A Search Engine Backed by Internet-Wide Scanning

Zakir Durumeric
University of Michigan

David Adrian
University of Michigan

Ariana Mirian
University of Michigan

Michael Bailey
University of Illinois

J. Alex Halderman
University of Michigan

ZMap

2013 A **1200x performance improvement** over Nmap for an Internet-wide single port TCP scan

2014 Scan the Internet in **under 5 minutes**.

2015 Popular in industry and academia, used by over **104** academic studies

ZMap Vision

Goals

Enable new and exciting research

Decrease the barriers to entry for Internet-wide surveys

Anyone can scan the entire Internet using a single host

ZMap Vision

Goals	Reality
Enable new and exciting research	Not all researchers can run ZMap
Decrease the barriers to entry for Internet-wide surveys	Negotiate with network administrators for bandwidth and address space
Anyone can scan the entire Internet using a single host	Maintain an opt-out list and respond to complaints

scans.io

A **public archive** of Internet-wide scan data

Data from University of Michigan, Rapid7, Fedora, and more

Over **35 TB** downloaded in July 2015

https://scans.io

scans.io

What is the impact of a particular vulnerability?

What types of cryptography are in use?

What version of software are most popular?

scans.io

What is the impact of a particular vulnerability?

What types of cryptography are in use?

What version of software are most popular?

Nearly **5TB** of HTTPS data in the last year

Data needs to be processed and annotated

Analysis is time-consuming, and error-prone

What if?

…we could answer questions with a **single query**?

…we always knew the **current state** of the Internet?

…we built a **search engine** on top of Internet-wide scan data?

censys

Search engine that allows researchers to **ask questions** about the *devices* and *networks* that compose the Internet

censys

Search

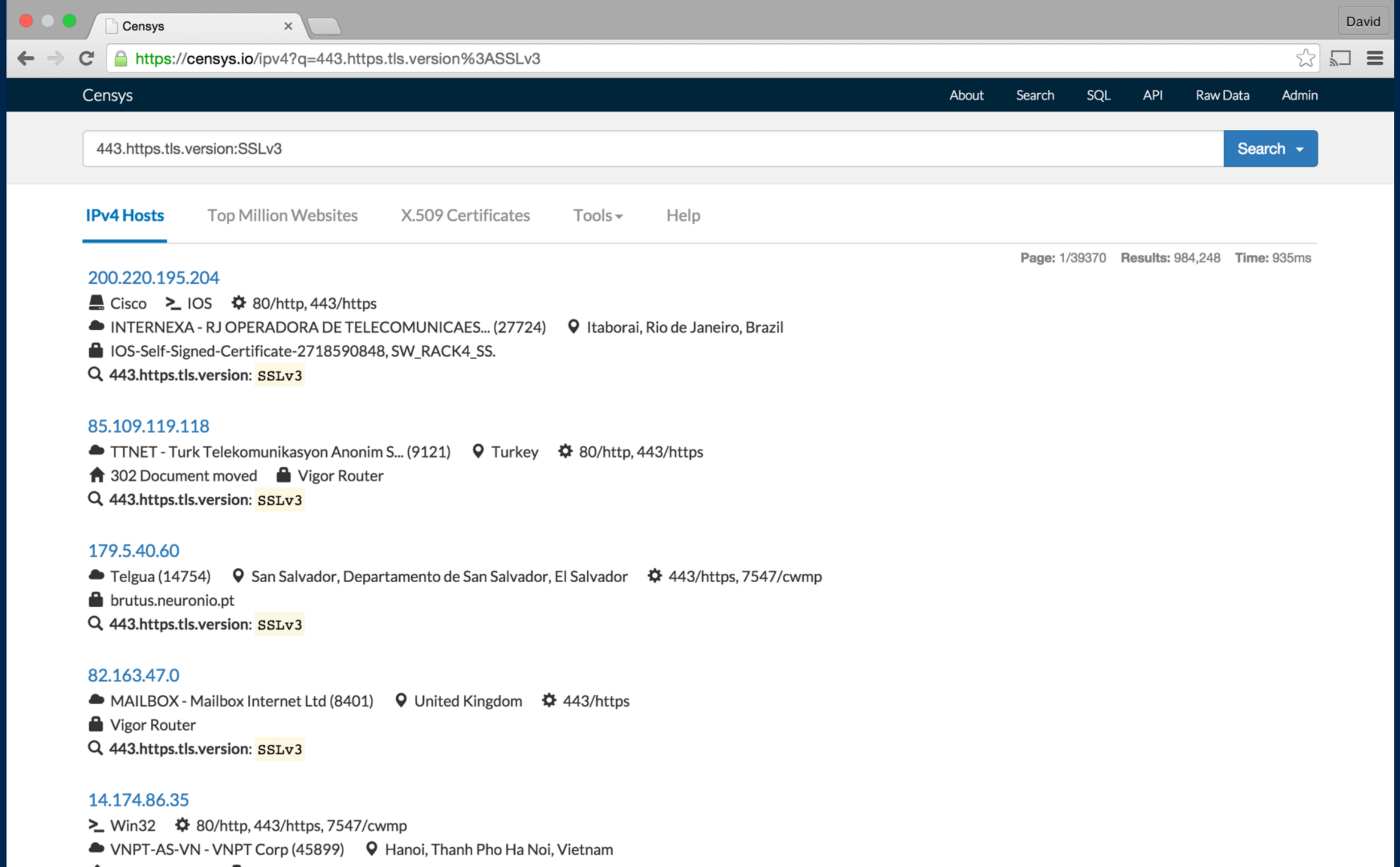
Example

What is the impact of disabling support for SSLv3?

