ENTRADA2

Maarten Wullink

ICANN DNS Symposium

Sep 25, 2024 – Santa Marta, Columbia (Remote)

SIDN LABS

Introduction

ENTRADA: **EN**hanced **T**op-level domain **R**esilience through **A**dvanced **D**ata **A**nalysis

Enables analysis of large volumes of DNS pcap data, by:

- Converting pcap data to a column-oriented data format.
- Matching and enrichment of DNS queries
- Automated handling of workflow steps

SIDN LABS

Introduction

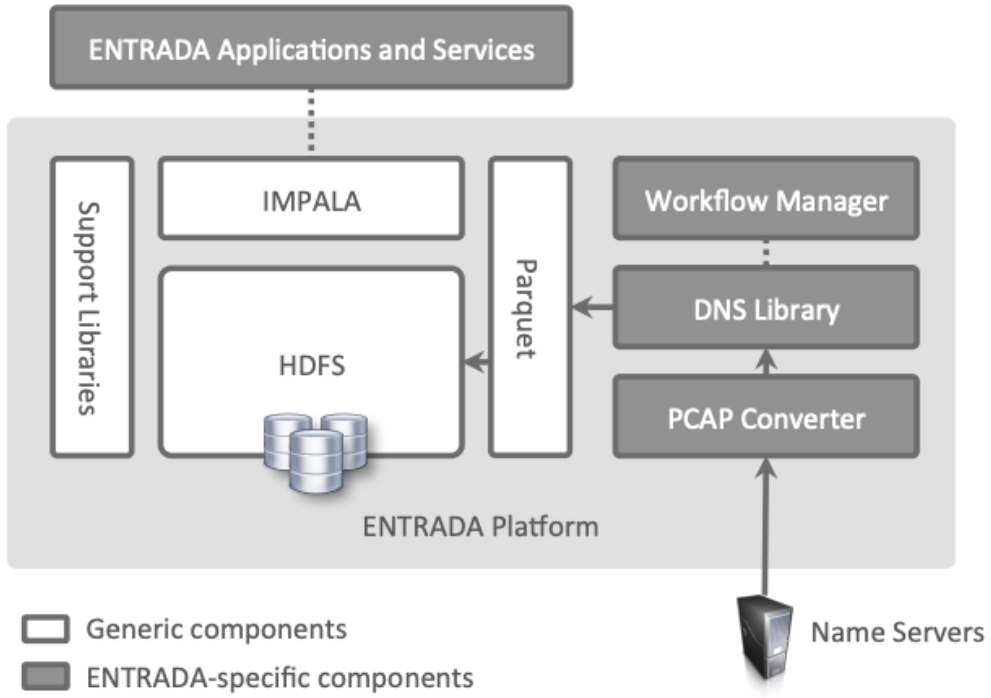
First released as open source 8 years ago, based on:

- Hadoop ecosystem
- Impala + Spark query engines

Mostly used by:

- Domain name (TLD) registry operators
- Researchers

Previous Architecture



Previous Architecture

Based on Hadoop ecosystem, obsoleted by newer technologies

- **HDFS** (Distribute storage) is replaced by s3 compatible Object Storage (AWS, MinIO)
- **YARN** (Process scheduling) is replaced by Kubernetes
- **Kerberos** is replaced by pluggable security providers

ENTRADA2

A new project, based on modern data architecture and cloud native technology

Improvements in area of:

- Scalability (cloud scaling)
- Performance (smaller data footprint)
- Usability (Easier to deploy and maintain than Hadoop)
- Security

SIDN LABS

ENTRADA2

Major changes:

- No longer depends on Hadoop
- Optimized Parquet data output, 40% size reduction
- Clustering support (Kubernetes, AWS and Docker)
- Use of a lakehouse architecture using open table format (Iceberg)
- Support for any JDBC compatible SQL query engine
- Data model is NOT compatible with original ENTRADA

New Functionality

- Optionally include RDATA of DNS response in output
- Support for decoding Extended rcode
- Support for Extended DNS Errors

SIDN LABS

ENTRADA2

Multiple deployment modes:

- Kubernetes (on-premise, cloud)
- AWS
- Docker

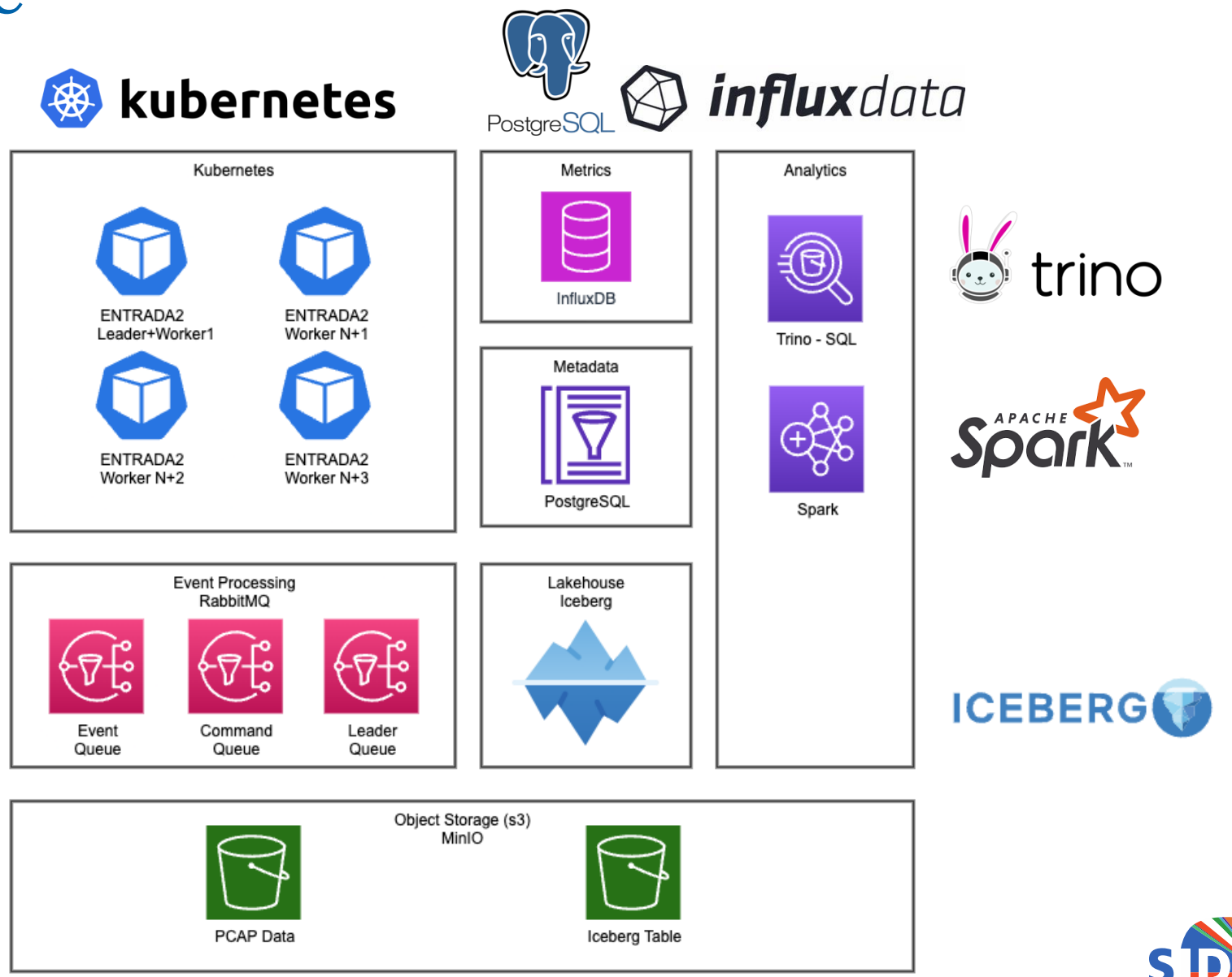
Easily deploy a cluster using multiple containers

- Auto create required resources (s3, queues, data table)
- Name server site no longer linked to specific container

