Cracking the Code: Unleashing CodeQL's Superpowers for Open Source Security

Dan Shanahan
@leftrightleft

Jose Palafox
@josepalafox

Where the world builds software

The world's largest developer platform

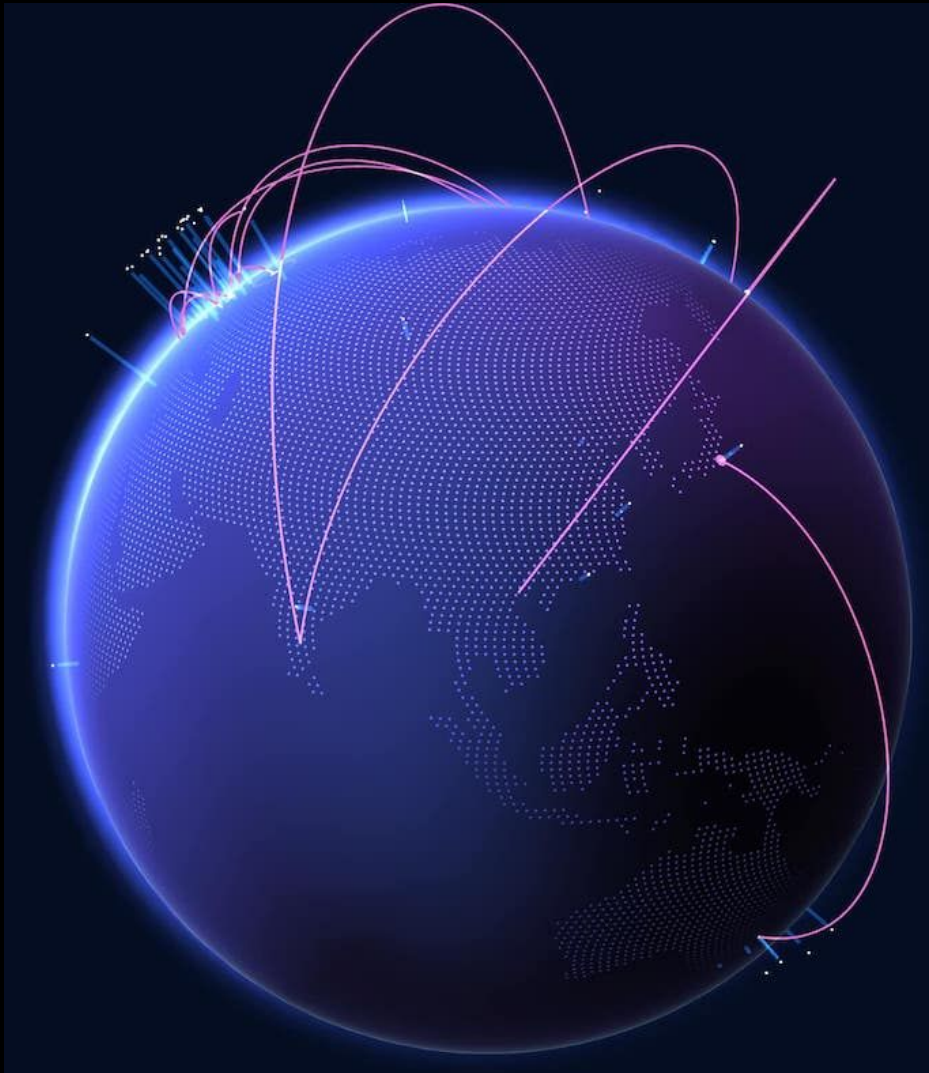
100M+
Developers

4M+
Organizations

3.5B+
Contributions per year

1,000s
Top open source communities

330M+
Private and public repositories



Application security is challenging

Applications continue to be a top attack vector

Applications are at the center of more than 40% of all data breaches

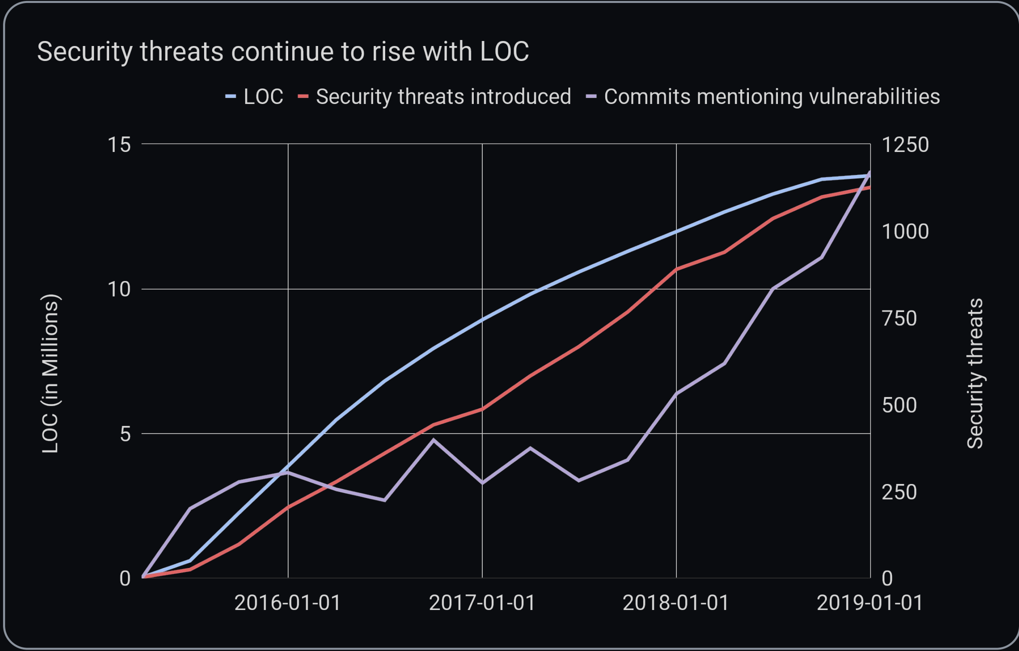
Supply Chain attacks are on the rise

45% of global organizations will be impacted in some way by a supply chain attack by 2025

