DSS

Data & Storage Services

CERN IT Department

Introduction to splunk>

Alex Iribarren
IT-DSS

CERN IT Department
CH-1211
Geneva 23
Switzerland

CERN

What is Splunk?

Search engine for machine data

- Machine data: log files, configuration, monitoring metrics, etc.
- Splunk collects and indexes data to enable:

  Searching • Reporting • Correlating • Trending • Alerting

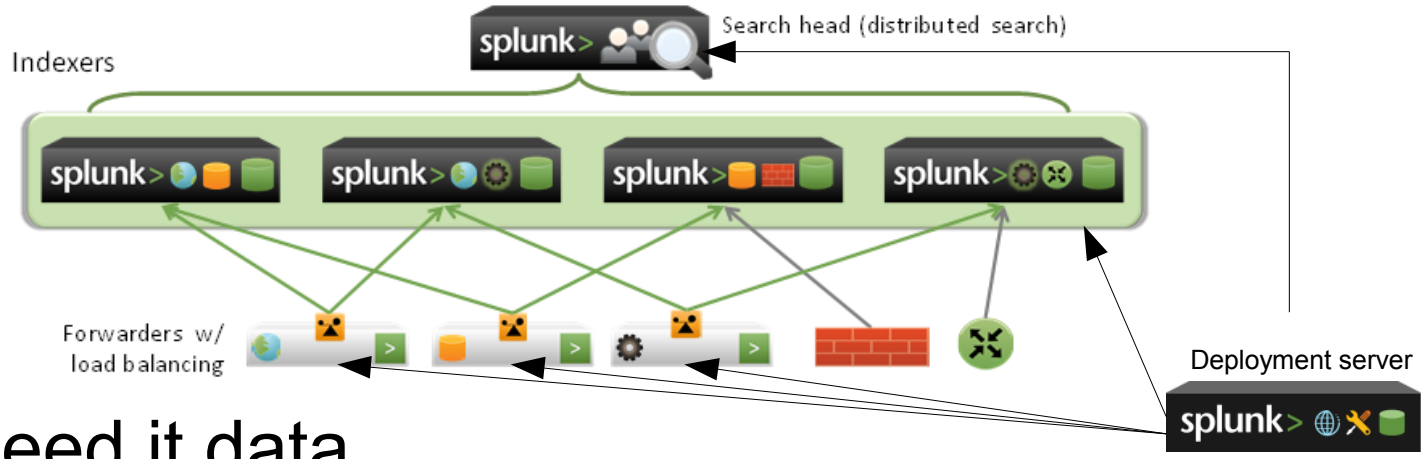
- Built to scale
  - Up to 100 GB/day on a single (modest) indexing server
  - Biggest installations are in the 4 PB/day range
- Enterprise-grade software
  - Great documentation, user community and support
  - SSO integration, LDAP support, access controls, etc.
  - Easy deployment and integration

CERN IT Department
CH-1211
Geneva 23
Switzerland

How does it work?

- 4 basic components:

  Indexer • Search head • Forwarder • Deployment server



- Feed it data

  local files/directories • TCP/UDP syslog • Splunk forwarders • custom scripts

- Splunk extracts/searches key-value pairs
  - You can help it out with regexps, training, etc.
  - Timestamps are essential

CERN IT Department
CH-1211
Geneva 23
Switzerland

Data examples

- TSM application logs

Nov 9 16:35:11 lxtsm610 dsmserv-TSM610: ANR0405I Session 22694 ended for administrator PROP (Linux86). (SESSION: 22694)

  - Easy data extraction through regular expressions:

    EXTRACT-admin = [Aa]dministrator (?<admin>[\w_\-&+]+)

- TSM accounting data

6,0,ADSM,11/09/2011,15:06:45,CERNDATA10,,WinNT,1,Tcp/Ip,1,0,0,0,0,0,0,0,0,0,3,0,0,0,0,5,0,0,0,0,2,2

- Lemon metrics

1351188709 eth0 270309624 0.18 6858092 0.30

- Data pulled from a DB by some scripts

[2011-11-09 16:40:55] name="db_status" event_id="db_status-1320853255" tsmserver="TSM65" db_used_mb="163840" db_utilization="82.6" db_last_backup="2011-11-09 06:31:24"

CERN IT Department
CH-1211
Geneva 23
Switzerland

Now... Search!

