ENTRADA: TLD Traffic Analysis goes Open Source

Moritz Müller | 5th CENTR Jamboree - 17 May 2016, Brussels, Belgium
DNS Data @SIDN

> 3.1 million distinct resolvers

> 1.3 billion queries daily

> 300 GB of PCAP data daily
ENTRADA
ENhanced Top-Level Domain Resilience through Advanced Data Analysis

- **Goal**: data-driven improved security & stability of .nl

- **Problem**: Existing solutions do not work well with large datasets and have limited analytical capabilities.
ENTRADA Architecture

**SQL on Hadoop (Impala + Parquet +HDFS)**

**Main components**

- Data sources
- Platform
- Applications and services
- Privacy framework
Workflow

Query data available for analysis within 10 minutes
SQL on Hadoop
Best fit for our requirements

- Hadoop Node N
  - IMPALA
  - PARQUET
- Hadoop Node N+1
  - IMPALA
  - PARQUET
- Hadoop Node N+2
  - IMPALA
  - PARQUET

HDFS
DEMO

• HUE Web Interface for basic DNS traffic exploration

• Python Notebook with Pandas Data Frames: Impact of TTL change at .nl
Use Cases
Focussed on increasing the security and stability of .nl

- Visualize DNS patterns
- Statistics (stats.sidnlabs.nl)
- Scientific research
- Support for operators
- Real-time Phishing detection
- Detect botnet infections
It’s open source!

- Since January 2016
- Project site: entrada.sidnlabs.nl
- GitHub: github.com/SIDN/ENTRADA/
It's open source!

- Since January 2016
- Project site: entrada.sidnlabs.nl
- GitHub: github.com/SIDN/ENTRADA/

Moritz Müller
moritz.muller@sidn.nl
@dhr_moe

Questions? Feedback?