

# Robinhood Reports: a new Robinhood web front end to help users find their data

Lustre User Group 2019

Houston, TX

May 16, 2019

Shawn Hall

HPC Systems Engineer



Experts in numerical software and  
High Performance Computing

# Link to the code

## ▶ In case you want to follow along

- <https://github.com/numericalalgorithmsgroup/robinhood-reports>

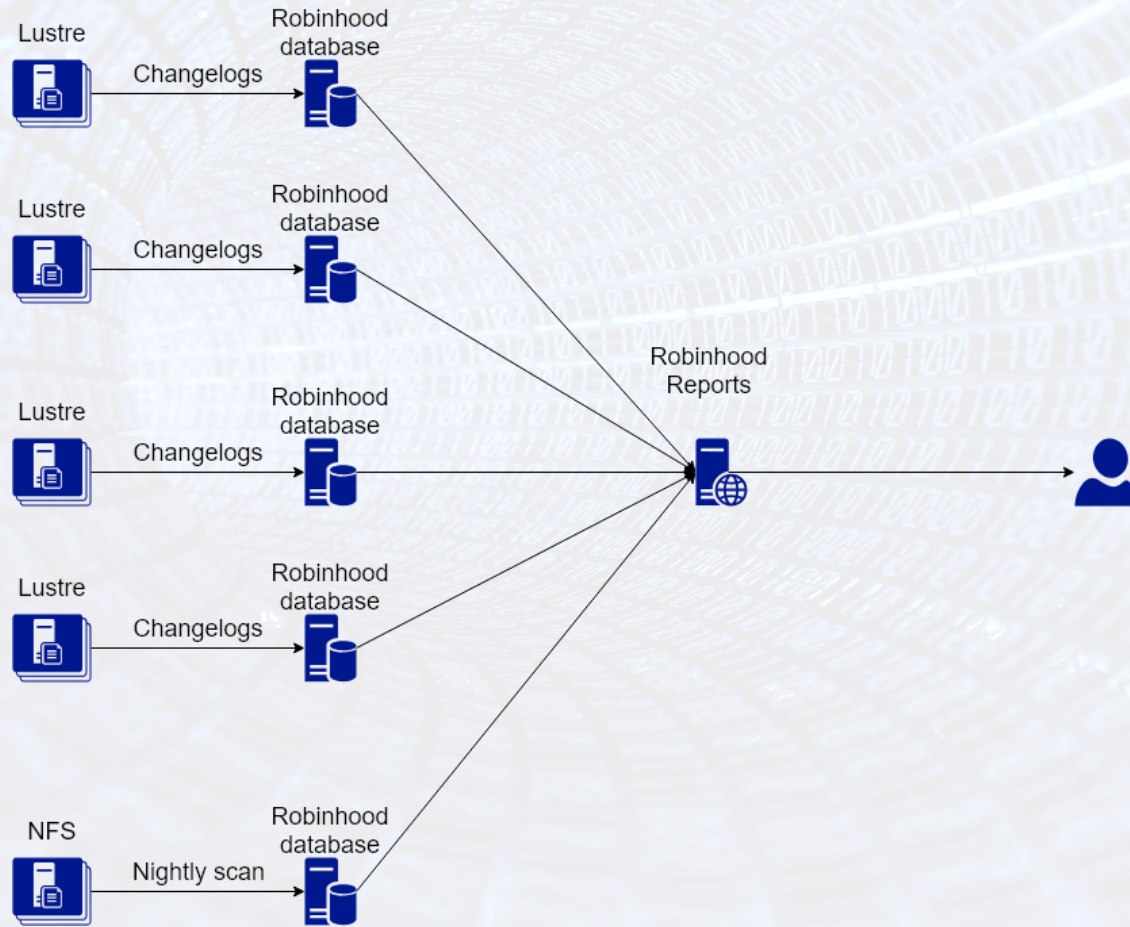
# Oil and Gas HPC Storage Environment

- ▶ Data on file systems is persistent
- ▶ Clean up is done by the user
  - Users don't clean up their data if they don't remember where it is
- ▶ As personnel change, administrators have to locate and reassign data
- ▶ Results in “parking lots” of old data

# Robinhood Reports History

- ▶ Previously, a recursive “ls” was run nightly to generate a list of files ☹️
- ▶ Raw file listing was parsed into text reports with custom scripts
- ▶ About 3 years ago, our “ls” scripts stopped working – luckily we had started setting up Robinhood and it soon became the replacement
- ▶ To replace the reports, we built a custom web front end for Robinhood

