Shareable Benchmarking Reports with Enhanced Filters and Dynamic Statistics for BenchExec

Bachelor Thesis by Dennis Simon

28.04.2021

Software and Computational Systems at the Ludwig-Maximilians-Universität München

Introduction

BenchExec:

- Benchmarking tool developed by the SoSy Lab of the LMU Munich
- One of its core tools is the Table-Generator
  - Used to create HTML result tables

Summary | Table | Quantile Plot | Scatter Plot | Info

Showing 46 of 46 tasks

Tool	CBMC	CPAchecker 1.4-svn 18912M
Limits	timelimit: 60 s, memlimit: 4000 MB, CPU core limit: 2	
Host	tortuga	
OS	Linux 3.13.0-71-generic x86_64	
System	CPU: Intel Core i7-2600 CPU @ 3.40GHz, cores: 8, frequency: 3401 MHz, Turbo Boost: enabled; RAM: 16783 MB	
Date of execution	2015-12-11 12:11:02 CET	2015-12-11 10:59:27 CET
Run set	cbmc	predicateAnalysis.ABEl

Options		`-heap 13000M`
		`-noout`
		`-disable-java-assertions`
		`-setprop cpa.predicate.memoryAllocationsAlwaysSucceed=true`
		`-predicateAnalysis-PredAbsRefiner-ABEl`

Properties	unreach-call	

Statistics

Fixed row title: ☑	CBMC 2015-12-11 12:11:02 CET cbmc				CPAchecker 2015-12-11 10:59:27 CET predicateAnalysis.ABEl			
Click here to select columns	status	cputime (s)	walltime (s)	memUsage (MB)	status	cputime (s)	walltime (s)	memUsage (MB)
total	46	1590	1600	28800	46	1160	691	38700
local summary	-	148	1620	-	-	1160	697	-
correct results	22	147	148	6040	40	903	518	25200
correct true	0	-	-	-	18	533	312	18200
correct false	22	147	148	6040	22	370	206	7030
incorrect results	0	-	-	-	0	-	-	-
incorrect true	0	-	-	-	0	-	-	-
incorrect false	0	-	-	-	0	-	-	-

