



# Summer Internship Program

JANUARY 2024

Hugo Watanuki  
LexisNexis Risk Solutions

# Contents

- The Business
- The Technology
- The Internship

# The Business

# LexisNexis Risk Solutions is Part of RELX



*RELX is a global provider of information-based analytics and decision tools for professional and business customers, enabling them to make better decisions, get better results and be more productive. The Group serves customers in more than 180 countries and has offices in about 40 countries.*

*It employs over 33,000 people, of whom almost half are in North America.*

*Learn more at [www.relx.com](http://www.relx.com)*

RELX operates in four major market segments:

Scientific, Technical  
& Medical



Risk



Exhibitions



Legal



# What we do

We leverage five main capabilities to provide end-to-end solutions that help customers assess risk and opportunity.



## Vast Data Resources

We maintain over 12 petabytes of content comprising billions of public and proprietary records.



## Big Data Technology

We designed our massively-scalable super-computing platform, HPCC Systems®, enabling us to process at very high speeds – over 270 million transactions per hour.



## Advanced Linking

We use our own unique identifier, LexID®, together with a proprietary linking technology. Our patented linking and clustering method is the engine behind many of our products.



## Sophisticated Analytics & Insight Capability

We apply data science and leverage patented algorithms, predictive modeling, machine learning and AI to provide data driven solutions and better decision intelligence.



## Industry-Specific Expertise & Delivery

The people in our businesses have deep industry experience and expertise – we employ professionals that worked in the industries we serve, so they have walked in the shoes of our customers.



## Customer-Focused Solutions

We connect the dots between public records and transactions, resulting in actionable information our customers use to advance their goals.



# Our Customers

We work with Fortune 1000 and mid-market clients globally across industries, and federal and state governments.

- Customers in more than 180 countries
- 9 out of 10 of the world's top 10 banks
- 78% of the Fortune 500 companies
- 98 of the top 100 personal lines insurance carriers
- More than 7,500 federal, state and local government agencies

# The Technology

# Open Source Technology for Big Data Processing



## HPCC Systems®

Born from the deep data analysis experience of LexisNexis Risk Solutions, HPCC Systems is a comprehensive, dedicated cloud-native platform that makes combining data stored in massive, mixed schema data lakes easier and faster. The platform scales very quickly as your data needs grow enabling companies of all sizes to save time and money, now and in the future.

[www.hpccsystems.com](http://www.hpccsystems.com) [video](#)



Single, Complete Platform

Two data engines (query and refinery) operating at a high level of speed and accuracy



Cloud-Native Architecture

Automation of Kubernetes makes it easy to set-up, manage and scale your data



Parallel Programming Language (ECL)

High level, data-centric declarative programming language



Data Management Tools

Data is easier to manage with robust profiling, cleansing, update and consolidation tools