- Real-time and historical searches
- ~120 search commands, plus your own
  - Manipulate the data
  - Add more data
  - Correlate it
  - Choose how to present it
- Dashboards
- Reports (ie. dashboards as PDFs)
- Alerts

CERN IT Department
CH-1211
Geneva 23
Switzerland

TSM Alerts

CERN IT Department

From TSM Monitor <tsmms@cern.ch>

Subject **[TSMMS] [HIGH] 2 events**

05/22/2012 01:05 PM

To tsm-admin

HIGH: Detected 2 events of type 'Generic Alert: High'.

May 22 13:03:18 lxtsm068 dsmserv-TSMLIB61: ANR8351E 021: Mount request for volume I18133 has timed out.
May 22 13:03:18 lxtsm068 dsmserv-TSMLIB61: ANR8313E Volume I18133 is not present in library IBMLIB0. (SESSION: 6009913)
May 22 13:03:18 lxtsm068 dsmserv-TSMLIB61: ANR9790W Request to mount volume I18133 for library client TSM610 failed. (SESSION: 6009913)
May 22 13:03:18 lxtsm610 dsmserv-TSM610: ANR1402W Mount request denied for volume I18133 - volume unavailable. (SESSION: 766412, PROCESS: 2028)
May 22 13:03:18 lxtsm610 dsmserv-TSM610: ANR1410W Access mode for volume I18133 now set to 'unavailable'. (SESSION: 766412, PROCESS: 2028)
May 22 13:03:18 lxtsm610 dsmserv-TSM610: ANR1081E Space reclamation is ended for volume I18133. Storage media is inaccessible. (SESSION: 766412, PROCESS: 2028)
May 22 13:03:18 lxtsm610 dsmserv-TSM610: ANR1893E Process 2028 for SPACE RECLAMATION completed with a completion state of FAILURE. (SESSION: 766412, PROCESS: 2028)
May 22 13:03:18 lxtsm610 dsmserv-TSM610: ANR1463E RUN: Command script DAILY_HK completed in error. (SESSION: 766412, PROCESS: 2028)
May 22 13:03:18 lxtsm610 dsmserv-TSM610: ANR2752E Scheduled command DAILY_HK failed. (SESSION: 766412, PROCESS: 2028)

See event c5432511677d07e91e5b1444ee68d75a and context

May 22 13:00:13 lxtsm052 dsmserv-TSM52: ANR1402W Mount request denied for volume I21086 - volume unavailable. (SESSION: 211740, PROCESS: 67)
May 22 13:00:13 lxtsm052 dsmserv-TSM52: ANR1410W Access mode for volume I21086 now set to 'unavailable'. (SESSION: 211740, PROCESS: 67)
May 22 13:00:13 lxtsm068 dsmserv-TSMLIB61: ANR8300E I/O error on library IBMLIB0 (OP=C0106C03, CC=314, KEY=05, ASC=3B, ASCQ=0E, SENSE= 70.00.05.00.00.00.00.0A.00.00.00.00.3B.0E.00.C0.00.04., Description=The source slot or drive was empty in an attempt to move a volume). Refer to the &STGM; documentation on I/O error code descriptions. (SESSION: 6009945)
May 22 13:00:13 lxtsm068 dsmserv-TSMLIB61: ANR8312E Volume I21086 could not be located in library IBMLIB0. (SESSION: 6009945)
May 22 13:00:13 lxtsm068 dsmserv-TSMLIB61: ANR8358E Audit operation is required for library IBMLIB0. (SESSION: 6009945)
May 22 13:00:13 lxtsm068 dsmserv-TSMLIB61: ANR8381E 3592 volume I21086 could not be mounted in drive IBM0409 (/dev/ibm0409). (SESSION: 6009945)
May 22 13:00:13 lxtsm068 dsmserv-TSMLIB61: ANR9790W Request to mount volume I21086 for library client TSM52 failed. (SESSION: 6009945)

See event 54288f93634c267a03738b219735df21 and context

- Configure alert

Lookups FTW!