Fixing vulnerabilities is hard

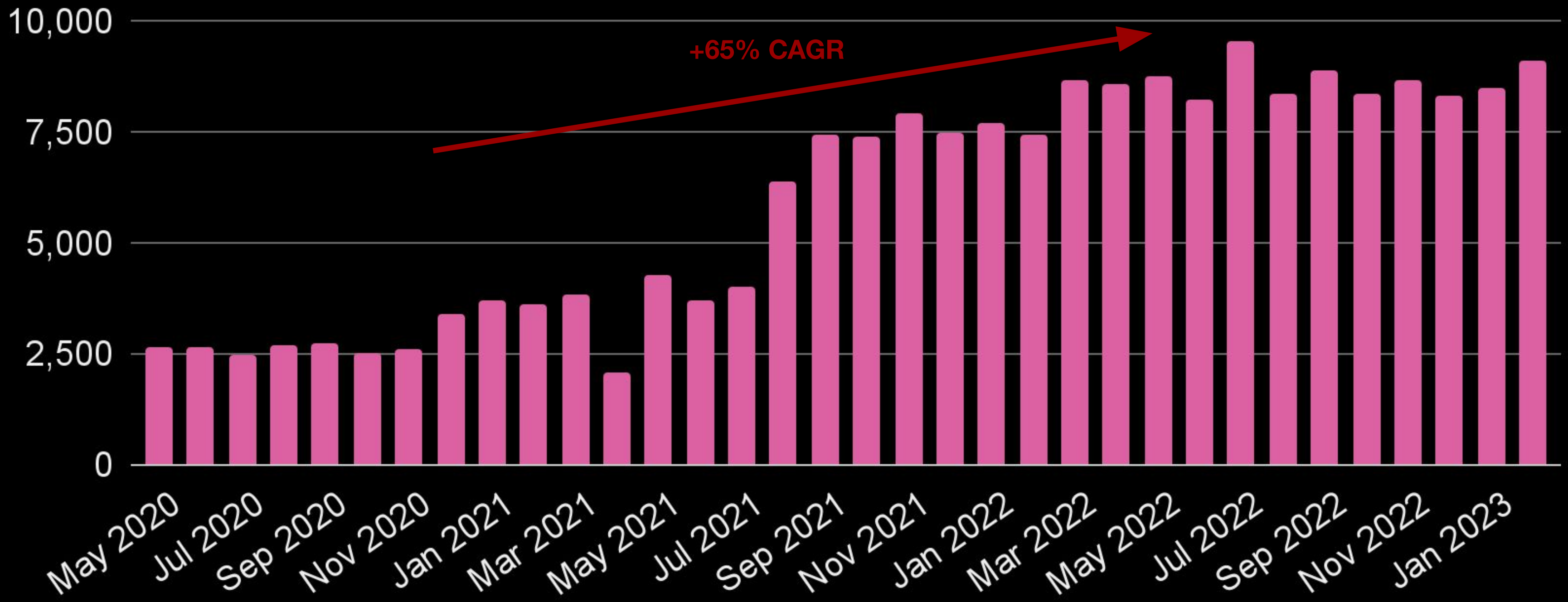
67% of vulnerabilities still exist after 3 months, and 81% of devs still choose to ship vulnerable code to meet deadlines.

Despite increasing developer awareness, security threats continue to rise



We’re seeing more credential leaks than ever

GitHub access tokens leaked in public repositories



Source: GitHub data

Code Security Improvements

Tools

Secret Scanning
CodeQL
Dependabot
2FA

Outreach

Security Lab
Github Advisories
Private Vulnerability Disclosure

Community

Improves overall security posture of everything

Code Security and Analysis

Code Scanning

Static analysis of every pull request, integrated into the developer workflow and powered by CodeQL

Secret Scanning

Automatic notifications of any API tokens or other secrets exposed anywhere in your git history

Supply Chain Security

Secure your open source project and secure the open source dependencies your applications rely on