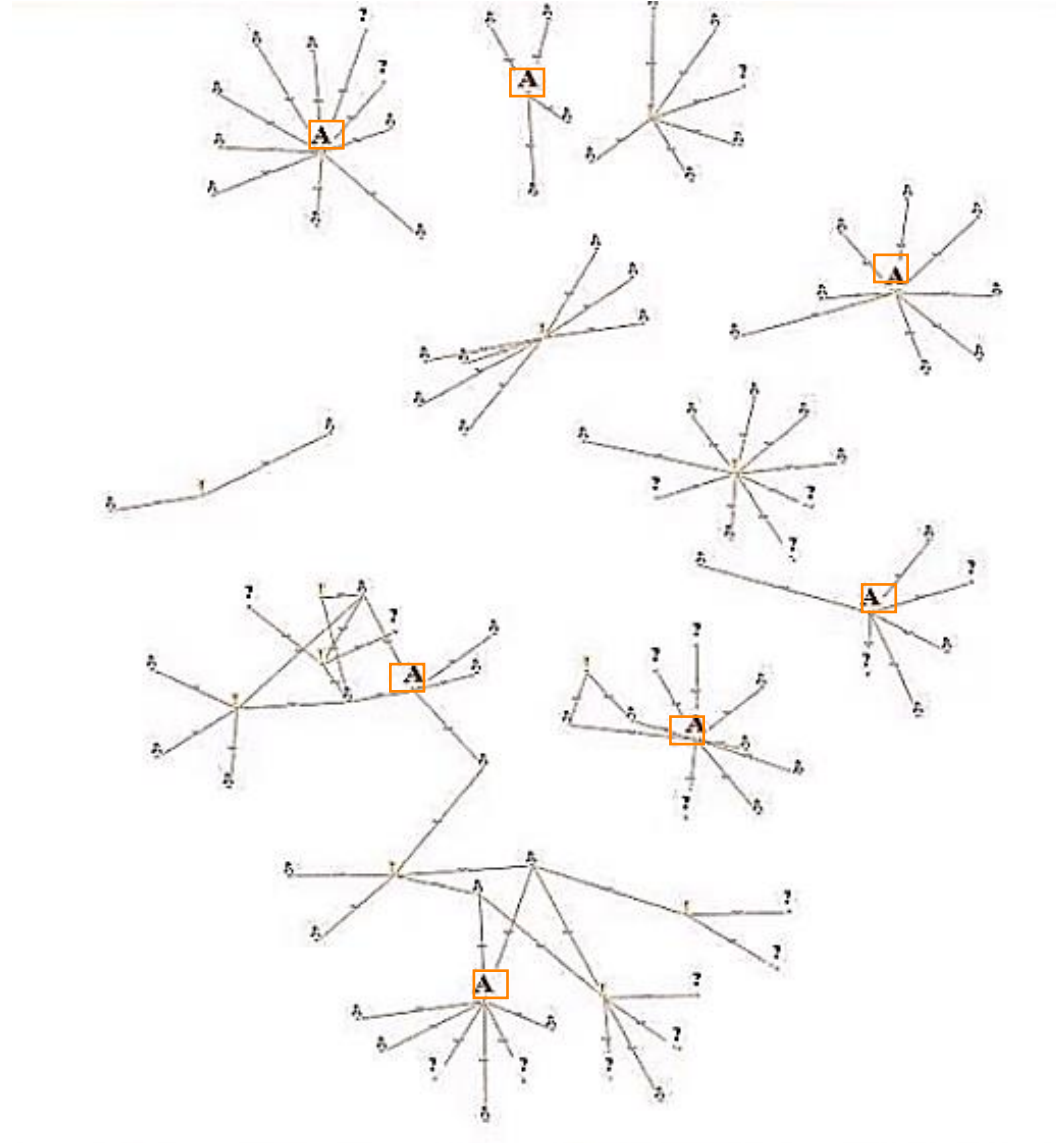
Machine Learning Library

Predictive modeling functionality for machine learning performance

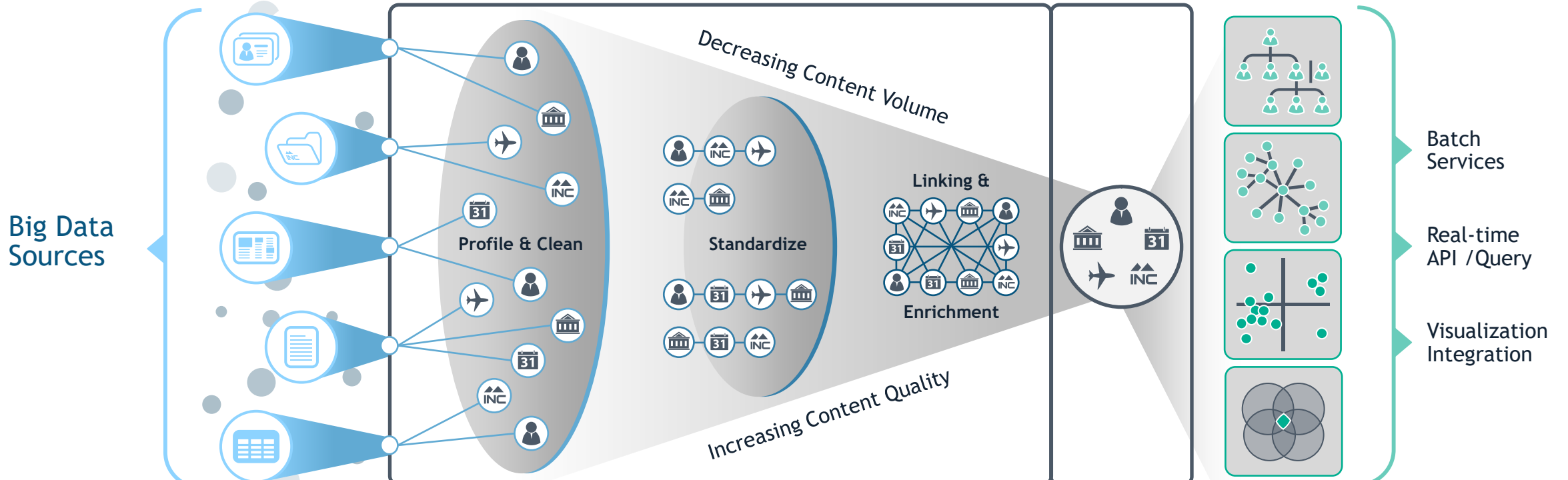




# Insurance Fraud Use Case



# HPCC Systems Data Enrichment Pipeline



**Unstructured and Structured Content**

- Over 12 petabytes of content
- More than 10,000 sources
- 7.5 billion unique name and address combinations
- 7.3 billion motor vehicle records

**Thor**

- Distributed Computing
- Data-centric language (ECL)
- An integrated delivery system that offers data plus analytics
- 86 billion records processed through Thor

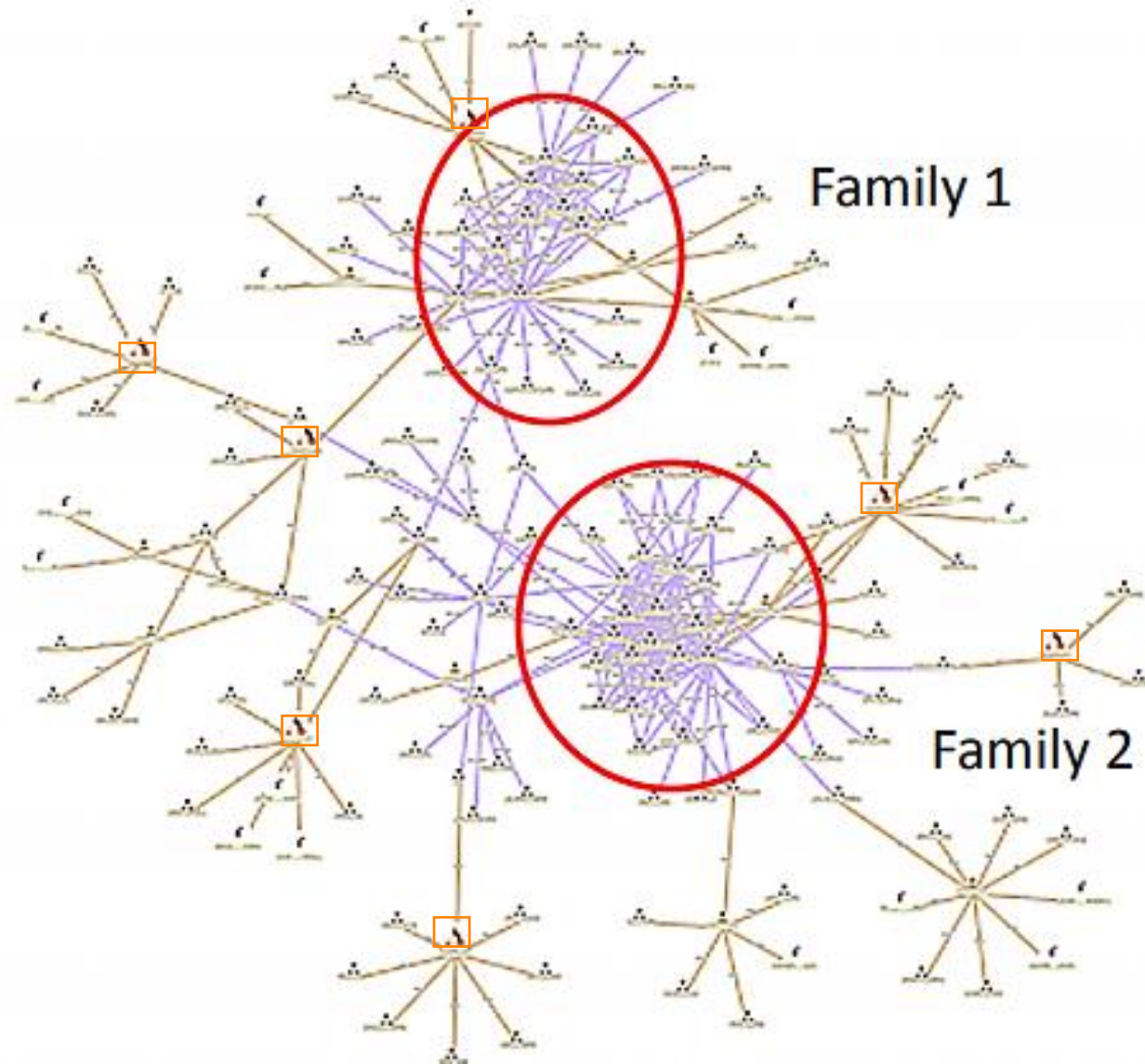
**ROXIE**

- Complex Searching and Matching
- Approximately 400+ models deployed
- Deliver 270 million ROXIE calls every hour

**Products**

- LexisNexis Risk Products

# Insurance Fraud Use Case



# The Internship

# HPCC Systems Internship

Are you interested in big data?

Coding in an open language?

Working with machine learning algorithms?

Or contributing to a cloud native platform?