443.https.tls.version:SSLv3 Search

Example

What is the impact of disabling support for SSLv3?


Censys
David
https://censys.io/ipv4?q=443.https.tls.version%3ASSLv3
Censys
About
Search
SQL
API
Raw Data
Admin
443.https.tls.version:SSLv3
Search
IPv4 Hosts
Top Million Websites
X.509 Certificates
Tools
Help
Page: 1/39370 Results: 984,248 Time: 935ms
200.220.195.204
Cisco IOS 80/http, 443/https
INTERNEXA - RJ OPERADORA DE TELECOMUNICAES... (27724) Itaborai, Rio de Janeiro, Brazil
IOS-Self-Signed-Certificate-2718590848, SW_RACK4_SS.
443.https.tls.version: SSLv3
85.109.119.118
TTNET - Turk Telekomunikasyon Anonim S... (9121) Turkey 80/http, 443/https
302 Document moved Vigor Router
443.https.tls.version: SSLv3
179.5.40.60
Telgua (14754) San Salvador, Departamento de San Salvador, El Salvador 443/https, 7547/cwmp
brutus.neuronio.pt
443.https.tls.version: SSLv3
82.163.47.0
MAILBOX - Mailbox Internet Ltd (8401) United Kingdom 443/https
Vigor Router
443.https.tls.version: SSLv3
14.174.86.35
Win32 80/http, 443/https, 7547/cwmp
VNPT-AS-VN - VNPT Corp (45899) Hanoi, Thanh Pho Ha Noi, Vietnam

Censys

David

https://censys.io/ipv4?q=443.https.tls.version%3ASSLv3

Censys About Search SQL API Raw Data Admin

443.https.tls.version:SSLv3 Search

IPv4 Hosts Top Million Websites X.509 Certificates Tools Help

Page: 1/39370 **Results:** 984,248 **Time:** 935ms

200.220.195.204
Cisco IOS 80/http, 443/https
INTERNEXA - RJ OPERADORA DE TELECOMUNICAES... (27724) Itaborai, Rio de Janeiro, Brazil
IOS-Self-Signed-Certificate-2718590848, SW_RACK4_SS.
443.https.tls.version: `SSLv3`

85.109.119.118
TTNET - Turk Telekomunikasyon Anonim S... (9121) Turkey 80/http, 443/https
302 Document moved Vigor Router
443.https.tls.version: `SSLv3`

179.5.40.60
Telgua (14754) San Salvador, Departamento de San Salvador, El Salvador 443/https, 7547/cwmp
brutus.neuronio.pt
443.https.tls.version: `SSLv3`

82.163.47.0
MAILBOX - Mailbox Internet Ltd (8401) United Kingdom 443/https
Vigor Router
443.https.tls.version: `SSLv3`

14.174.86.35
Win32 80/http, 443/https, 7547/cwmp
VNPT-AS-VN - VNPT Corp (45899) Hanoi, Thanh Pho Ha Noi, Vietnam

censys

Search

Example

What is the impact of disabling support for SSLv3?

censys

443.https.tls.validation.browser_trusted:true Search

Example

What is the impact of disabling support for SSLv3?

Censys

David

https://censys.io/ipv4/report?q=443.https.tls.validation.browser_trusted%3Atrue&field=443.https.tls.version&max_buckets=

Censys About Search SQL API Raw Data Admin

443.https.tls.validation.browser_trusted:true Search

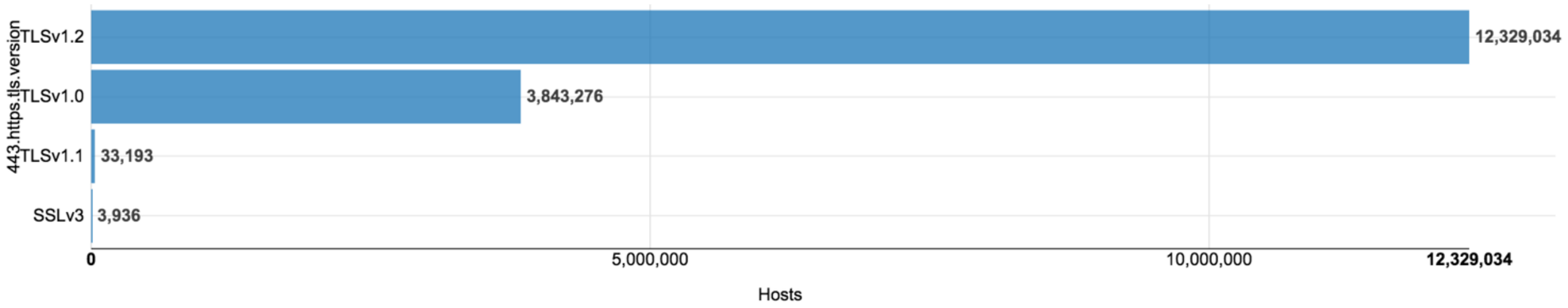
IPv4 Hosts | Top Million Websites | X.509 Certificates | Tools | Help

This tool allows you to generate a report on the breakdown of a value present on the ipv4s returned by your query. For example, to generate a report on the cipher suites chosen by HTTPS servers in the United States, you could query for `location.country_code: US AND protocols:443/https` and then generate a report on the breakdown of the field `443.https.tls.cipher_suite.name`. A list of reportable fields is available here.