- Add fields to your events with data from external sources
  - CSV files or external scripts
  - Static or time-based

```
"TSMSERVER","BUILDING","SERVER_TYPE","SERVER_STATUS","SERVER_VERSION","HOST"
"TSM62","613","user","production","5.5","lxtsm062.cern.ch"
"TSM58","513","user","retired","5.5","lxtsm059.cern.ch"
"TSM54","513","user","production","5.5","lxtsm054.cern.ch"
"TSM61","613","user","production","5.5","lxtsm006.cern.ch"
"TSM51","513","user","production","5.5","lxtsm057.cern.ch"
"TSM91","513","user","production","5.5","lxtsm009.cern.ch"
"TSM53","513","user","production","5.5","lxtsm053.cern.ch"
"TSM64","613","user","production","5.5","lxtsm064.cern.ch"
"TSM65","613","user","production","5.5","lxtsm067.cern.ch"
```

  - Gives you very powerful abstraction:
    - “Tape errors affecting user backups by client OS”
    - “Daily traffic by department and group”

CERN IT Department
CH-1211
Geneva 23
Switzerland

Demo

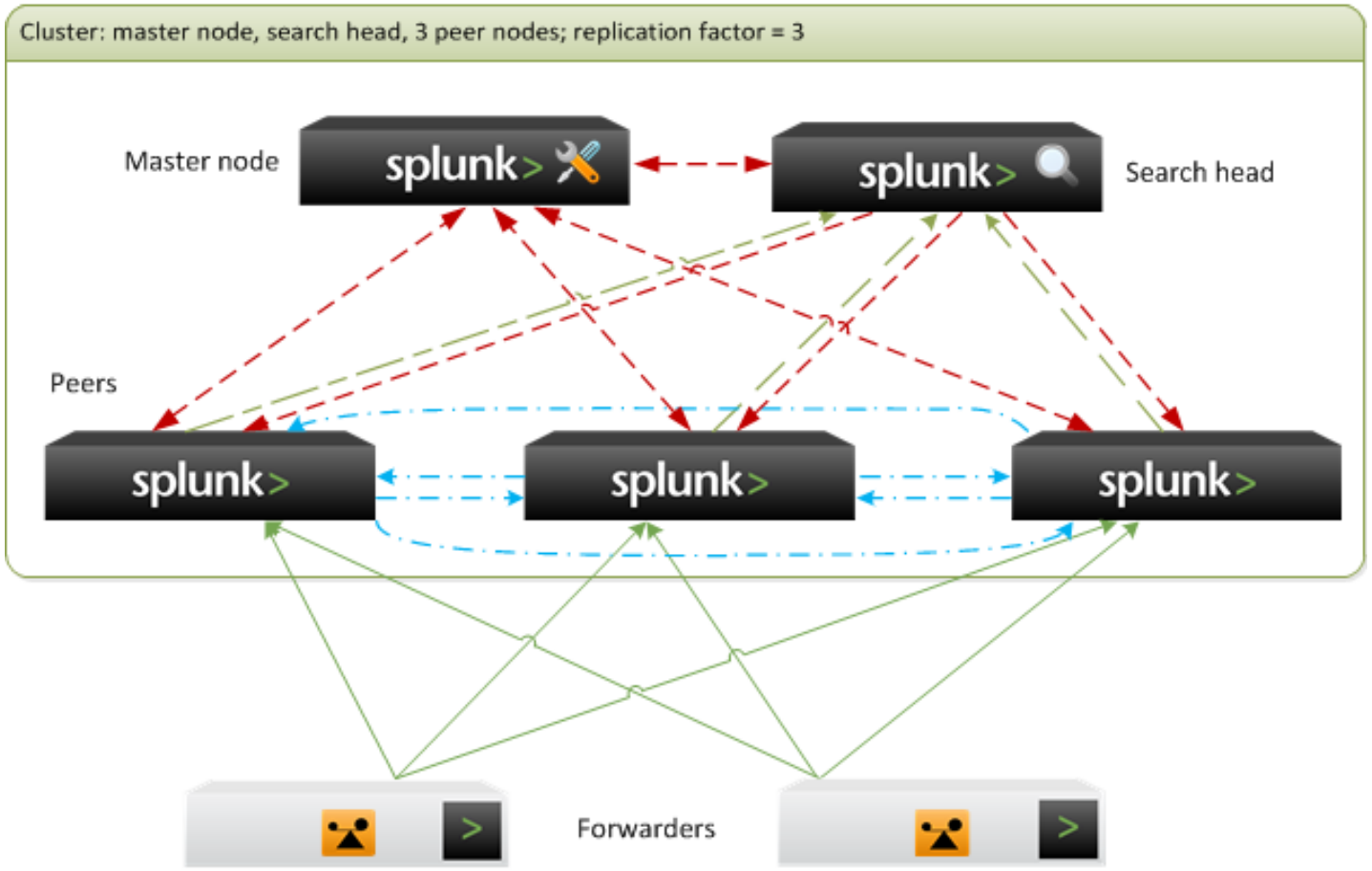
CERN IT
Department
CH-1211
Geneva 23
Switzerland

