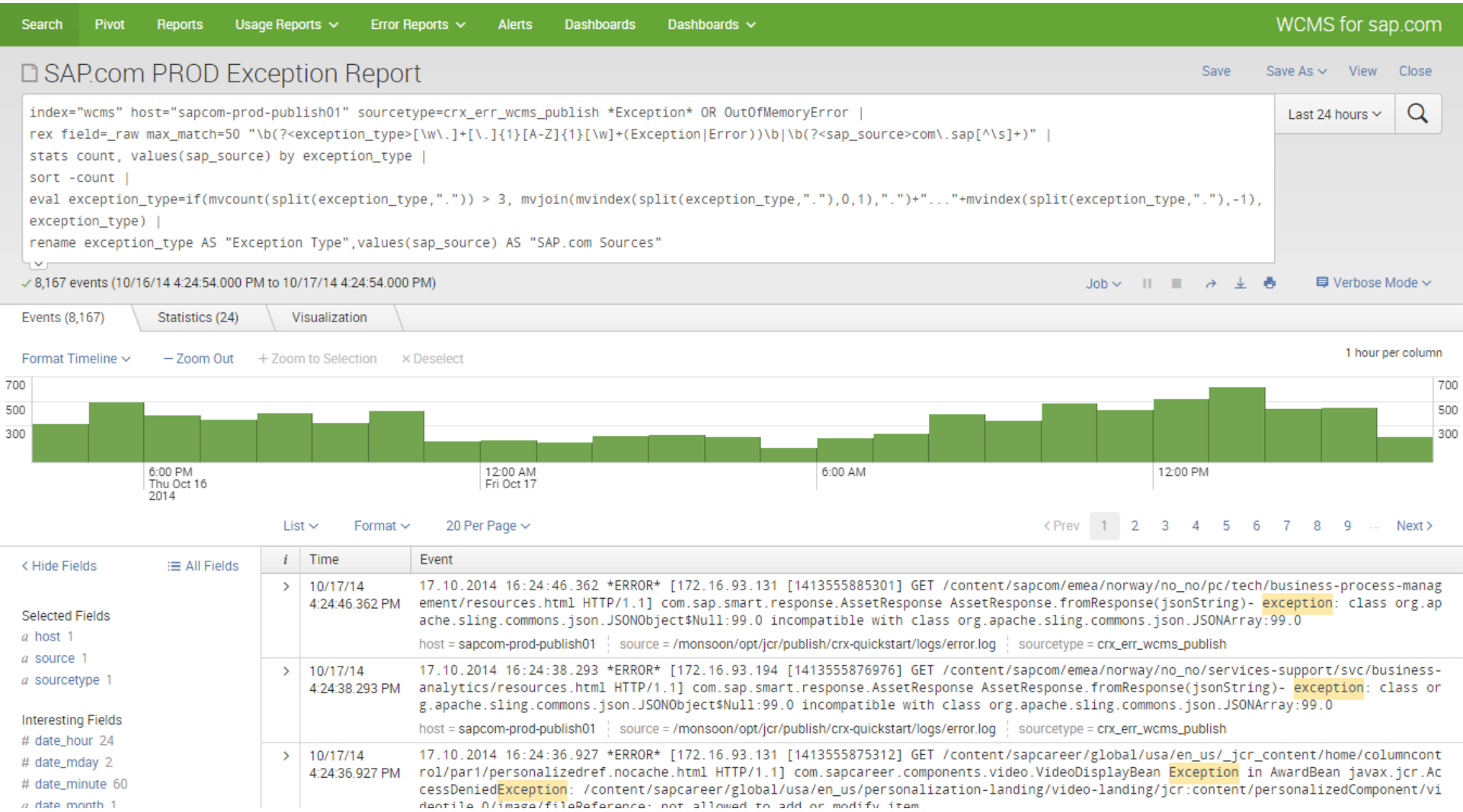
Any Data From Any Source

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



Statistics

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\.]+[\.]{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 | 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\.]