What can we learn from 15 million websites?

Kevin Farrugia

DevFest 2022 - Malta

A brief intro...

Hi, I'm Kevin Farrugia

- Consultant on Web Performance & Frontend Architecture.
- HTTP Archive & Web Almanac contributor.
- Author of the Resource Hints chapter in 2021 Web Almanac.

@imkevdev | @kevinfarrugia@webperf.social | imkev.dev

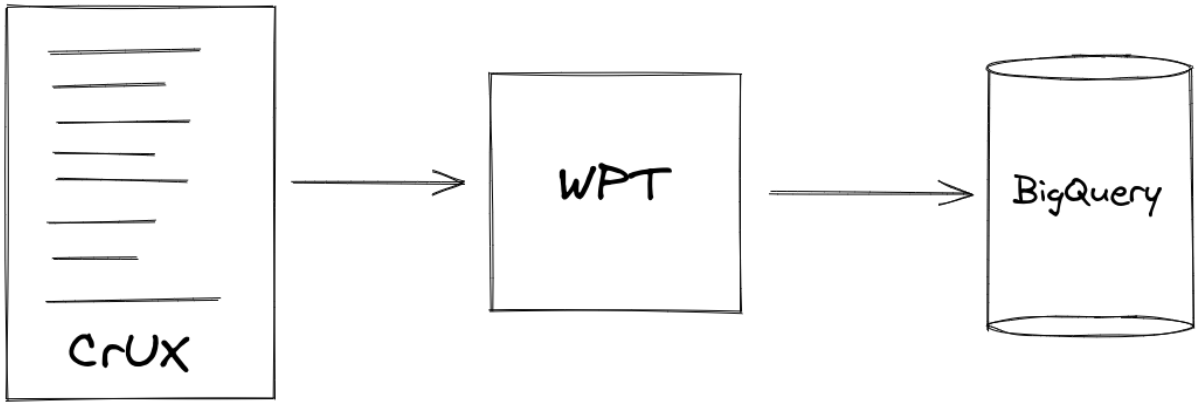
HTTP Archive

“We periodically crawl the top sites on the web and record detailed information about fetched resources, used web platform APIs and features, and execution traces of each page.”

Source: https://httparchive.org/

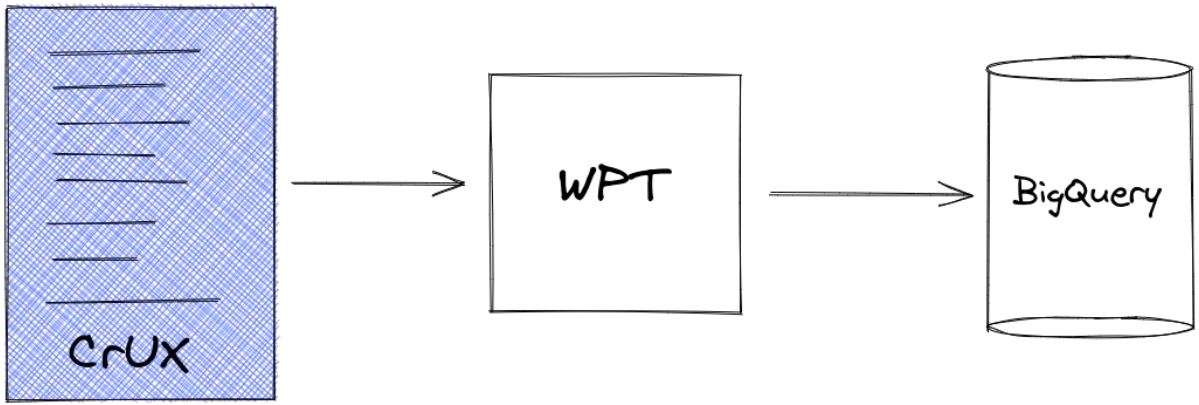
HTTP Archive

“We periodically crawl the top sites on the web and record detailed information about fetched resources, used web platform APIs and features, and execution traces of each page.”