Many fields have both both parsed and raw values available (e.g., `80.http.get.headers.server` and `80.http.get.headers.server.raw`. In these cases, the raw value will represent the *exact* string (e.g., `Apache/2.2.22 (Debian)`) and the parsed version will bucket on individual terms (e.g., `Apache` and `Debian`). Incidentally, in this case, you likely want to aggregate on a parsed out version of the web server, `metadata.web_server`.

443.https.tls.version | Max Buckets | Build Report

Host Report



censys

Full-text search

SQL

Current and historical data

API

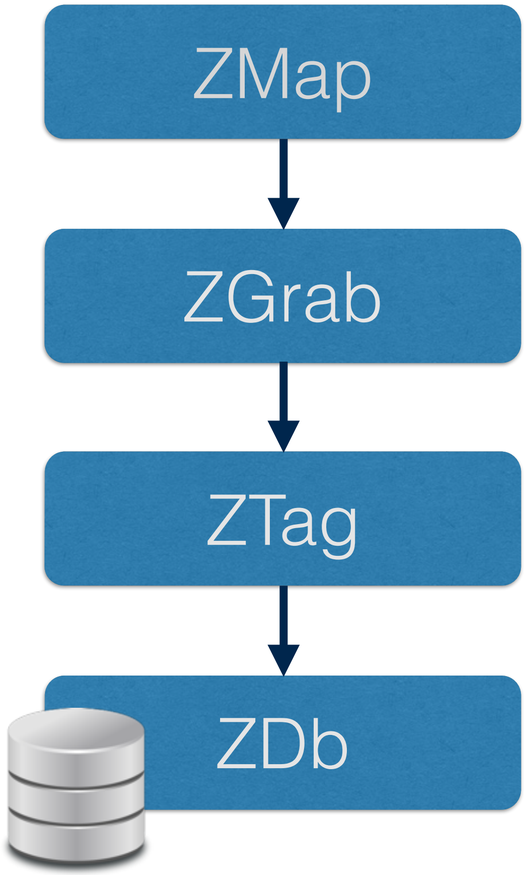
Motivation

Architecture

Looking Forward

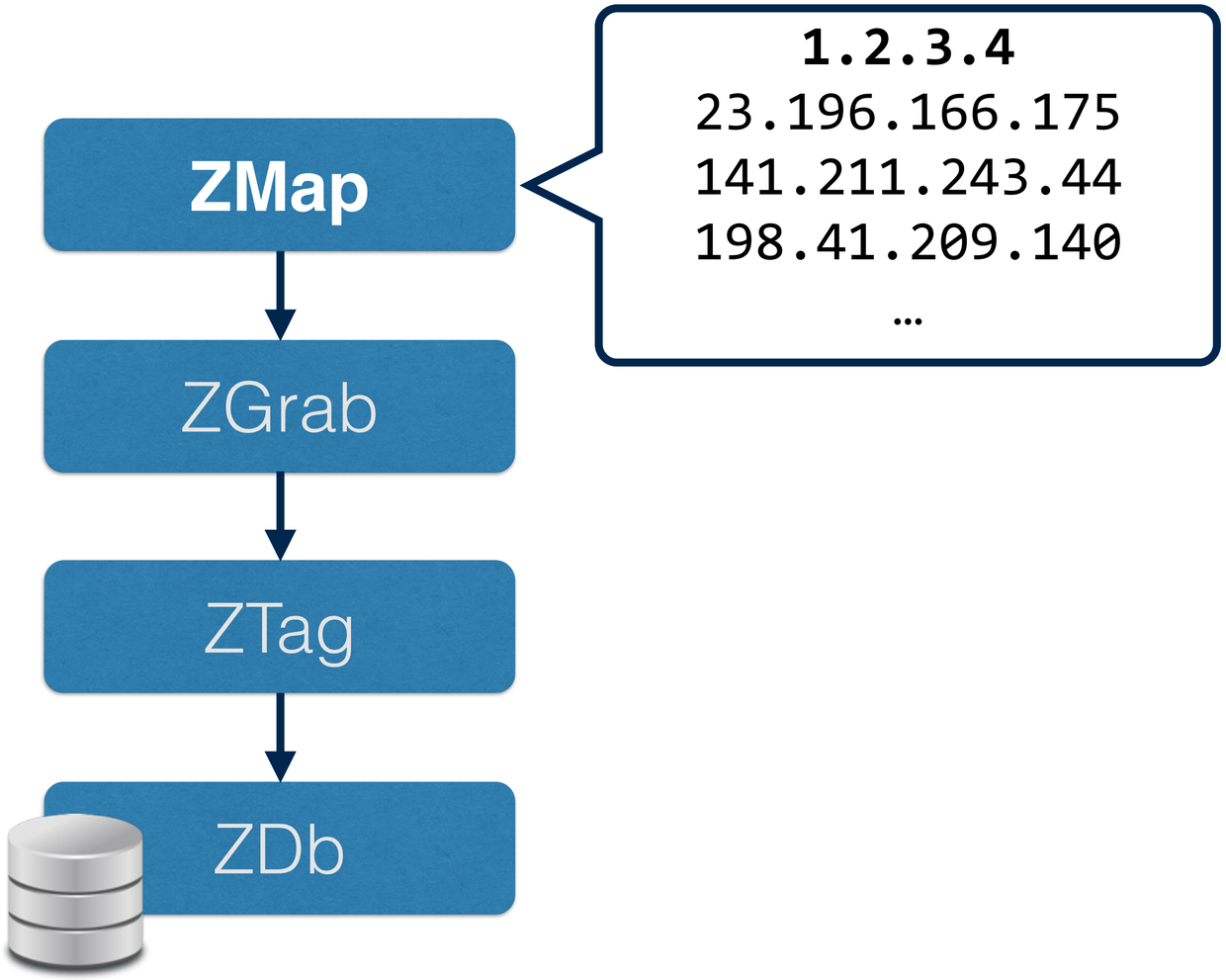
Data Collection

1. Identify listening hosts
2. Gather application-layer data
3. Annotate with additional metadata
4. Aggregate by host



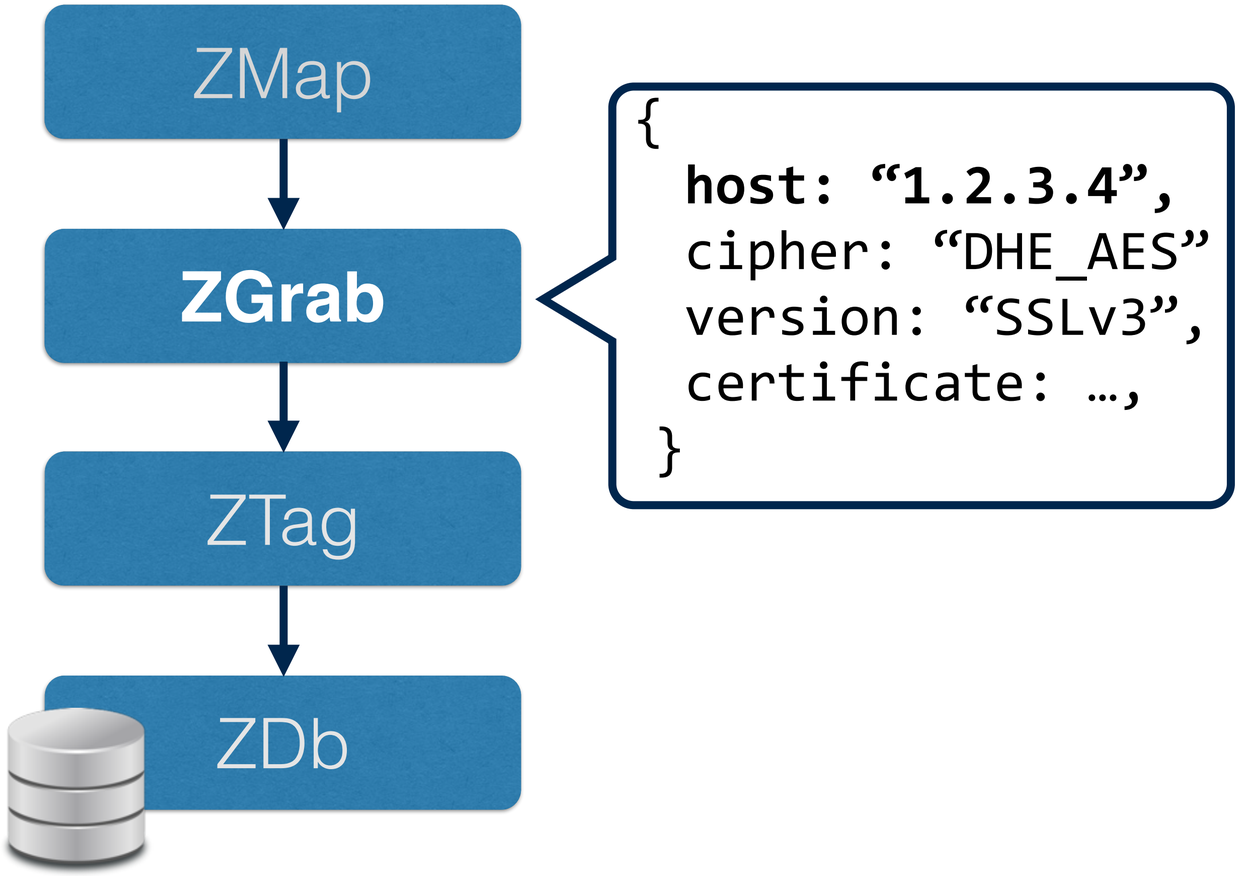
Data Collection

1. **Identify listening hosts**
2. Gather application-layer data
3. Annotate with additional metadata
4. Aggregate by host