+[\.]{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"
```

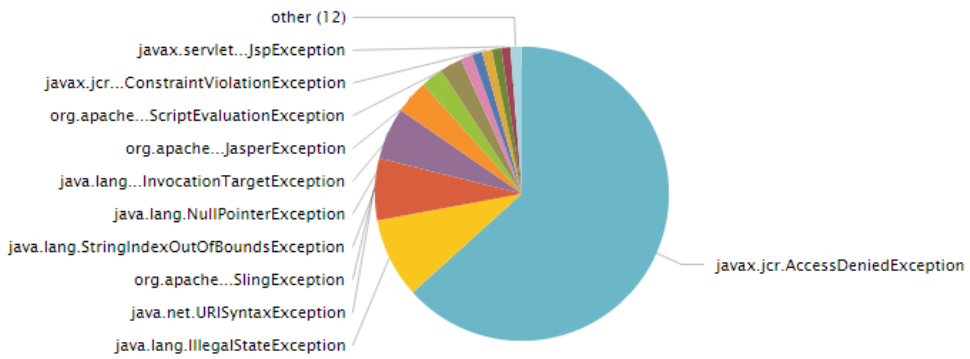
Last 24 hours

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

Job | Verbose Mode

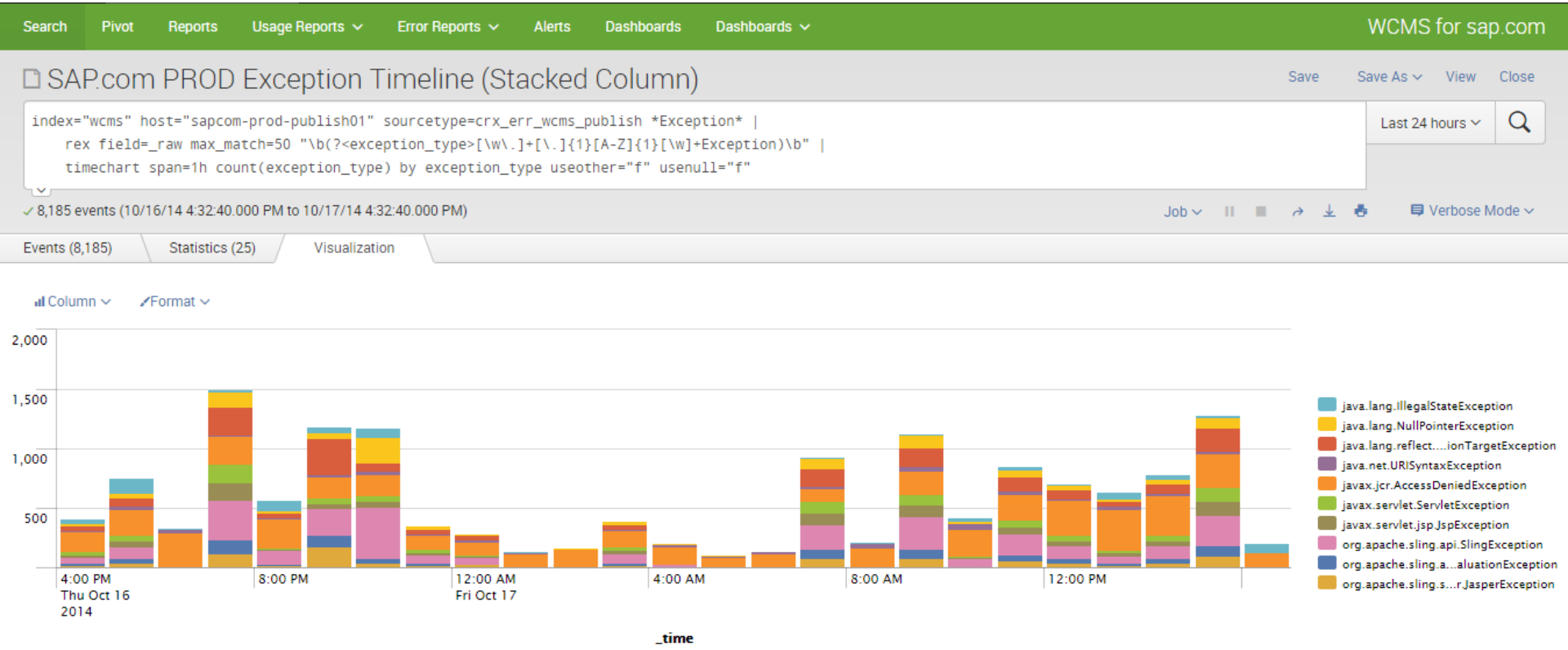
Events (8,167) | Statistics (24) | Visualization

Pie | Format