Security Impact

Code Scanning

Code vulnerability fix rate of

72%

compared to

15%

observed norm after 7 days

Dependabot

OSS vulnerabilities MTTR decreased from 180 days to 40 days

Secret Scanning

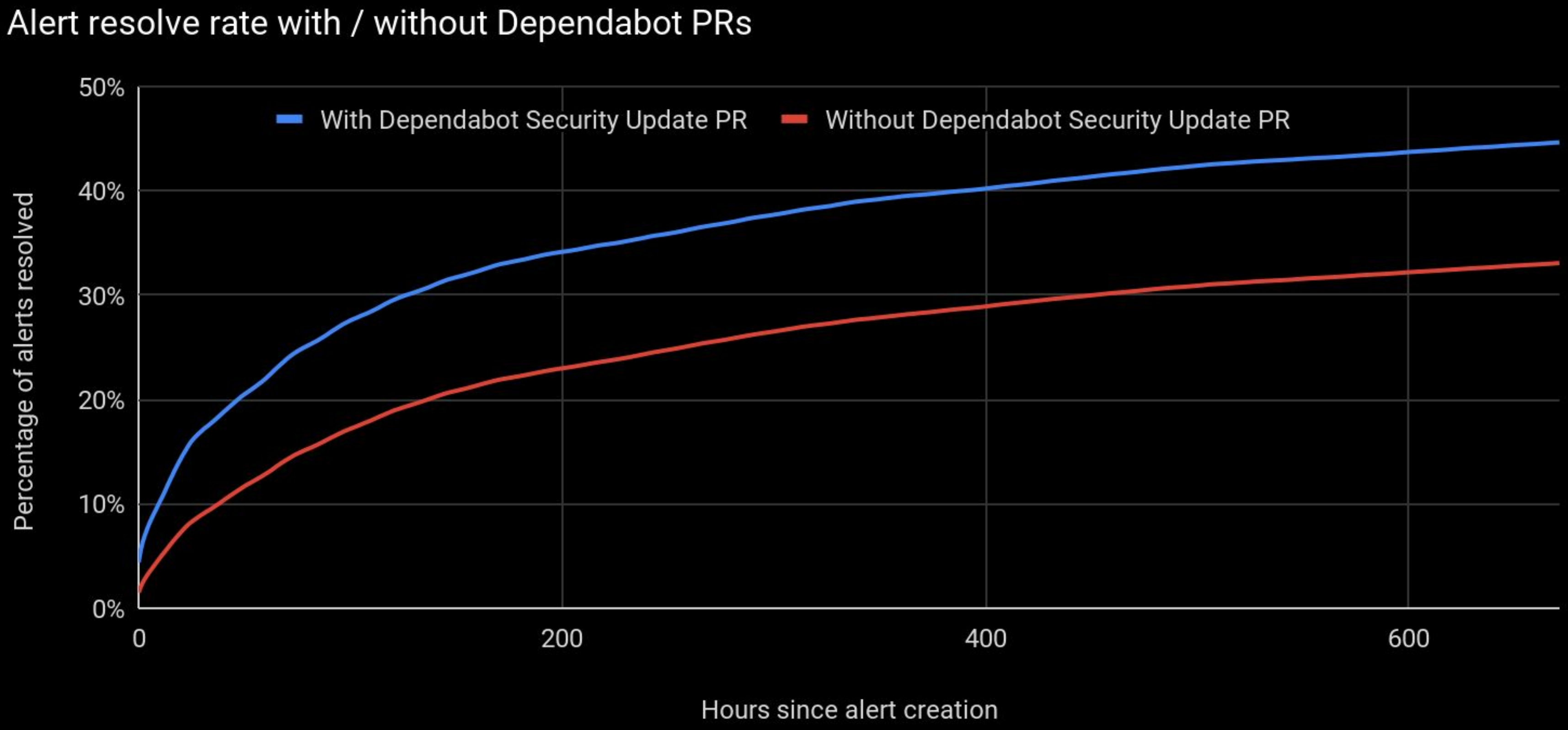
17K

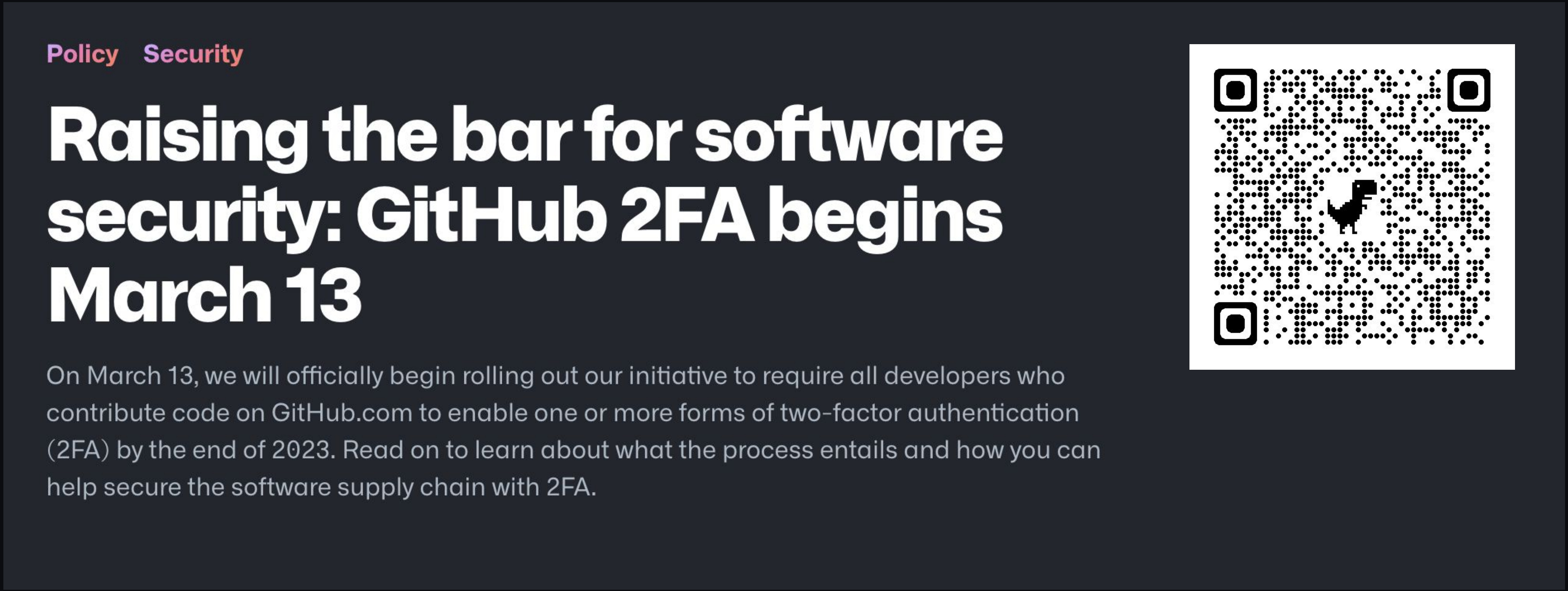
Potential secret leaks prevented by push protection (since Apr 2022)

3.5M

Secrets detected on public repos

Dependabot security updates increase the resolve rate and speed that vulnerable dependencies are addressed





Policy Security

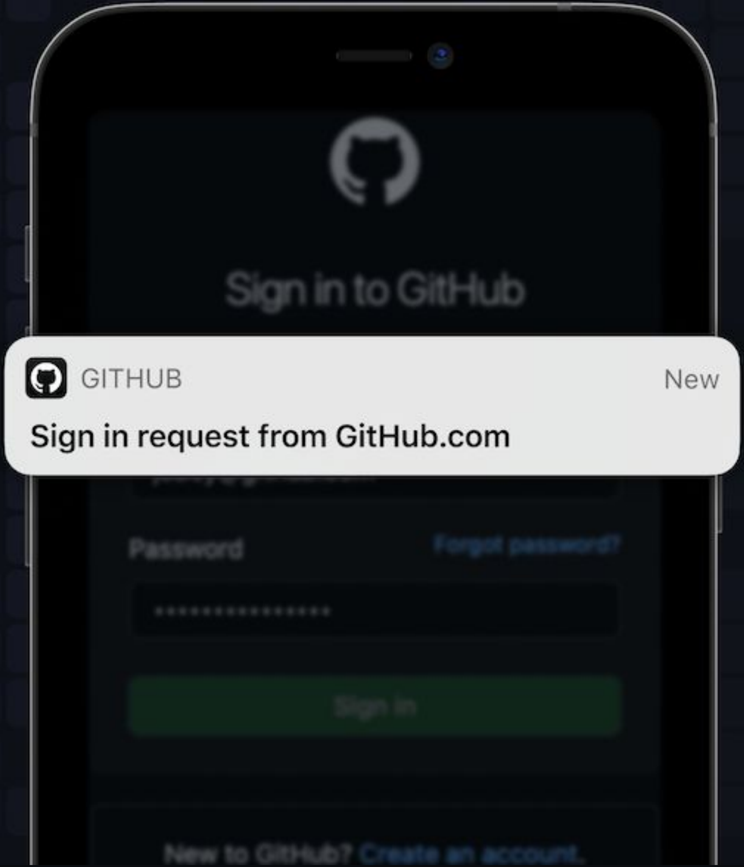
Raising the bar for software security: GitHub 2FA begins March 13

