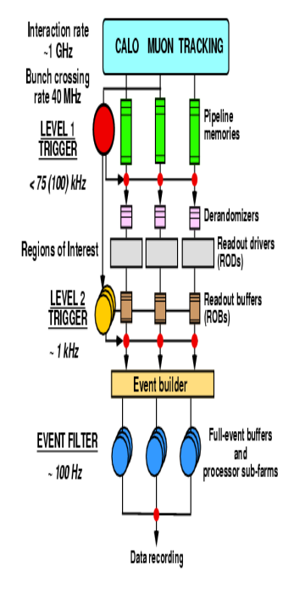
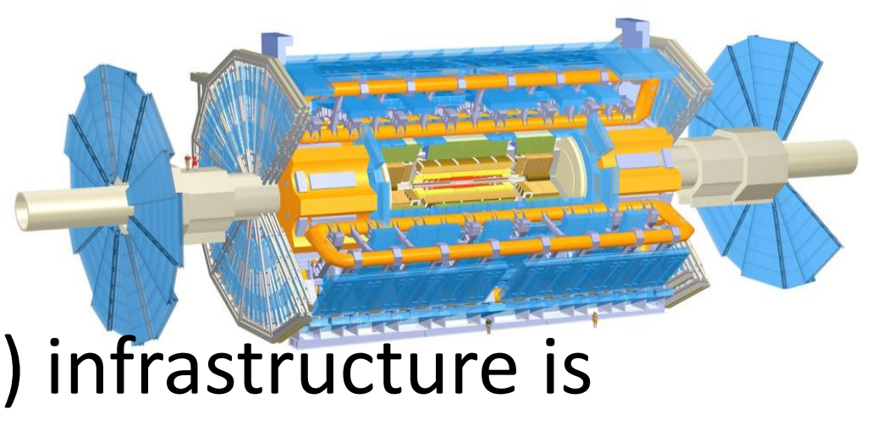


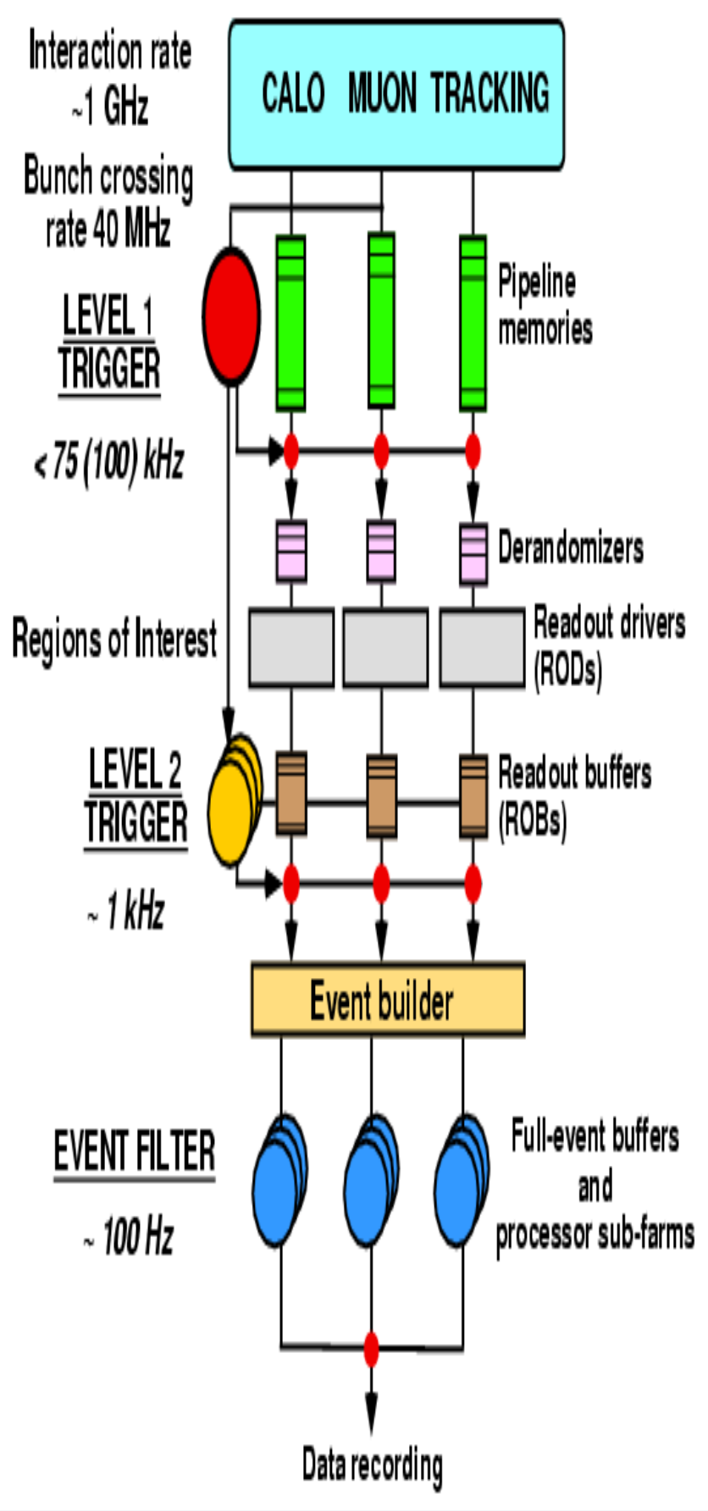
Splunk for ATLAS TDAQ: towards a generic framework for operational monitoring