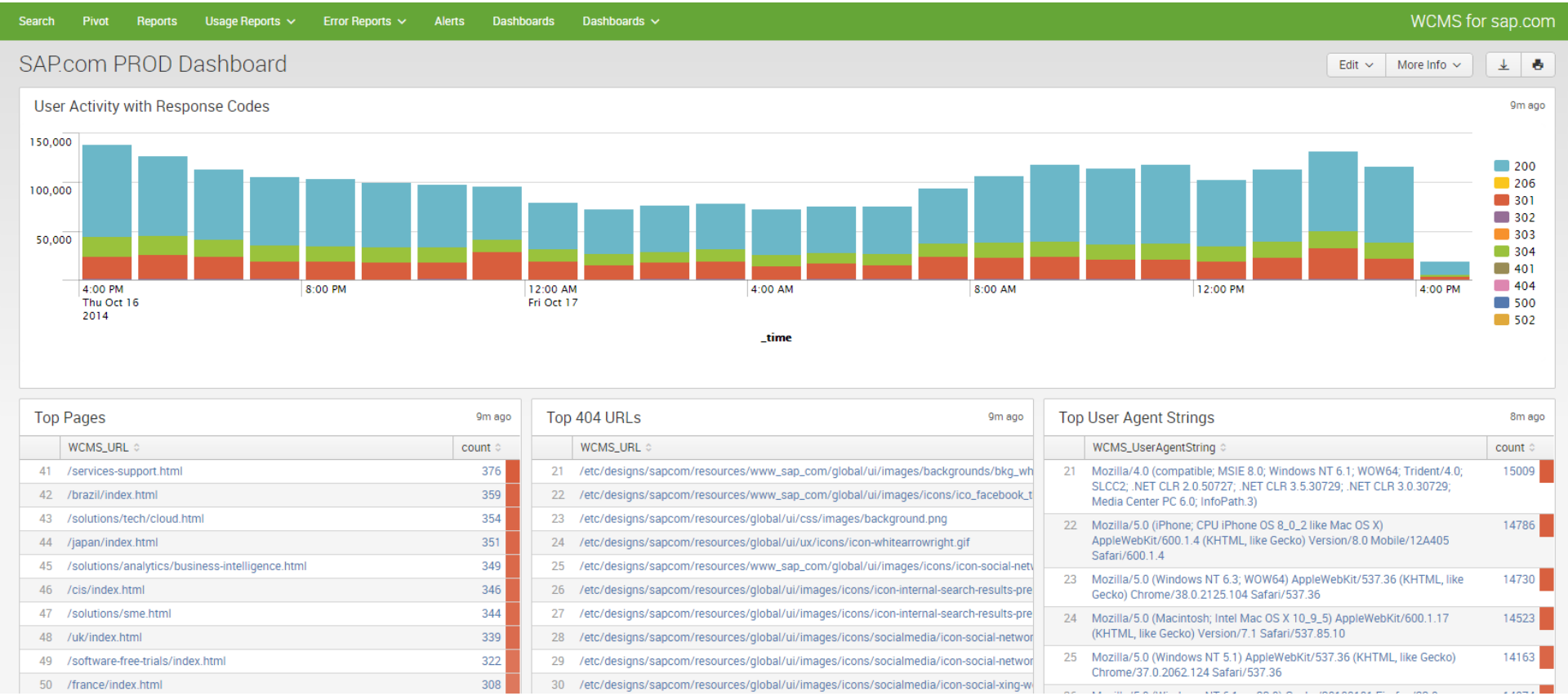


	Exception Type	count	SAP.com Sources
1	javax.jcr.AccessDeniedException	4743	com.sap.components.tags.IncludeTag.doEndTag(IncludeTag.java:79) com.sap.wcms.assetmgmt.downloadhandler.DownloadHandlerServlet com.sap.wcms.assetmgmt.downloadhandler.VideoDisplayBean