Generated by BenchExec 3.7

Exemplary HTML table generated by the Table Generator tool – Summary tab

Summary | Table | Quantile Plot | Scatter Plot | Info

Showing 46 of 46 tasks

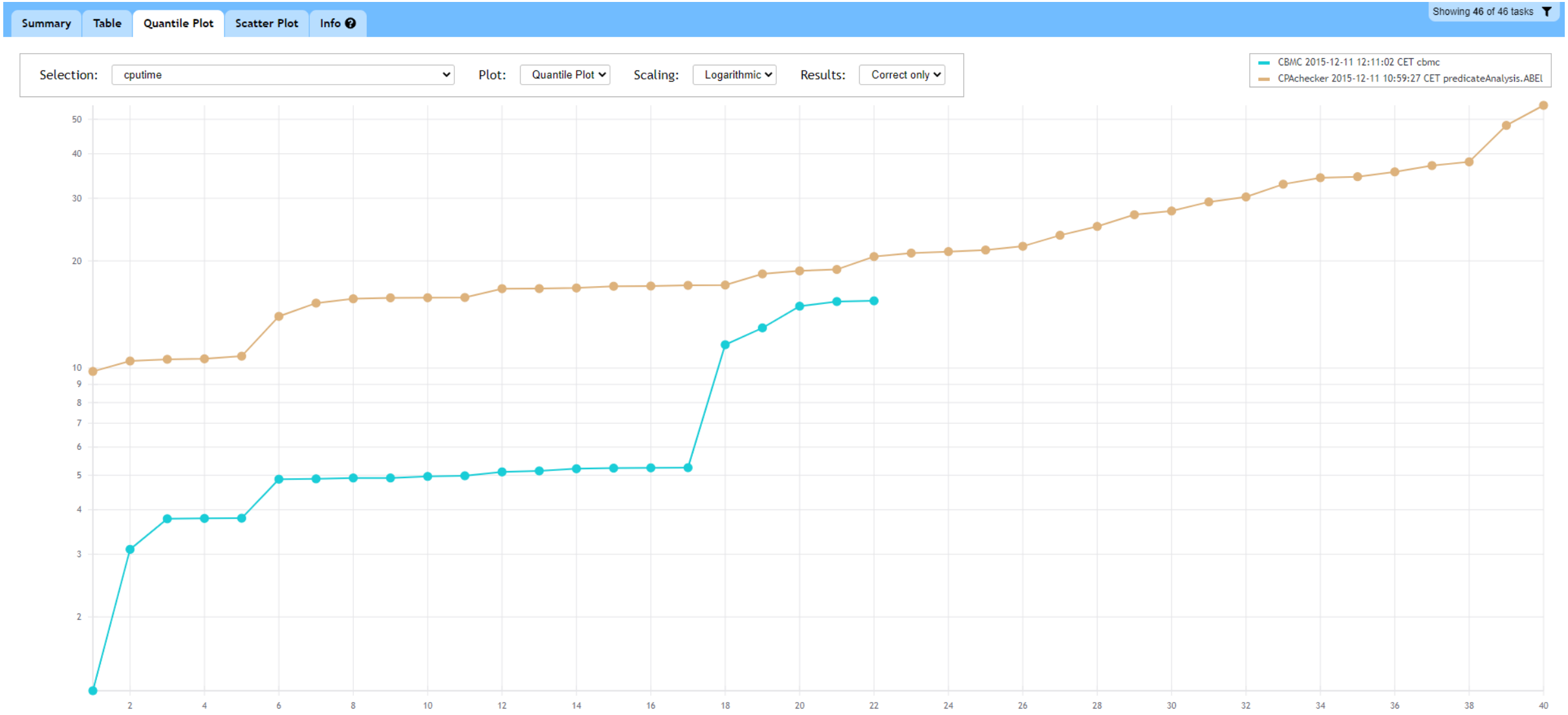
Fixed task: ☑

Click here to select columns

	CBMC 2015-12-11 12:11:02 CET cbmc				CPAchecker 2015-12-11 10:59:27 CET predicateAnalysis.ABEl			
	status	cputime (s)	walltime (s)	memUsage (MB)	status	cputime (s)	walltime (s)	memUsage (MB)
text	Show all	Min:Max	Min:Max	Min:Max	Show all	Min:Max	Min:Max	Min:Max
ntdrivers/cdaudio_false-unreach-call.i.cil.c	TIMEOUT	60.1	60.1	987	TIMEOUT	60.7	38.8	3070
ntdrivers/diskperf_false-unreach-call.i.cil.c	false(reach)	1.24	1.26	104	false(reach)	15.8	9.00	256
ntdrivers/floppy_false-unreach-call.i.cil.c	false(reach)	3.10	3.11	140	false(reach)	22.0	11.9	439
ntdrivers/kbfiltr_false-unreach-call.i.cil.c	TIMEOUT	60.1	60.1	573	false(reach)	10.8	6.08	219
ntdrivers/parport_false-unreach-call.i.cil.c	false(reach)	13.0	13.3	676	ERROR (recursion)	10.2	5.94	249
ntdrivers/cdaudio_true-unreach-call.i.cil.c	TIMEOUT	60.1	60.4	993	true	17.0	9.18	295
ntdrivers/diskperf_true-unreach-call.i.cil.c	TIMEOUT	60.2	60.1	3500	true	15.7	8.89	273
ntdrivers/floppy2_true-unreach-call.i.cil.c	TIMEOUT	60.1	60.0	989	TIMEOUT	60.9	39.7	3010
ntdrivers/floppy_true-unreach-call.i.cil.c	TIMEOUT	60.0	60.4	391	true	35.6	20.2	1260
ntdrivers/parport_true-unreach-call.i.cil.c	TIMEOUT	61.1	61.7	2180	ERROR (recursion)	9.85	6.13	251
ssh/s3_clnt.blast.01_false-unreach-call.i.cil.c	false(reach)	11.6	12.0	411	false(reach)	15.2	8.35	268
ssh/s3_clnt.blast.02_false-unreach-call.i.cil.c	false(reach)	3.79	4.15	182	false(reach)	17.1	9.61	270
ssh/s3_clnt.blast.03_false-unreach-call.i.cil.c	false(reach)	3.78	3.82	182	false(reach)	16.7	9.32	269
ssh/s3_clnt.blast.04_false-unreach-call.i.cil.c	false(reach)	3.77	3.79	180	false(reach)	15.8	8.98	269
ssh/s3_srvr.blast.01_false-unreach-call.i.cil.c	false(reach)	4.91	4.94	242	false(reach)	10.5	6.20	197
ssh/s3_srvr.blast.02_false-unreach-call.i.cil.c	false(reach)	4.87	4.90	229	false(reach)	10.6	5.98	253
ssh/s3_srvr.blast.03_false-unreach-call.i.cil.c	false(reach)	4.89	4.90	229	false(reach)	9.79	5.56	249
ssh/s3_srvr.blast.04_false-unreach-call.i.cil.c	false(reach)	4.91	4.96	228	false(reach)	10.6	6.24	256
ssh/s3_srvr.blast.06_false-unreach-call.i.cil.c	false(reach)	5.25	5.27	231	false(reach)	16.7	9.38	275
ssh/s3_srvr.blast.07_false-unreach-call.i.cil.c	false(reach)	5.11	5.14	234	false(reach)	17.0	9.30	272
ssh/s3_srvr.blast.08_false-unreach-call.i.cil.c	false(reach)	15.4	15.4	461	false(reach)	29.3	16.4	939
ssh/s3_srvr.blast.09_false-unreach-call.i.cil.c	false(reach)	5.14	5.17	233	false(reach)	15.8	8.87	263
ssh/s3_srvr.blast.10_false-unreach-call.i.cil.c	false(reach)	14.9	15.0	459	false(reach)	25.0	13.6	419
ssh/s3_srvr.blast.11_false-unreach-call.i.cil.c	false(reach)	5.24	5.26	234	false(reach)	21.2	11.3	302

