

Automated benchmarks for  
FIDO, DROID and siegfried

# Demo!

- <https://www.itforarchivists.com/siegfried/benchmarks> (Release benchmarks)
- <https://www.itforarchivists.com/siegfried/develop> (Development benchmarks)

## iPRES Systems Showcase

A corpus created for the 2014 iPRES conference comprising 2,206 files (5GB). Represents a range of formats, including AV and some uncommon types. Sourced from <http://www.webarchive.org.uk/datasets/ipres.ds.1/>

## RESULTS

Tool	Description	Duration
sf	Siegfried in its default mode (-multi 1).	1m50.055122438s
sf multi32	Siegfried with 'sf -multi 32' (sf scans multiple files simultaneously).	47.671004384s
sf 64k	Siegfried with 'sf -multi 32 -sig pronom64.sig' (a buffer of 64k).	1.601876243s
sf 128k	Siegfried with 'sf -multi 32 -sig pronom128.sig' (a buffer of 128k).	1.828308775s
droid 64k	DROID in its default mode (a buffer of 64K).	24.533826478s
droid 128k	DROID with FIDO's default buffer (a buffer of 128K).	23.193571689s
droid nolimit	DROID without a buffer limit (a buffer of -1).	36.949580584s
fido 64k	FIDO with DROID's default buffer (a -bufsize of 64K).	2m17.196926149s
fido 128k	FIDO in its default mode (a -bufsize of 128K).	2m27.335909229s
fido 1mb	FIDO with Archivemata's default buffer (a -bufsize of 1mb).	3m10.118521713s

The tools differed in output for 370 files in the corpus. 18 of those differences are because of siegfried's use of a text identification algorithm and the following chart excludes those files.

Show 10 entries

Search: [illegible]

# Unzipping

This benchmark checks the `sf -z` command (scans within zip files and other container formats) when run against the iPres corpus.