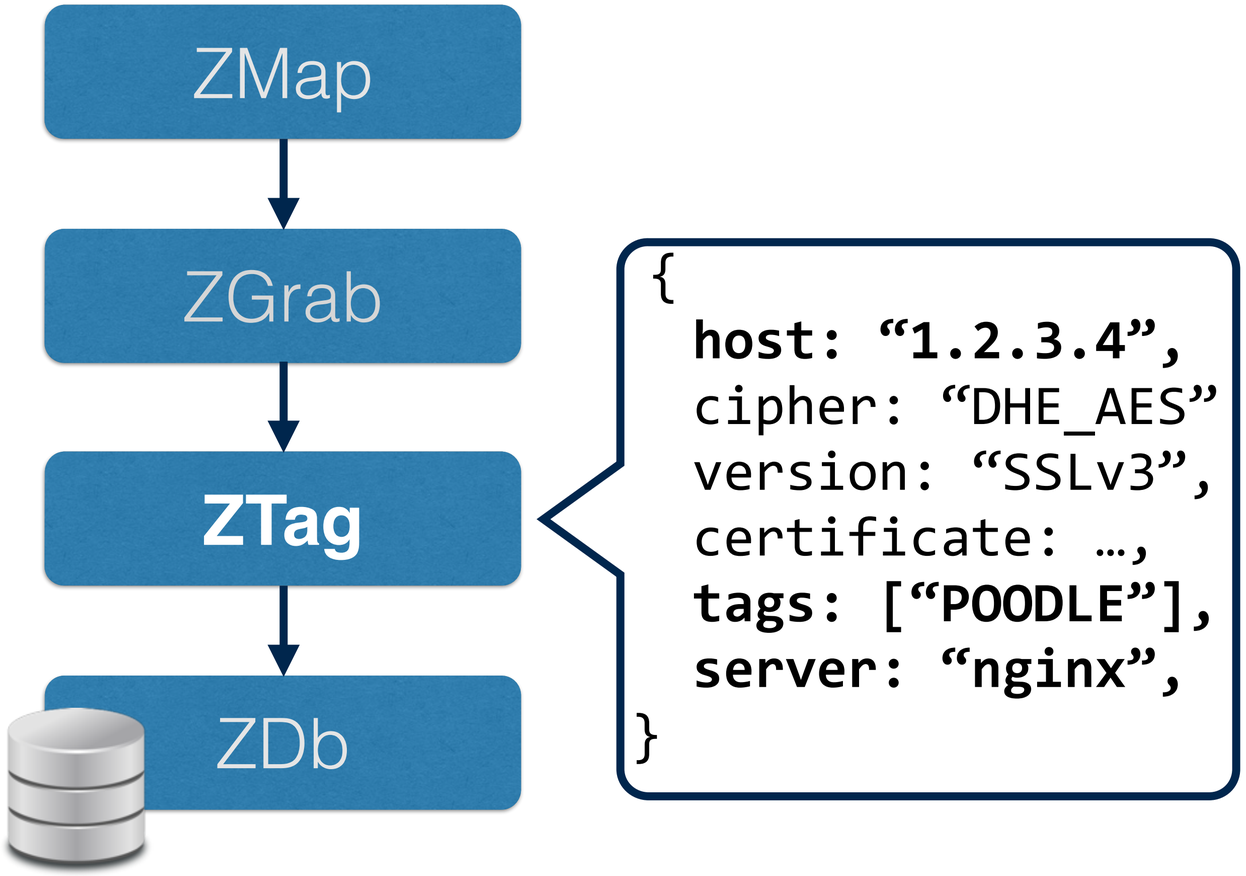
Data Collection

1. Identify listening hosts
2. **Gather application-layer data**
3. Annotate with additional metadata
4. Aggregate by host



Data Collection

1. Identify listening hosts
2. Gather application-layer data
3. **Annotate with additional metadata**
4. Aggregate by host