Splunk 5

- Many enhancements
  - Dynamic drilldowns, enhancements to charts, integrated SDKs, PDF reporting, etc.
- Modular inputs
  - Easier way of getting custom data into Splunk
- Report acceleration
  - Super-easy speedups of queries by precomputing results
- Index replication, aka. Splunk clusters

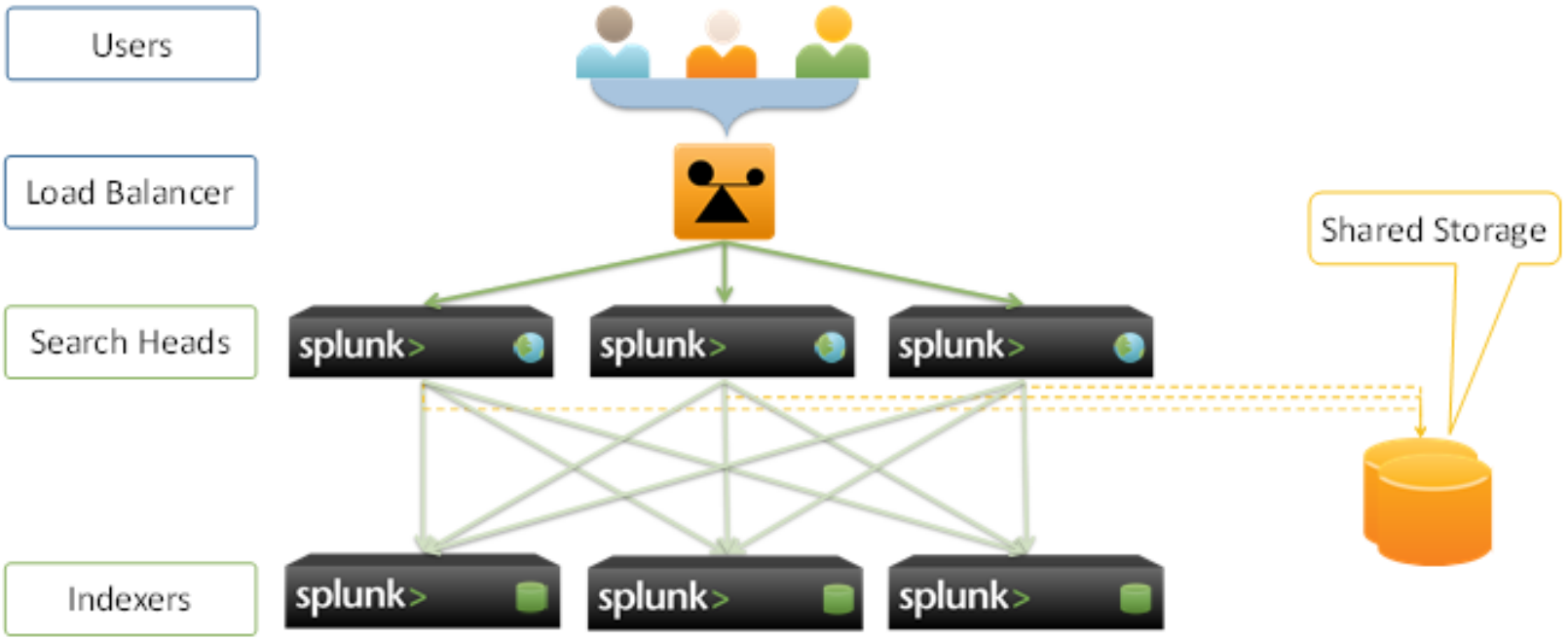
CERN IT Department
CH-1211
Geneva 23
Switzerland

Splunk 5: Clusters



CERN IT Department
CH-1211
Geneva 23
Switzerland

Scaling up



CERN IT Department
CH-1211
Geneva 23
Switzerland

What if?

- What if there was a central Splunk service?
  - Clustered instance, load-balanced search heads
- Like an Oracle DB:
  - User requests a “project”, gets a quota
  - User sends data, writes searches and configures dashboards, shares them with colleagues
  - Service managers handle the infrastructure (hardware, maintenance, backups, etc.)
- Interest from IT and experiments
  - Many details to sort out

CERN IT Department
CH-1211
Geneva 23
Switzerland

Thanks!

Splunk user's e-group:

- **splunk-users@cern.ch**

Exchange experiences, help each other out, keeping in touch with developments, etc.

CERN IT Department
CH-1211
Geneva 23
Switzerland