ELVis

Extensible Log Visualization

christopher.humphries @ inria.fr
nicolas.prigent @ supelec.fr
christophe.bidan @ supelec.fr
frederic.majorczyk @ supelec.fr

Hello.

My name is Christopher Humphries.

I'm a PhD student from Rennes, France.

Working at INRIA/Supélec, in team CIDre.

On a research grant from DGA-MI.

CIDre

We try to solve security problems.

INTRUSION DETECTION

ADHOC NETWORKS

PRIVACY

I work on security visualization.

Why visualize ?

Generally

Help ourselves.

Make more sense of data.

Understand the big picture.

Regain control.

In security

Bridge a gap.

Manual analysis is slow.

Automatic analysis is dumb.

However…

Extra knowledge required!

statistics	*for the numbers*
design	*for the colors*
psychology	*for brain compatibility*

Solution

Move knowledge into software.

“Pshaw! It's been done before!”

Autovis

Statistically automatic and unopinionated visualization.

Tableau

Assisted and facilitated creation of general visualizations.

ELVis

Parses logs

Uses security semantics

Assists exploration by selecting and matching visualisations

Log Organization

Each log has a specific format

APACHE STANDARD SNORT …

Log entries in one file have the same fields

TIME SOURCE IP DESTINATION PORT …

Every log field has types

ORDINAL CATEGORICAL TIME GEOGRAPHICAL …

Log Augmentation

Log datasets are augmented

Horizontally fields of certain types spawn extra fields

IP spawns GEOLOCATION(IP)

Vertically statistical summary for each field

MAX MIN DISTRIBUTION ...

Automated Selection of Representations

Informed decision based on stats and types.

NOMINAL fields use distributions *Pie charts, bar charts*

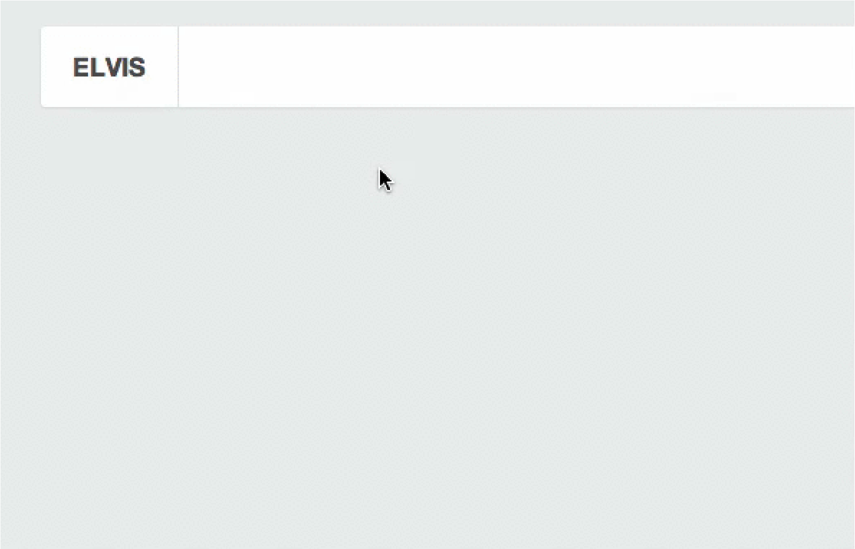
TIME requires trend visualizations *Line charts, gantt charts*