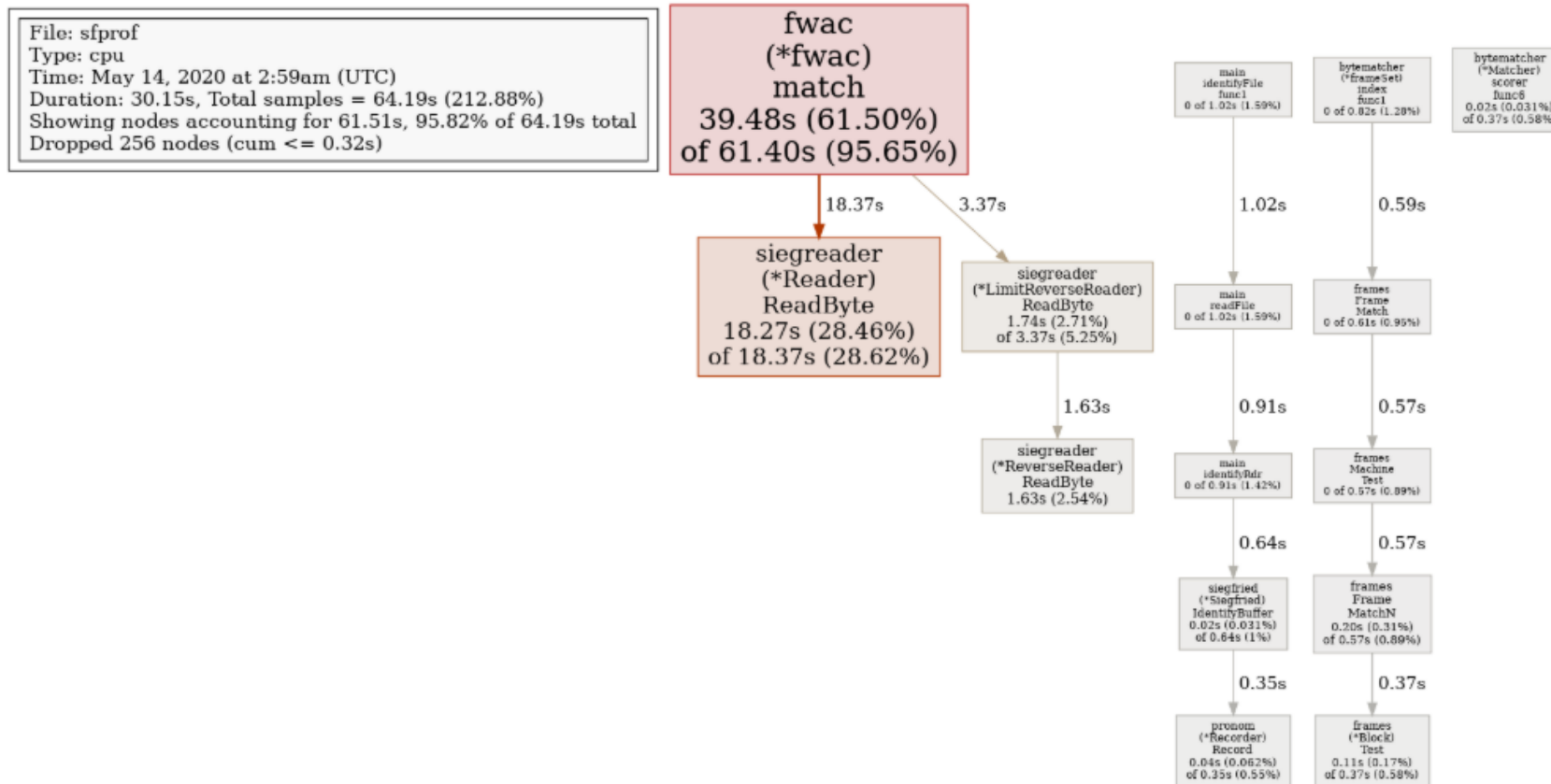
## RESULTS

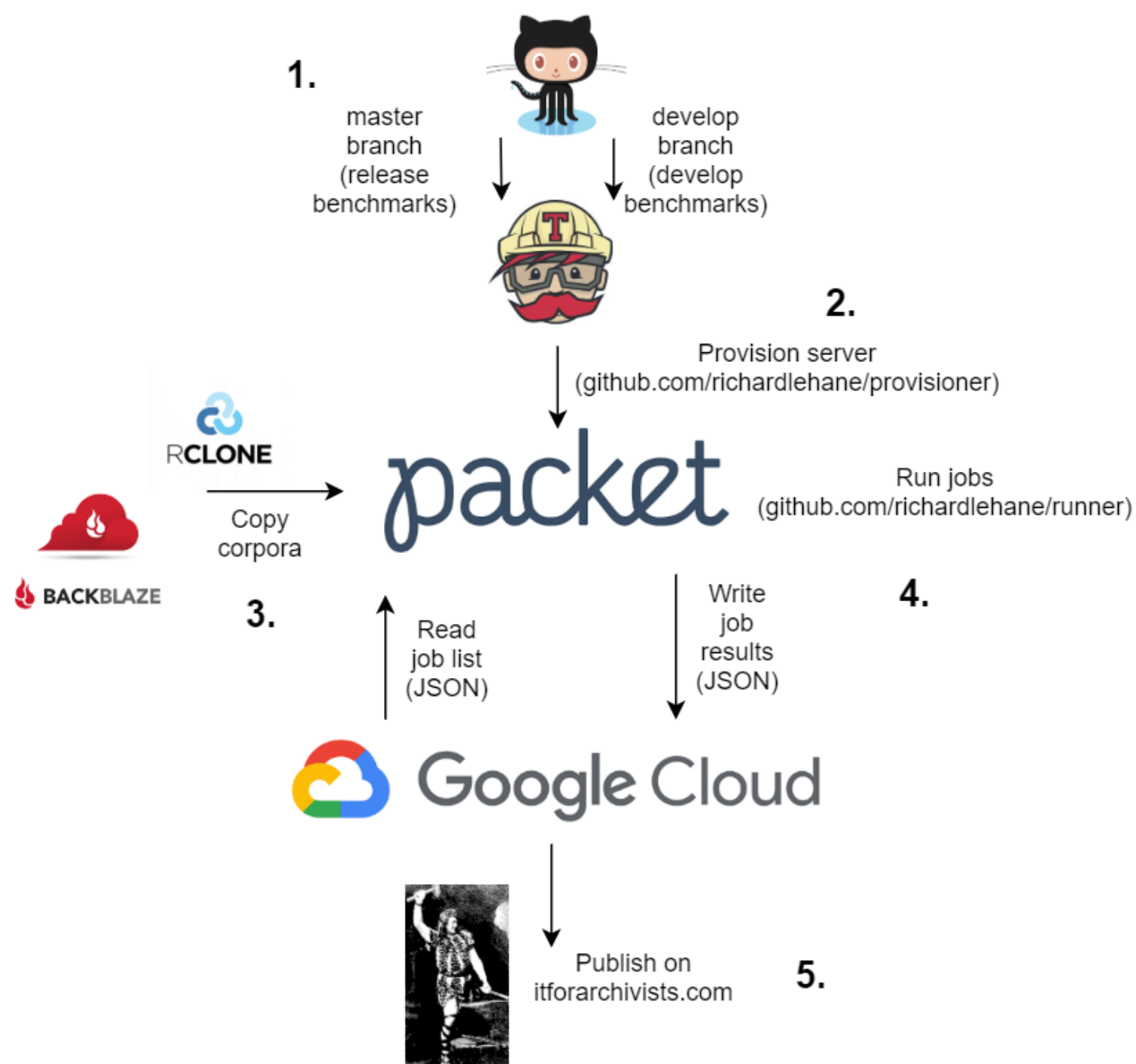
Tool	Description	Duration
master	Master branch of <a href="https://github.com/richardlehane/siegfried">github.com/richardlehane/siegfried</a> . Corresponds to latest production release.	2m25.716604378s
develop	Develop branch of <a href="https://github.com/richardlehane/siegfried">github.com/richardlehane/siegfried</a> . Tip of development and potentially unstable.	2m25.805805317s

The tools differed in output for 0 files in the corpus.

[Raw output](#)

## Profile





# Pros

- Transparent
  - Server specs available on [packet.com](https://packet.com)
  - Code on [github inc. server config](https://github.com/github/provisioner) in provisioner repo
  - Tasks published as JSON on [itforarchivists.com](https://itforarchivists.com)
  - Unfiltered results immediately available in raw JSON and views
  - Shows history too
- Automated
  - Not dependent on local setup
  - Doesn't use local compute
  - Collaborators
  - More that is automated, simpler it is to maintain
- Cheap
- Confidence

# Cons

- Fragile
  - Tools can break it
  - Cloud services can break it