# Robinhood Reports Diagram



# Robinhood Reports Installation

- ▶ Simply requires a LAMP stack (Linux, Apache, MySQL/MariaDB (client), PHP)
  - Drop the code on your web server
  - Edit the config file to point at your Robinhood databases
  - Visit the website
  
- ▶ Provides several out of the box reports

# Robinhood Reports Installation – Best Practices

## ▶ Create a read only database user

- `CREATE USER 'rorbh'@'%' IDENTIFIED BY 'password';`
- `GRANT SELECT,EXECUTE ON `robinhood%`. * TO 'rorbh'@'%';`
- `FLUSH PRIVILEGES;`

## ▶ Create database index on frequently accessed columns

- `USE robinhood_lustre0x;`
- `CREATE INDEX idx_uid ON ENTRIES (uid);`
- `CREATE INDEX idx_last_access ON ENTRIES (last_access);`

## ▶ Create any applicable Robinhood file classes (more later)

# Robinhood Reports Installation – Best Practices

▶ Added config to Robinhood config to improve database consistency vs. file system state

- <https://sourceforge.net/p/robinhood/mailman/message/35637636/>

```
##### as suggested by Thomas Leibovici

define_policy update {
    default_action = common.log; # simply update entries, run no action
    scope { type == file }
    status_manager = none;
    default_lru_sort_attr = none;
}

update_parameters {
    report_actions = no;
}

update_rules {
    # if file was modified in past hour don't run
    ignore { last_mod < 1h }
    # if file hasn't been modified in more than a day don't check
    ignore { last_mod > 1d }

    rule default {
        condition = true; # apply to all entries (except above exclusions)
    }
}

# run every 12h
update_trigger {
    trigger_on = scheduled;
    check_interval = 12h;
}

#####
```



# Robinhood Reports - Summary

HPC Diskstats **Summary** Detailed Special ▾

22/24 File Systems

File_System ▲	Number_of_Users	Total_Space	Used_Space	Available_Space	Number_of_Files	Size_of_Old_Files	Number_of_Old_Files	Percent_Old_Space
	27	3.75 TB	759.04 GB	3.01 TB	789,356	736.78 GB	745,248	97.07
	15	30.00 TB	2.70 TB	27.30 TB	169,812	2.70 TB	169,780	100
	14	30.00 TB	595.02 GB	29.42 TB	314,992	594.97 GB	314,937	99.99
	126	4.76 PB	3.91 PB	870.49 TB	17,766,928	2.63 PB	16,241,822	67.28
	93	4.72 PB	3.88 PB	861.68 TB	12,979,510	1.18 PB	3,774,407	30.29
	28	7.02 PB	2.65 PB	4.38 PB	8,448,388	0 bytes	0	0
	202	4.72 PB	4.06 PB	681.42 TB	13,835,106	2.83 PB	6,059,585	69.83
	229	6.68 PB	5.89 PB	816.65 TB	50,489,484	1.98 PB	15,716,408	33.55
	21	1.90 TB	11.97 GB	1.89 TB	79,171	10.39 GB	78,272	86.82
	22	19.80 TB	9.83 TB	9.97 TB	5,943,889	3.05 TB	2,906,130	31.01
	20	1.98 TB	208.06 GB	1.78 TB	1,078,510	15.11 MB	15	0.01
	9	1.90 TB	24.53 GB	1.88 TB	117	7.04 GB	60	28.69
	12	9.90 TB	3.31 TB	6.59 TB	4,043,254	1.79 TB	505,870	54.15
	9	1.98 TB	36.00 KB	1.98 TB	0	0 bytes	0	0
	63	1.90 TB	174.36 GB	1.73 TB	131,071	110.72 GB	86,664	63.5
	61	1.98 TB	655.77 GB	1.34 TB	55,850	70.23 GB	6,546	10.71
	19	1.90 TB	53.79 GB	1.85 TB	210,917	35.82 GB	169,448	66.59
	26	9.50 TB	12.51 TB	0 bytes	39,850,394	9.44 TB	5,359,199	75.47