Previous | Page 1 of 1 | 250 rows | Next

Exemplary HTML table generated by the Table Generator tool – Table tab



Exemplary HTML table generated by the Table Generator tool – Quantile plot tab

Motivation

Filters:

- Filters only accessible in table tab
- No multiselect for enumerable values

Statistics:

- not reacting to change

Motivation

Usability:

- No navigation history is being kept
- No possibility to link to specific tab

Solution (Overview)

Filtering

- Creation of new filter algorithm
- **Introduction of a new, globally accessible UI**
- **Enable multiselect**
- More granular task ID filters

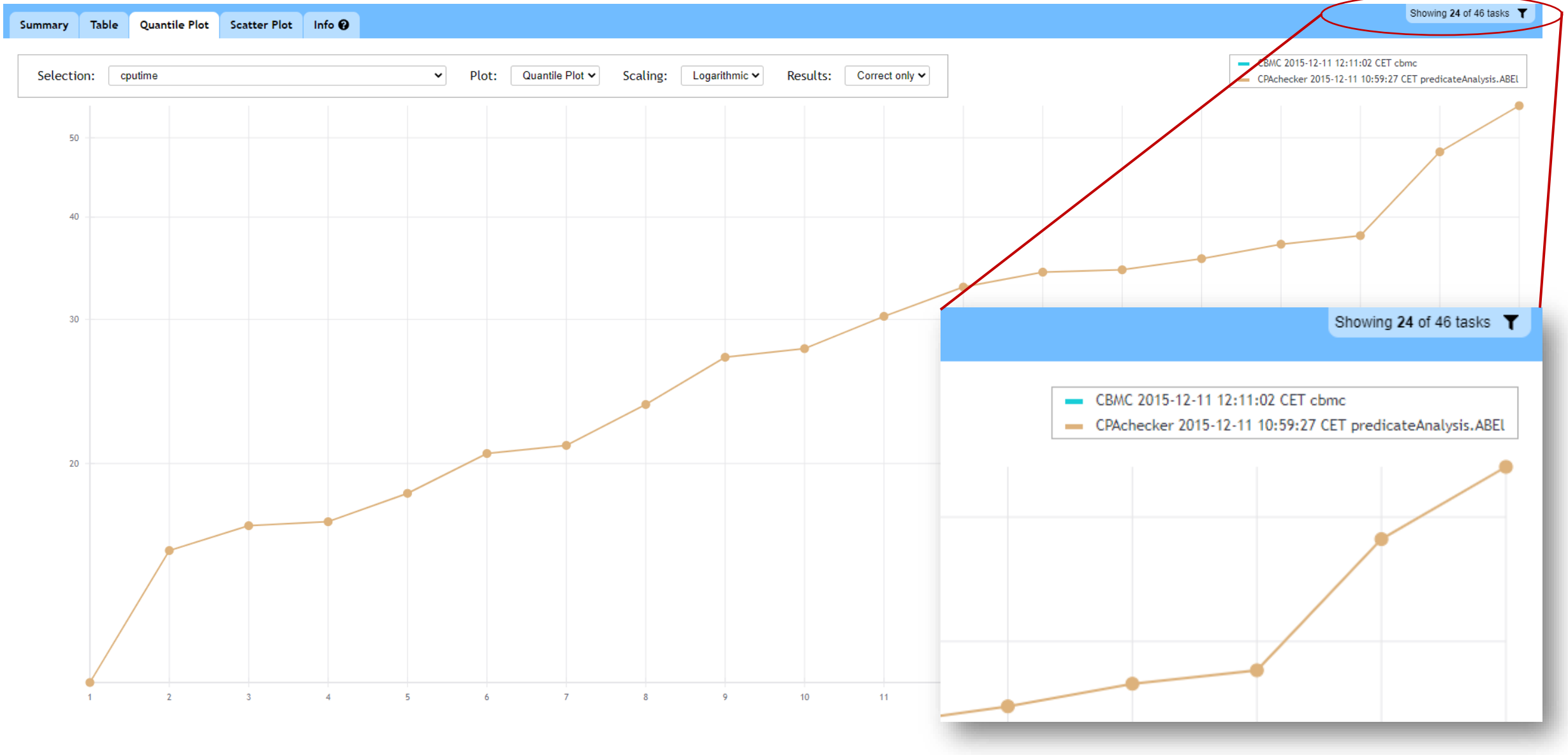
Statistics

- Automatic (re)calculation on filtering
- **Calculation task delegation to pooled web workers**

