

Indexer Clustering Fixups

Cluster recovery process

Da Xu | Engineering | Splunk

splunk

Forward-Looking Statements

During the course of this presentation, we may make forward-looking statements regarding future events or the expected performance of the company. We caution you that such statements reflect our current expectations and estimates based on factors currently known to us and that actual events or results could differ materially. For important factors that may cause actual results to differ from those contained in our forward-looking statements, please review our filings with the SEC.

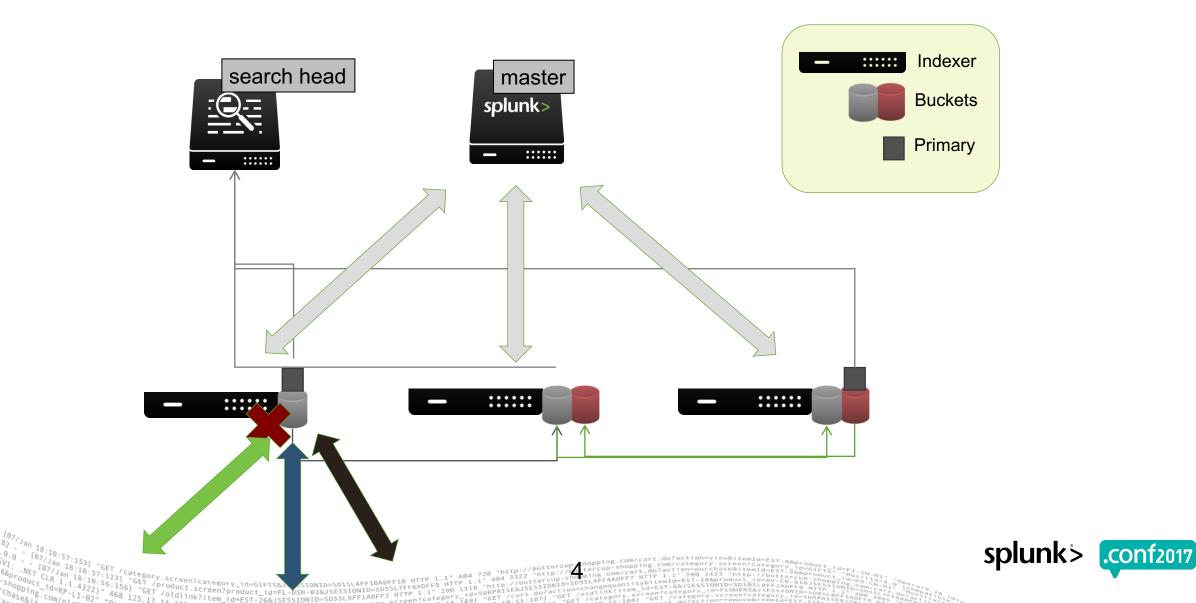
The forward-looking statements made in this presentation are being made as of the time and date of its live presentation. If reviewed after its live presentation, this presentation may not contain current or accurate information. We do not assume any obligation to update any forward looking statements we may make. In addition, any information about our roadmap outlines our general product direction and is subject to change at any time without notice. It is for informational purposes only and shall not be incorporated into any contract or other commitment. Splunk undertakes no obligation either to develop the features or functionality described or to include any such feature or functionality in a future release.

Splunk, Splunk>, Listen to Your Data, The Engine for Machine Data, Splunk Cloud, Splunk Light and SPL are trademarks and registered trademarks of Splunk Inc. in the United States and other countries. All other brand names, product names, or trademarks belong to their respective owners. © 2017 Splunk Inc. All rights reserved.

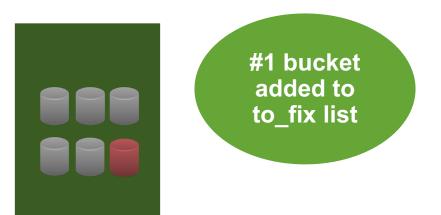
Clustering Fixup Cycle



Clustering Overview - Indexer Failure



Fixup Lifecycle – to_fix lists



to_fix lists

► The Cluster Master (CM) remembers a list of buckets it needs to fix.

Every time a bucket is modified, the CM will add the bucket into any to_fix lists that are needed

There are multiple to_fix lists, each corresponding to a different fixup

 Generation, Replication Factor, Search Factor, Excess Buckets, Summary, etc.