Abdul Rahim Umar
Summer Student 2012/ASP 2014
CERN PH-ADT-CC group

ATLAS Trigger and DAQ

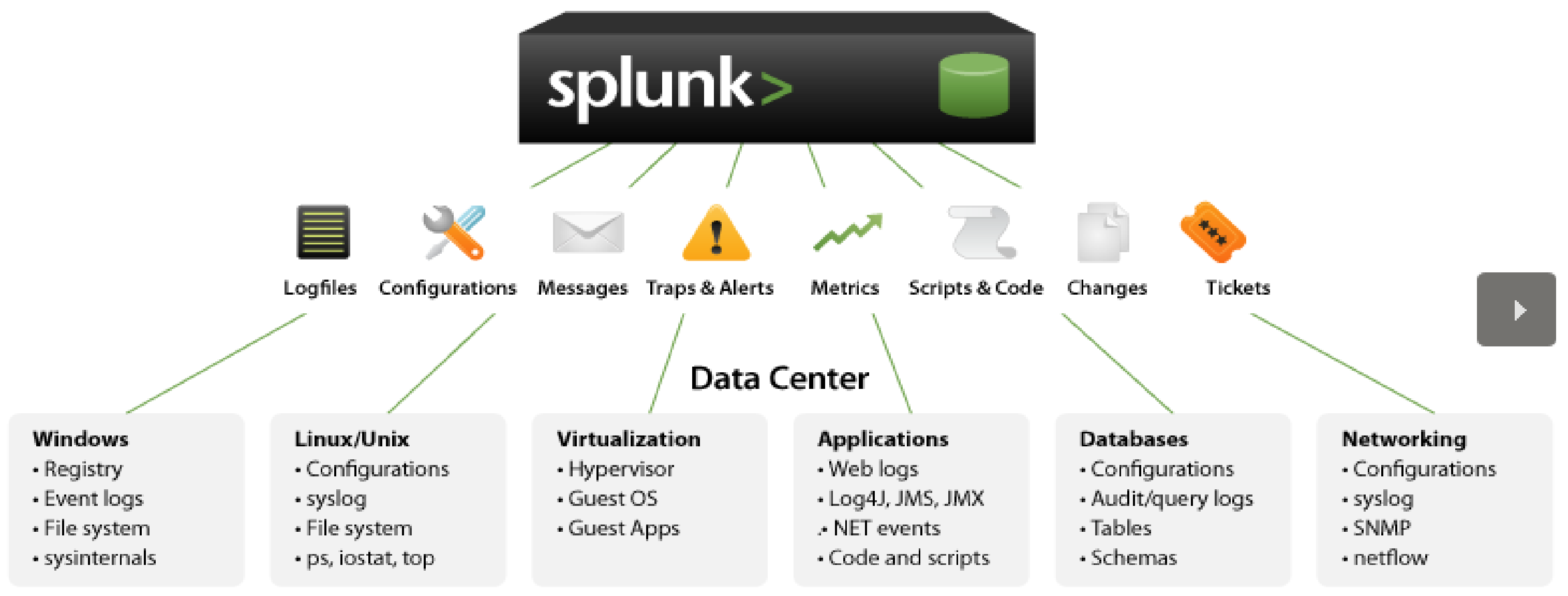


The ATLAS Trigger and Data Acquisition (TDAQ) infrastructure is responsible for filtering and transferring ATLAS experimental data from detectors to mass storage systems. It relies on a large, distributed computing environment composed by thousands of software applications running concurrently. In such a complex environment, information sharing is fundamental for controlling applications behaviour, error reporting and operational monitoring.



Operational monitoring

During data taking runs, the streams of messages sent by each of the 20000 applications and data published via information services are constantly monitored by experts to verify correctness of running operations and to understand problematic situations. To simplify and improve system analysis and errors detection tasks, this project evaluates the Splunk framework to collect, correlate