On March 13, we will officially begin rolling out our initiative to require all developers who contribute code on GitHub.com to enable one or more forms of two-factor authentication (2FA) by the end of 2023. Read on to learn about what the process entails and how you can help secure the software supply chain with 2FA.

GitHub

Mobile 2FA

Fast and secure two-factor authentication.



Sign in to GitHub

GITHUB New

Sign in request from GitHub.com

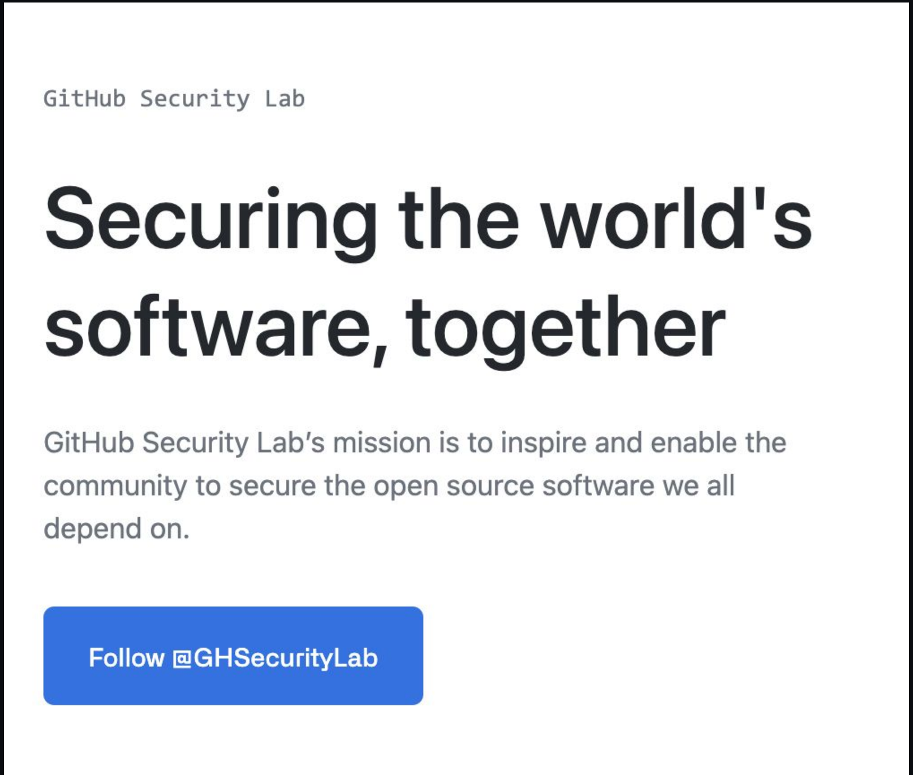
Password Forgot password?

Sign in

New to GitHub? Create an account.

Security Research

Vulnerabilities disclosed every year	~200
CVEs credited to GitHub Security Lab every year	~100
Fix rate	95%



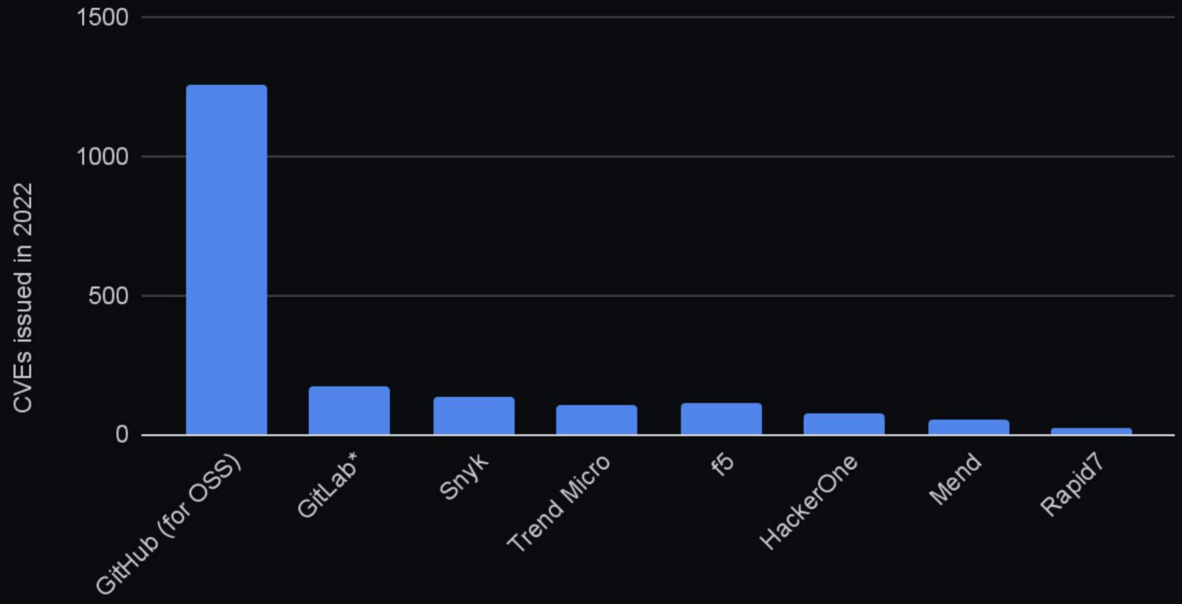
GitHub Security Lab

Securing the world's software, together

GitHub Security Lab's mission is to inspire and enable the community to secure the open source software we all depend on.