These lists are an internal list of all the buckets that **POTENTIALLY** need fixing

 Very possible that nothing needs to be done - in which case they'll be removed on the service() call (part 2)

splunk> .conf2017

var/log/splunk/metrics.log

08-25-2017 15:02:37.915 -0700 INFO Metrics - group=subtask_counts, name=cmmaster_service, to_fix_streaming=100, to_fix_data_safety=2, to_fix_gen=4, to_fix_rep_factor=104, to_fix_search_factor=57, to_fix_summary=0, to_fix_rebalance=0, to_fix_excess=0, to_fix_sync=0, to_fix_added=7, to_fix_removed =7, to_fix_total=104, count=31

08-25-2017 15:02:06.915 -0700 INFO Metrics - group=subtask_counts, name=cmmaster_service, to_fix_streaming=101, to_fix_data_safety=2, to_fix_gen=4, to_fix_rep_factor=104, to_fix_search_factor=57, to_fix_summary=2, to_fix_rebalance=0, to_fix_excess=0, to_fix_sync=0, to_fix_added=4, to_fix_removed =0, to_fix_total=104, count=31

08-25-2017 14:30:04.917 -0700 INFO Metrics - group=subtask_counts, name=cmmaster_service, to_fix_streaming=102, to_fix_data_safety=2, to_fix_gen=2, to_fix_rep_factor=102, to_fix_search_factor=55, to_fix_summary=2, to_fix_rebalance=0, to_fix_excess=0, to_fix_sync=0, to_fix_added=154, to_fix_remov ed=156, to_fix_total=102, count=31

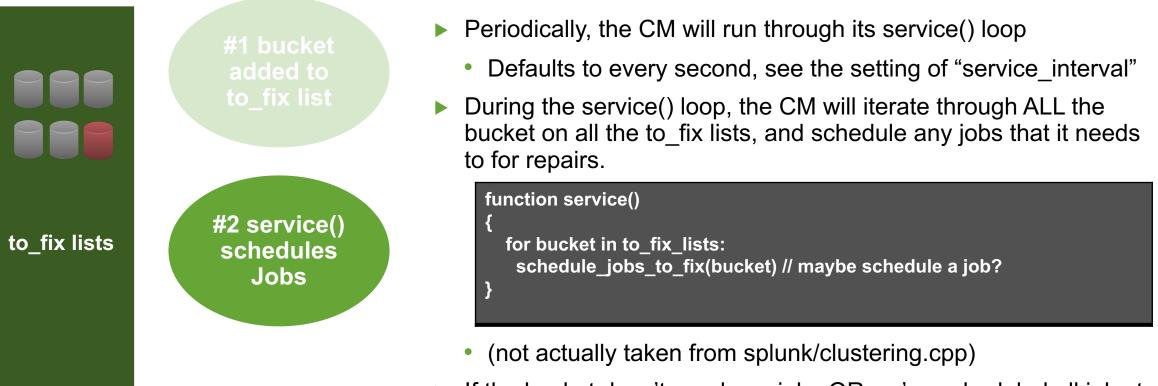
08-25-2017 14:29:33.916 -0700 INFO Metrics - group=subtask_counts, name=cmmaster_service, to_fix_streaming=102, to_fix_data_safety=2, to_fix_gen=3, to_fix_rep_factor=103, to_fix_search_factor=55, to_fix_summary=2, to_fix_rebalance=0, to_fix_excess=0, to_fix_sync=0 to_fix_added=116084, to_fix_re moved=116160, to_fix_total=104, count=22

to_fix_rep_factor=104

- buckets to check if we need/can schedule replications
- to_fix_added=116k to_fix_removed=116k
 - likely a re-add operation a peer resynced its buckets (restart) and most buckets were added/removed from the to_fix list with little jobs scheduled



Fixup Lifecycle – service



Screen?product_id=FL-DSM-01&JS

If the bucket does't need any jobs OR we've scheduled all jobs to fix a bucket, its removed from the to_fix list

splunk

Fixup Lifecycle – jobs

- Once service() schedules jobs, the CM has a threadpool running and picking up all the scheduled jobs.
- These jobs are generally REST POSTs to the cluster Indexers, telling them to run some command