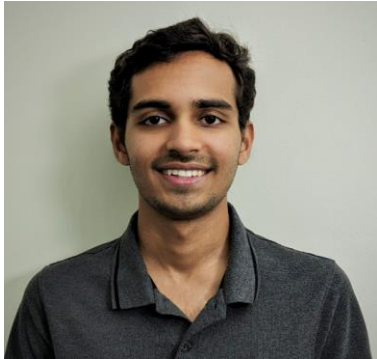
- 12-week paid and remote program over the summer
- Mentor-based and focused on open source HPCC Systems projects
- Open to students from high school to PhD levels that are authorized to work
- Proposal-based application, either original or from our list
- Visit our student wiki: [hpccsystems.com/student-wiki](https://hpccsystems.com/student-wiki)



Questions can be emailed to [students@hpccsystems.com](mailto:students@hpccsystems.com)



# MSc Student Project



NC STATE  
UNIVERSITY

## Applying Causality Toolkit to Real-world Datasets

Arun Gaonkar  
Mentor : Roger Dev



### Introduction

Everything in this universe happens for a reason and every action has a reaction.  
Analyzing causality can help in medical diagnostic analysis, time series analysis, and strategic planning. In real-world datasets, variables are inter-related, implying subtle correlations, which makes causal analysis difficult.

### Causal Toolkits

1. HPCC\_Causality
2. Because
  - Visualization bundle
  - Dependence & Independence tests
  - Causal Direction Tests

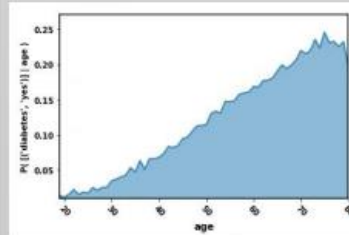
### Analysis Steps

1. Finding & Analyzing Dataset
2. Pre-processing the dataset
3. Propose a Causal Hypothesis
4. Applying causal toolkits & analyzing
5. Interpretation & Causal Model
6. Hypothesis & Model Verification

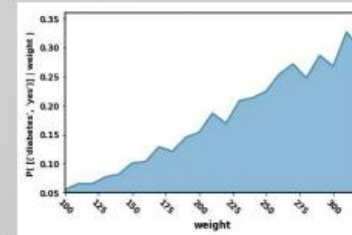
### Causal Hypothesis Question

Proposed causal hypothesis question:  
"What factors can influence the likelihood of a person having Diabetes?"

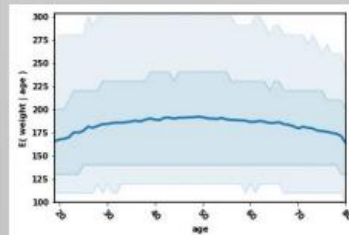
### Analysis LLCP-CDC Dataset



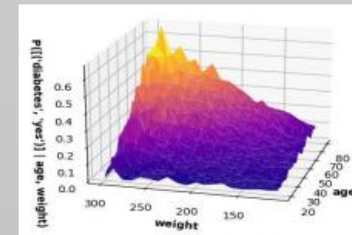
Diabetes vs Age



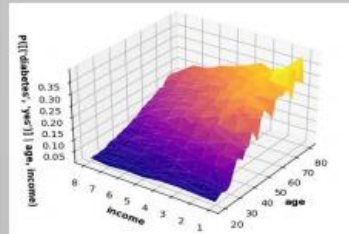
Diabetes vs Weight



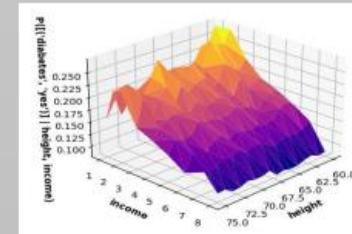
Weight vs Age



Diabetes vs Age, Weight

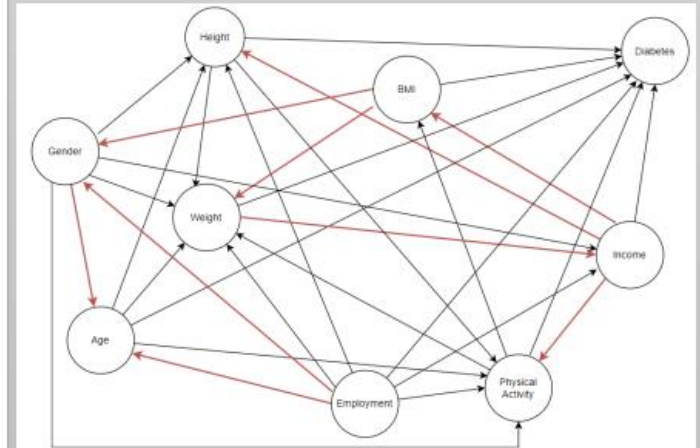


Diabetes vs Age, Income



Diabetes vs Height, Income

### Causal Model




### Conclusion

- Factors like Age, Height, Weight, Income, type of Employment, Physical Activity, and gender have their effect on Diabetes.
- Most of the relations are practically and analytically correct.
  - Some relations are unexpected, but probable valid proof can be generated.
  - For some other relations, any explanation is rationally invalid.
- Causality toolkit can be applied to analyze real-world datasets. But the cause-effect of latent variables cannot be incorporated into the causal model.

# Undergraduate Student Project






Georgiastate  
University

## TEST SUITE FOR A ROXIE CLUSTER ON KUBERNETES

NIVEDHA SIVAKUMAR | GEORGIA STATE UNIVERSITY |  
MENTORS: KRISHNA TURLAPATHI & ATTILA VAMOS



HPCC™  
SYSTEMS

### ABSTRACT

HPCC Systems is an open-source platform offering other businesses high-performance data processing and analytics.

- Thor cluster is responsible for manipulating massive amounts of data
- Roxie cluster supports high-performance data delivery applications using indexed data files.

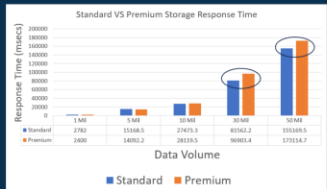
The motivation of my project is to create a test suite for Roxie designed to provide a more in-depth understanding of how different query, cluster, and infrastructure configurations can affect the functionality and performance of Roxie in the cloud.

### PROJECT GOALS

The primary goal for the project is to develop a test suite for Roxie in Kubernetes

The results provided by the test suite can give indications or guidelines as to what configuration will be suitable for each use case of Roxie in the future. The test cases will be executed to simulate different usage patterns and can help identify potential problems or areas for improvement within the cluster which can be particularly useful during new releases of HPCC Systems or changes in the infrastructure

### RESULTS



The graph displays the results of response time for each query in msecs. You may notice premium was slightly better than standard with smaller datasets, but had higher response time in larger datasets. As of right now, we are not sure why it is happening, but it will be discussed further with Microsoft.