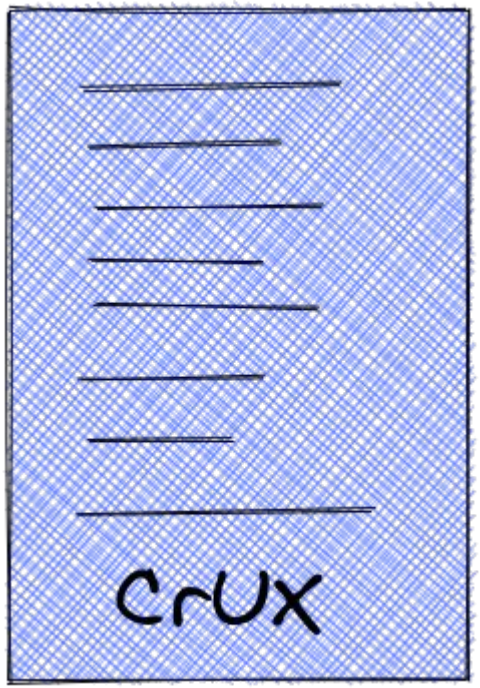


CrUX

“We periodically crawl the top sites on the web and record detailed information about fetched resources, used web platform APIs and features, and execution traces of each page.”



Chrome User Experience Report

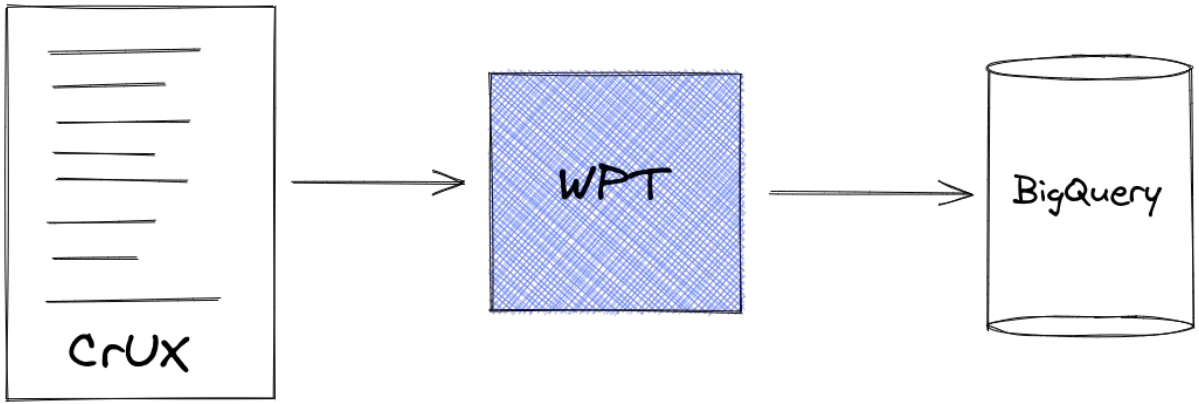


Collected from *real-world* Chrome users.

- BigQuery
- Dashboard
  - E.g. https://timesofmalta.com
- API
  - ```
    curl -s --request POST "https://chromeuxreport.googleapis.com/v1/records:queryRecord?key=${CRUX_API_KEY}" --header 'Accept: application/json' --header 'Content-Type: application/json' --data '{"formFactor":"PHONE","origin":"https://timesofmalta.com","metrics":["largest_contentful_paint"]}'
    ```
    ```

WPT

“We periodically crawl the top sites on the web and record detailed information about fetched resources, used web platform APIs and features, and execution traces of each page.”

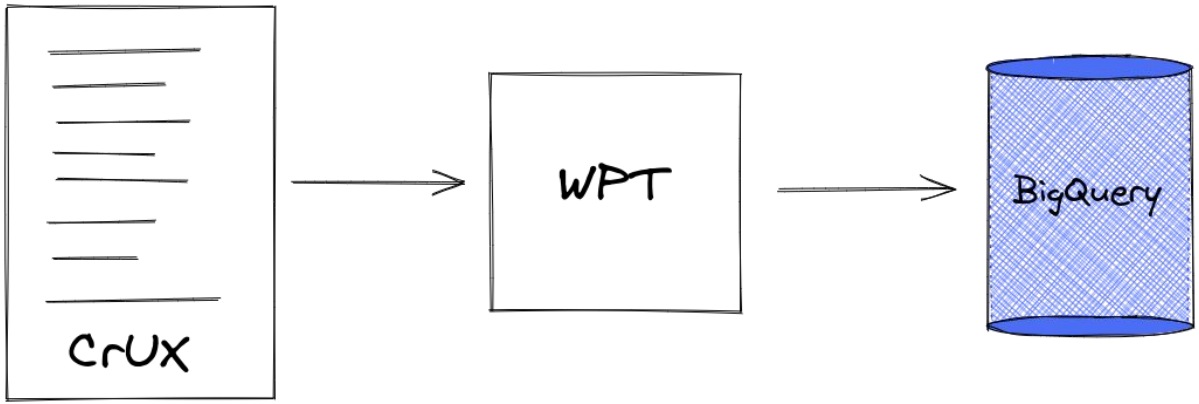


WebPageTest

- Private instance of WebPageTest
  - E.g. https://timesofmalta.com
- Data is augmented using Wappalyzer, Lighthouse, custom metrics and other tools.

BigQuery

“We periodically crawl the top sites on the web and record detailed information about fetched resources, used web platform APIs and features, and execution traces of each page.”



BigQuery