Visualization – Timechart

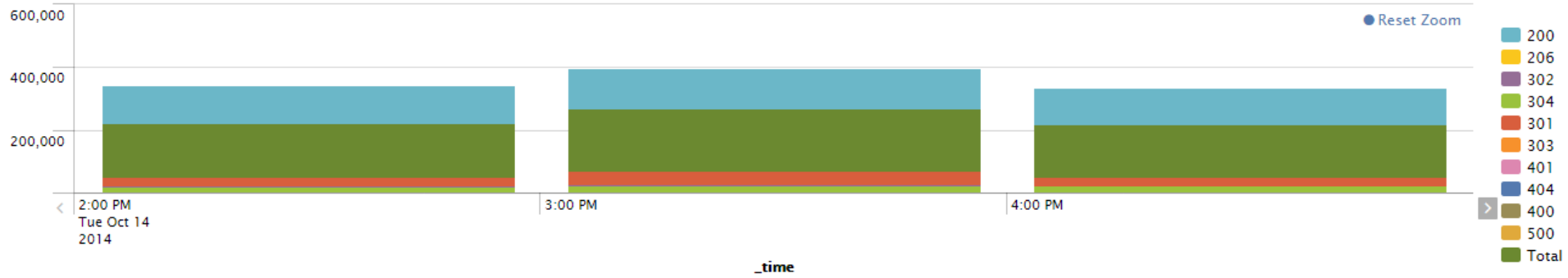


Real-Time Dashboards

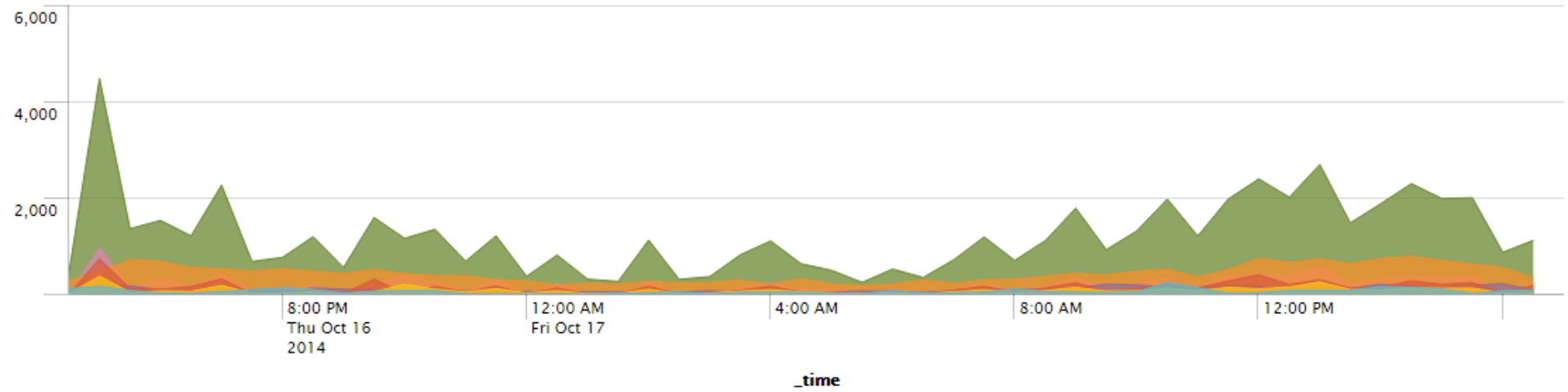


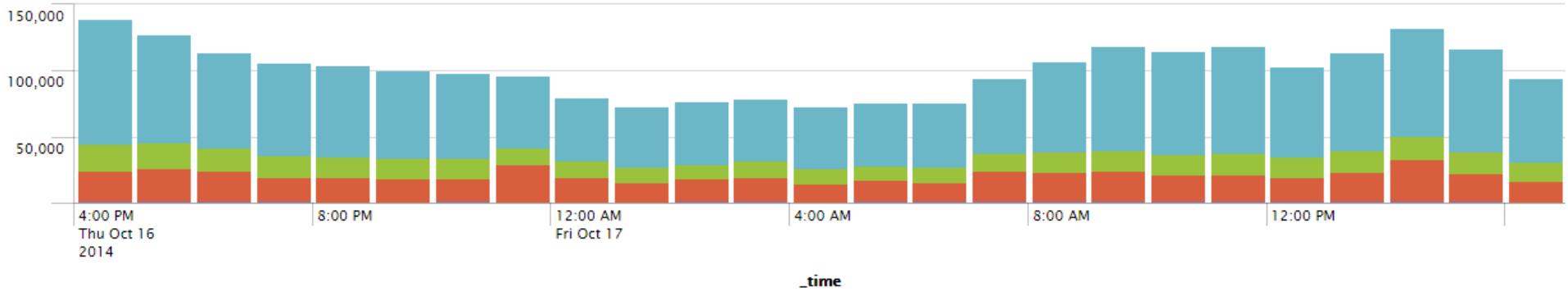
Pan and Zoom

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`

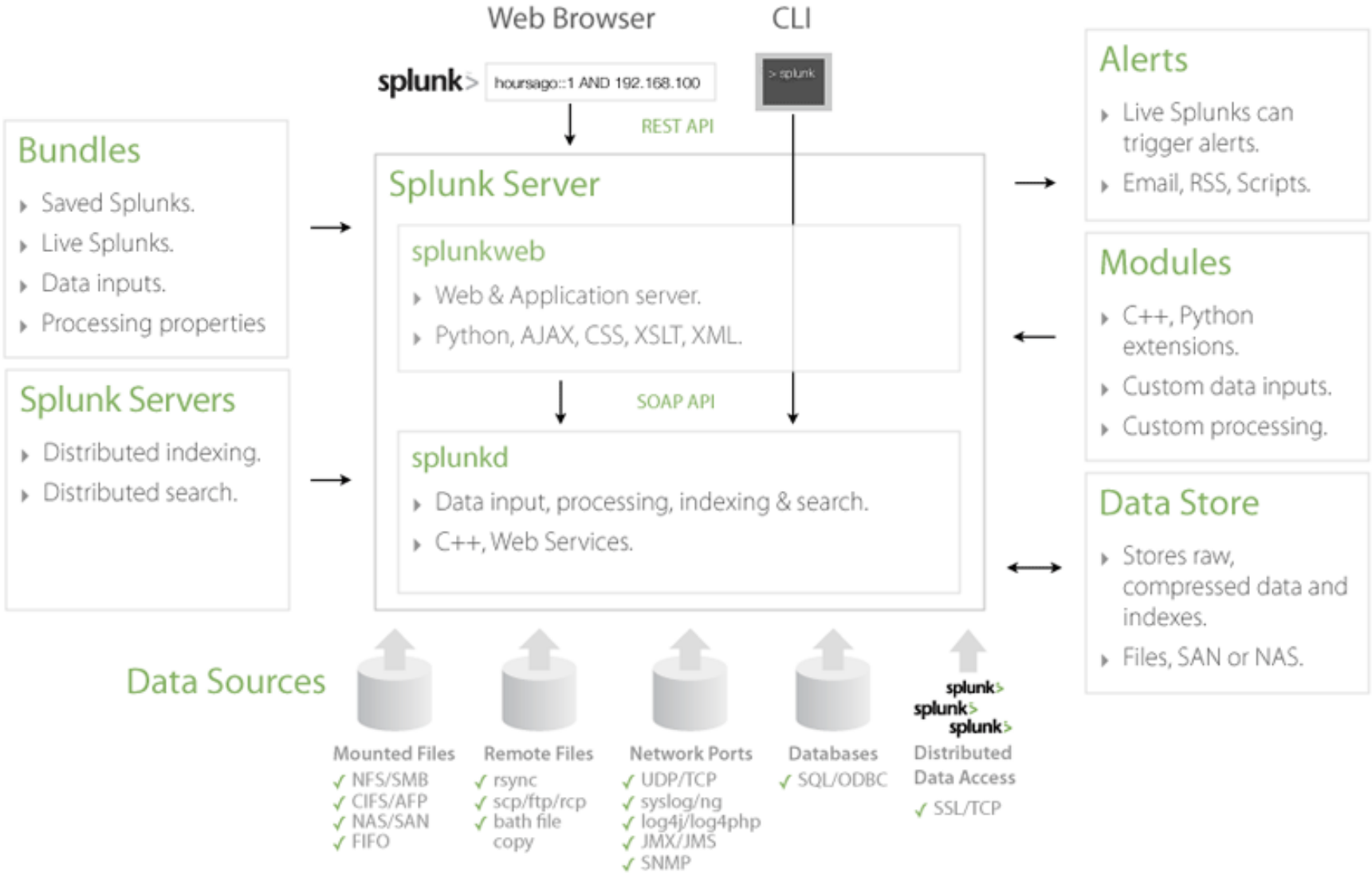
Search Processing Language