### METHODOLOGY

#### Test Execution

- Cloned terraform repo made by Godji, made changes in the esp.yaml file. This portion of the Roxie configuration needs to be added to the YAML file and used the default Roxie configuration which was 4 CPUs, and 3 nodes Roxie.
- Started with running the HPCC System cluster on AKS with two different storage types (Premium & Standard).
- Execute `kubectl get svc` command to obtain the ECL Watch and the ECL Queries page info
- `ecl.exe publish --target=roxie --server=<external IP> --no-reload --no-files --wait=7200000 -v "query file directory" --name=name of the file --comment="Test" -l "query folder directory"`

```

130 router
131   - name: roxie
132   labels:
133     app: roxie
134     tier: front
135   service:
136     ports:
137       - port: 80
138         protocol: TCP
139     selector:
140       app: roxie
141       tier: front
142   deployment:
143     replicas: 3
144     strategy:
145       type: RollingUpdate
146     selector:
147       app: roxie
148     template:
149       metadata:
150         labels:
151           app: roxie
152           tier: front
153       spec:
154         containers:
155           - name: roxie
156             image: hpccsystems/roxie:latest
157             ports:
158               - containerPort: 80
159             resources:
160               requests:
161                 cpu: 100m
162                 memory: 256Mi
163               limits:
164                 cpu: 100m
165                 memory: 256Mi
166             volumeMounts:
167               - name: roxie-data
168                 mountPath: /data
169             livenessProbe:
170               httpGet:
171                 path: /health
172                 port: 80
173               initialDelaySeconds: 30
174               timeoutSeconds: 5
175             readinessProbe:
176               httpGet:
177                 path: /health
178                 port: 80
179               initialDelaySeconds: 30
180               timeoutSeconds: 5
181             securityContext:
182               runAsUser: 1000
183               allowPrivilegeEscalation: false
184             volumeMounts:
185               - name: roxie-data
186                 mountPath: /data
187             volumeMounts:
188               - name: roxie-data
189                 mountPath: /data
190             volumeMounts:
191               - name: roxie-data
192                 mountPath: /data
193             volumeMounts:
194               - name: roxie-data
195                 mountPath: /data
196             volumeMounts:
197               - name: roxie-data
198                 mountPath: /data
199             volumeMounts:
200               - name: roxie-data
201                 mountPath: /data
202             volumeMounts:
203               - name: roxie-data
204                 mountPath: /data
205             volumeMounts:
206               - name: roxie-data
207                 mountPath: /data
208             volumeMounts:
209               - name: roxie-data
210                 mountPath: /data
211             volumeMounts:
212               - name: roxie-data
213                 mountPath: /data
214             volumeMounts:
215               - name: roxie-data
216                 mountPath: /data
217             volumeMounts:
218               - name: roxie-data
219                 mountPath: /data
220             volumeMounts:
221               - name: roxie-data
222                 mountPath: /data
223             volumeMounts:
224               - name: roxie-data
225                 mountPath: /data
226             volumeMounts:
227               - name: roxie-data
228                 mountPath: /data
229             volumeMounts:
230               - name: roxie-data
231                 mountPath: /data
232             volumeMounts:
233               - name: roxie-data
234                 mountPath: /data
235             volumeMounts:
236               - name: roxie-data
237                 mountPath: /data
238             volumeMounts:
239               - name: roxie-data
240                 mountPath: /data
241             volumeMounts:
242               - name: roxie-data
243                 mountPath: /data
244             volumeMounts:
245               - name: roxie-data
246                 mountPath: /data
247             volumeMounts:
248               - name: roxie-data
249                 mountPath: /data
250             volumeMounts:
251               - name: roxie-data
252                 mountPath: /data
253             volumeMounts:
254               - name: roxie-data
255                 mountPath: /data
256             volumeMounts:
257               - name: roxie-data
258                 mountPath: /data
259             volumeMounts:
260               - name: roxie-data
261                 mountPath: /data
262             volumeMounts:
263               - name: roxie-data
264                 mountPath: /data
265             volumeMounts:
266               - name: roxie-data
267                 mountPath: /data
268             volumeMounts:
269               - name: roxie-data
270                 mountPath: /data
271             volumeMounts:
272               - name: roxie-data
273                 mountPath: /data
274             volumeMounts:
275               - name: roxie-data
276                 mountPath: /data
277             volumeMounts:
278               - name: roxie-data
279                 mountPath: /data
280             volumeMounts:
281               - name: roxie-data
282                 mountPath: /data
283             volumeMounts:
284               - name: roxie-data
285                 mountPath: /data
286             volumeMounts:
287               - name: roxie-data
288                 mountPath: /data
289             volumeMounts:
290               - name: roxie-data
291                 mountPath: /data
292             volumeMounts:
293               - name: roxie-data
294                 mountPath: /data
295             volumeMounts:
296               - name: roxie-data
297                 mountPath: /data
298             volumeMounts:
299               - name: roxie-data
300                 mountPath: /data
301             volumeMounts:
302               - name: roxie-data
303                 mountPath: /data
304             volumeMounts:
305               - name: roxie-data
306                 mountPath: /data
307             volumeMounts:
308               - name: roxie-data
309                 mountPath: /data
310             volumeMounts:
311               - name: roxie-data
312                 mountPath: /data
313             volumeMounts:
314               - name: roxie-data
315                 mountPath: /data
316             volumeMounts:
317               - name: roxie-data
318                 mountPath: /data
319             volumeMounts:
320               - name: roxie-data
321                 mountPath: /data
322             volumeMounts:
323               - name: roxie-data
324                 mountPath: /data
325             volumeMounts:
326               - name: roxie-data
327                 mountPath: /data
328             volumeMounts:
329               - name: roxie-data
330                 mountPath: /data
331             volumeMounts:
332               - name: roxie-data
333                 mountPath: /data
334             volumeMounts:
335               - name: roxie-data
336                 mountPath: /data
337             volumeMounts:
338               - name: roxie-data
339                 mountPath: /data
340             volumeMounts:
341               - name: roxie-data
342                 mountPath: /data
343             volumeMounts:
344               - name: roxie-data
345                 mountPath: /data
346             volumeMounts:
347               - name: roxie-data
348                 mountPath: /data
349             volumeMounts:
350               - name: roxie-data
351                 mountPath: /data
352             volumeMounts:
353               - name: roxie-data
354                 mountPath: /data
355             volumeMounts:
356               - name: roxie-data
357                 mountPath: /data
358             volumeMounts:
359               - name: roxie-data
360                 mountPath: /data
361             volumeMounts:
362               - name: roxie-data
363                 mountPath: /data
364             volumeMounts:
365               - name: roxie-data
366                 mountPath: /data
367             volumeMounts:
368               - name: roxie-data
369                 mountPath: /data
370             volumeMounts:
371               - name: roxie-data
372                 mountPath: /data
373             volumeMounts:
374               - name: roxie-data
375                 mountPath: /data
376             volumeMounts:
377               - name: roxie-data
378                 mountPath: /data
379             volumeMounts:
380               - name: roxie-data
381                 mountPath: /data
382             volumeMounts:
383               - name: roxie-data
384                 mountPath: /data
385             volumeMounts:
386               - name: roxie-data
387                 mountPath: /data
388             volumeMounts:
389               - name: roxie-data
390                 mountPath: /data
391             volumeMounts:
392               - name: roxie-data
393                 mountPath: /data
394             volumeMounts:
395               - name: roxie-data
396                 mountPath: /data
397             volumeMounts:
398               - name: roxie-data
399                 mountPath: /data
400             volumeMounts:
401               - name: roxie-data
402                 mountPath: /data
403             volumeMounts:
404               - name: roxie-data
405                 mountPath: /data