and visualize effectively this real time flow of information. The project is composed by a set of gatherer to collect the monitoring data and by the Splunk tool to performs aggregation, processing and filtering of real time data stream and computes statistical correlation on sliding windows of time.



Powerful and versatile IT search software that takes the pain out of tracking and utilizing the information in your data center.

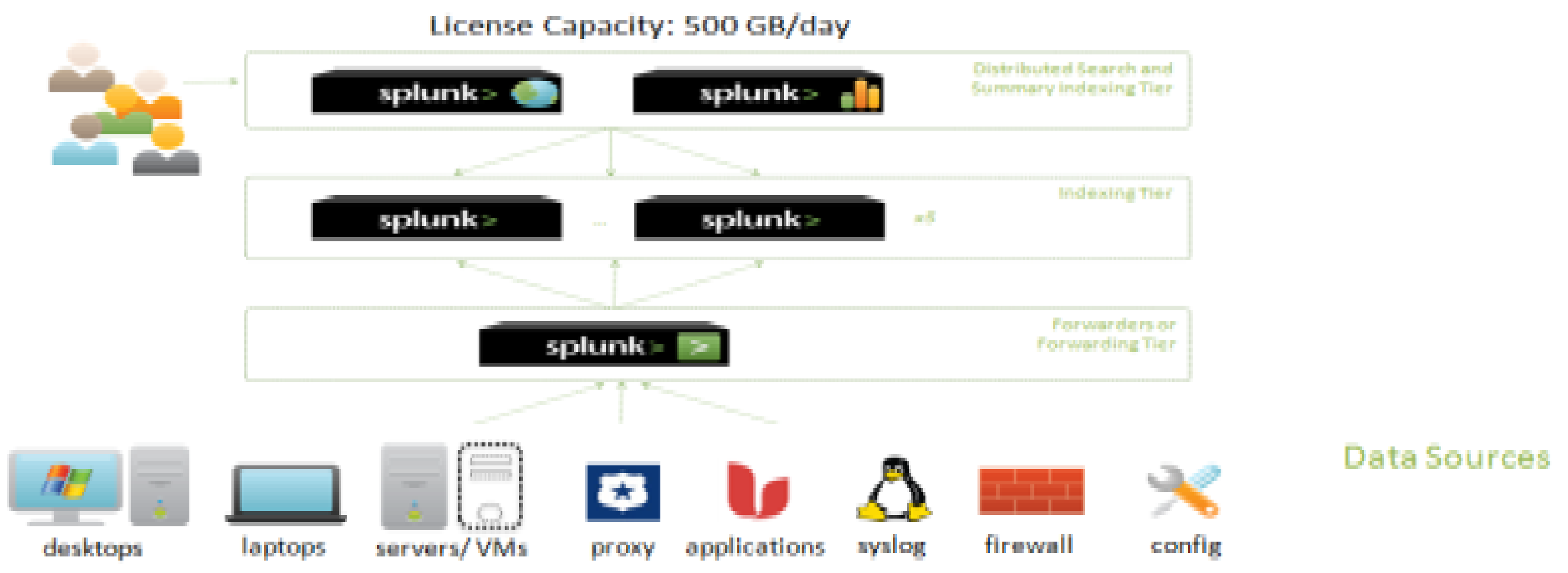
Splunk's Solution

- The core of Splunk's software is a proprietary machine data engine, comprised of collection, indexing, search and data management capabilities. The software can collect and index terabytes of information daily, irrespective of format or source.
- The machine data engine uses an innovative data architecture that enables dynamic, schema creation on the fly, allowing users to run queries on data without having to understand the structure of the data prior to collection and indexing.