Job name	Function		
CMRepJob	Replicate a bucket (includes recipient info)		
CMChangeBucketJob_build	Make bucket searchable (RF copy -> SF copy)		
CMChangeMasks	Change bucket primaries (batch job, can change maybe buckets)		
CMRollHotBucket	Roll a bucket from hot->warm		
CMTruncJob	Truncate a bucket's size (remove bucket will use this with size=0)		



splunk

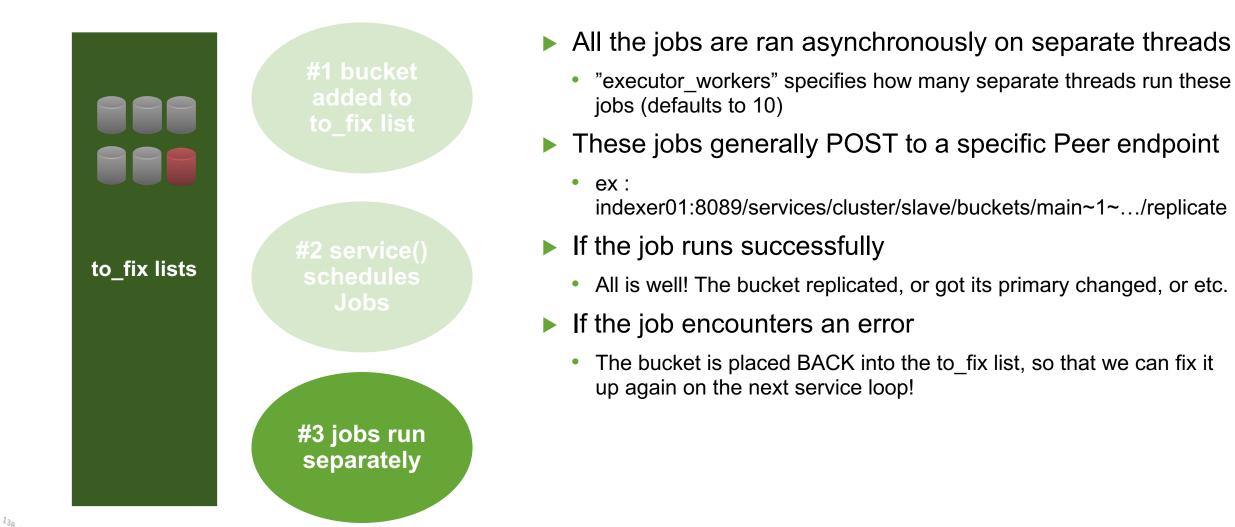
var/log/splunk/metrics.log

INFO	<pre>Metrics - group=jobs, name=cmmaster,</pre>	CMChangeMasksJob=57, CMPeerJob=14, CMRepJob=1, CMRollHotBucketJob=2
INFO	<pre>Metrics - group=jobs, name=cmmaster,</pre>	CMChangeBucketJob=8, CMChangeBucketJob_build=3, CMChangeMasksJob=71, CMPeerJob=5, CMRepJob=25, CMRollHotBucketJob=23
INFO	<pre>Metrics - group=jobs, name=cmmaster,</pre>	CMChangeBucketJob_build=32, CMChangeBucketJob_removePrimary=1, CMChangeMasksJob=16, CMRepJob=40, CMRollHotBucketJob=1
INFO	<pre>Metrics - group=jobs, name=cmmaster,</pre>	CMRepJob=1
INFO	<pre>Metrics - group=jobs, name=cmmaster,</pre>	CMChangeBucketJob_build=1, CMRepJob=1
INFO	<pre>Metrics - group=jobs, name=cmmaster,</pre>	CMRepJob=2, CMTruncJob=2
INFO	<pre>Metrics - group=jobs, name=cmmaster,</pre>	CMChangeBucketJob_build=1, CMRepJob=1
INFO	<pre>Metrics - group=jobs, name=cmmaster,</pre>	CMChangeBucketJob_build=3, CMChangeMasksJob=5, CMPeerJob=4, CMRepJob=7
INFO	<pre>Metrics - group=jobs, name=cmmaster,</pre>	CMChangeBucketJob_build=1, CMRepJob=2

CMChangeMasksJob=57

- The CM ran 57 jobs to change bucket primaries amongst the indexers
- CMRepJob=40
 - The CM ran 40 jobs to start bucket replication jobs amongst the indexers (fixup RF)