Data Collection

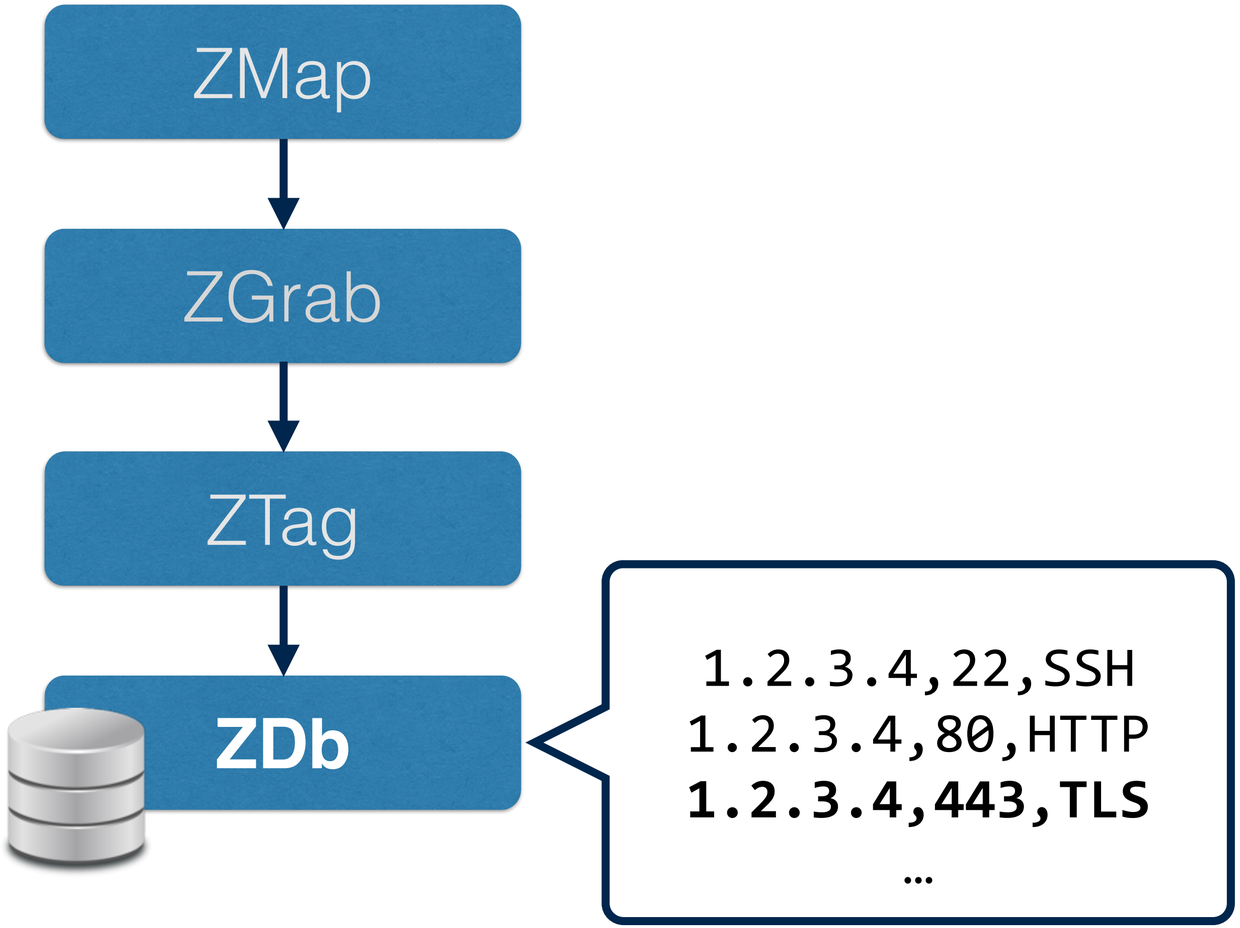
Annotations are simple Python functions

```
class CiscoServer(Annotation):

  protocol = protocols.HTTP

  def process(self, obj, meta):
      server = obj["headers"]["server"]
      if "cisco" in server.lower():
          meta.manufacturer = "Cisco"
          return meta
```

https://github.com/zmap/ztag

Data Collection

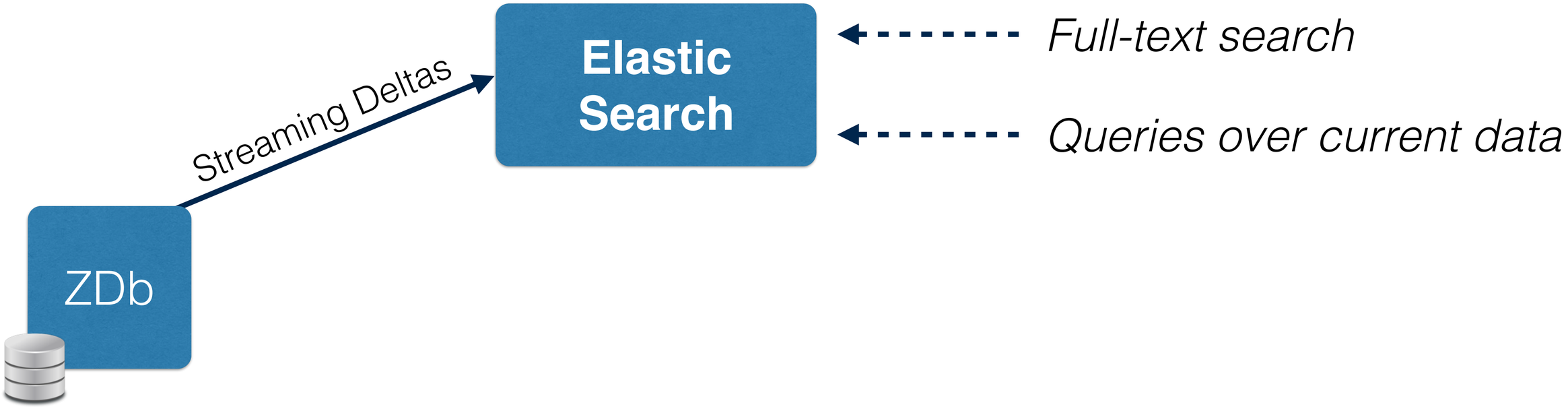
1. Identify listening hosts
2. Gather application-layer data
3. Annotate with additional metadata
4. **Aggregate by host**



