

Query Broker: An Efficient Cybersecurity Data Access Platform

Enabling Efficient, Secure Data Retrieval for Enhanced Security and AI Services.

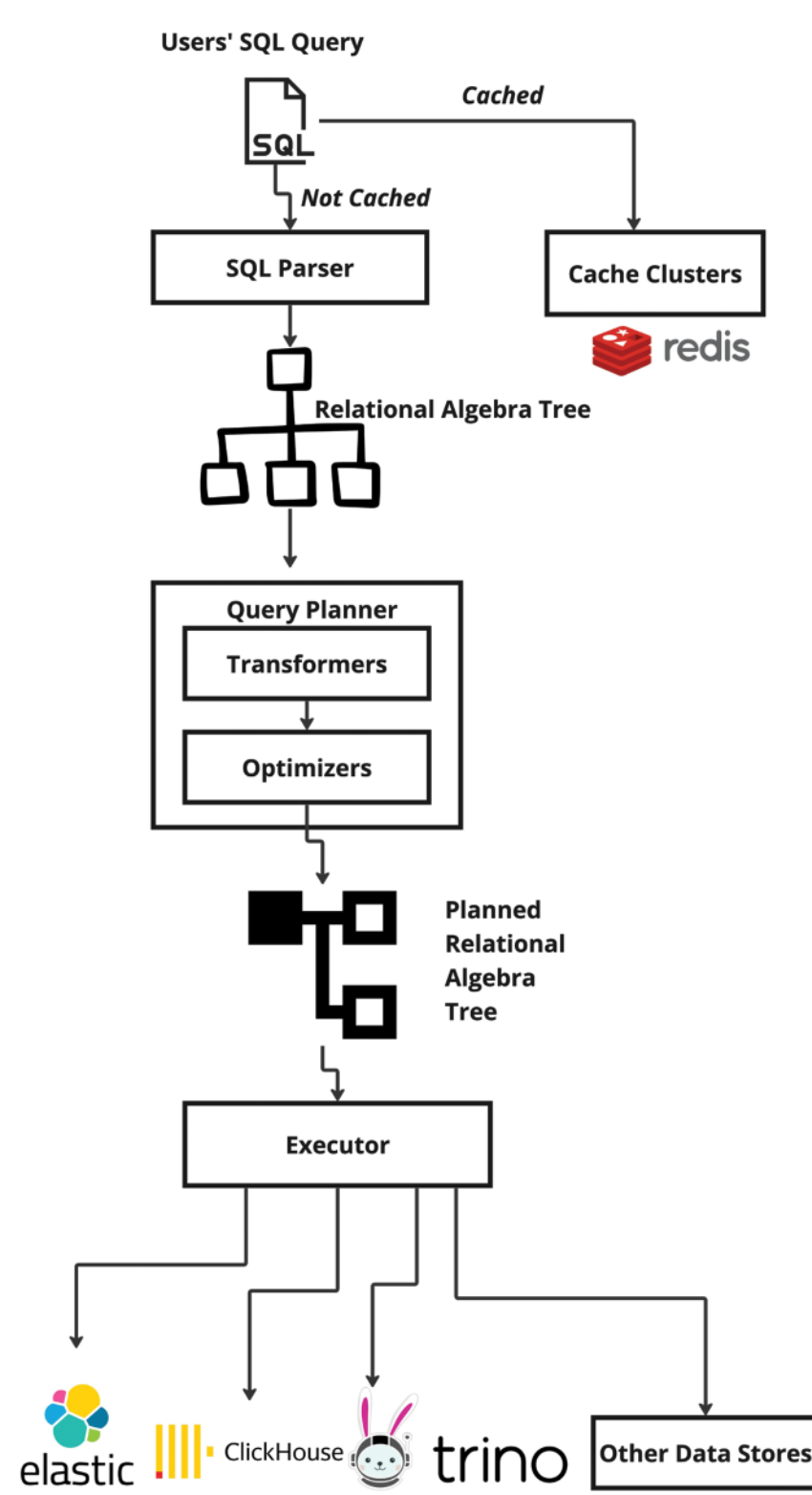
Cijie Xia

Shurui Zhou

ACADEMIC SUPERVISOR

Dmitry Kaminski

INDUSTRY SUPERVISOR



Relational Algebra

```
1 SELECT event_id
2 FROM observations
3 WHERE classification = 'internal'
```

Parse into

```
1 PROJECT(event_id)
2 FILTER(=(classification, 'internal'))
3 SCAN(observations)
```

Relational algebra provides a formal and precise way to express queries and transformations on relational databases.

Predicate Pushdown

```
1 FILTER(=(user, 'Bob'))
2 PROJECT(user, someExpensiveOperation(groups))
3 SCAN(table)
```

Optimize into

```
1 PROJECT(user, someExpensiveOperation(groups))
2 FILTER(=(user, 'Bob'))
3 SCAN(table)
```

Pushing filter conditions as close to the data source as possible to reduce the amount of data that needs to be processed.

PROJECT SUMMARY

Efficiently and securely retrieving observation and incident data collected from customers poses a significant challenge, due to the daily processing of a large volume of data in its security monitoring services and the substantial data requirements of security research team's AI services.

In response to this challenge, Arctic Wolf Networks has initiated the Query Broker project. It serves as a data access layer (DAL) that offers a layer of abstraction, security control, portability and performance optimization, which aims to provide a consistent and controlled way how the customers will interact with the data.

The benefits of using Query Broker include but not are limited to:

- Simplified Complexity: Abstract the underlying database structure and complexity.
- Security Control: Control data access through authentication, authorization, and encryption.
- Database Portability: Allow switching to a different database system with minimal code change.
- Logging and Auditing: Enable log database interactions, providing a trail of actions for auditing and debugging purposes.

Furthermore, query transformation and optimization constitute two critical phases that underscore how the Query Broker delivers advantages to its users. Through the automated enhancement of queries via transformation, users are empowered to craft straightforward queries capable of executing intricate tasks, such as generating a histogram distribution of data points. Moreover, by optimizing queries using relational algebra, the Query Broker aspires to attain optimal query execution speeds, supported by mathematical assurances.