Fixup Lifecycle – jobs



.screen?product_id=FL-DSH-01&JSH



Clusteirng Fixup Cycle - examples



to_fix lists metrics

Q New Search					Save As ∽ Close
index=_internal sourc =30s	e=*metrics.log name=cmmaster_ser	<pre>vice timechart max(to_fix_ge</pre>	en) max(to_fix_rep_factor) max(to_fix_search_factor)	Date time range ~ Q
-505					
✓ 18 events (9/5/17 1:41:00.)	000 AM to 9/5/17 1:45:30.000 AM) No Ev	rent Sampling 🗸		Job 🗸 🔢 🔳 🤌 🖶	⊥ • Smart Mode ∽
Events Patterns	Statistics (9) Visualization				
■ Column Chart > F 2,000 1,500 - 1,000 - 500 -	Format V				max(to_fix_gen) max(to_fix_rep_factor) max(to_fix_search_factor)
1:41 AM Tue Sep 5	1:42 AM	1:43 AM	1:44 AM	1:45 AM	
2017					
130,60,4 S	earchable			RF+SF Met	
317 27:20 (07/Jan 18:10:57:153] "GET / 0 ws N7 5:10:82 - (07/Jan 18:10:57:153] "GET / 0 '/butsE57:00 - (07/Jan 18:10:57:123] '/butsE57:16&product CLR 18:10:57:123] opation=shopset int 1:4220:56:163	Sategory.screen?category_id=GIFTS&JSESSIONID=SDISL4F	19ADFF19 HTTP 1.1" 404 720 "http://buttercup.ht 19ADFF19 HTTP 1.1" 404 3322 "http://but 19ADF19TE6ADFF9 HTTP .10" http://buttercup.shou	2 ping.com/cart.do?action=view&itemId=Est.G∏ 2 p.shopping.com/category.screen?category_ided 2 ng.com/cart.do?action=pur23 ************************************	Splui	nk> .conf2017
Outton/chasesit.com/cli-02" 468	"GET'''.Screen?category_id=GIFT5&lSESSIONID=SDISLAF 5) "GET /oldit.Screen?product_id=FL-DSH=01&JSESSIONI 125.17 14 controls/item_id=EST-26&JSESSIONID=SD55L9FF1	D=50352,TTP 1.1" 200 1314 ADFF3 HTTP 1.1" 200 id=SURPRISE&JS550 (0?action=C) ADFF3 HTTP 1.1" 200 id=SURPRISE&JS551871 (0?action=C) GET /0 GET /0 (0) (0) (0) (0) (0) (0) (0) (0) (0) (0	Idlink?item_10=category_id=FLOWERS&JSESSIONID=S Idlink?item_10=category.screen?category_id=SU /category.screen?category_id=SU /category.screen?category_id=SU	ITP 1. ADDESS SELENCE GOVERNOUS SO 38518FF1ADF SADE SELENCE GOVERNOUS SO RPRISEESSEESE SELENCE GOVERNUESSE RPRISEESSEESE	

jobs metrics

Q New Search					Save As ∽ Close
index=_internal source	=*metrics.log group=jobs n	name=cmmaster timechart max	(*Job*) span=10s		Date time range ~ Q
✓ 8 events (9/5/17 1:41:00.000	0 AM to 9/5/17 1:46:00.000 AM)	No Event Sampling 🗸		Job 🗸 🔲 🛋 🤌	🖶 🛓 🍷 Smart Mode 🗸
Events Patterns	Statistics (30) Visualiz	ation			
	rmat 🗸				
1,000					
100					
10					max(CMChangeBucketJob_build) max(CMChangeMasksJob) max(CMRepJob)
					тах(Смкерјор)
1:41 AM Tue Sep 5 2017	1:42 AM	1:43 AM	1:44 AM	1:45 AM	
				RF+SF M	et
122 60 4 228 241 317 22 220 107/1 317 22 220 107/1	chable				
(-N7-7.16602 'an 18:10:57:153) "GET /cat temId=E5.1:50.0 ' [07/Jan 18:10:57:153] "GET /cat //butte=E5.1682p. NET (c/Jan 10:57.15.1) practice.ce.c.	egory -		uttercup a Sping.com/cart.do?action=view&i http://lukteScup.shopping.com/category.scre http://lukteScup/cart.do?action=purchase& scionFF7 WITP 1.1. 2006 2423 */	temid=Est-G&productid=F1 enfcategory_id=cif_E:d=F1 itemid=cory_id=cif_E:d=F1 ttp://buriz_S&productMo_SW-91.	lunk'> .conf2017
pping.com/caseseit.com/call.021 "G	"BOTY.SCreen?category_id=GIFTS&JSESSIONIC "GET /product.screen?product_id=FL-DSH-01& "GET /oldlink?item_id=EST-26&JSESSIONID=")=SD1SL4FF10ADFF10 HTTP 1.1" 404 720 JSESSIONID=SD5SL7FF6ADFF0 HTTP 1.1" 404 3322 JSESSIONID=SD5SL7FF6ADFF0 HTTP 1.1" 200 1318 "http://butl SD5SL9FF1ADFF3 HTTP 1.1" 200 1318 "http://cat. SD5SL9FF1ADFF3 HTTP 1.1" 200 1318 "http://cat.	ercup-snc+id+F4ADFF7 Hill SIONID-SD95L4FF4ADFF7 Hill ND=SD95L4Ff4ADFF7 Hill ND2action=changequantity&itemId=EST-6&JSESIONID=S/ ND2action=changequantid=EST-6&JSESIONID=S/ ND2action=changequantid=ELOW 77 GET /category.scree GET /category.scree GET /category.scree	ct_ld=AV-CD_SUB_sDODC=1d=KF11a/a_box DISSLSF7ADFF8 HSS5100B_CDWW2018 HERSAUSESSTONICOSDBS_LS11 - 23055171; cm/00080430 HERSAUSESSTONICOSDBS_LS11 - 23055171; cm/00080430 HERSAUSESTONICOSDBS_LS11 - 23055171; cm/00080400000000000000000000000000000000	-