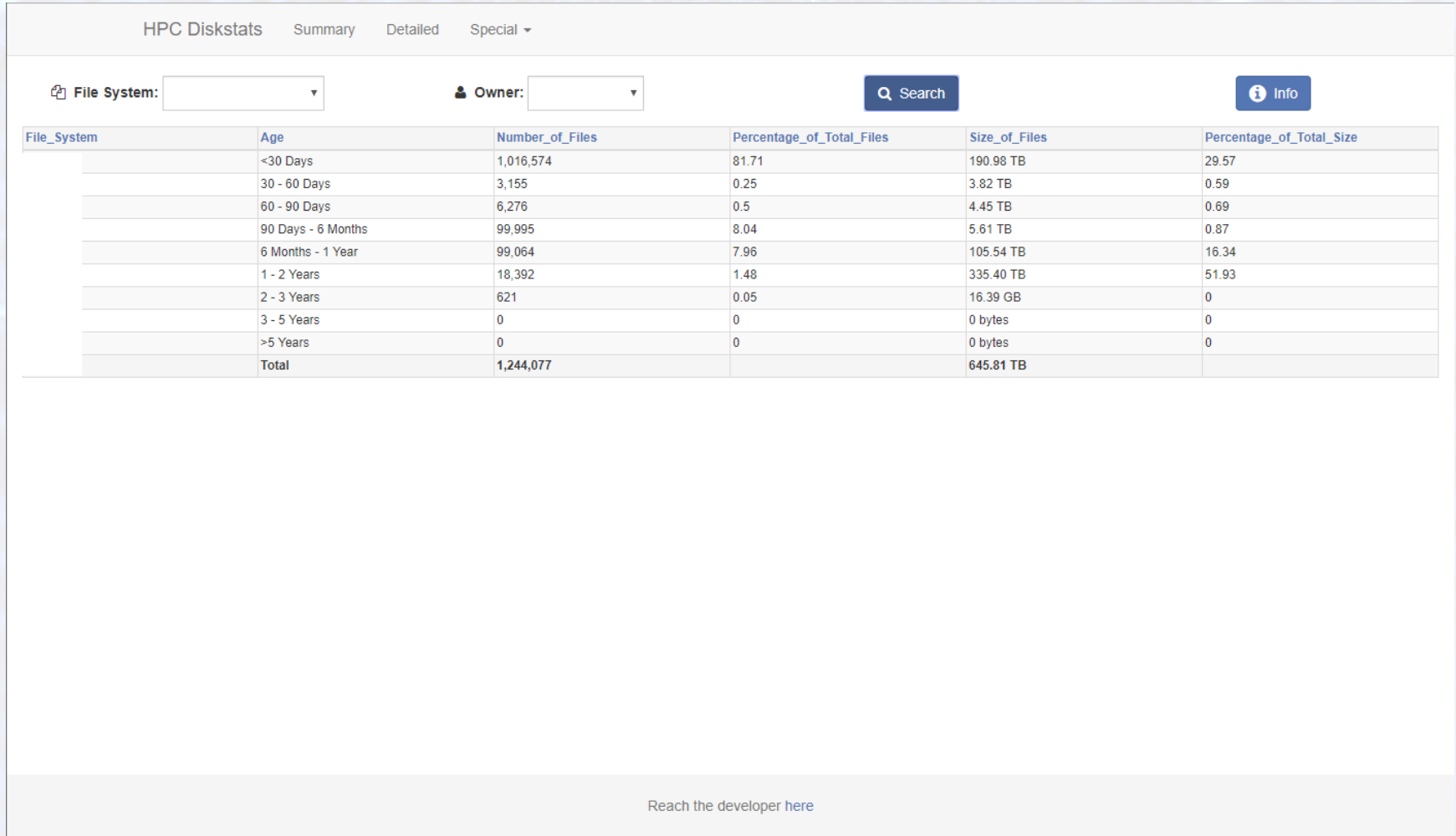
# Robinhood Reports - Detailed

HPC Diskstats Summary Detailed Special

File System: ALL Owner: ALL Summarize by owner:  Info

File_System	Owner	Number_of_Files	Size_of_Files	Number_of_Old_Files	Size_of_Old_Files	Percent_Old_Space
		0	0 bytes	0	0 bytes	0
		21	11.41 MB	15	7.95 MB	69.66
		584,614	214.65 GB	572,879	194.93 GB	90.81
		3	9.09 MB	2	6.09 MB	66.96
		2	4.02 MB	1	4.02 MB	100
		108,321	525.91 GB	108,260	525.82 GB	99.98
		5	3.92 MB	5	3.92 MB	100
		170	157.40 MB	39	106.66 MB	67.76
		2	36.50 KB	0	0 bytes	0
		1	10.01 MB	1	10.01 MB	100
		11,165	3.72 GB	11,162	3.72 GB	100
		3	3.33 KB	2	110 bytes	3.22
		2	24.02 KB	2	24.02 KB	100
		2	20.88 KB	0	0 bytes	0
		75,842	4.04 GB	43,736	2.37 GB	58.56
		8,298	4.42 GB	8,298	4.42 GB	100
		861	254.53 MB	802	67.94 MB	26.69
		44	1.27 MB	44	1.27 MB	100
		21,453	125.46 GB	21,450	125.46 GB	100
		118,586	2.41 TB	118,580	2.41 TB	100
		3,949	295.70 MB	3,948	295.70 MB	100
		254	18.52 MB	252	18.52 MB	100
		65	1.91 MB	55	1.86 MB	97.01
		14,278	37.63 GB	14,275	37.63 GB	100
		2,022	310.71 MB	2,021	310.71 MB	100
		56	495.00 KB	56	495.00 KB	100
		4,523	116.65 GB	4,520	116.65 GB	100
		11	30.78 KB	11	30.78 KB	100
		10	410.24 KB	10	410.24 KB	100
		46	3.08 MB	45	3.08 MB	100

# Robinhood Reports – Age Histogram



# Robinhood Reports – Big Directories

HPC Diskstats Summary Detailed Special

File System: ALL Owner: ALL Info

Directory	Number_of_Files	Owner	Group
	629,668		
	467,132		
	467,132		
	357,943		
	277,971		
	230,043		
	203,540		
	203,449		
	202,599		
	200,907		
	197,687		
	182,590		
	182,157		
	174,889		
	155,032		
	151,538		
	140,319		
	130,273		
	125,925		
	124,201		
	117,670		
	114,179		
	112,675		
	111,296		
	102,926		
	102,764		
	102,390		
	102,390		
	102,390		
	101,691		

# Robinhood Reports – Interesting Files

HPC Diskstats   Summary   Detailed   Special ▾

