Chair of Network Architectures and Services
School of Computation, Information, and Technology
Technical University of Munich

SLICES GINO PLATFORM

A platform for easy Internet measurement data analysis for the **G**lobal **IN**ternet **O**bservatory

Project GINO – Global INternet Observatory

- Our chair performs regular Internet measurements in order to better understand the Internet
- We conduct Internet-wide DNS, TLS, QUIC and port scans for common protocols
- We publish an IPv6 hitlist service
- Downloads of Internet-related data like
  - Zone files of many gTLDs and some ccTLDs
  - Top and block lists
  - CT logs
  - BGP data
  - ...
- Huge amount of data available for analysis but
  - as files on our storage servers
  - in lots of different file formats and complex folder structures

More information on project GINO can be found under
https://net.in.tum.de/projects/gino/

Challenges of the Current Setup and Proposed Solutions

Challenges of the current Setup:

Time	Privacy	Restrictions
Long setup times until one can start with research	Data contains sensitive, private information	License restrictions may restrict sharing

Fine-grained access control of the data is essential for data sharing

Proposed Solutions:

1. ASQ-GINO
Answering pre-defined queries

2. General purpose data access
Access to data provided on general-purpose computation infrastructure

Goals of the GINO platform

High-level goals:

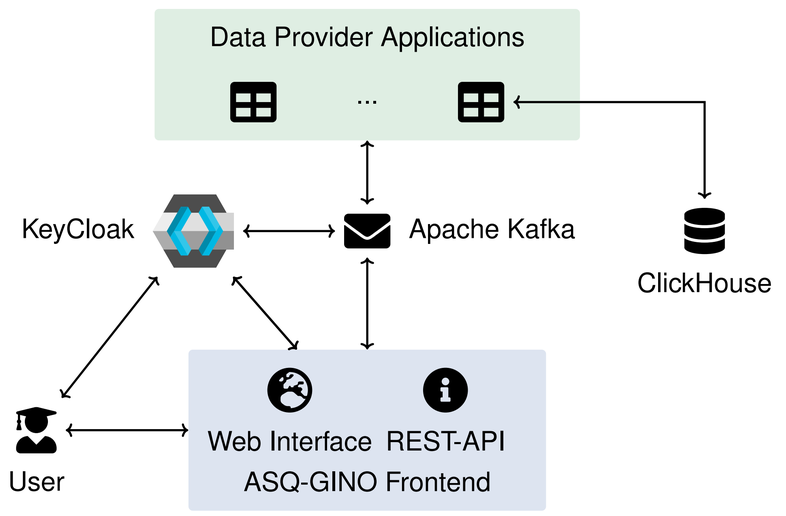
- Allow external researchers and students to quickly start working with the available data they are allowed to access
- Create a platform for easy and quick analysis of the available data

Base goals:

- Have commonly used data sources already loaded in a database ready for analysis
- Allow quick importation of non-imported datasets on demand
- Enable fine-grained access control of
  - the available data
  - the computation resources
- User authentication for internal and external researchers using ORCID
- Offer separate computation resources for analysis
- Allow sharing of workspaces
- Offer services with pre-defined analysis and queries on user request

ASQ-GINO: Answering Pre-Defined Queries

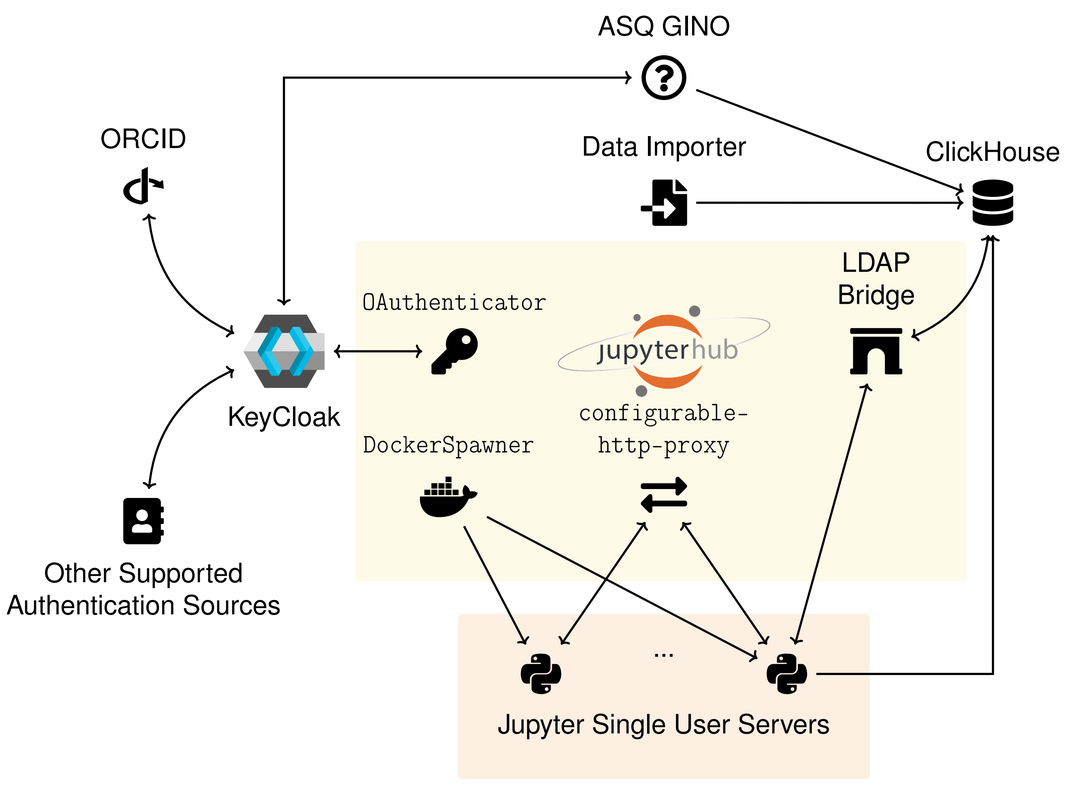
Goal: Users should be able to get answers to pre-defined queries
Architecture:



Example: When have we observed a certificate for a specified domain?

ASQ-GINO is currently being implemented by Tobias Wothge in his Master's Thesis

Overview over System Components



Key Components

KeyCloak An openID connect and OAuth2 capable open source identity and access management application
- Connects to different authentication sources, like ORCID or the chair's LDAP
- Manages authorization for ASQ-GINO

JupyterHub A platform to serve Jupyter single user notebook servers to multiple users
- Manages spawning and access to Jupyter single user servers
- Authentication and authorisation using openID connect/OAuth2 with KeyCloak

ClickHouse A high-performance columnar database
- Used as database for storing the analysis data
- Access control implemented using ClickHouse roll-based access control over the LDAP external authentication and directory functionality

LDAP Bridge Connector between KeyCloak, Jupyterhub and ClickHouse generating access tokens for a valid KeyCloak Java Web Token
- Map suitable KeyCloak roles to ClickHouse roles
- Generate an access token for the single user servers to authenticate with ClickHouse

Data Importer Import the data from our storage servers into ClickHouse
- Perform scheduled and on-demand data imports
- Perform scheduled and manual data deletion from the database

Christian Dietze

diec@net.in.tum.de