GitHub issues the most CVEs for OSS

CVE count by numerating authority



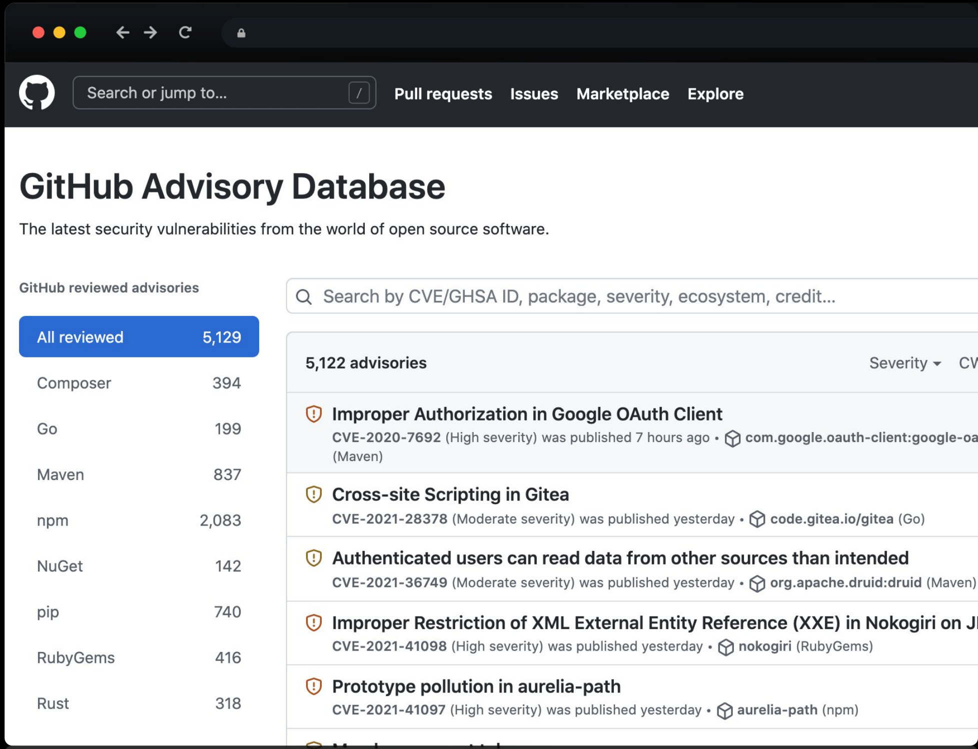
* Includes CVEs issued for the vendor’s own software
Source: The CVE Program (via https://github.com/CVEProject/cvelist)

Advisory data direct from maintainers, curated by GitHub

Free, open source
database of advisories with a creative commons license

Maintained by GitHub
with a full-time, dedicated curation team (part of the Security Lab)

Built into the disclosure workflow
for maintainers who can request CVEs from GitHub



GitHub Advisory Database

The latest security vulnerabilities from the world of open source software.

GitHub reviewed advisories

All reviewed	5,129
Composer	394
Go	199
Maven	837
npm	2,083
NuGet	142
pip	740
RubyGems	416
Rust	318

Search by CVE/GHSA ID, package, severity, ecosystem, credit...

5,122 advisories

Improper Authorization in Google OAuth Client
CVE-2020-7692 (High severity) was published 7 hours ago • **com.google.oauth-client:google-oa** (Maven)

Cross-site Scripting in Gitea
CVE-2021-28378 (Moderate severity) was published yesterday • **code.gitea.io/gitea** (Go)

Authenticated users can read data from other sources than intended
CVE-2021-36749 (Moderate severity) was published yesterday • **org.apache.druid:druid** (Maven)

Improper Restriction of XML External Entity Reference (XXE) in Nokogiri on J
CVE-2021-41098 (High severity) was published yesterday • **nokogiri** (RubyGems)

Prototype pollution in aurelia-path
CVE-2021-41097 (High severity) was published yesterday • **aurelia-path** (npm)