```
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

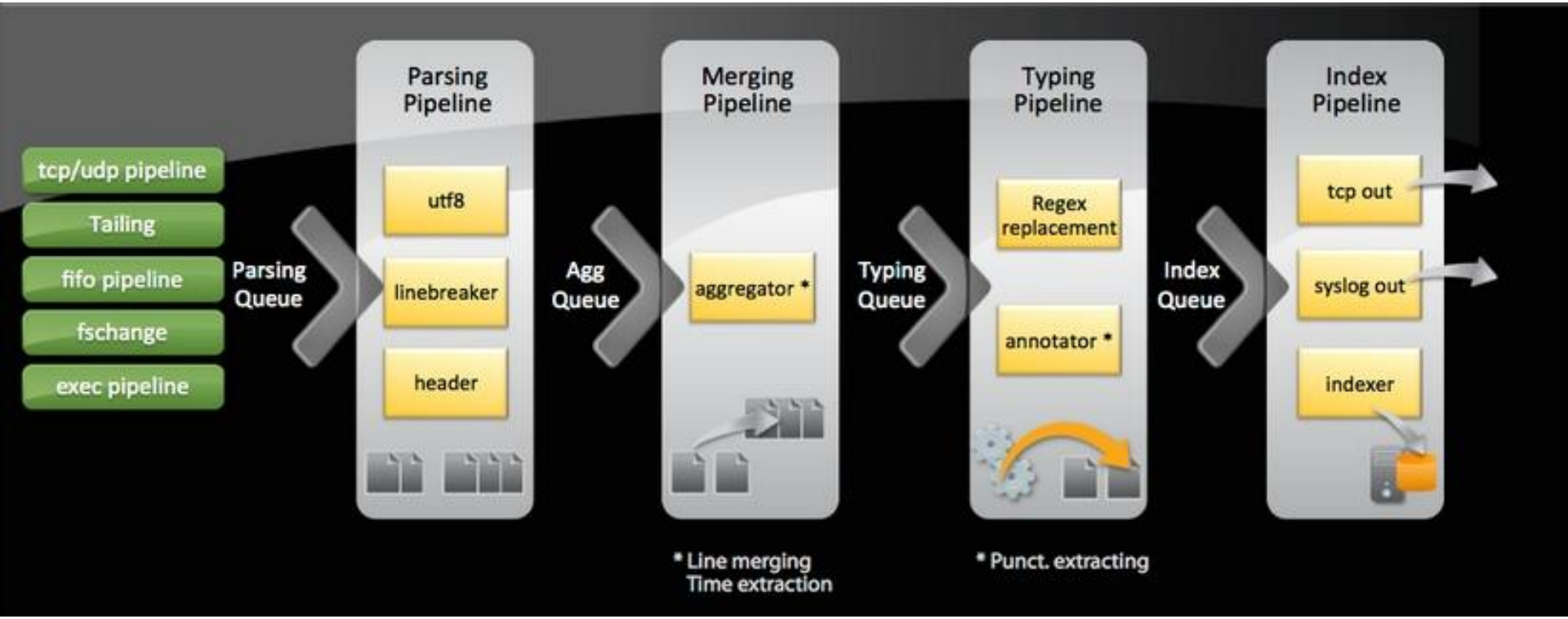
```
index="wcms" host="sapcom*"
source="/error.log" *Exception* |
rex field=_raw max_match=50
"\b(?<exc_type>[\w\.]+[\.]{1}[A-
Z]{1}[\w]+(Exception|Error))\b" |
timechart span=1h count(exc_type) by
exc_type useother="f" usenull="f"
```