406             volumeMounts:
407               - name: roxie-data
408                 mountPath: /data
409             volumeMounts:
410               - name: roxie-data
411                 mountPath: /data
412             volumeMounts:
413               - name: roxie-data
414                 mountPath: /data
415             volumeMounts:
416               - name: roxie-data
417                 mountPath: /data
418             volumeMounts:
419               - name: roxie-data
420                 mountPath: /data
421             volumeMounts:
422               - name: roxie-data
423                 mountPath: /data
424             volumeMounts:
425               - name: roxie-data
426                 mountPath: /data
427             volumeMounts:
428               - name: roxie-data
429                 mountPath: /data
430             volumeMounts:
431               - name: roxie-data
432                 mountPath: /data
433             volumeMounts:
434               - name: roxie-data
435                 mountPath: /data
436             volumeMounts:
437               - name: roxie-data
438                 mountPath: /data
439             volumeMounts:
440               - name: roxie-data
441                 mountPath: /data
442             volumeMounts:
443               - name: roxie-data
444                 mountPath: /data
445             volumeMounts:
446               - name: roxie-data
447                 mountPath: /data
448             volumeMounts:
449               - name: roxie-data
450                 mountPath: /data
451             volumeMounts:
452               - name: roxie-data
453                 mountPath: /data
454             volumeMounts:
455               - name: roxie-data
456                 mountPath: /data
457             volumeMounts:
458               - name: roxie-data
459                 mountPath: /data
460             volumeMounts:
461               - name: roxie-data
462                 mountPath: /data
463             volumeMounts:
464               - name: roxie-data
465                 mountPath: /data
466             volumeMounts:
467               - name: roxie-data
468                 mountPath: /data
469             volumeMounts:
470               - name: roxie-data
471                 mountPath: /data
472             volumeMounts:
473               - name: roxie-data
474                 mountPath: /data
475             volumeMounts:
476               - name: roxie-data
477                 mountPath: /data
478             volumeMounts:
479               - name: roxie-data
480                 mountPath: /data
481             volumeMounts:
482               - name: roxie-data
483                 mountPath: /data
484             volumeMounts:
485               - name: roxie-data
486                 mountPath: /data
487             volumeMounts:
488               - name: roxie-data
489                 mountPath: /data
490             volumeMounts:
491               - name: roxie-data
492                 mountPath: /data
493             volumeMounts:
494               - name: roxie-data
495                 mountPath: /data
496             volumeMounts:
497               - name: roxie-data
498                 mountPath: /data
499             volumeMounts:
500               - name: roxie-data
501                 mountPath: /data
502             volumeMounts:
503               - name: roxie-data
504                 mountPath: /data
505             volumeMounts:
506               - name: roxie-data
507                 mountPath: /data
508             volumeMounts:
509               - name: roxie-data
510                 mountPath: /data
511             volumeMounts:
512               - name: roxie-data
513                 mountPath: /data
514             volumeMounts:
515               - name: roxie-data
516                 mountPath: /data
517             volumeMounts:
518               - name: roxie-data
519                 mountPath: /data
520             volumeMounts:
521               - name: roxie-data
522                 mountPath: /data
523             volumeMounts:
524               - name: roxie-data
525                 mountPath: /data
526             volumeMounts:
527               - name: roxie-data
528                 mountPath: /data
529             volumeMounts:
530               - name: roxie-data
531                 mountPath: /data
532             volumeMounts:
533               - name: roxie-data
534                 mountPath: /data
535             volumeMounts:
536               - name: roxie-data
537                 mountPath: /data
538             volumeMounts:
539               - name: roxie-data
540                 mountPath: /data
541             volumeMounts:
542               - name: roxie-data
543                 mountPath: /data
544             volumeMounts:
545               - name: roxie-data
546                 mountPath: /data
547             volumeMounts:
548               - name: roxie-data
549                 mountPath: /data
550             volumeMounts:
551               - name: roxie-data
552                 mountPath: /data
553             volumeMounts:
554               - name: roxie-data
555                 mountPath: /data
556             volumeMounts:
557               - name: roxie-data
558                 mountPath: /data
559             volumeMounts:
560               - name: roxie-data
561                 mountPath: /data
562             volumeMounts:
563               - name: roxie-data
564                 mountPath: /data
565             volumeMounts:
566               - name: roxie-data
567                 mountPath: /data
568             volumeMounts:
569               - name: roxie-data
570                 mountPath: /data
571             volumeMounts:
572               - name: roxie-data
573                 mountPath: /data
574             volumeMounts:
575               - name: roxie-data
576                 mountPath: /data
577             volumeMounts:
578               - name: roxie-data
579                 mountPath: /data
580             volumeMounts:
581               - name: roxie-data
582                 mountPath: /data
583             volumeMounts:
584               - name: roxie-data
585                 mountPath: /data
586             volumeMounts:
587               - name: roxie-data
588                 mountPath: /data
589             volumeMounts:
590               - name: roxie-data
591                 mountPath: /data
592             volumeMounts:
593               - name: roxie-data
594                 mountPath: /data
595             volumeMounts:
596               - name: roxie-data
597                 mountPath: /data
598             volumeMounts:
599               - name: roxie-data
600                 mountPath: /data
601             volumeMounts:
602               - name: roxie-data
603                 mountPath: /data
604             volumeMounts:
605               - name: roxie-data
606                 mountPath: /data
607             volumeMounts:
608               - name: roxie-data
609                 mountPath: /data
610             volumeMounts:
611               - name: roxie-data
612                 mountPath: /data
613             volumeMounts:
614               - name: roxie-data
615                 mountPath: /data
616             volumeMounts:
617               - name: roxie-data
618                 mountPath: /data
619             volumeMounts:
620               - name: roxie-data
621                 mountPath: /data
622             volumeMounts:
623               - name: roxie-data
624                 mountPath: /data
625             volumeMounts:
626               - name: roxie-data
627                 mountPath: /data
628             volumeMounts:
629               - name: roxie-data
630                 mountPath: /data
631             volumeMounts:
632               - name: roxie-data
633                 mountPath: /data
634             volumeMounts:
635               - name: roxie-data
636                 mountPath: /data
637             volumeMounts:
638               - name: roxie-data
639                 mountPath: /data
640             volumeMounts:
641               - name: roxie-data
642                 mountPath: /data
643             volumeMounts:
644               - name: roxie-data
645                 mountPath: /data
646             volumeMounts:
647               - name: roxie-data
648                 mountPath: /data
649             volumeMounts:
650               - name: roxie-data
651                 mountPath: /data
652             volumeMounts:
653               - name: roxie-data
654                 mountPath: /data
655             volumeMounts:
656               - name: roxie-data
657                 mountPath: /data
658             volumeMounts:
659               - name: roxie-data
660                 mountPath: /data
661             volumeMounts:
662               - name: roxie-data
663                 mountPath: /data
664             volumeMounts:
665               - name: roxie-data
666                 mountPath: /data
667             volumeMounts:
668               - name: roxie-data
669                 mountPath: /data
670             volumeMounts:
671               - name: roxie-data
672                 mountPath: /data
673             volumeMounts:
674               - name: roxie-data
675                 mountPath: /data
676             volumeMounts:
677               - name: roxie-data
678                 mountPath: /data
679             volumeMounts:
680               - name: roxie-data
681                 mountPath: /data
682             volumeMounts:
683               - name: roxie-data
684                 mountPath: /data
685             volumeMounts:
686               - name: roxie-data
687                 mountPath: /data
688             volumeMounts:
689               - name: roxie-data
690                 mountPath: /data
691             volumeMounts:
692               - name: roxie-data
693                 mountPath: /data
694             volumeMounts:
695               - name: roxie-data
696                 mountPath: /data
697             volumeMounts:
698               - name: roxie-data
699                 mountPath: /data
700             volumeMounts:
701               - name: roxie-data
702                 mountPath: /data
703             volumeMounts:
704               - name: roxie-data
705                 mountPath: /data
706             volumeMounts:
707               - name: roxie-data
708                 mountPath: /data
709             volumeMounts:
710               - name: roxie-data
711                 mountPath: /data
711             volumeMounts:
712               - name: roxie-data
713                 mountPath: /data
713             volumeMounts:
714               - name: roxie-data
715                 mountPath: /data
715             volumeMounts:
716               - name: roxie-data
717                 mountPath: /data
717             volumeMounts:
718               - name: roxie-data
719                 mountPath: /data
719             volumeMounts:
720               - name: roxie-data
721                 mountPath: /data
721             volumeMounts:
722               - name: roxie-data
723                 mountPath: /data
723             volumeMounts:
724               - name: roxie-data
725                 mountPath: /data
725             volumeMounts:
726               - name: roxie-data
727                 mountPath: /data
727             volumeMounts:
728               - name: roxie-data
729                 mountPath: /data
729             volumeMounts:
730               - name: roxie-data
731                 mountPath: /data
731             volumeMounts:
732               - name: roxie-data
733                 mountPath: /data
733             volumeMounts:
734               - name: roxie-data
735                 mountPath: /data
735             volumeMounts:
736               - name: roxie-data
737                 mountPath: /data
737             volumeMounts:
738               - name: roxie-data
739                 mountPath: /data
739             volumeMounts:
740               - name: roxie-data
741                 mountPath: /data
741             volumeMounts:
742               - name: roxie-data
743                 mountPath: /data
743             volumeMounts:
744               - name: roxie-data
745                 mountPath: /data
745             volumeMounts:
746               - name: roxie-data
747                 mountPath: /data
747             volumeMounts:
748               - name: roxie-data
749                 mountPath: /data
749             volumeMounts:
750               - name: roxie-data
751                 mountPath: /data
751             volumeMounts:
752               - name: roxie-data
753                 mountPath: /data
753             volumeMounts:
754               - name: roxie-data
755                 mountPath: /data
755             volumeMounts:
756               - name: roxie-data
757                 mountPath: /data
757             volumeMounts:
758               - name: roxie-data
759                 mountPath: /data
759             volumeMounts:
760               - name: roxie-data
761                 mountPath: /data
761             volumeMounts:
762               - name: roxie-data
763                 mountPath: /data
763             volumeMounts:
764               - name: roxie-data
765                 mountPath: /data
765             volumeMounts:
766               - name: roxie-data
767                 mountPath: /data
767             volumeMounts:
768               - name: roxie-data
769                 mountPath: /data
769             volumeMounts:
770               - name: roxie-data
771                 mountPath: /data
771             volumeMounts:
772               - name: roxie-data
773                 mountPath: /data
773             volumeMounts:
774               - name: roxie-data
775                 mountPath: /data
775             volumeMounts:
776               - name: roxie-data
777                 mountPath: /data
777             volumeMounts:
778               - name: roxie-data
779                 mountPath: /data
779             volumeMounts:
780               - name: roxie-data
781                 mountPath: /data
781             volumeMounts:
782               - name: roxie-data
783                 mountPath: /data
783             volumeMounts:
784               - name: roxie-data
785                 mountPath: /data
785             volumeMounts:
786               - name: roxie-data
787                 mountPath: /data
787             volumeMounts:
788               - name: roxie-data
789                 mountPath: /data
789             volumeMounts:
790               - name: roxie-data
791                 mountPath: /data
791             volumeMounts:
792               - name: roxie-data
793                 mountPath: /data
793             volumeMounts:
794               - name: roxie-data
795                 mountPath: /data
795             volumeMounts:
796               - name: roxie-data
797                 mountPath: /data
797             volumeMounts:
798               - name: roxie-data
799                 mountPath: /data
799             volumeMounts:
800               - name: roxie-data
801                 mountPath: /data
801             volumeMounts:
802               - name: roxie-data
803                 mountPath: /data
803             volumeMounts:
804               - name: roxie-data
805                 mountPath: /data
805             volumeMounts:
806               - name: roxie-data
807                 mountPath: /data
807             volumeMounts:
808               - name: roxie-data
809                 mountPath: /data
809             volumeMounts:
810               - name: roxie-data
811                 mountPath: /data
811             volumeMounts:
812               - name: roxie-data
813                 mountPath: /data
813             volumeMounts:
814               - name: roxie-data
815                 mountPath: /data
815             volumeMounts:
816               - name: roxie-data
817                 mountPath: /data
817             volumeMounts:
818               - name: roxie-data
819                 mountPath: /data
819             volumeMounts:
820               - name: roxie-data
821                 mountPath: /data
821             volumeMounts:
822               - name: roxie-data
823                 mountPath: /data
823             volumeMounts:
824               - name: roxie-data
825                 mountPath: /data
825             volumeMounts:
826               - name: roxie-data
827                 mountPath: /data
827             volumeMounts:
828               - name: roxie-data
829                 mountPath: /data
829             volumeMounts:
830               - name: roxie-data
831                 mountPath: /data
831             volumeMounts:
832               - name: roxie-data
833                 mountPath: /data
833             volumeMounts:
834               - name: roxie-data
835                 mountPath: /data
835             volumeMounts:
836               - name: roxie-data
837                 mountPath: /data
837             volumeMounts:
838               - name: roxie-data
839                 mountPath: /data
839             volumeMounts:
840               - name: roxie-data
841                 mountPath: /data
841             volumeMounts:
842               - name: roxie-data
843                 mountPath: /data
843             volumeMounts:
844               - name: roxie-data
845                 mountPath: /data
845             volumeMounts:
846               - name: roxie-data
847                 mountPath: /data
847             volumeMounts:
848               - name: roxie-data
849                 mountPath: /data
849             volumeMounts:
850               - name: roxie-data
851                 mountPath: /data
851             volumeMounts:
852               - name: roxie-data
853                 mountPath: /data
853             volumeMounts:
854               - name: roxie-data
855                 mountPath: /data
855             volumeMounts:
856               - name: roxie-data
857                 mountPath: /data
857             volumeMounts:
858               - name: roxie-data
859                 mountPath: /data
859             volumeMounts:
860               - name: roxie-data
861                 mountPath: /data
861             volumeMounts:
862               - name: roxie-data
863                 mountPath: /data
863             volumeMounts:
864               - name: roxie-data
865                 mountPath: /data
865             volumeMounts:
866               - name: roxie-data
867                 mountPath: /data
867             volumeMounts:
868               - name: roxie-data
869                 mountPath: /data
869             volumeMounts:
870               - name: roxie-data
871                 mountPath: /data
871             volumeMounts:
872               - name: roxie-data
873                 mountPath: /data
873             volumeMounts:
874               - name: roxie-data
875                 mountPath: /data
875             volumeMounts:
876               - name: roxie-data
877                 mountPath: /data
877             volumeMounts:
878               - name: roxie-data
879                 mountPath: /data
879             volumeMounts:
880               - name: roxie-data
881                 mountPath: /data
881             volumeMounts:
882               - name: roxie-data
883                 mountPath: /data
883             volumeMounts:
884               - name: roxie-data
885                 mountPath: /data
885             volumeMounts:
886               - name: roxie-data
887                 mountPath: /data
887             volumeMounts:
888               - name: roxie-data
889                 mountPath: /data
889             volumeMounts:
890               - name: roxie-data
891                 mountPath: /data
891             volumeMounts:
892               - name: roxie-data
893                 mountPath: /data
893             volumeMounts:
894               - name: roxie-data
895                 mountPath: /data
895             volumeMounts:
896               - name: roxie-data
897                 mountPath: /data
897             volumeMounts:
898               - name: roxie-data
899                 mountPath: /data
899             volumeMounts:
900               - name: roxie-data
901                 mountPath: /data
901             volumeMounts:
902               - name: roxie-data
903                 mountPath: /data
903             volumeMounts:
904               - name: roxie-data
905                 mountPath: /data
905             volumeMounts:
906               - name: roxie-data
907                 mountPath: /data
907             volumeMounts:
908               - name: roxie-data
909                 mountPath: /data
909             volumeMounts:
910               - name: roxie-data
911                 mountPath: /data
911             volumeMounts:
912               - name: roxie-data
913                 mountPath: /data
913             volumeMounts:
914               - name: roxie-data
915                 mountPath: /data
915             volumeMounts:
916               - name: roxie-data
917                 mountPath: /data
917             volumeMounts:
918               - name: roxie-data
919                 mountPath: /data
919             volumeMounts:
920               - name: roxie-data
921                 mountPath: /data
921             volumeMounts:
922               - name: roxie-data
923                 mountPath: /data
923             volumeMounts:
924               - name: roxie-data
925                 mountPath: /data
925             volumeMounts:
926               - name: roxie-data
927                 mountPath: /data
927             volumeMounts:
928               - name: roxie-data
929                 mountPath: /data
929             volumeMounts:
930               - name: roxie-data
931                 mountPath: /data
931             volumeMounts:
932               - name: roxie-data
933                 mountPath: /data
933             volumeMounts:
934               - name: roxie-data
935                 mountPath: /data
935             volumeMounts:
936               - name: roxie-data
937                 mountPath: /data
937             volumeMounts:
938               - name: roxie-data
939                 mountPath: /data
939             volumeMounts:
940               - name: roxie-data
941                 mountPath: /data
941             volumeMounts:
942               - name: roxie-data
943                 mountPath: /data
943             volumeMounts:
944               - name: roxie-data
945                 mountPath: /data
945             volumeMounts:
946               - name: roxie-data
947                 mountPath: /data
947             volumeMounts:
948               - name: roxie-data
949                 mountPath: /data
949             volumeMounts:
950               - name: roxie-data
951                 mountPath: /data
951             volumeMounts:
952               - name: roxie-data
953                 mountPath: /data
953             volumeMounts:
954               - name: roxie-data
955                 mountPath: /data
955             volumeMounts:
956               - name: roxie-data
957                 mountPath: /data
957             volumeMounts:
958               - name: roxie-data
959                 mountPath: /data
959             volumeMounts:
960               - name: roxie-data
961                 mountPath: /data
961             volumeMounts:
962               - name: roxie-data
963                 mountPath: /data
963             volumeMounts:
964               - name: roxie-data
965                 mountPath: /data
965             volumeMounts:
966               - name: roxie-data
967                 mountPath: /data
967             volumeMounts:
968               - name: roxie-data
969                 mountPath: /data
969             volumeMounts:
970               - name: roxie-data
971                 mountPath: /data
971             volumeMounts:
972               - name: roxie-data
973                 mountPath: /data
973             volumeMounts:
974               - name: roxie-data
975                 mountPath: /data
975             volumeMounts:
976               - name: roxie-data
977                 mountPath: /data
977             volumeMounts:
978               - name: roxie-data
979                 mountPath: /data
979             volumeMounts:
980               - name: roxie-data
981                 mountPath: /data
981             volumeMounts:
982               - name: roxie-data
983                 mountPath: /data
983             volumeMounts:
984               - name: roxie-data
985                 mountPath: /data
985             volumeMounts:
986               - name: roxie-data
987                 mountPath: /data
987             volumeMounts:
988               - name: roxie-data
989                 mountPath: /data
989             volumeMounts:
990               - name: roxie-data
991                 mountPath: /data
991             volumeMounts:
992               - name: roxie-data
993                 mountPath: /data
993             volumeMounts:
994               - name: roxie-data
995                 mountPath: /data
995             volumeMounts:
996               - name: roxie-data
997                 mountPath: /data
997             volumeMounts:
998               - name: roxie-data
999                 mountPath: /data
999             volumeMounts:
1000              - name: roxie-data
1001                mountPath: /data
1001             volumeMounts:
1002              - name: roxie-data
1003                mountPath: /data
1003             volumeMounts:
1004              - name: roxie-data
1005                mountPath: /data
1005             volumeMounts:
1006              - name: roxie-data
1007                mountPath: /data
1007             volumeMounts:
1008              - name: roxie-data
1009                mountPath: /data
1009             volumeMounts:
1010              - name: roxie-data
1011                mountPath: /data
1011             volumeMounts:
1012              - name: roxie-data
1013                mountPath: /data
1013             volumeMounts:
1014              - name: roxie-data
1015                mountPath: /data
1015             volumeMounts:
1016              - name: roxie-data
1017                mountPath: /data
1017             volumeMounts:
1018              - name: roxie-data
1019                mountPath: /data
1019             volumeMounts:
1020              - name: roxie-data
1021                mountPath: /data
1021             volumeMounts:
1022              - name: roxie-data
1023                mountPath: /data
1023             volumeMounts:
1024              - name: roxie-data
1025                mountPath: /data
1025             volumeMounts:
1026              - name: roxie-data
1027                mountPath: /data
1027             volumeMounts:
1028              - name: roxie-data
1029                mountPath: /data
1029             volumeMounts:
1030              - name: roxie-data
1031                mountPath: /data
1031             volumeMounts:
1032              - name: roxie-data
1033                mountPath: /data
1033             volumeMounts:
1034              - name: roxie-data
1035                mountPath: /data
1035             volumeMounts:
1036              - name: roxie-data
1037                mountPath: /data
1037             volumeMounts:
1038              - name: roxie-data
1039                mountPath: /data
1039             volumeMounts:
1040              - name: roxie-data
1041                mountPath: /data
1041             volumeMounts:
1042              - name: roxie-data
1043                mountPath: /data
1043             volumeMounts:
1044              - name: roxie-data
1045                mountPath: /data
1045             volumeMounts:
1046              - name: roxie-data
1047                mountPath: /data
1047             volumeMounts:
1048              - name: roxie-data
1049                mountPath: /data
1049             volumeMounts:
1050              - name: roxie-data
1051                mountPath: /data
1051             volumeMounts:
1052              - name: roxie-data
1053                mountPath: /data
1053             volumeMounts:
1054              - name: roxie-data
1055                mountPath: /data
1055             volumeMounts:
1056              - name: roxie-data
1057                mountPath: /data
1057             volumeMounts:
1058              - name: roxie-data
1059                mountPath: /data
1059             volumeMounts:
1060              - name: roxie-data
1061                mountPath: /data
1061             volumeMounts:
1062              - name: roxie-data
1063                mountPath: /data
1063             volumeMounts:
1064              - name: roxie-data
1065                mountPath: /data
1065             volumeMounts:
1066              - name: roxie-data
1067                mountPath: /data
1067             volumeMounts:
1068              - name: roxie-data
1069                mountPath: /data
1069             volumeMounts:
1070              - name: roxie-data
1071                mountPath: /data
1071             volumeMounts:
10
```