Community-powered security

Secure the Supply Chain

Open Source projects are enabled to improve security posture

Developer Experience

Automated code scanning,

Largest vulnerability database

Automated security updates via dependabot

Virtuous Cycle

Community of top security experts

World’s most advanced code analysis

Vulnerability hunting interface for open source

CodeQL

Code represented as data

Data models representing your source code in a relational database

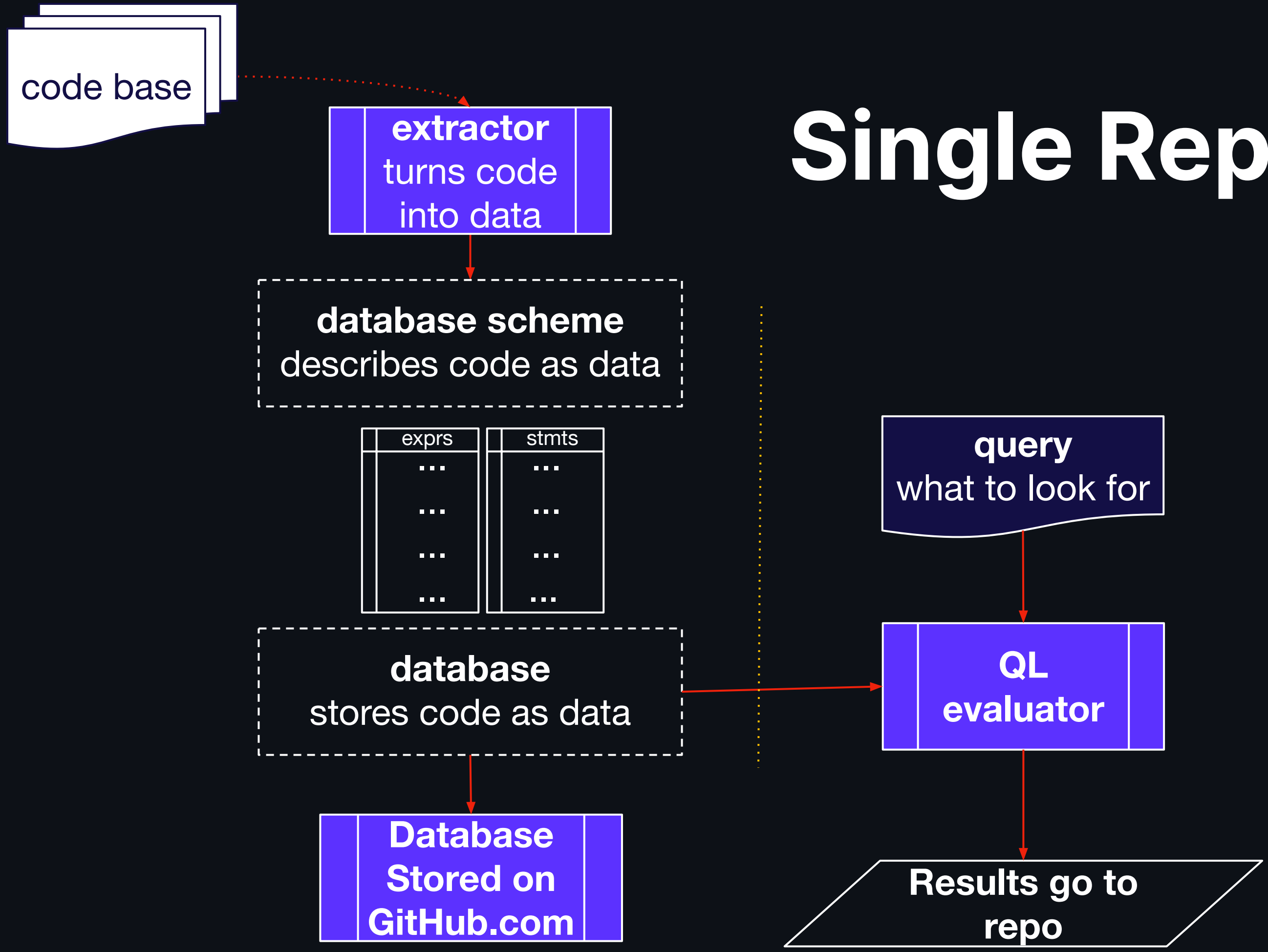
An expressive query language

Expressive language that allows you ask complex questions of your code

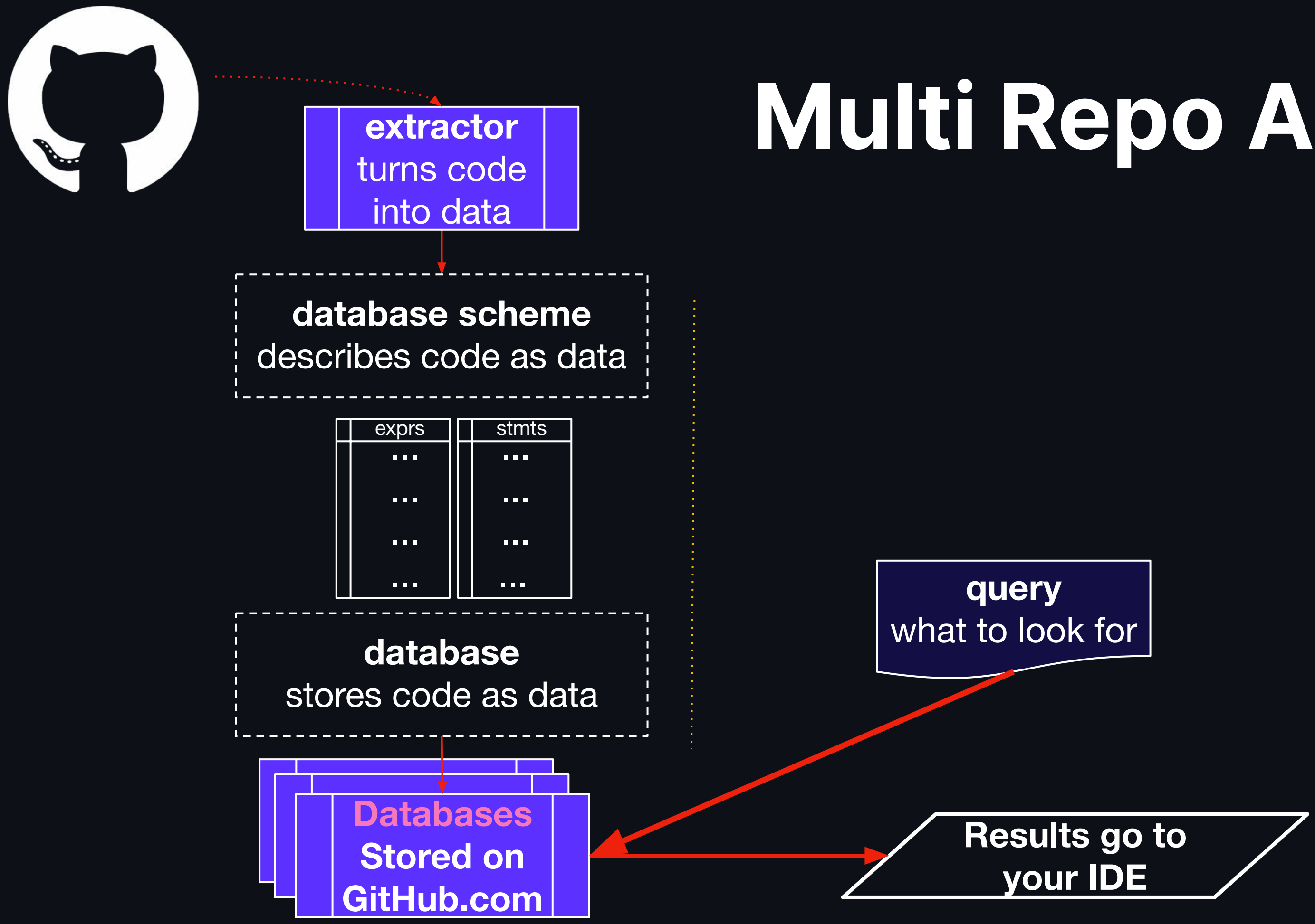
A suite of utilities and helpers

CLI utilities, IDE plugins, and management capabilities all baked into the GitHub platform.