GEOGRAPHIC fields require spatial charts *Maps, real world layout*

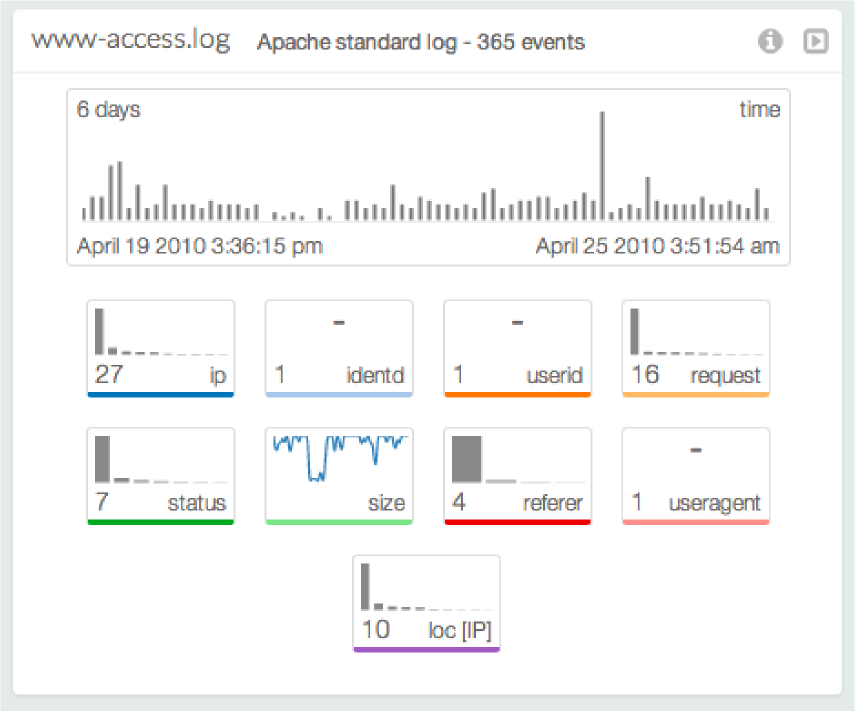
Log Acquisition

Logs are parsed using the right format

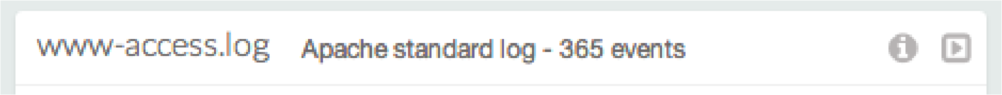
Files can be dragged in straight from the system



Summary View



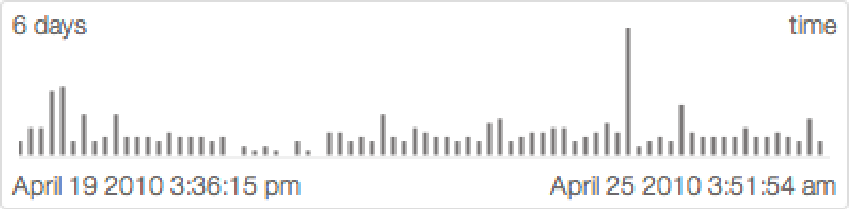
Top Bar



www-access.log Apache standard log - 365 events

Dataset name, brief info, tools.

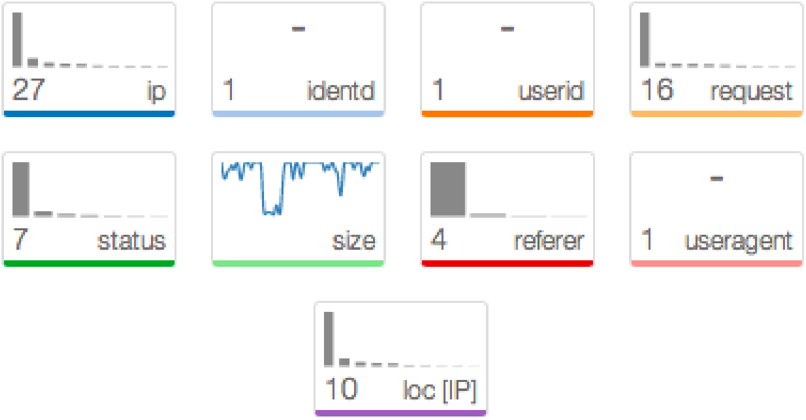
Key Field



Basic visualization, filtering is already available.