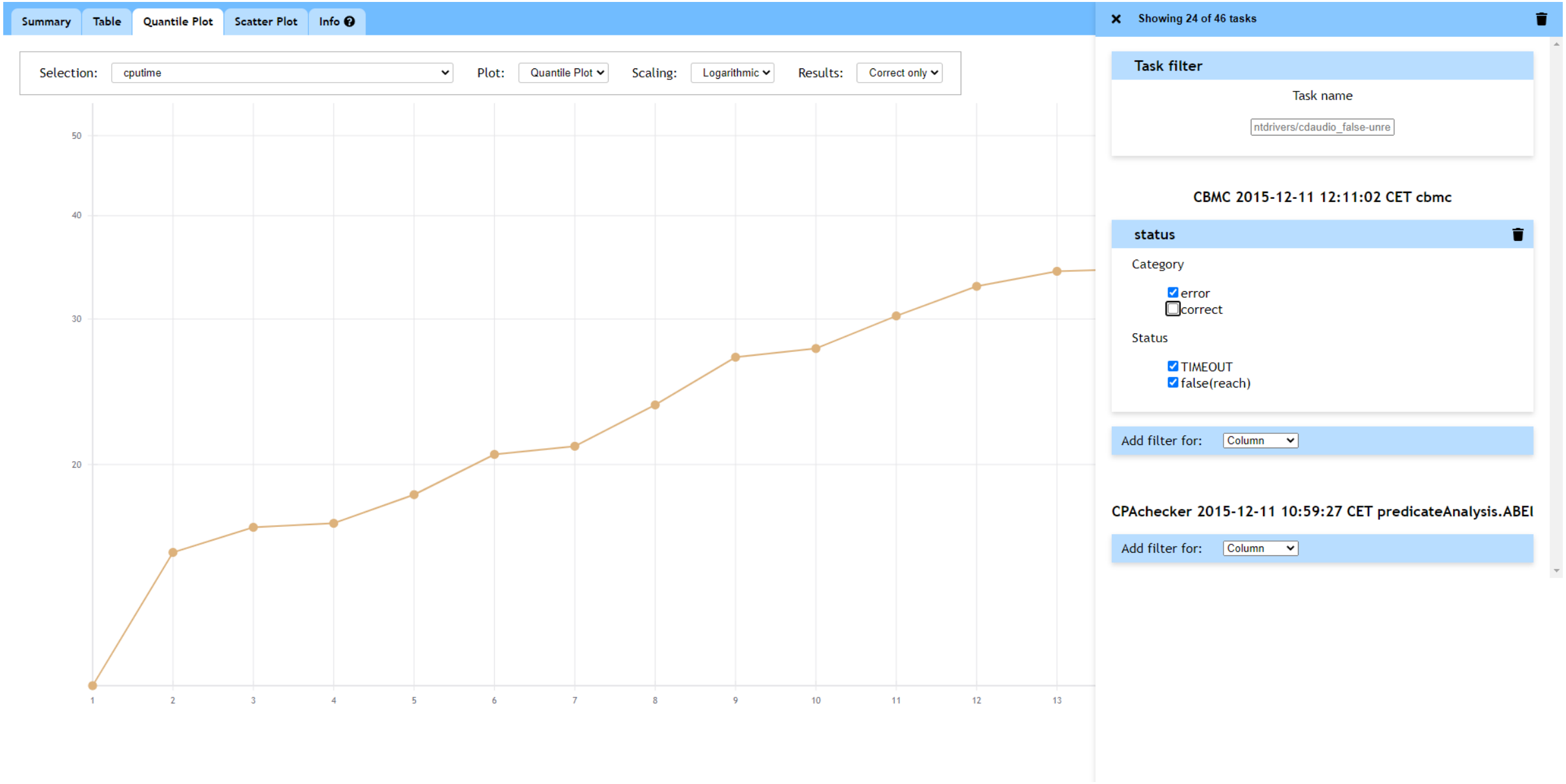
Usability

- Introduction of navigation history handling
- **Navigation via hash routing**
- **Application state serialized in the URL**

Solution in detail (Filters)