File System:    Owner:    Search   Info

File ▲	Size	Owner	Group
	1.25 KB		
	1.24 KB		
	1.25 KB		
	1.23 KB		
	427 bytes		
	255 bytes		
	259 bytes		
	182 bytes		
	169 bytes		
	139 bytes		
	125 bytes		
	177 bytes		
	178 bytes		
	184 bytes		
	176 bytes		
	139 bytes		
	150 bytes		
	150 bytes		
	111 bytes		
	177 bytes		
	171 bytes		
	139 bytes		
	508 bytes		
	177 bytes		
	171 bytes		
	139 bytes		
	125 bytes		
	177 bytes		
	176 bytes		

# Robinhood Reports – Print Files

HPC Diskstats   Summary   Detailed   Special ▾

File System:    Owner:      

File ▲	Size	Owner	Group
	1.25 KB		
	1.24 KB		
	1.25 KB		
	1.23 KB		
	427 bytes		
	255 bytes		
	259 bytes		
	182 bytes		
	169 bytes		
	139 bytes		
	125 bytes		
	177 bytes		
	178 bytes		
	184 bytes		
	176 bytes		
	139 bytes		
	150 bytes		
	150 bytes		
	111 bytes		
	177 bytes		
	171 bytes		
	139 bytes		
	508 bytes		
	177 bytes		
	171 bytes		
	139 bytes		
	125 bytes		
	177 bytes		
	176 bytes		

# Robinhood Reports – Largest Directories

HPC Diskstats Summary Detailed Special ▾

File System:  Owner:  [Search](#) [Info](#)

Directory	Size_of_Files ▾	Number_of_Files	Size_of_Old_Files	Number_
	57.49 TB	17	57.49 TB	16
	57.49 TB	7	57.49 TB	7
	27.21 TB	2,411	27.05 TB	2,397
	19.95 TB	1,701	19.95 TB	1,701
	19.27 TB	1,932	0 bytes	0
	18.59 TB	1,864	0 bytes	0
	17.99 TB	103	9.48 TB	60
	17.81 TB	13	17.81 TB	13
	17.55 TB	36	17.55 TB	36
	16.51 TB	45	16.51 TB	45
	13.65 TB	41	11.37 TB	39
	13.10 TB	10,032	13.10 TB	10,032
	12.70 TB	93	12.70 TB	93
	10.31 TB	25	10.31 TB	25
	10.31 TB	15	10.31 TB	15
	10.31 TB	12	10.31 TB	12
	10.31 TB	8	10.31 TB	8
	9.64 TB	2,833	0 bytes	0
	9.24 TB	2,704	0 bytes	0
	9.15 TB	55	6.87 TB	51
	8.92 TB	24	8.92 TB	24
	7.87 TB	11	7.87 TB	11
	7.78 TB	4,057	7.78 TB	4,057
	7.51 TB	2,416	5.15 TB	980
	6.90 TB	70	6.90 TB	70
	6.60 TB	70	6.60 TB	70
	5.95 TB	34	5.95 TB	34
	5.60 TB	6,481	5.60 TB	6,481