The key field is time so distribution of events is displayed.

Other Fields



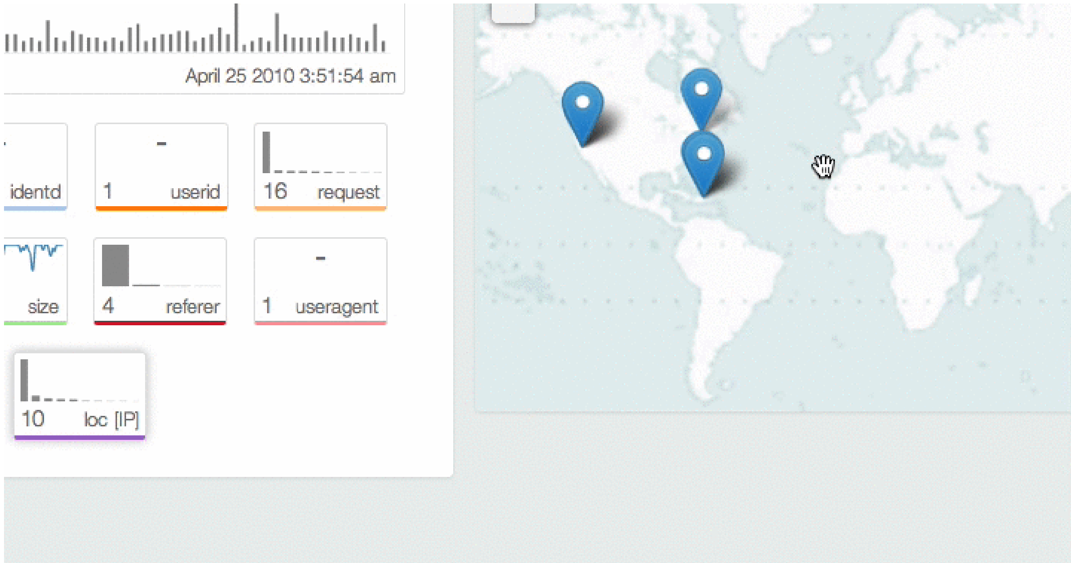
Displayed as small multiples according to type and stats.

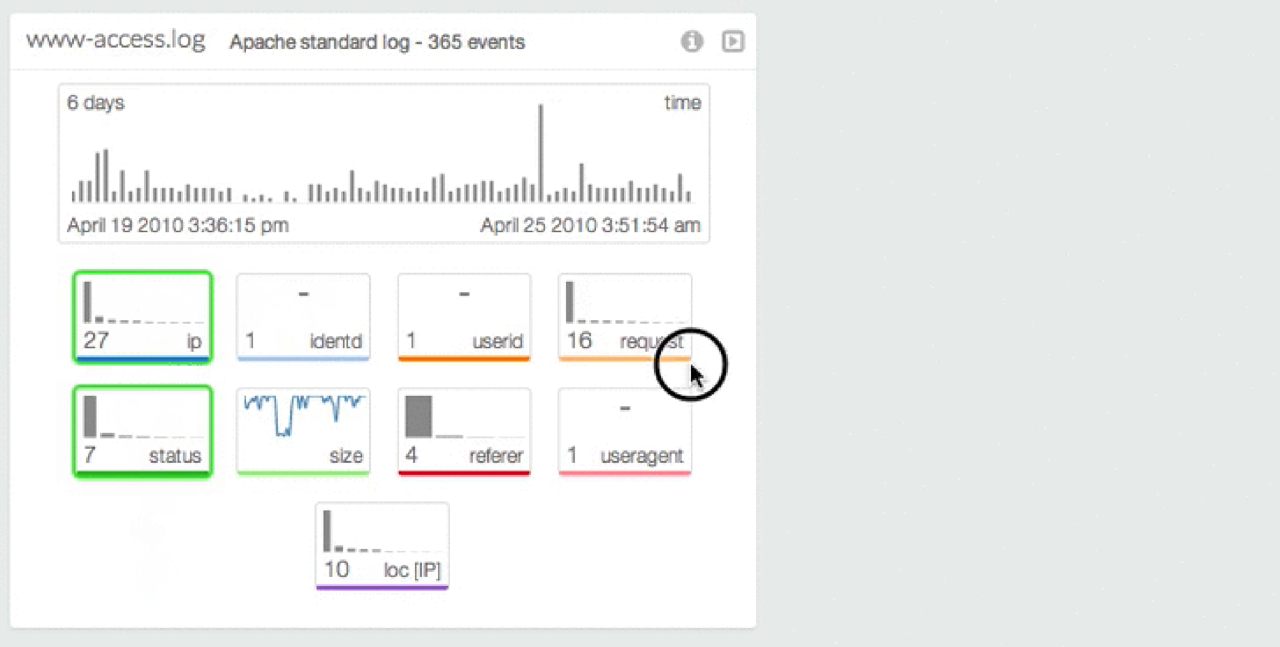
Distribution histogram for NOMINAL fields.