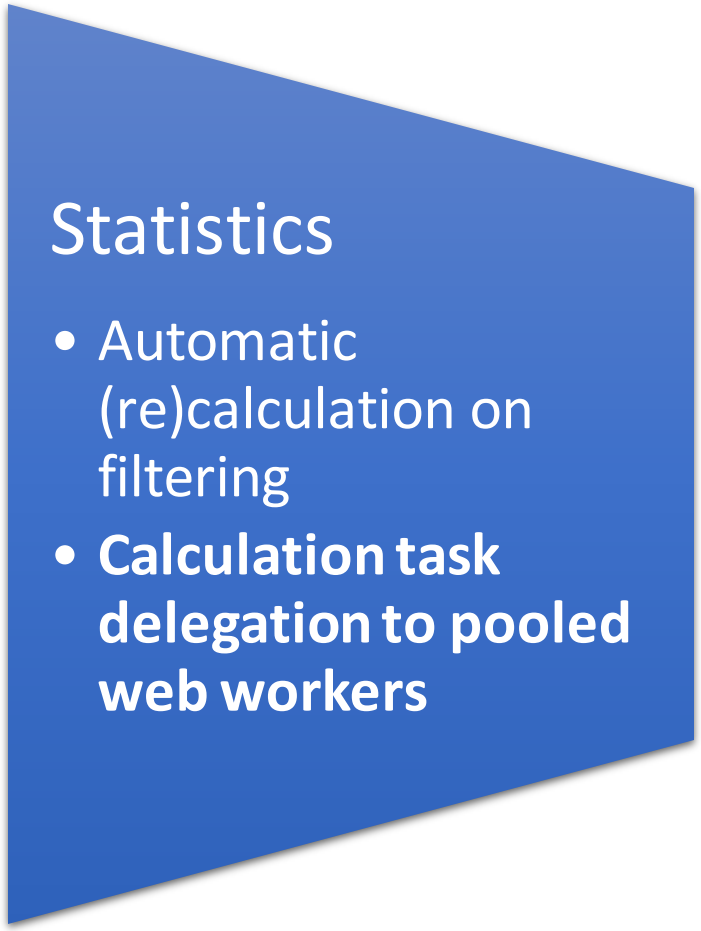


Permanently accessible filter button



Multi-select enabled filters

Solution in detail (Statistics)



Solution in detail (Statistics)

Challenge:

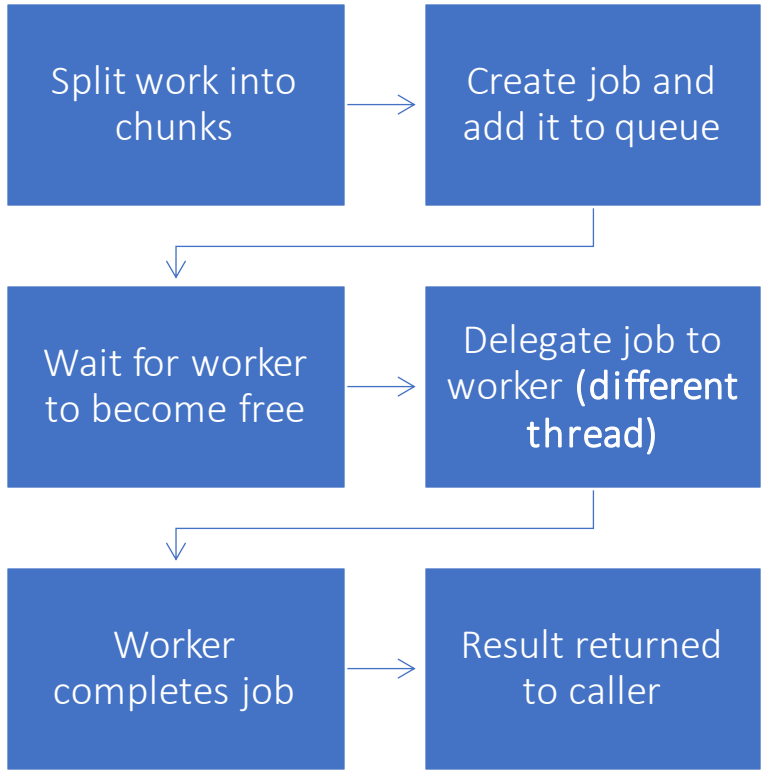
JavaScript is single-threaded
→ UI might freeze under high workload

