Scaling the Data Infrastructure @ Spotify

matti@spotify.com



Matti Pehrs

> 25 years in IT
Emacs since 1987
Java since 1995
Spotify since 2013



Agenda

- 1. Data at Spotify
- 2. Summer of 2015
- **3.** Challenges & Victory
 - Datamon
 - o Styx
 - GABO

Data at Spotify

Hadoop



Data history through Hadoop

- → Started with 5 servers in the office 2007
- → Tried Amazon Elastic Map/Reduce 2010
- → In 2012 we started using Hortonworks HDP
- → We now run a 2000+ node cluster in London
- → We are out of physical space!





Spotify is moving to the cloud



It's all about s focus.

Spotify's core business is to <u>serve music</u> not operate data centers.



•

۲

Spotify big-data context

- Over 100 million monthly active users
- Over 30 million song
- Over 2 billion playlists
- Active in 60 markets



Our growth in Data



Data is at the heart of Spotify

In 2007

- Reporting



In 2017

- Reporting
- All features use big data in some form











Summer of 2015



EN PRE-TASK PL N N ALWAYS REME BER:

· Colonify Potential Ha

Identify Required Tools/Education



• A strain of incidents



- A strain of incidents
- War-room



- A strain of incidents
- War-room
- Hadoop on it's knees



- A strain of incidents
- War-room
- Hadoop on it's knees
- Event Delivery Catch up



- A strain of incidents
- War-room
- Hadoop on it's knees
- Event Delivery Catch up
- Reprocessing of data



- A strain of incidents
- War-room
- Hadoop on it's knees
- Event Delivery Catch up
- Reprocessing of data
- Hard to debug data issues



1. Early Warning

Datamon - Data monitoring

1. Early Warning

Datamon - Data monitoring

2. Debuggability & Control

Styx - Scheduling and control

1. Early Warning

Datamon - Data monitoring

2. Debuggability & Control

Styx - Scheduling and control

3. Automate Capacity

GABO - Event Delivery

1. Early Warning

Datamon - Data monitoring

2. Debuggability & Control

Styx - Scheduling and control

3. Automate Capacity

GABO - Event Delivery

Early Warning - Datamon



Early Warning - Datamon

• Unified view

• Alignment between teams

• Ownership

- Clear ownership of data
- SLA
 - Alert on late data

Early Warning - Datamon

• Define terminology

• Provide metadata language

Implement a Datamon service



7 Nov

Data Endpoint

coredata.activation.Dataset1 coredata.activation.Dataset2 coredata activation Dataset3 coredata.activation.Dataset4 coredata.activation.Dataset5 coredata activation. Dataset6 coredata.activation.Dataset7 coredata.activation.Dataset8 coredata.activation.Dataset9 coredata.activation.Dataset10 coredata.activation.Dataset11 coredata.activation.Dataset12 coredata.activation.Dataset13 coredata.activation.Dataset14 coredata.activation.Dataset15 coredata.activation.Dataset16 coredata.activation.Dataset17 coredata.activation.Dataset18 coredata.activation.Dataset19

1. Early Warning

Datamon - Data monitoring

2. Debuggability & Control

Styx - Scheduling and control

3. Automate Capacity

GABO - Event Delivery

- Execution control

- Self service for data users
- Execution information
 - Expose debug information
- Execution isolation
 - Docker for data jobs



• Execution control

Centralized execution API

E DATA ENDPOINTS Icoredata.activation.Dataset1hourlycoredata.activation.Dataset2hourlycoredata.activation.Dataset3hourlycoredata.activation.Dataset4hourlycoredata.activation.Dataset5hourlycoredata.activation.Dataset6