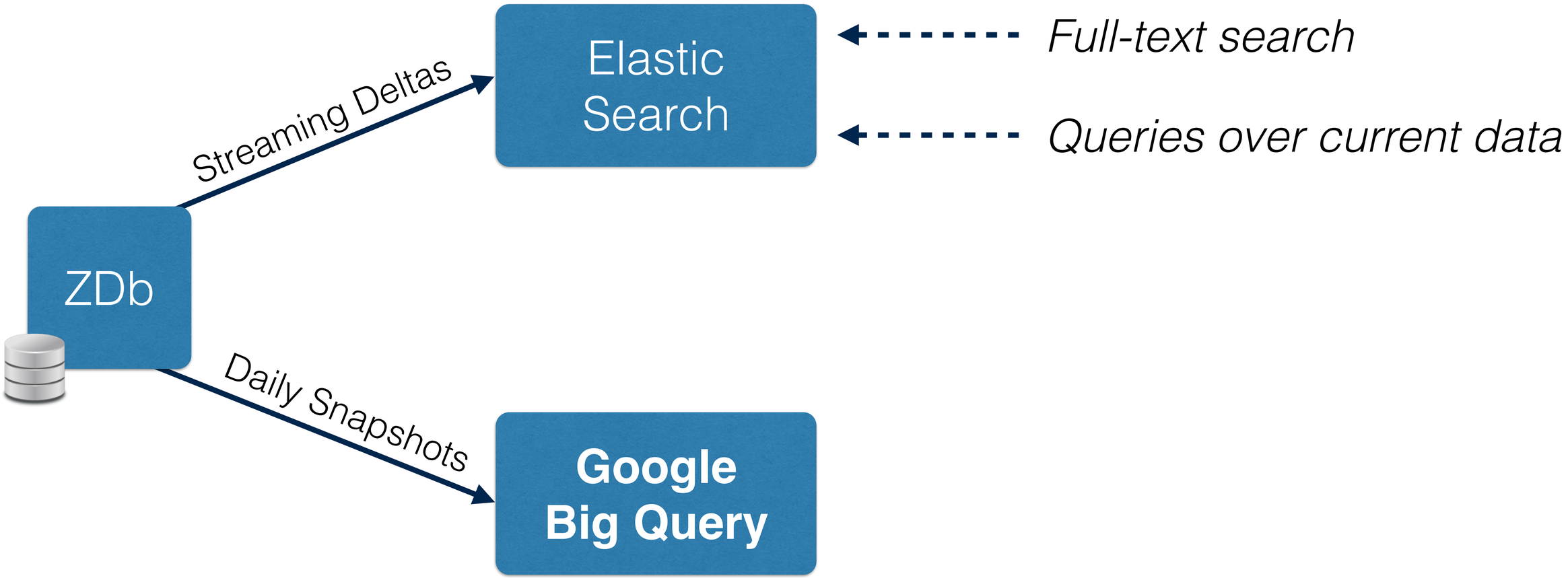
Querying

Querying

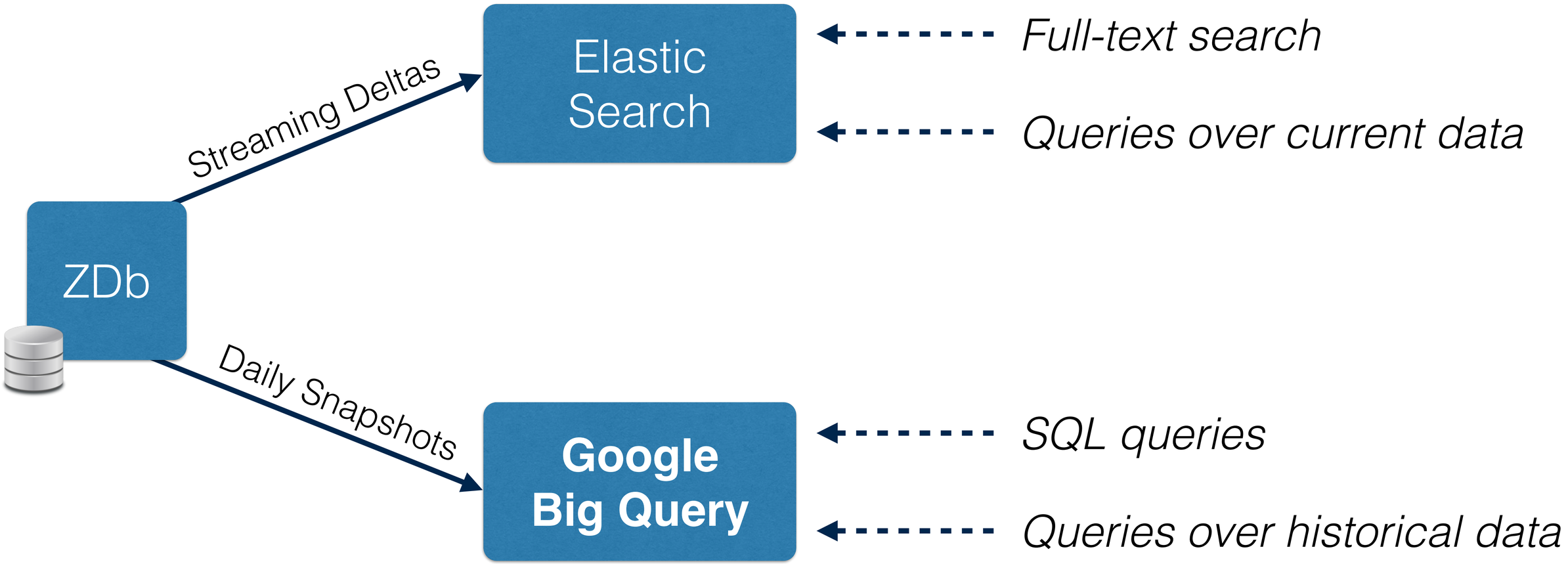
Querying


Elastic Search
Streaming Deltas
ZDb
Full-text search
Queries over current data

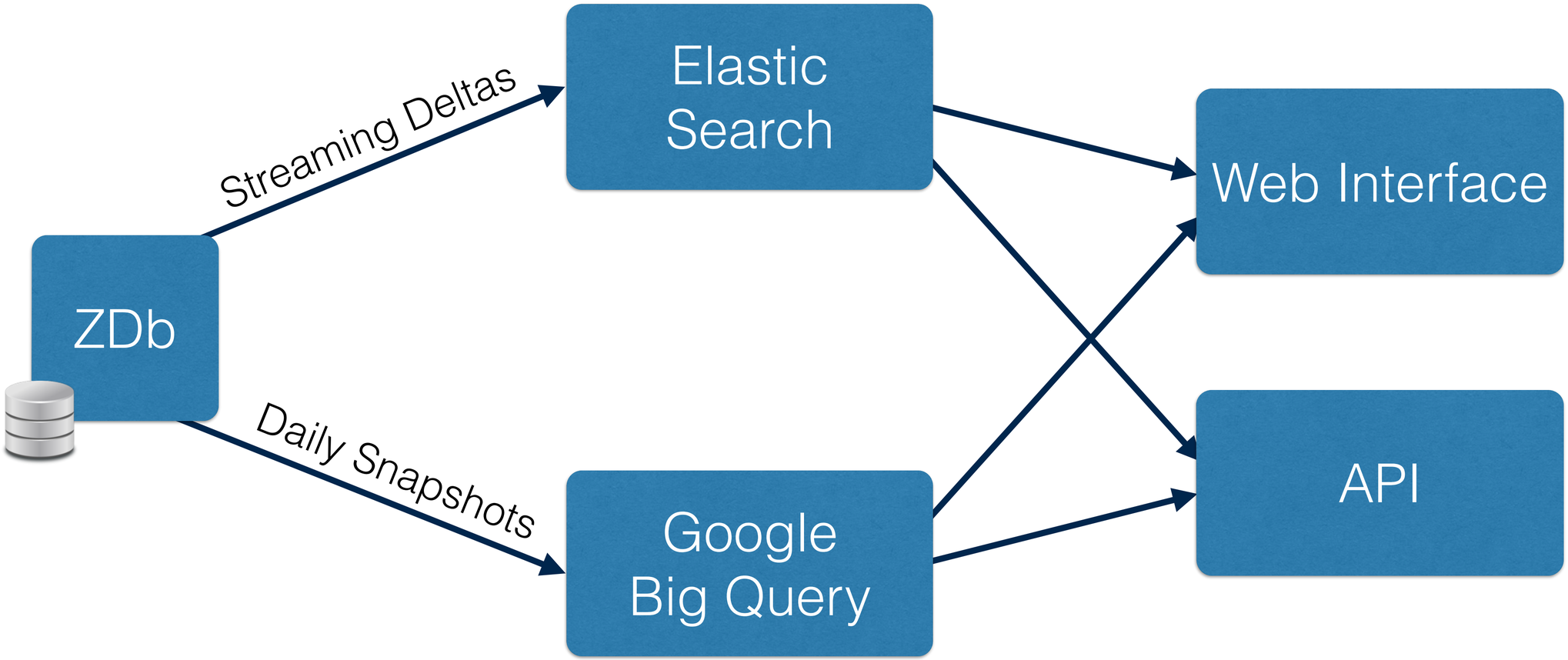
Querying


Elastic Search
Full-text search
Queries over current data
Streaming Deltas
ZDb
Daily Snapshots
Google Big Query

Querying


Elastic Search
Full-text search
Queries over current data
Streaming Deltas
ZDb
Daily Snapshots
Google Big Query
SQL queries
Queries over historical data

Querying


Streaming Deltas
Elastic Search
Web Interface
ZDb
Daily Snapshots
Google Big Query
API

Motivation

Architecture

Looking Forward

Censys aims to be **open** and **community-driven**

Contributing

Are you extending ZMap, ZGrab, or another scanner with a new protocol?

Do you have annotations to add to our framework?

We'll work with researchers to add new scan modules to Censys

https://github.com/zmap/zmap

https://github.com/zmap/zgrab

https://github.com/zmap/ztag

Future Research

Censys strives to be **research enabling more research**

Contribute back scanners and annotations — we do the heavy lifting

Bring **measurement-driven security** to a wider audience

Acknowledgements

Google, for providing much of the infrastructure that runs Censys

Ben Burgess, Alishah Chator, Harsha Gotur, Drew Springall

Elie Bursztein, Brad Campbell, Aleksander Durumeric, James Kasten, Kyle Lady, Adam Langley, HD Moore, Pat Pannuto, Paul Pearce, Niels Provos, Mark Schloesser, Eric Wustrow

The many contributors to the ZMap and ZGrab open source projects

censys

https://www.censys.io

team@censys.io

@censysio @davidcadrian

A Search-Engine Backed by Internet-Wide Scanning.

Zakir Durumeric, **David Adrian**, Ariana Mirian, Michael Bailey, J. Alex Halderman