Single Repo Analysis

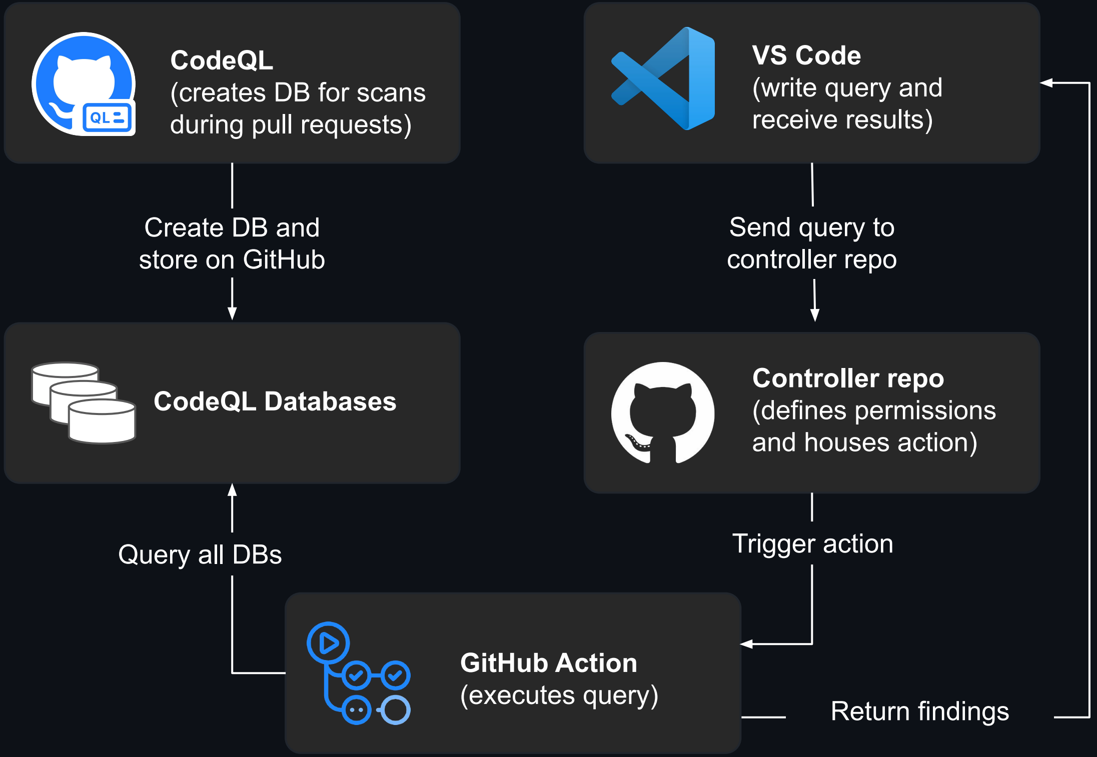


Multi Repo Analysis



Demo

How MRVA Works



CodeQL
(creates DB for scans during pull requests)

Create DB and store on GitHub

CodeQL Databases

VS Code
(write query and receive results)

Send query to controller repo

Controller repo
(defines permissions and houses action)

Trigger action

Query all DBs

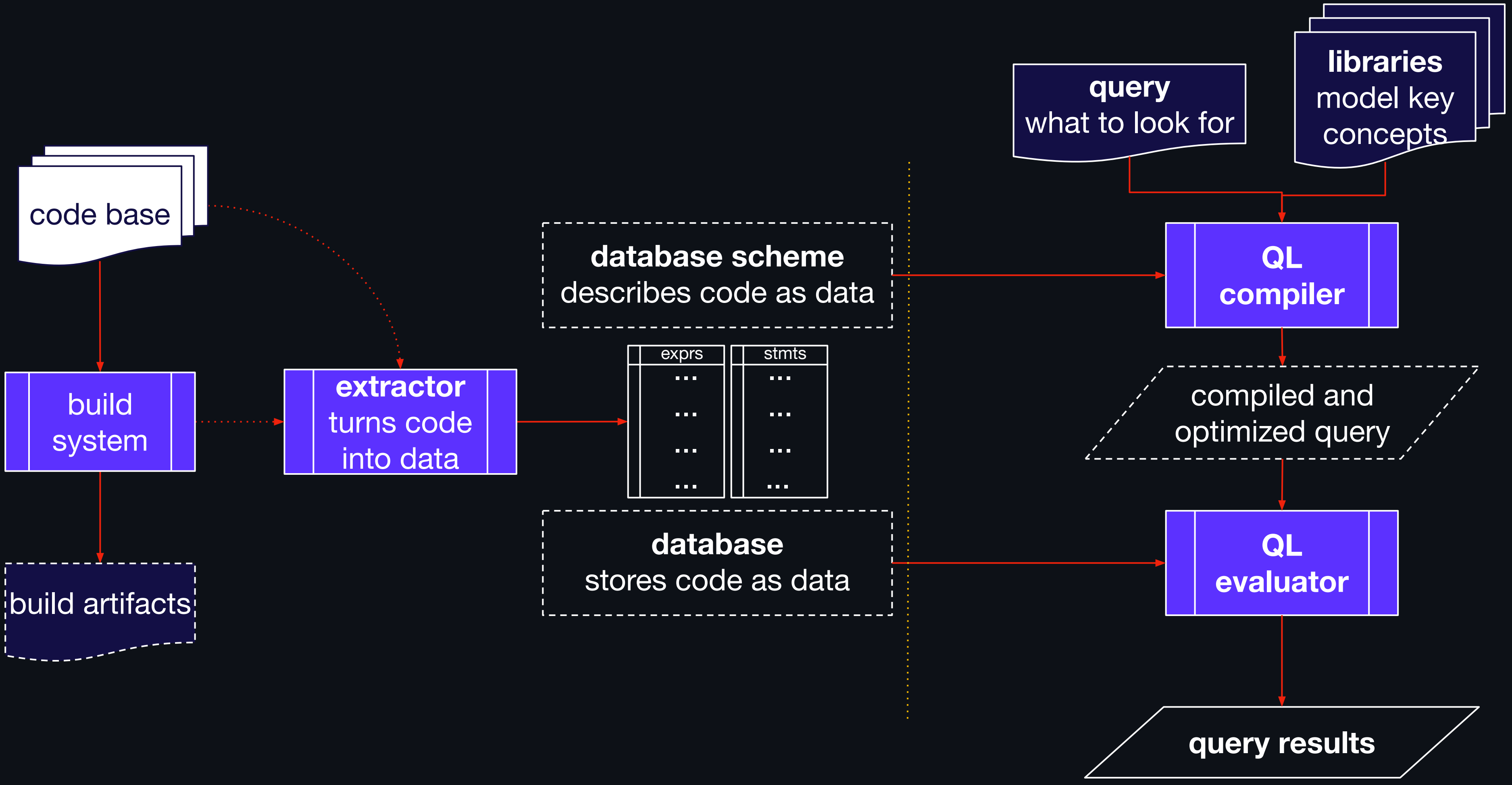
GitHub Action
(executes query)