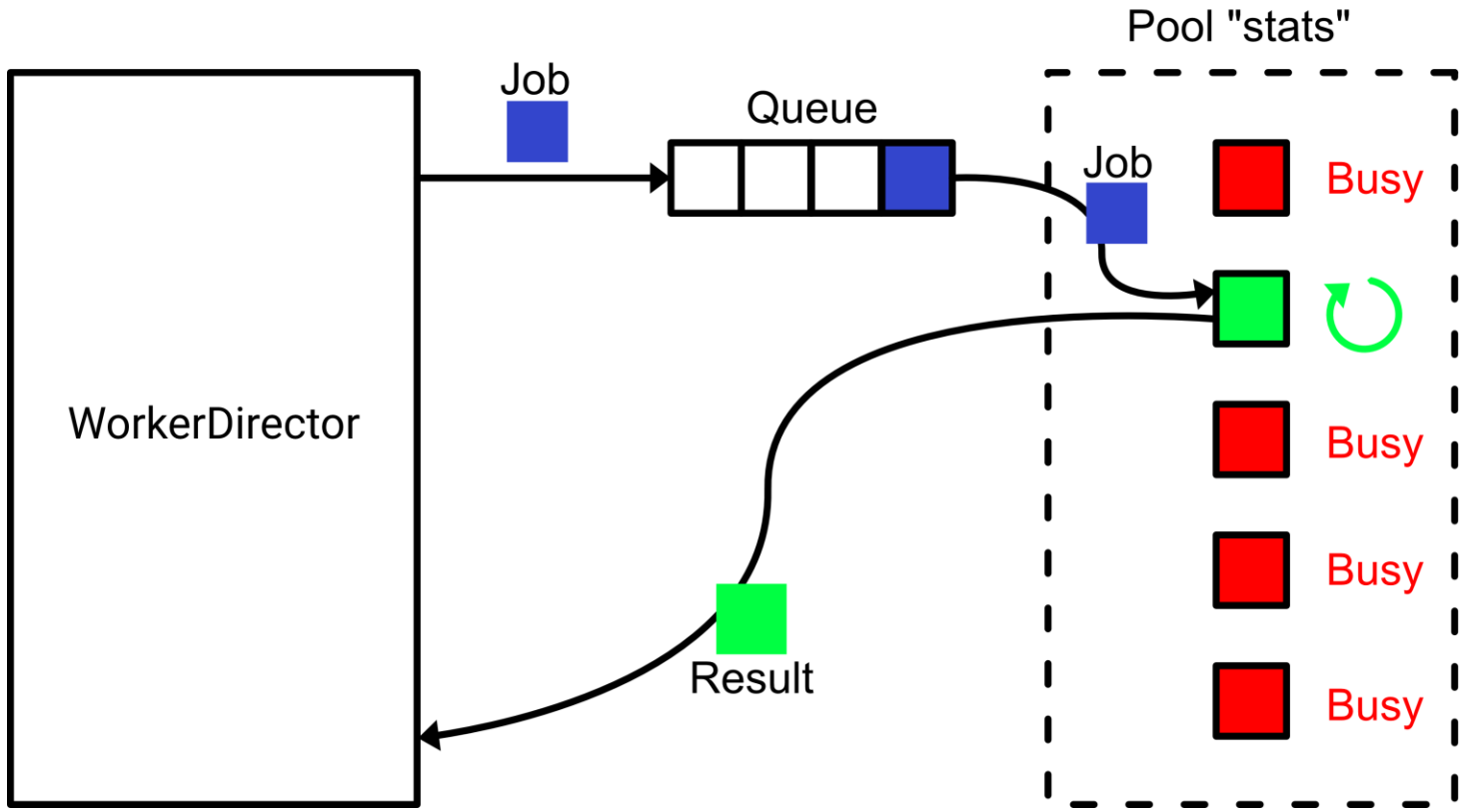
Solution in detail (Statistics)

Solution:

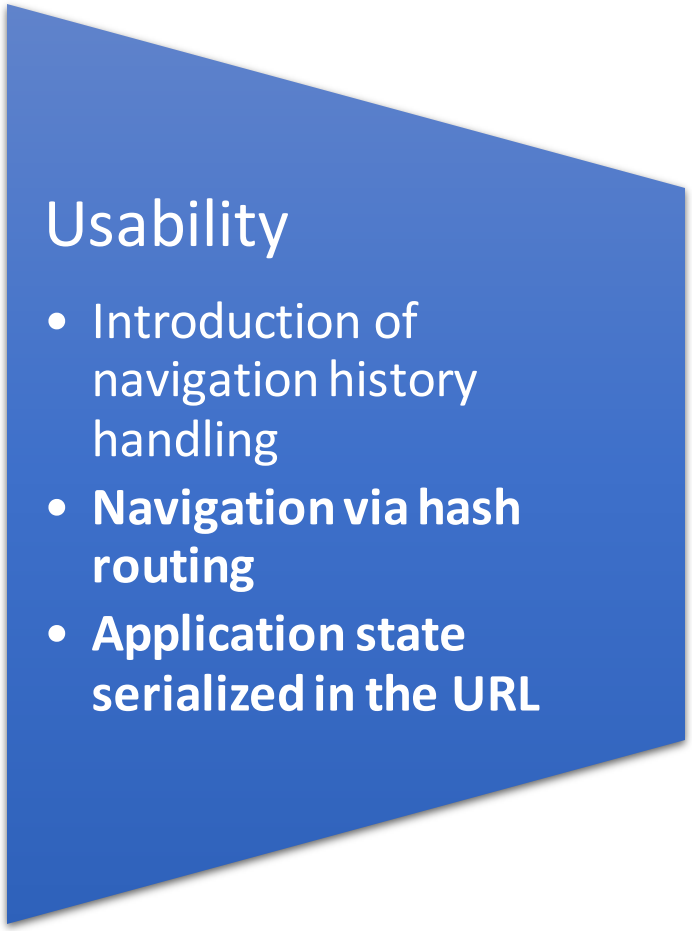
Offloading via pooled **Web-Workers**

Solution in detail (Statistics)





Solution in detail (Usability)



Solution in detail (Usability)

/table.html

navigation

/table.html**#/table**

active tab

Hash routing

Solution in detail (Usability)

#/quantile*?scaling=Linear*

active tab Quantile plot configuration

State serialization

Solution in detail (Usability)