SIDN LABS

New Architecture

Kubernetes mode



SDN LABS

Lakehouse Architecture

Warehouse-like capabilities on top of a Data Lake provided by Apache Iceberg

- A **Data Lake** is a centralised repository that allows you to store structured, semi-structured, and unstructured data at any scale, using open file formats

- The **Lakehouse** architecture extends the Data Lake by integrating a metadata management layer to provide warehouse-like capabilities

Iceberg

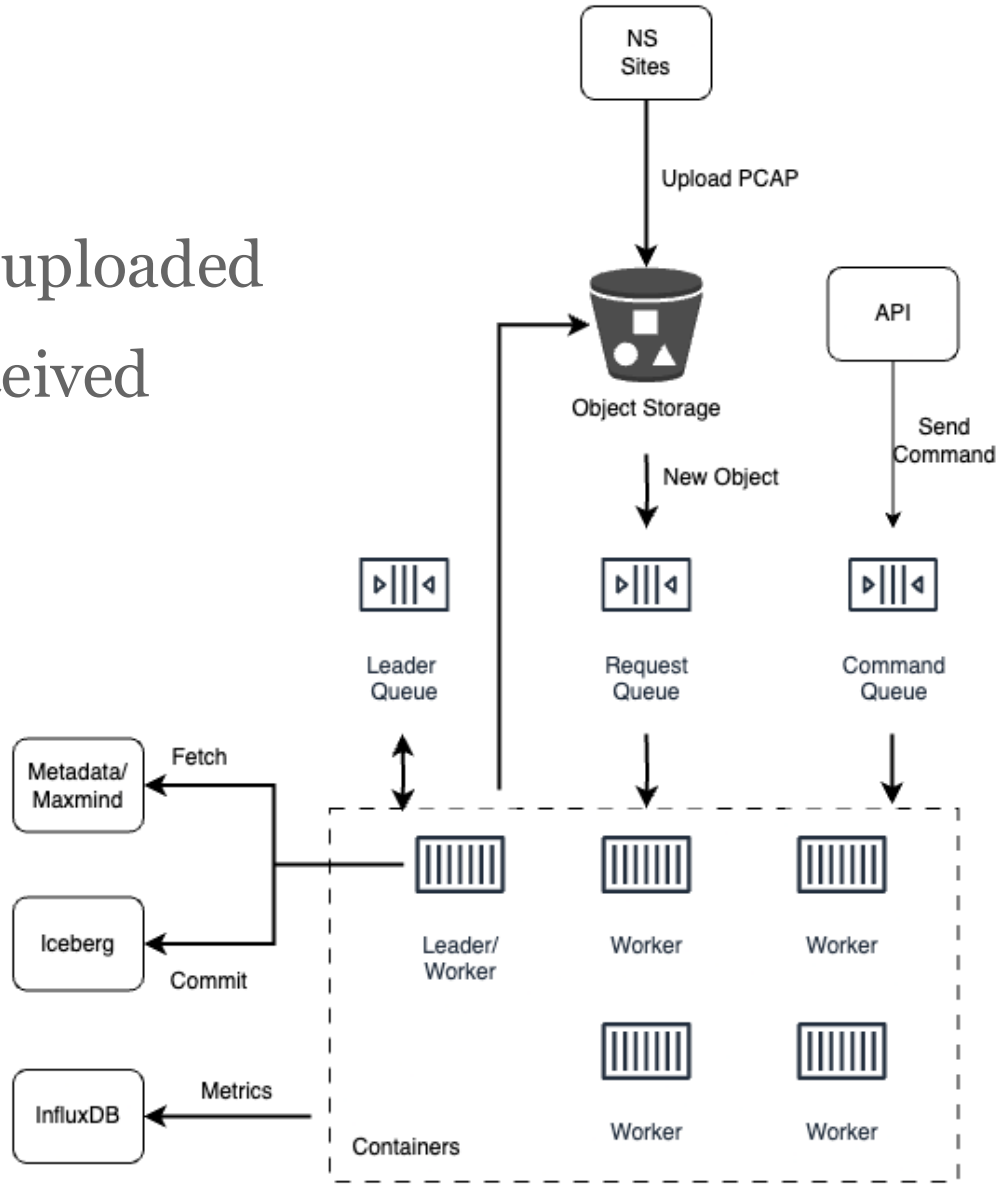
Iceberg Lakehouse features:

- Schema Evolution
- Hidden Partitioning
- Time Travel
- Transactions
- Data Compaction

Workflow

Event-driven

- New PCAP object uploaded
- API command received



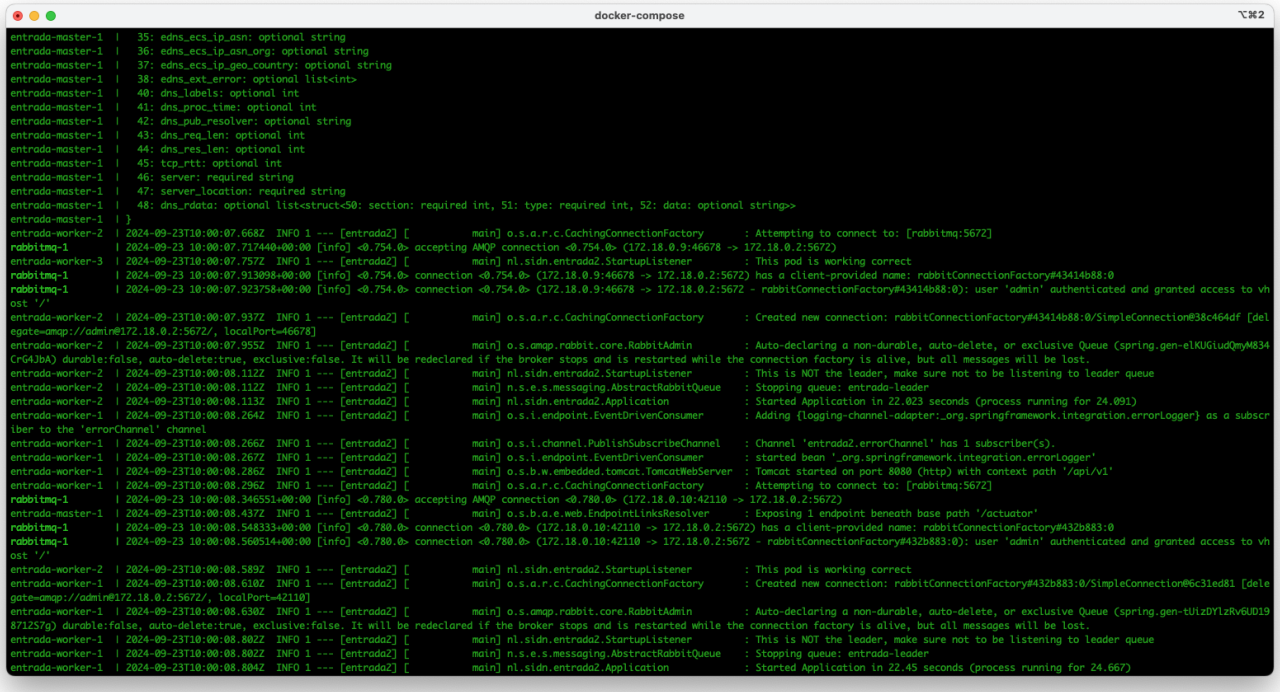
Demo

Run docker-compose.yml to start new 2-node cluster

- 1 Leader/worker
- 1 Worker

and dependencies:

- MinIO (s3)
- Trino (SQL engine)
- InfluxDB (metrics)
- RabbitMQ (messaging)
- PostgreSQL (meta data)



Demo

Data processing

- Upload new PCAP to s3 bucket
- Monitor s3 event processing
- Check results
  - Data in s3 bucket
  - Metrics in InfluxDB

Analyze data using

- Trino command line
- Dbeaver SQL-client

More details

For more information and getting started details, see:

https://github.com/SIDN/entrada2


SIDN.nl


@SIDN


SIDN

Thank You

www.sidnlabs.nl | stats.sidnlabs.nl