Return findings

CodeQL
Community driven

Benefit from thousands of analyses created by GitHub's team of security researchers and language experts and contributions from other GitHub users- such as leading security researchers at Microsoft, Google, Uber, and others.

- Collaborate and learn with a dedicated security team that also leverages the world's expert community of developers
- Open source foundations encourage collaboration and community engagement
- Benefit from user driven design and frequent innovation



code base

build system

build artifacts

extractor turns code into data

exprs

…

…

…

…

stmts

…

…

…

…

database scheme describes code as data

database stores code as data

query what to look for

libraries model key concepts

QL compiler

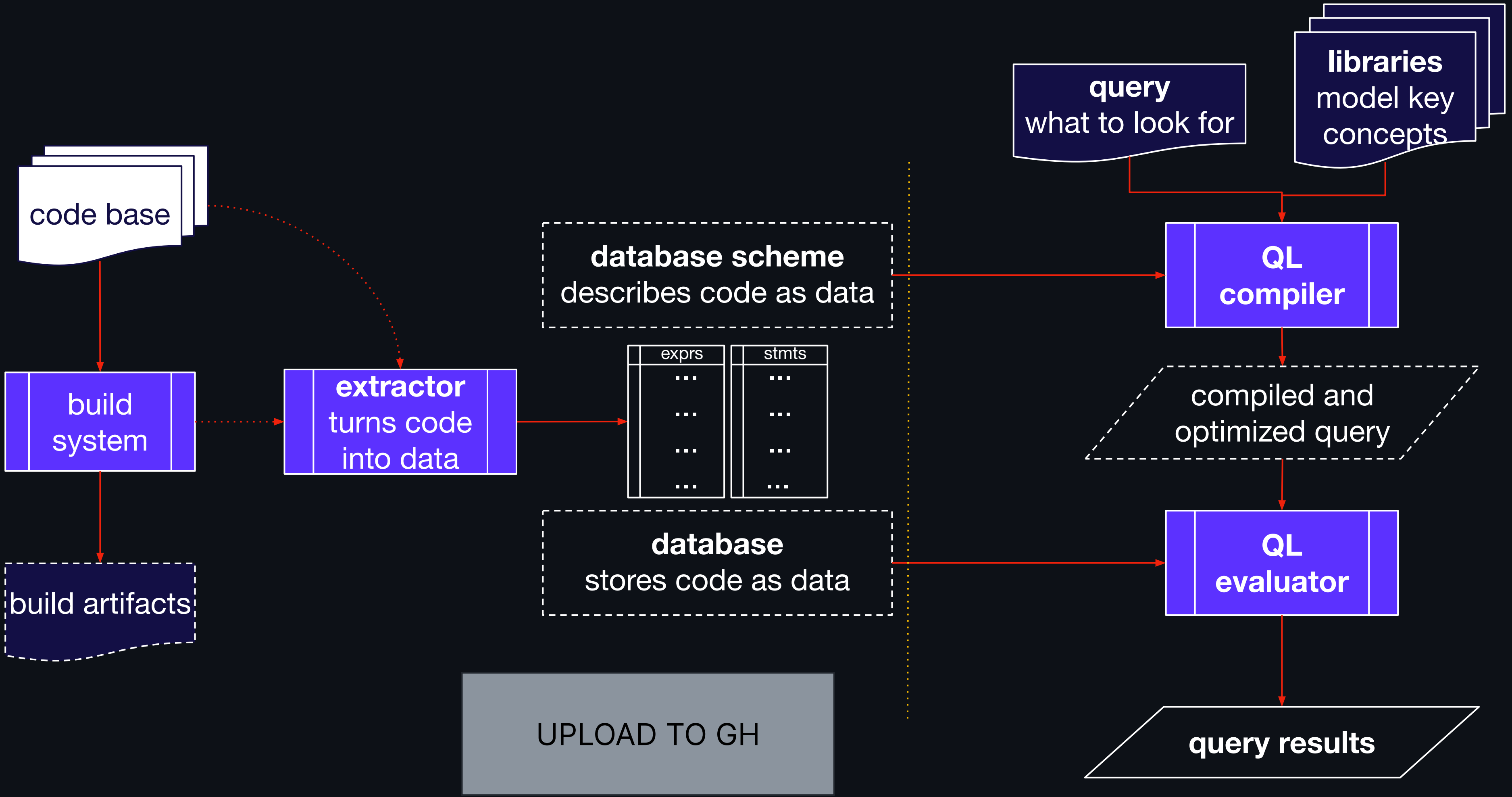
compiled and optimized query

QL evaluator

query results

The CodeQL Language

- A **logic language** based on first-order logic
- A **declarative language** allowing us to focus on what, not how
- An **object-oriented language**
- A **query language** working on a read-only snapshot database
- Rich **standard libraries** for program analysis



code base

build system

build artifacts

extractor turns code into data

database scheme describes code as data

exprs	stmts
…	…
…	…
…	…
…	…

database stores code as data

UPLOAD TO GH

query what to look for

libraries model key concepts

QL compiler

compiled and optimized query

QL evaluator

query results

Import: lets us reuse logic defined in another module.

```
import java
```

```
from CatchClause clause, BlockStmt block
where clause.getBlock() = block
and block.getNumStmt() = 0
select clause, “Exception swallowed."
```

Query clause: describes what we are trying to find.

Responsible Disclosure