Message Analysis

Message analysis is fundamental for controlling applications behavior, error reporting and operational monitoring.

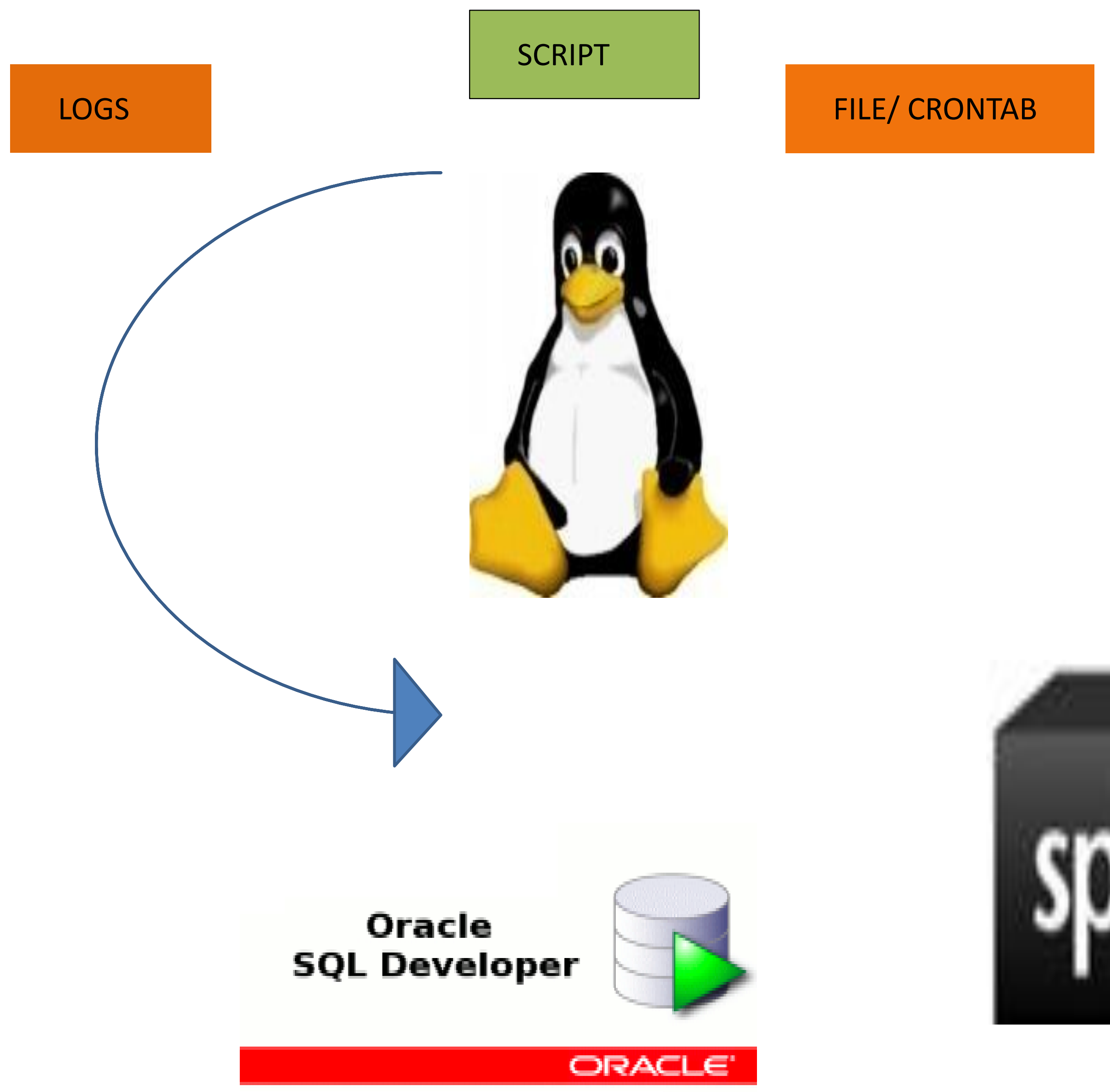
Splunk architecture