# Disclaimer

- ▶ I'm a systems engineer, not a software developer
- ▶ This software was built specifically for a certain environment. Portability was desired, but not guaranteed
- ▶ This is not professionally supported software
- ▶ Contributions are welcome



# Robinhood Reports Software Architecture

## ▶ Development principles

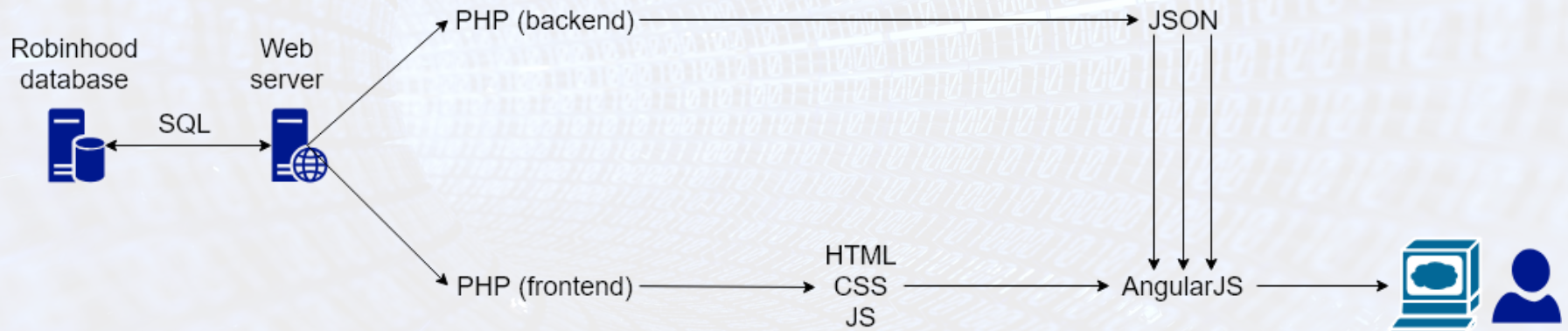
- Help users find their data, especially their old data!
- Produce a scalable design
- Produce a modular design, so creating new reports is simple
- Give users results as soon as they're available, and give them an idea of progress along the way since queries can take time
- Return reports within the attention span of a normal user

# Robinhood Reports Software Architecture

## ▶ Built with

- Bootstrap – makes it look pretty
- AngularJS – interactive pages
- Angular-ui-bootstrap – glue between Bootstrap and AngularJS, provides interactive features such as table sorting, pagination, progress bars
- jQuery - dependency
- Font Awesome – user friendly icons
- Humanize – converts numbers into user friendly values
- Angularjs-humanize – glue between Humanize and AngularJS

# Robinhood Reports Software Architecture



# Robinhood Reports vs. Robinhood Web GUI

Robinhood Reports	Robinhood Web GUI
Built specifically to find data	Generalized reporting tool
Can do anything you can do in SQL	Leverages Robinhood REST API
Single pane of glass	One site per file system
Built by yours truly	Built by CEA
Open source Github project	A part of the Robinhood distribution
No access control	Fine grained access control

# Robinhood Reports Future Improvements

- ▶ Allow front end pages to accept HTTP GET queries
- ▶ Make age user-selectable
- ▶ Make the results downloadable
- ▶ Show file size distribution
- ▶ Add a report calling out bad permissions

# Experts in High Performance Computing, Algorithms and Numerical Software Engineering

[www.nag.com](http://www.nag.com) | [blog.nag.com](http://blog.nag.com) | [@NAGtalk](https://twitter.com/NAGtalk)

Questions?

<https://github.com/numericalalgorithmsgroup/robinhood-reports>

[Shawn.Hall@nag.com](mailto:Shawn.Hall@nag.com)