Architecture



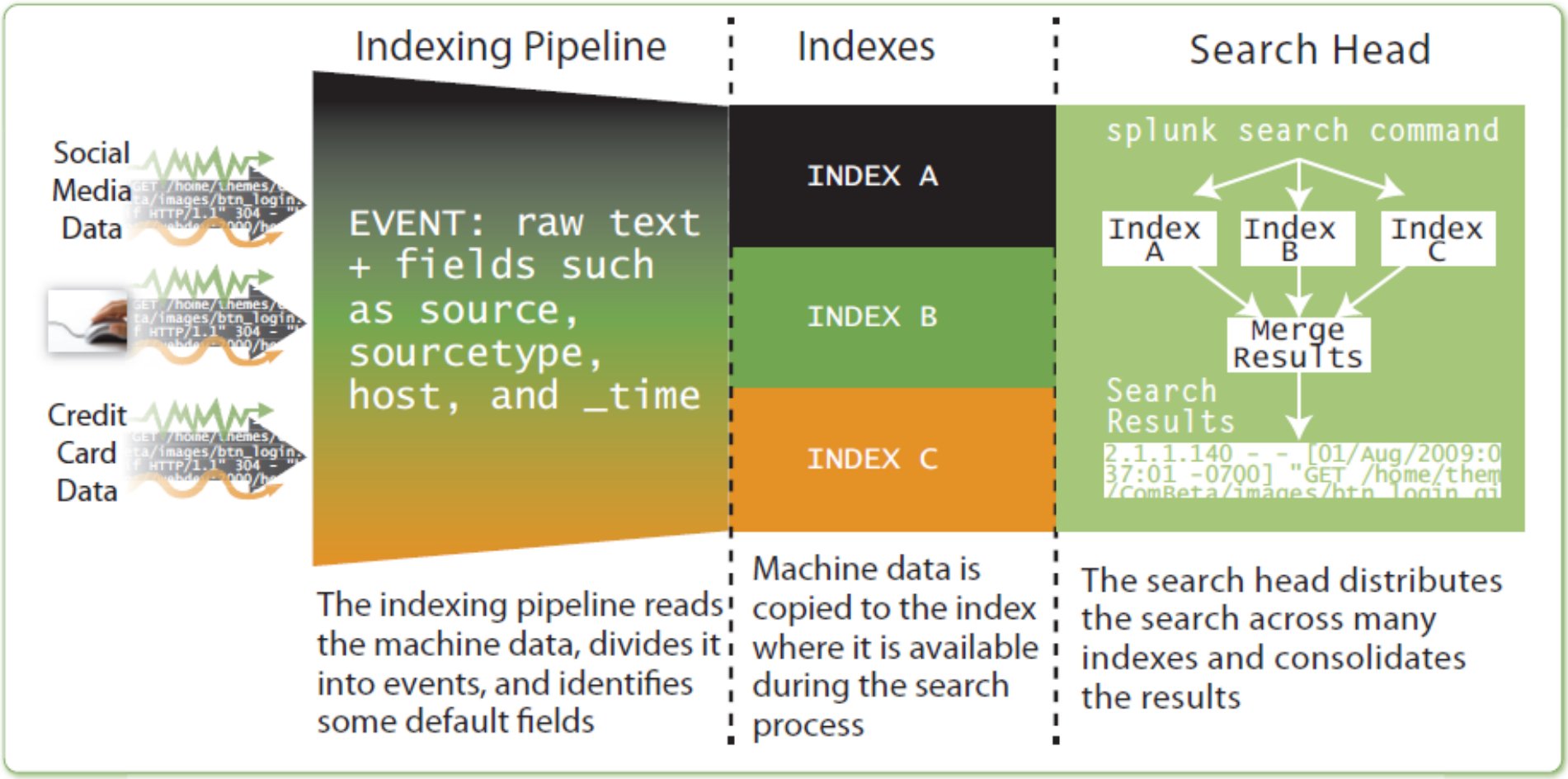
http://www.splunk.com/themes/splunk_com/img/assets/images/developers/splunkarchitecture.gif

Data Pipeline



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

Indexing



http://www.splunk.com/web_assets/v5/book/Exploring_Splunk.pdf

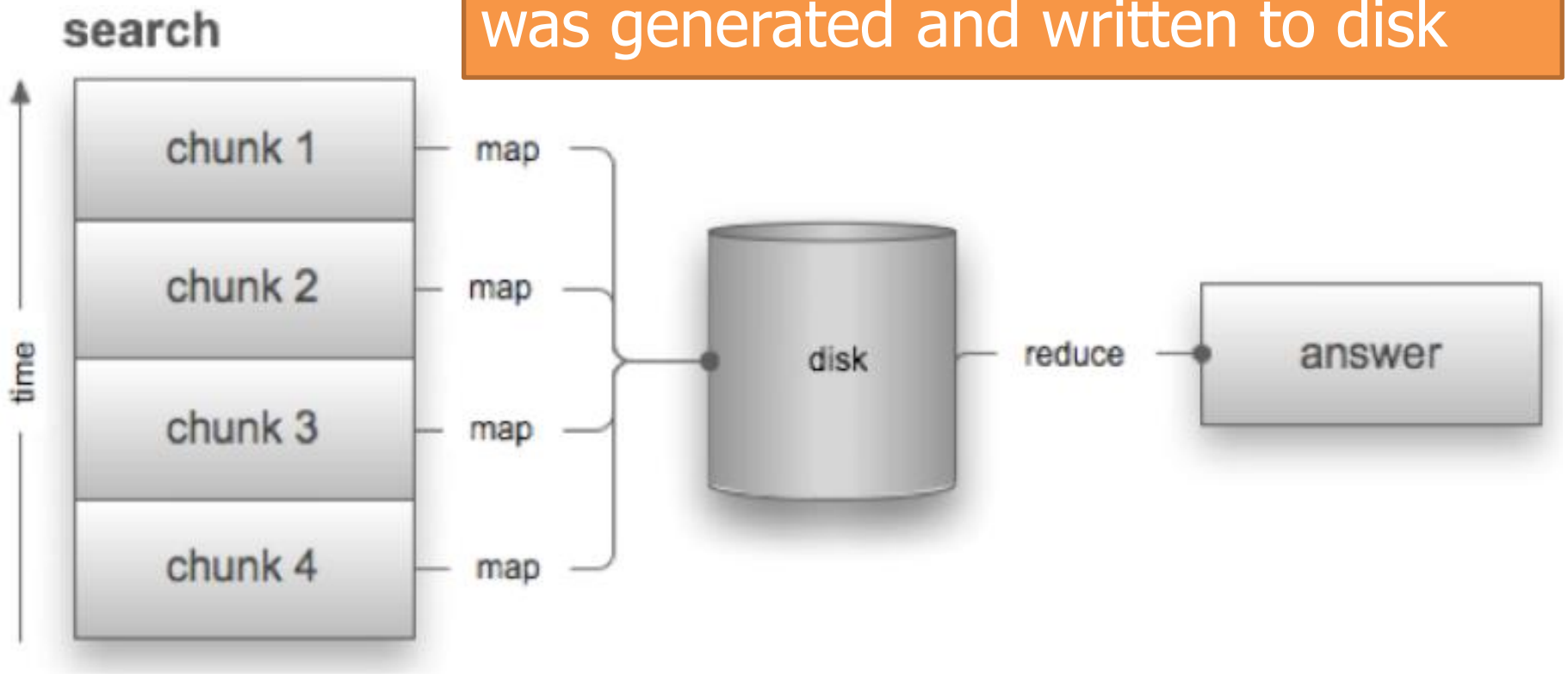
MapReduce (1)