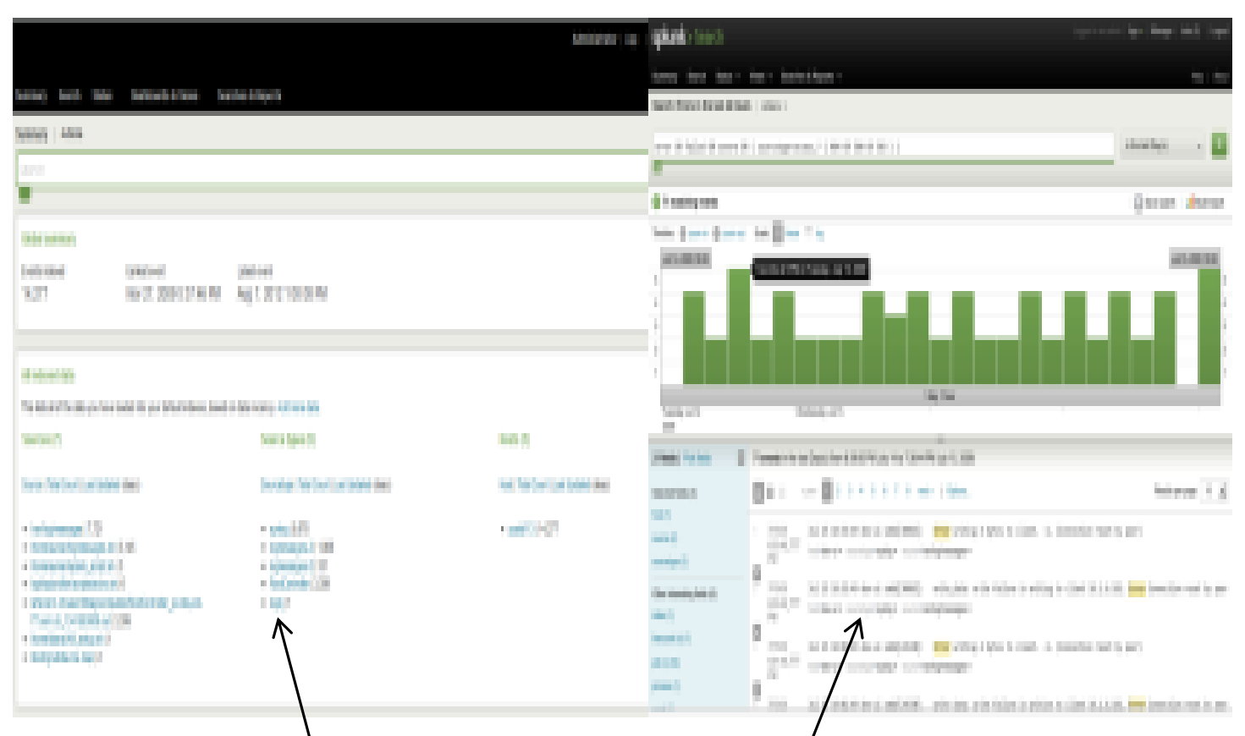
The TDAQ Computing Infrastructure

>12.000 cores
>20.000 software applications running on SLC linux
>3 independent GbE networks
> 100.000 channels