#/quantile*?filter=0(0*status*(category(in(correct))))*

active tab

Serialized filter

State serialization

Solution in detail (Usability)

State is **read** from URL **during navigation**

Shareable application state

Solution in detail (Usability)

Navigation, filtering, configuring now result in changed browser history

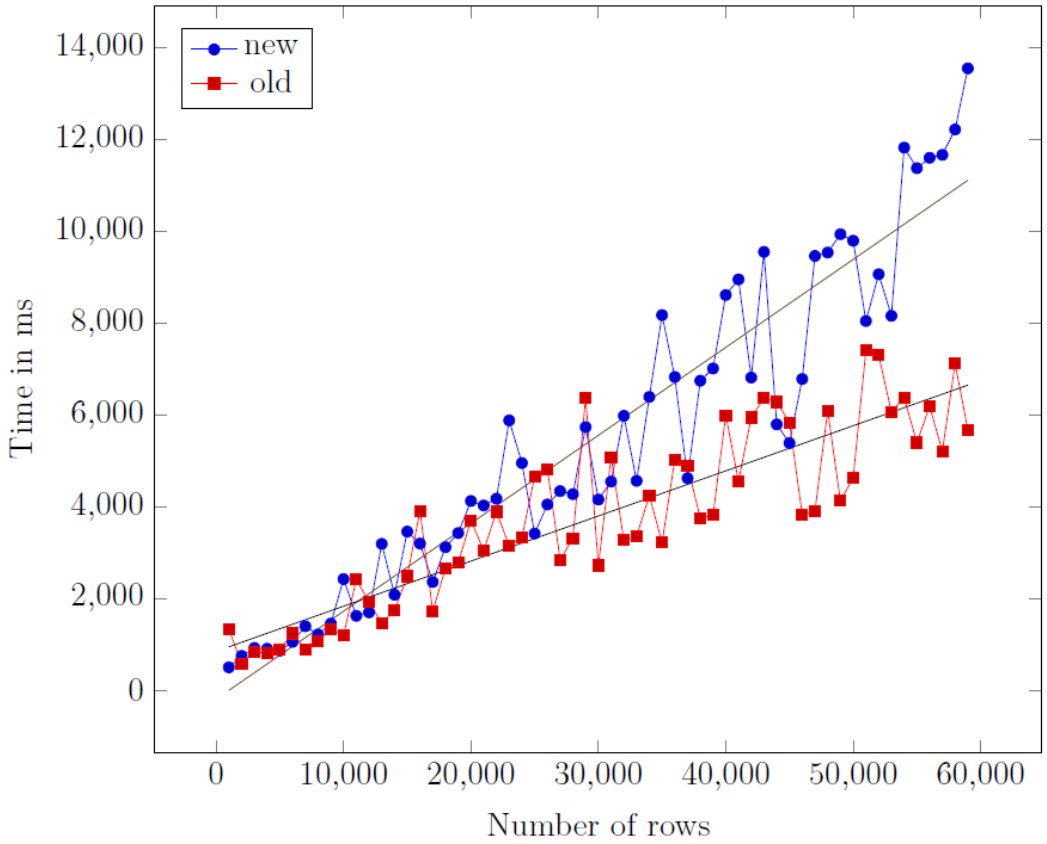
Results

Result was measured using benchmarks:

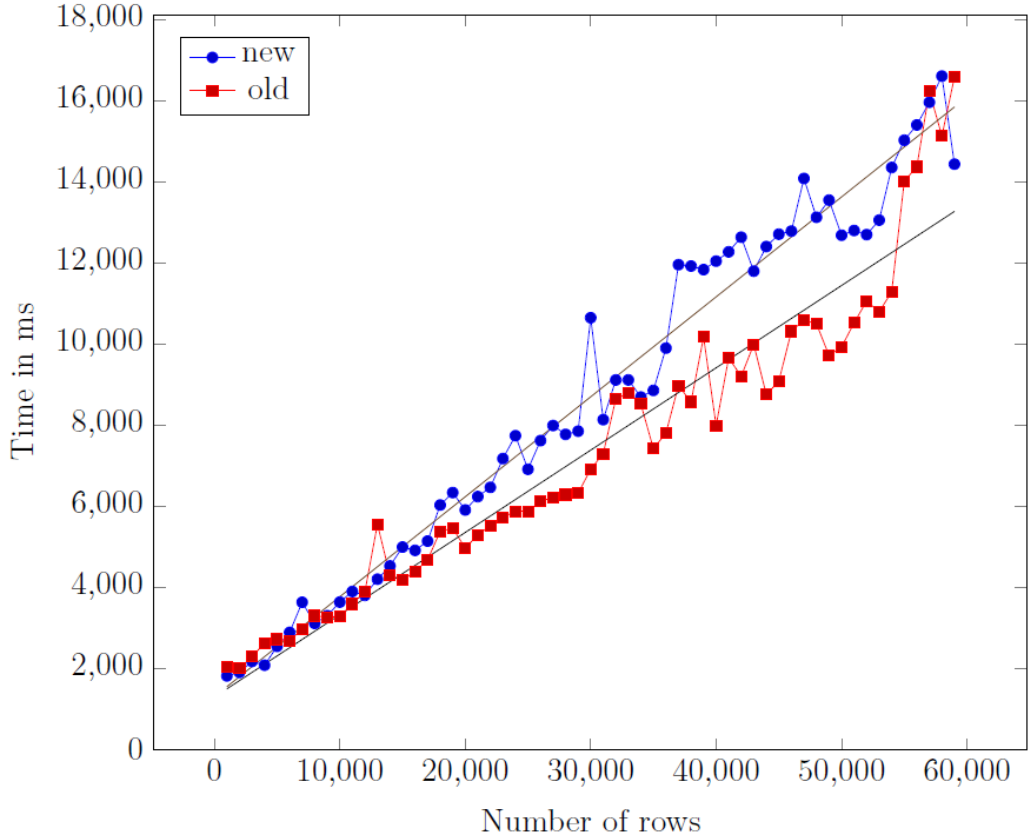
- Comparison: before and after changes
- Tables sized 10 columns * number of rows
  - Number of rows: from 1000 to 60 000
- Timings averaged over 10 executions
- Executed using **Cypress**