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

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

Q & A

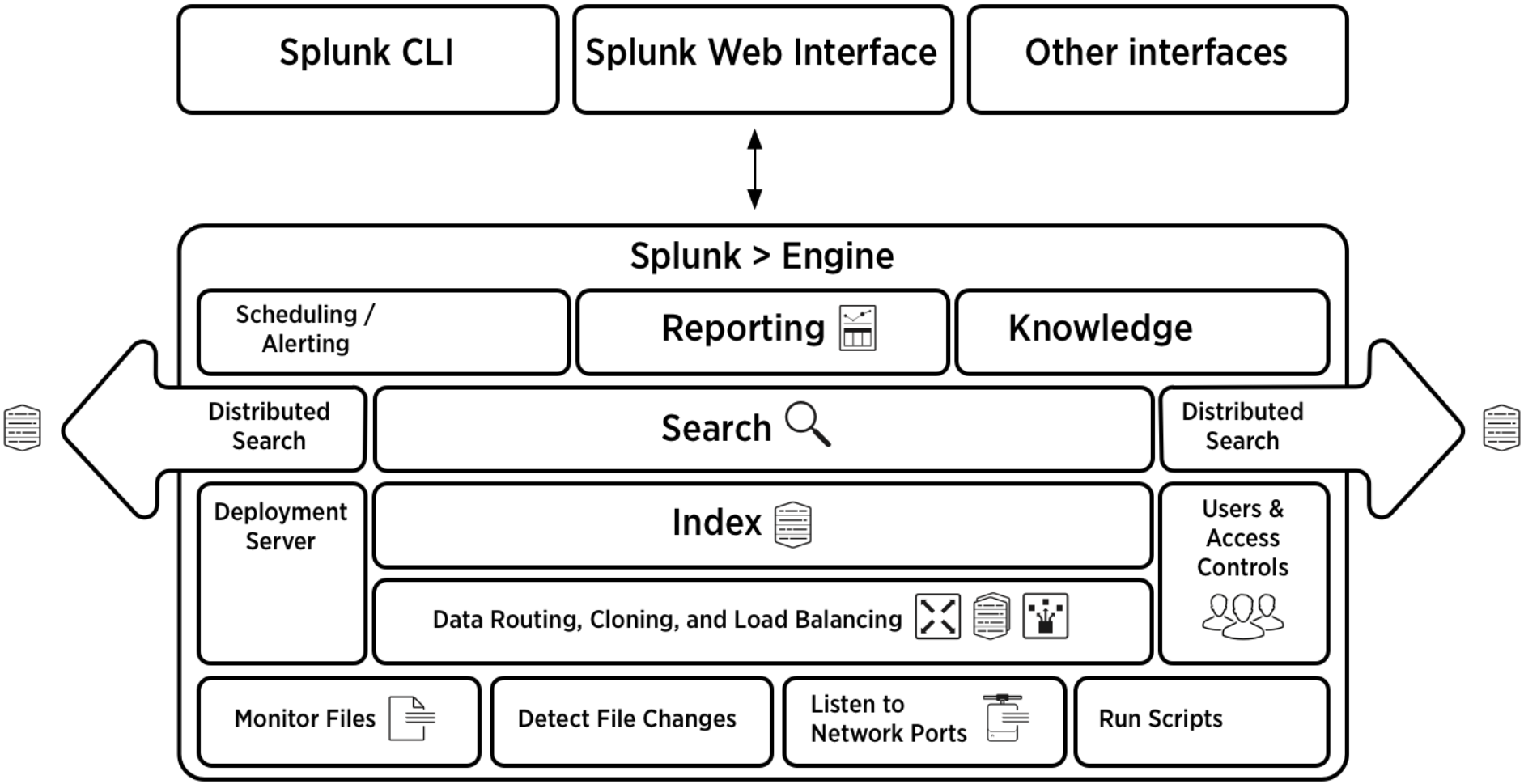
Use Cases

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

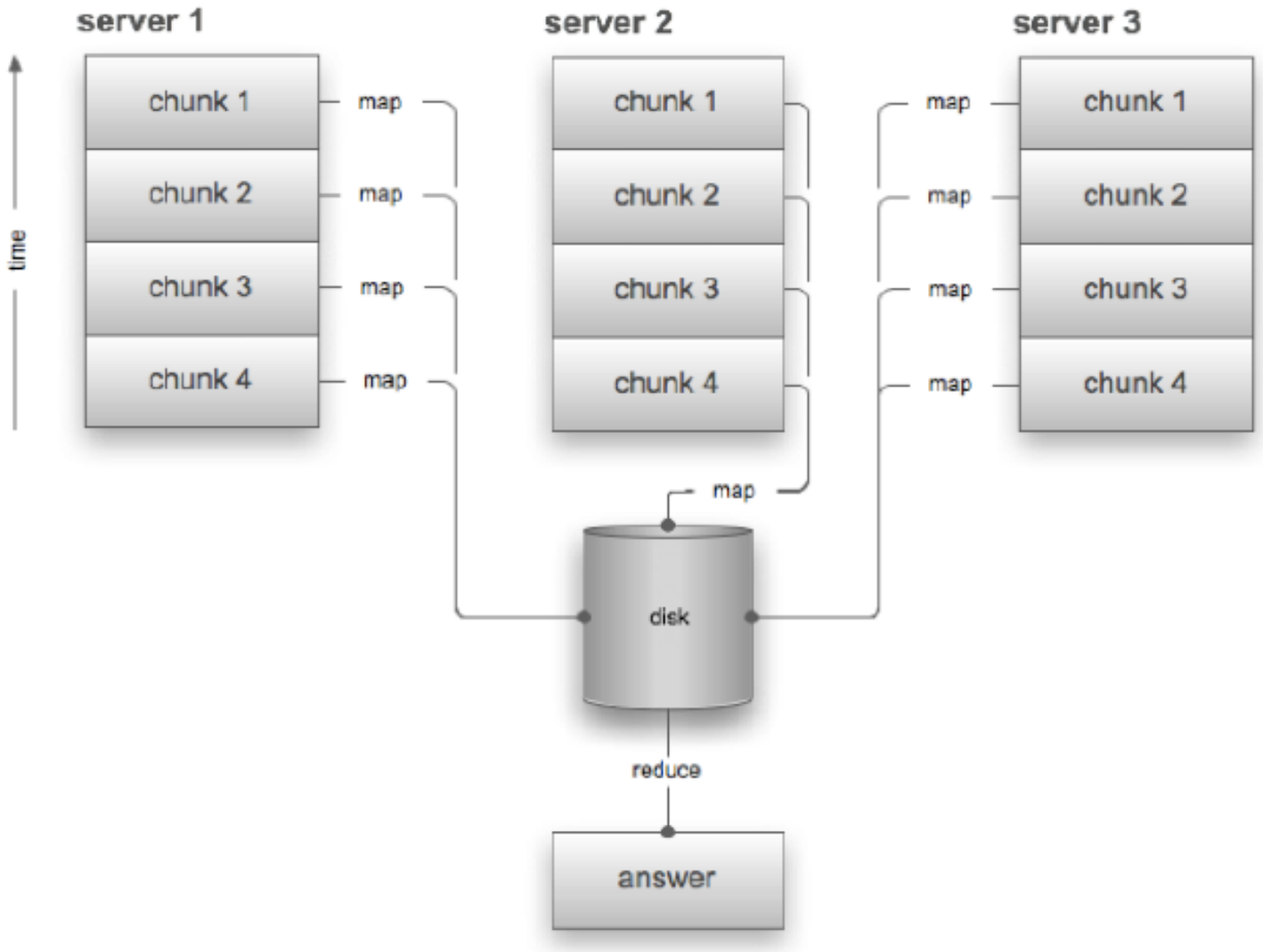
- **Pipelines** – single threads in the splunkd process with XML config, can pass data to one another via **queues**