Data gatherer: linking TDAQ with Splunk via custom plug-in



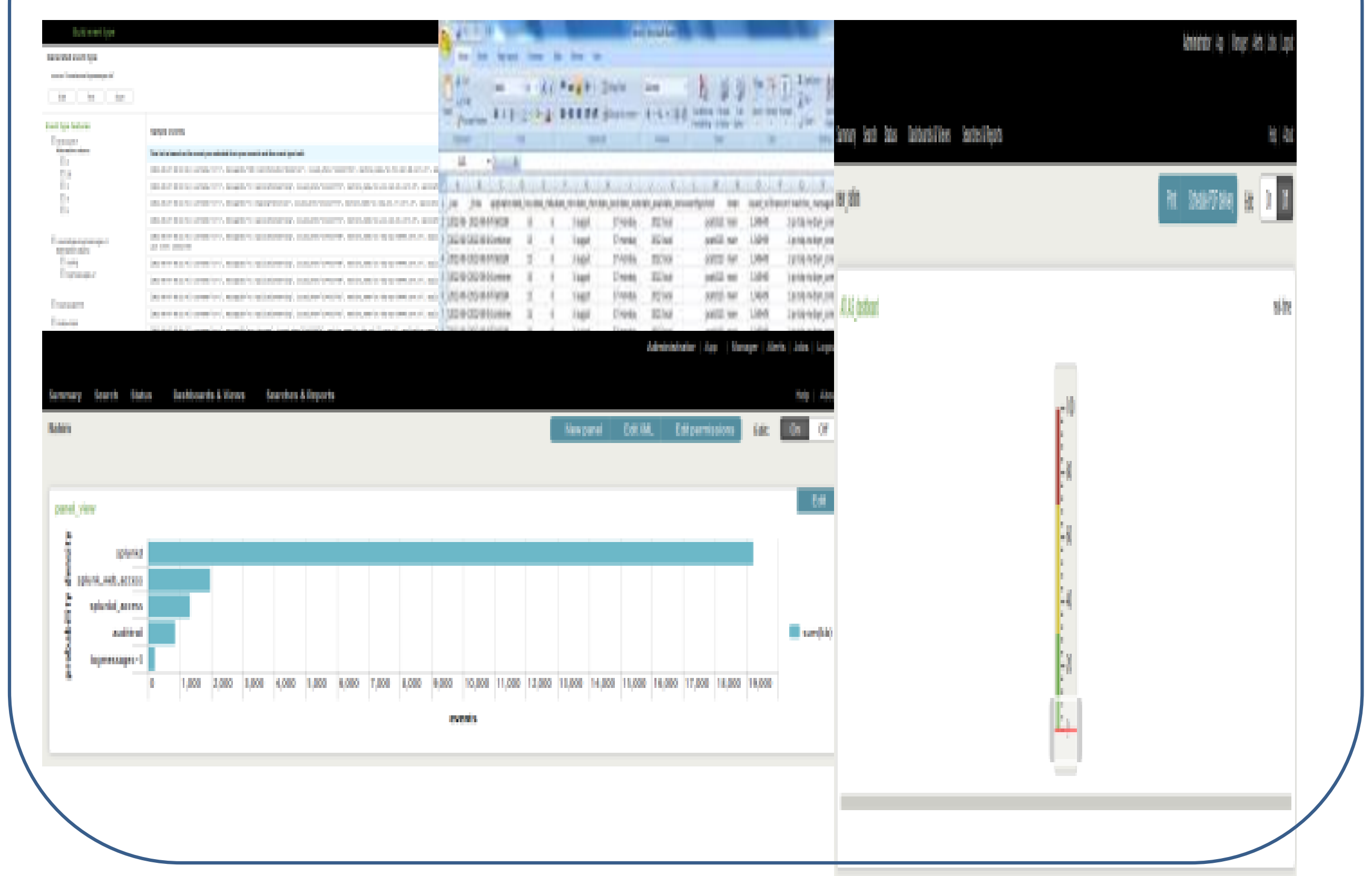
The Splunk framework



- Splunk indexes the operational data gathered by the TDAQ plug-in
- It provides users and experts with a simple SQL-like syntax to query collected data
- Many visualization options are supported to effectively present operational data

Difficulties:
To follow flow of messages
To extract meaningful information
To detect global behavior
What we can do (splunk):
To help system analysis and error detection
To visualize the flow of messages effectively
Show real time information as well as historical data
Challenge: *Data visualization*

Results



Conclusion

This work shows that the Splunk framework is a viable and effective solution to build a generic framework to monitor TDAQ operational data. The plug-in approach allow to easily gather information form custom monitoring services, and the visualization option can be easily customized with a dashboard approach. Acceptable performance as long as the indexers are not currently heavily loaded and do not have more than a few concurrent real-time searches.