# High School Student Project



## HPCC Systems Storage Support with Container Storage Interface (CSI)



Ryan Rao, HPCC Systems intern, American Heritage Palm Beach | Mentors: Xiaoming Wang, Godji Fortil

### Introduction

AWS EFS and AWS FSx for Lustre are two powerful storage services that offer different use cases. EFS is similar to a standard NFS and has a more general usage, while FSx for Lustre is designed specifically for HPC applications due to its fast speed. As it is important to provide users a versatile range of storage options when deploying HPCC Systems, my project focused on implementing these two storages through CSI to allow HPCC Systems to utilize their different features. In the first phase of my project, I focused on developing a storage lifecycle beyond Kubernetes for EFS, and in the second phase, I focused on developing a storage lifecycle within Kubernetes for FSx for Lustre.

### Project Goals

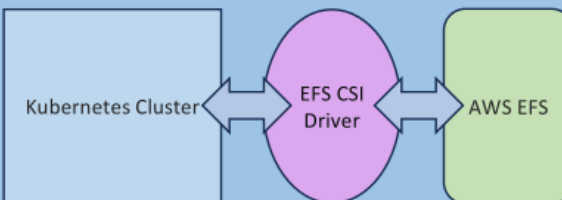
- Investigate and understand the three storage lifecycles
- Design and build the helm chart for the EFS lifecycle beyond Kubernetes
- Explore FSx for Lustre and any limitations
- Design and build the helm chart for the FSx for Lustre lifecycle within Kubernetes

### Three Storage Lifecycles

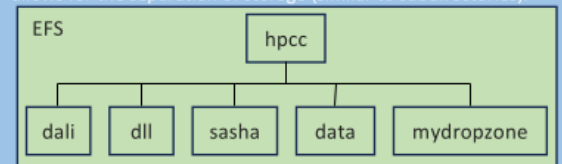
1. Storage lives on the HPCC Systems cluster level. It depends on the HPCC Systems cluster.
2. Storage lives within the Kubernetes cluster. It is independent of the HPCC Systems cluster.
3. Storage lives beyond the Kubernetes cluster. It is independent of the Kubernetes cluster.

### Phase 1: EFS

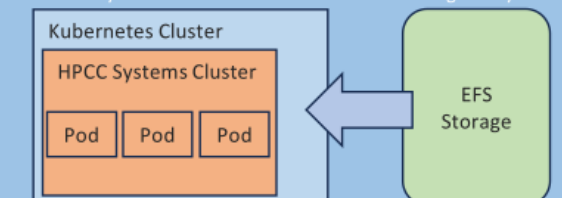
The user installs the EFS CSI driver by running a script. The EFS CSI driver allows the Kubernetes cluster communicate and interact with AWS resources, in this case, EFS.



The pre-created EFS file system uses **five manually created access points**, which act as isolated entry points into the file system. This allows for the separation of storage (similar to subdirectories).

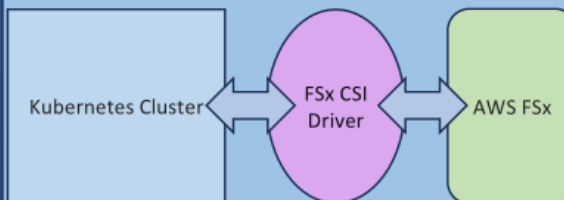


The `hpc-efs-static-pv` helm chart deploys the required Kubernetes components, connecting the external EFS storage to the cluster. The HPCC Systems cluster can now use the 3rd storage lifecycle.

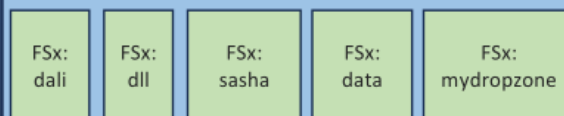


### Phase 2: FSx for Lustre

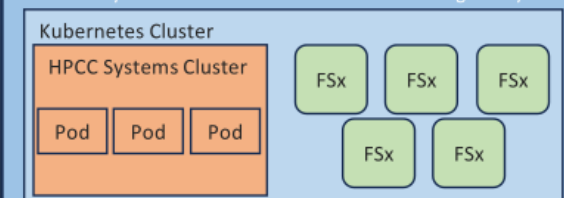
The user installs the FSx CSI driver by running a script. The FSx CSI driver allows the Kubernetes cluster to communicate and interact with AWS resources, in this case, FSx.



FSx doesn't offer an access point feature like EFS, so there is no way to have subdirectories within a single FSx file system. Instead, **five separate FSx file systems** are created, allowing for the separation of storage.



The `hpc-fsx-dynamic-pv` helm chart deploys the required Kubernetes components to set up the FSx storage in Kubernetes. The HPCC Systems cluster can now use the 2nd storage lifecycle.



### Results

- My PR for the EFS storage implementation has been approved for merging. It includes a new helm chart, improved documentation, and simplified code.
- In the next few weeks, I plan to create a PR for the FSx storage implementation, including a new helm chart, documentation, and various other files.
- FSx integrates with S3. After the data is imported from FSx to S3, the FSx volume can be temporarily shut down. This is something to explore in the future.

### Conclusion

As new storage services emerge, it is important for HPCC Systems to integrate with these storage solutions, offering its users a range of choices. To this end, I hope my work with EFS and FSx for Lustre and any new features I have added serve to enhance the platform's capabilities.

### Resources

[Blog](#) | [EFS](#) | [FSx for Lustre](#)





# Internship Project Life Cycle

Onboarding and welcome	Development phase	Completion phase	HPCC Systems Community Day
<ul style="list-style-type: none"><li>• Welcome meeting</li><li>• Infrastructure and access</li><li>• Training</li></ul>	<ul style="list-style-type: none"><li>• Daily meetings</li><li>• Code development</li><li>• Testing</li><li>• Blog reports</li></ul>	<ul style="list-style-type: none"><li>• Check-in code</li><li>• Documentation</li><li>• Team presentations</li><li>• Poster submission</li></ul>	<ul style="list-style-type: none"><li>• Annual conference</li><li>• Poster competition</li><li>• Talk (encouraged!)</li></ul>

# Application Process

## 1) Select your individual project:

- Available projects list
- Suggest your own project

## 2) Write a proposal:

- Liaise with a mentor
- Highlight the deliverables
- Timeline of work for each week

## 3) Submit the proposal and your CV:

- Apply online: [RELX careers website](#)



# Important Dates

- Deadline for applications: March 22, 2024
- Notification of acceptance: up to mid-April, 2024
- Hiring and onboarding: late April-early May, 2024
- Start date: May 20, 2024
- End date: Aug 9, 2024



# Get in Touch

---

[Hugo.Watanuki@lexisnexisrisk.com](mailto:Hugo.Watanuki@lexisnexisrisk.com)

[linkedin.com/in/hugowatanuki](https://www.linkedin.com/in/hugowatanuki)