Line chart for trends in CARDINAL fields.

User Interaction

Fields are selected and dragged to construct visualizations




www-access.log Apache standard log - 365 events

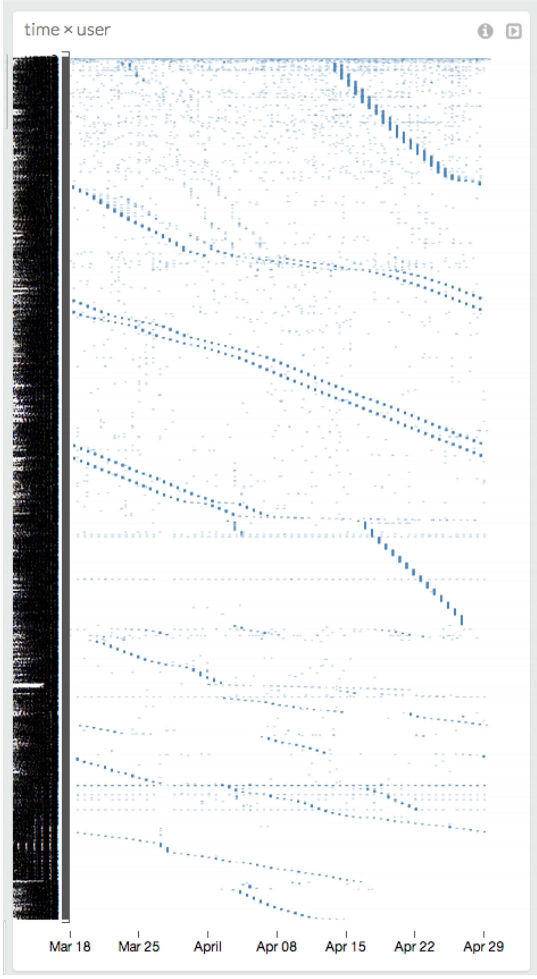
Testing

Exploring the HoneyViz dataset.

Patterns found! Ideas formed!

Promising…

Some logs were a strain to load...


time × user
Mar 18
Mar 25
April
Apr 08
Apr 15
Apr 22
Apr 29

Roadmap

User Experience

Better exploration.

Brushing and Filtering.

Chained visualizations.

Scalability

Larger and multiple datasets.

Load more data.

Compare and reference datasets.

Server integration ***Splunk, Hive?***

Sharing and collaboration.

Recording

Take notes, save configurations.

Record datamining scenarios.

Save effective dashboard arrangements.

Inform the datasets.

Help generate reports.

Technical Stuff

Web based

D3.js • Miso Chart • Miso Dataset

Server prototypes in Node.js.

Thank you.

Questions?