```
SELECT
  COUNT(*)
FROM
  `httparchive.urls.latest_crux_mobile`
LIMIT 1
```

BigQuery


```
SELECT
  COUNT(*)
FROM
  `httparchive.urls.latest_crux_mobile`
LIMIT 1
```

16,784,417

Queries

- Usage:
  - Which JavaScript technology is the most popular?
- Comparison:
  - Which websites have a better LCP - those built using React or those built using Svelte? *
- Correlation:
  - How does the number of preload hints correlate with good LCP? *

* correlation does not imply causation

Hypothesis

- Lighthouse Audits
- Opportunities: new ideas, directives or frameworks
- Recommendations
- The unusual

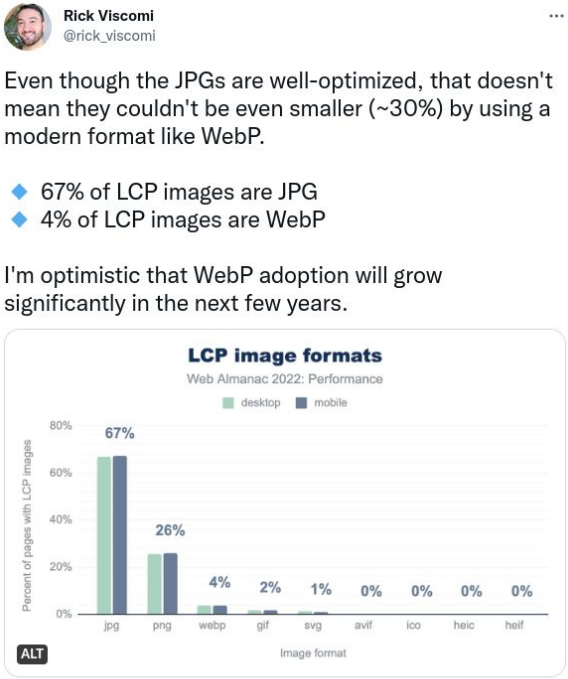
Hypothesis - Preload LCP image

- Preload Largest Contentful Paint image
- Query
  - https://www.anandfurnishers.in/
    - PageSpeed Insights
    - WebPageTest
    - Experiment

Hypothesis - fetchpriority

- Demo
  - Render-blocking scripts
  - fetchpriority=”high”
  - Opportunity: when there is more than one high priority inflight request AND render-blocking scripts
- Query
  - https://greenenergy.nus.edu.sg/
  - WebPageTest
  - Experiment

Hypothesis - WebP vs JPG



Source: @rick_viscomi

Hypothesis - WebP vs JPG

- Query

Hypothesis - Unusual

- Websites downloading React *and* AngularJS
- Query
  - https://www.goneforarun.com/
  - App (AngularJS)
  - ZenDesk's Web Widget (React)

Performance is Accessibility

- *“The mission of web performance is to expand access to information and services on the web.”*

  Source: Alex Russell

Contribute

- HTTP Archive Forums
- Web Almanac
- Web Performance Calendar

Resources

- DevFest 2022
- HTTP Archive
- 2022 Web Almanac
- CrUX documentation
- GitHub - kevinfarrugia/crux_csv
- GitHub - kevinfarrugia/bq-query