jobs by time

Q New Search				Save As ∨	Close
<pre>index=_internal source=*master*splunkd.log CMRepJob</pre>	timechart span=15s count by job	5		Date time range 🗸	Q
✓ 3,264 events (9/5/17 1:41:00.000 AM to 9/5/17 1:46:00.000 AM)	Io Event Sampling 🗸		Job 🗸 🔢 🔳 🤌 🍓	上 🥊 Smart M	lode 🗸
Events Patterns Statistics (20) Visualization					
∎ Column Chart ✓					
1,000	CMChangeBucketJob_build CMC	ChangeMasksJob 📕 CMRepJob			
1,000					
100					
10					
1:41 AM Tue Sep 5 2017	1:43 AM	1:44 AM	1:45 AM		
searchable			RF+	-SF Met	
3/241 (%) (%) (%) (%) (%) (%) (%) (%) (%) (%)	10ADFF10 HTTP 1.1" 404 720 "http://buttercup.shopping. =SDS5L7FF6ADFF9 HTTP 1.1" 404 3322 "http://buttercup-shdbimme.com =SDS5L7FF6ADFF9 HTTP 1.1" entric/buttercup-shdbimme.com DFF3 HTTP 1.1" entric/id=SURFEISEAJSESSIONID=SDS1AFFAAD FF3 HTTP / 1.1" entric/id=SURFEISEAJSESSIONID=SDS1AFFAAD	com/cart.do?action=viou&itomId=EST-6&production hopping.com/category.scroon?category_id=GIF5_id=Fi /cart.do?il.iv.gog2423 "http://bir26&product_imoz /FF7 MitemId=EST-18&product_id=Av-CB_60B_5hoppind=k ittmid=fr64JSESIONID=S1081.8FESIONES_1081.5 itemid=fr64JSESIONID=S1081.8FESIONED=s081.5 itemid=fr64JSESIONID=S081.8FESIONED=s081.5 itemid=fr64JSESIONID=s081.8FESIONED=s081.5 itemid=fr64JSESIONID=s081.8FESIONED=s081.5 /068] "GET /category_id=FLOWERS4JSESSIONID=s081.5 /068] "GET /category_id=FLOW	splu	nk'> .conf2	017



Don't forget to rate this session in the .conf2017 mobile app



APPENDIX Searches

- 1. index=_internal host=MASTER source=*splunkd.log* CMRepJob running job | timechart count by job
 - Master jobs ran
- 2. index=_internal source=*metrics.log* name=cmmaster group=jobs | timechart max(CM*)
 - Master jobs metrics
- 3. index=_internal source=*metrics.log* *fix* host=MASTER | timechart max(to_fix_*)
 - to_fix list sizes
- 4. index=_internal source=*metrics.log* group=subtask_seconds name=cmmaster | timechart max(service)
 - Master time spent calling service() in between previous log to metrics.log (every 30s)