Results

Numeric filters

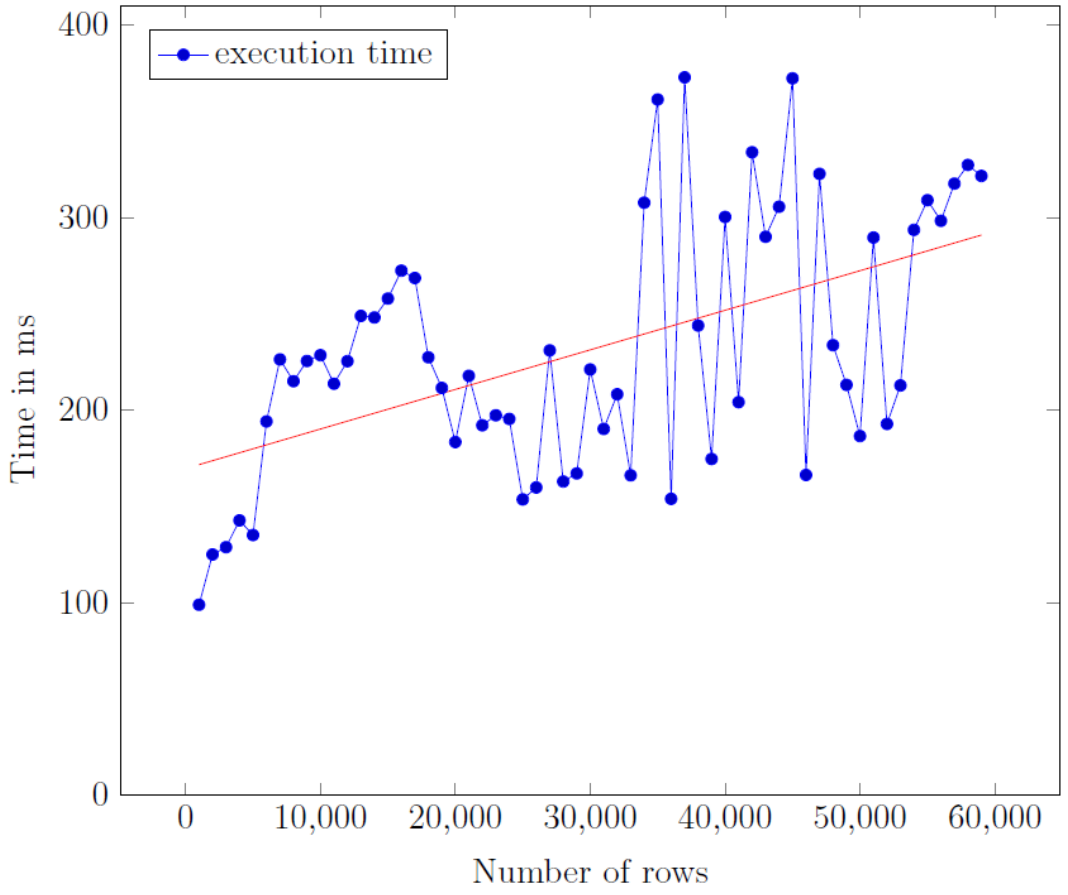


Enumerable filters



Average filter application times per number of rows
(including render time)

Results



Average computation times of statistics (including render time)

Results

Additional time penalty added to filtering

but

- Enabling multiselect
- Allowing for improvements and extensions in the future

Conclusion

- ✔ New way to interact with the application
- ✔ Implementation of requested features
- ✔ Implemented filter and worker modules
- ✔ ↳ Can be extended or built on

Conclusion

In the future:

- Optimization of table rendering
- Offloading of additional work to web-workers
- Extension of filter logic