- **Processors** – individual, reusable C/C++ functions that act on the stream of IT data passing through a pipeline

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

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

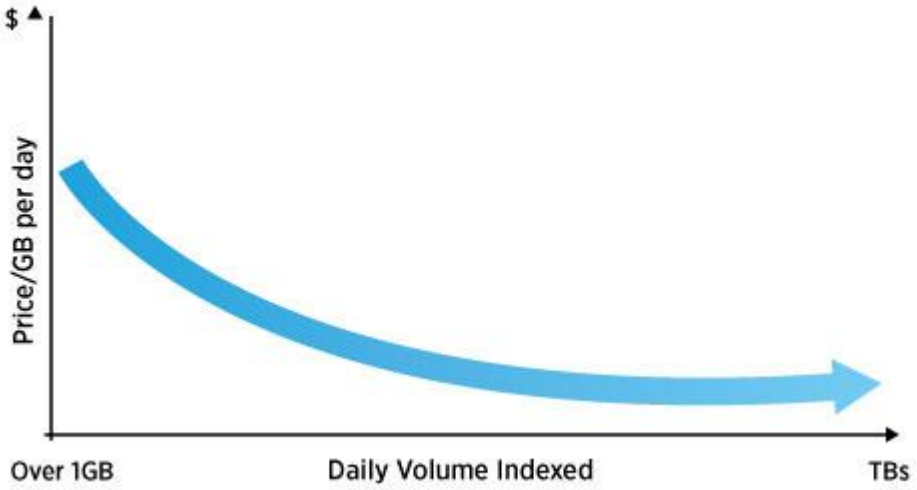
http://docs.splunk.com/Documentation/Splunk/6.1.4/Installation/Systemrequirements

Free vs Enterprise

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