• Execution control

- Centralized execution API
- Backfilling and reprocessing

💹 Terminal - matt	i@matti-ThinkPad-T5	40p: /opt/matti/src/tmp	- • ×
File Edit View Terminal	Tabs Help		
2016-10-28T12:08:38	submit	Execution description: ExecutionDescription{dockerImage=regis	try.spotify.net/spotify
/some-job:bc25f14-1477	653357111, dockerArgs=[wr	ap-luigi,hydra,module, tst, matti.SampleJob,datehour, {},	local-scheduler], se
cret=Optional[Secret{n	ame=pipeline-core-secret,	mountPath=/etc/gcp-keys}], commitSha=Optional[bc25f141e00e9ff20ea	1f953cc8c91ad/9/f54d5]}
2016-10-28112:08:38	submitted	Execution id: styx-run-y9gd/	
2016-10-28112:08:39	started		
2016-10-28112:08:55	terminate	Exit code: 0	
2016-10-28112:08:55	success	T-/// -//	
2016-10-28115:41:42	triggerExecution	Trigger id: ad-hoc-cli-14/7669302657-07468	
2016-10-28115:41:42	SUDMIT	Execution description: ExecutionDescription{dockerimage=regis	try_spotity.net/spotity
/some-job:21a3469-14//	668/49515, dockerArgs=[Wr	ap-luigi,nydra,module, tst, matti.SampleJob,datenour, {},	local-scheduler], se
cret=Optional[Secret{n	ame=pipeline-core-secret,	mountPath=/etc/gcp-keys}], commitSha=Optional[2fa3469/88abbb16e94	8c5b/a9590e51/003a630]}
2016-10-28115:41:43	submitted	Execution id: styx-run-oevn9	
2016-10-28115:41:44	started		
2016-10-28115:43:03	terminate	Exit code: 35	
2016-10-28115:43:03	retryAtter	Delay (seconds): 180	
2016-10-28115:46:04	retry		
2016-10-28115:46:04	Submit	Execution description: ExecutionDescription{dockerimage=regis	try.spotity.net/spotity
/some-job:21a3469-14//	668/49515, dockerArgs=[Wr	ap-luigi,nydra,module, tst, matti.SampleJob,datenour, {},	local-schedulerj, se
cret=Uptional[Secret{n	ame=pipeline-core-secret,	mountPath=/etc/gcp-keys}], commitSha=uptional[21a3469/88abbb16e94	8C20/93230621/0039930]}
2010-10-28115:40:04	submitted	Execution id: styx-run-tcasd	
2016-10-20115:40:00	started	Full the sector 25	
2016-10-28115:47:27	terminate	Exit code: 35	
2016 10 20115:47:27	retryAtter	Delay (seconds): 180	
2016-10-20115:50:20	retry		
2010-10-28115:50:28	SUDMIT	Execution description: ExecutionDescription(dockerimage=regis	try.spotity.net/spotity
/some-job:2183469-14//	666749515, dockerargs=[wr	ap-luigi,nydra,module, tst, matti.SampleJob,datenour, {},	LOCAL-SCHEduler], Se
	ame=pipetine-core-secret,	mountPath=/etc/gcp-keys], commitsha=0ptionat[21a3469/88a00016e94	9020199290621100290201}
2016-10-20115:50:20	submitted	Execution Id: Styx-Tun-Tvj15	
2010-10-20115:50:50	started	Evit rade, 35	
2010-10-20113:31:19	retrutter	Delaw (seconds): 190	
2010-10-20115:51:19	retry	beray (seconds): 180	
2010-10-20115:54:20	retry	Everytics description, EveryticsDescription[deckerTesso_resis	ter contific not (contific
/comp_iob+920242f_1477	660071257 dockortrac-fu	an luigi budra modulo tet matti Camploloh datohour []	local schodulorl so
cret=Ontional [Secret in	ame-nineline.core.secret	mountPath=/etc/acn-keysll_commitSha=Ontional[8a0242fe7cA0f5dea8A	27d559f4426364aa1a98b11
2016-10-20115-54-20	submitted	Execution id: stux run llm24	210333144203040013040013
2016-10-20115-54-20	started	Excedence of a style for comment	
2010-10-20115.34.27	terminate	Exit code: A	
2016-10-20110-33-44	SUCCESS	LAIL COUC. V	
stvx retry nineline-	core matti Sampleloh GCP	2016-10-27700	

• Execution control

- Execution information
 - Timeline



 \equiv

 \equiv

Google Cl

Stackdriver

Logging

- Execution control
- Execution information
 - Timeline
 - Google Cloud Logging



oud Platfo	Drm Spotify - EmployeeGameday -				
	Logs				
etrics	label:container.googleapis.com/pod_name:styx-run-cb208aa6-04a6-4bc9-a85c-70a8a11b836b ×				
	Container Engine, All cluster IDs, All nam 👻 All logs 💌 Any log level 👻 Jump				
	2016-11-04				
	III 10:29:44.000 INFU: INFU [Inread-0] (AsyncLonnection.java:154) - Shutdo				
	III 10:30:44.000 /src/luigi/luigi/target.py:182: UserWarning: File system				
	III 10:30:44.000 warnings.warn("File system {} client doesn't support atom				
	III 10:30:44.000 DEBUG: Running file existence check: hadoop jar /data/gcs				
	II 10:31:01.000 DEBUG: Running file existence check: hadoop jar /data/gcs				
	III 10:31:04.000 DEBUG: Running file existence check: hadoop jar /data/gcs				
	III 10:31:16.000 INFO: [pid 1] Worker Worker(salt=637160908, workers=1, ho				
	II:31:16.000 Warning: Cannot read schema for gs://pipeline-core/test/a				
	II:31:18.000 DEBUG: 1 running tasks, waiting for next task to finish				
	B 10:31:18.000 INFO: Informed scheduler that task coredata.CreateUserMap				
	10:31:18.000 DEBUG: Asking scheduler for work				
	10:31:18.000 DEBUG: Pending tasks: 1				
	III 10:31:18.000 INFO: [pid 1] Worker Worker(salt=637160908, workers=1, ho				
	III 10:31:18.000 DEBUG: Running file existence check: hadoop jar /data/gcs				
	III 10:31:22.000 DEBUG: Running file existence check: hadoop jar /data/gcs				
	II 10:31:27.000 INFO: Publishing dataset to path: gs://pipeline-core/test				
	III 10:31:27.000 DEBUG: Running file existence check: hadoop jar /data/gcs				
	The second secon				

- Execution control
- Execution information
- Execution isolation
 - Docker



1. Early Warning

Datamon - Data monitoring

2. Debuggability & Control

Styx - Scheduling and control

3. Automate Capacity

GABO - Event Delivery

• Complex and manual config



- Complex and manual config
- Pubsub & Dataflow streaming



- Complex and manual config
- Pubsub & Dataflow streaming
- Pubsubs at scale



- Complex and manual config
- Pubsub & Dataflow streaming
- Pubsubs at scale
- Dataflow streaming



- Complex and manual config
- Pubsub & Dataflow streaming
- Pubsubs at scale
- Dataflow streaming :-(
- 2 micro services + 1 Map/Reduce job



- Complex and manual config
- Pubsub & Dataflow streaming
- Pubsubs at scale
- Dataflow streaming :-(
- 2 micro services + 1 Map/Reduce job
- Autoscaling & The Stuffer



Summary

• Make sure you have the right tools to deal with data incidents

 Make sure you have time to implement the tools you need

- Remember that your capacity model can fail at larger scale
 - Keep track of your scale and Automate, automate, automate...



Thank you!

matti@spotify.com

Want to join the band? http://spoti.fi/jobs

