

Thanks for coming early!

# What to do when your job goes OOM in the night (Flowcharts!)

Anya Bida, Jacek Laskowski, Holden Karau

While you wait :D

<https://holdenk.github.io/spark-flowchart>

# What to do when your job goes OOM in the night (Flowcharts!)

Anya Bida, Jacek Laskowski, Holden Karau

Meet Holden



OSS Engineer, queer AF,  
co-author of some books

Anya



Tech Evangelist

Jacek\*



Freelance Awesomeness

Author: [Internals of Apache Spark](#) & other books

(\*sadly not onstage, they only let us have two people :( but he is \*awesome\* and he'll be at the conference)

...we all have spent waaay too much time debugging Spark jobs

# What are we talking about?

- What to do when your Spark job fails
- Or it's slow
- Or you have some spare time to contribute to open source
- Or... the magic of flowcharts (insert jazz-hands here)
- And of course: high pressure timeshare sales. Wait no low pressure book sales.

# Ok but more seriously:

- [OG Spark Flowchart \(video\)](#)
- [New Spark Flowchart \(source\)](#)
- How you can contribute (aka do our jobs for us)
- Something something timeshares. Wait books.
- Some quick hands up very unbiased surveys :p

# Dedication

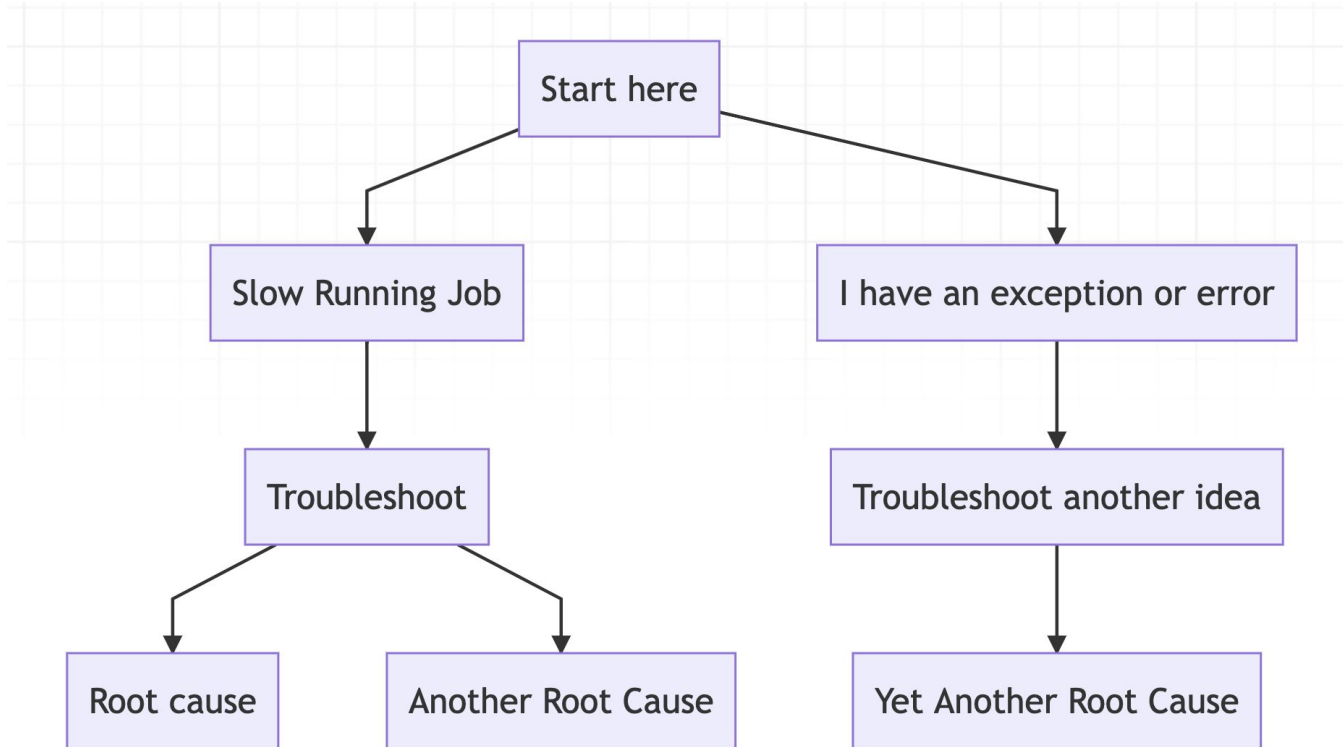
To all the engineers tasked with troubleshooting  
your (team's) spark jobs

```
org.apache.spark.scheduler.TaskSetManager: Starting task 119.0 in stage 2.0 (TID 422) in 62903 ms on 24-8a-07-d6-7f-30.pa4.hpc.i
org.apache.spark.scheduler.TaskSetManager: Finished task 119.0 in stage 2.0 (TID 422) in 62903 ms on 24-8a-07-d6-7f-30.pa4.hpc.i
org.apache.spark.scheduler.TaskSetManager: Starting task 132.0 in stage 2.0 (TID 435, 24-8a-07-d6-7f-30.pa4.hpc.i
org.apache.spark.scheduler.TaskSetManager: Finished task 132.0 in stage 2.0 (TID 435, 24-8a-07-d6-7f-30.pa4.hpc.i
org.apache.spark.scheduler.TaskSetManager: Starting task 146.0 in stage 2.0 (TID 442, 24-8a-07-d6-7f-30.pa4.hpc.i
org.apache.spark.scheduler.TaskSetManager: Finished task 146.0 in stage 2.0 (TID 442, 24-8a-07-d6-7f-30.pa4.hpc.i
org.apache.spark.scheduler.TaskSetManager: Starting task 122.0 in stage 2.0 (TID 429) in 62903 ms on 24-8a-07-d6-7f-30.pa4.hpc.i
org.apache.spark.scheduler.TaskSetManager: Finished task 122.0 in stage 2.0 (TID 429) in 62903 ms on 24-8a-07-d6-7f-30.pa4.hpc.i
org.apache.spark.scheduler.TaskSetManager: Starting task 146.0 in stage 2.0 (TID 442, 24-8a-07-d6-7f-30.pa4.hpc.i
org.apache.spark.scheduler.TaskSetManager: Finished task 146.0 in stage 2.0 (TID 442, 24-8a-07-d6-7f-30.pa4.hpc.i
```

**stressed**  
I'm too busy to find a root cause quickly.  
**sleepy**  
**distracted**

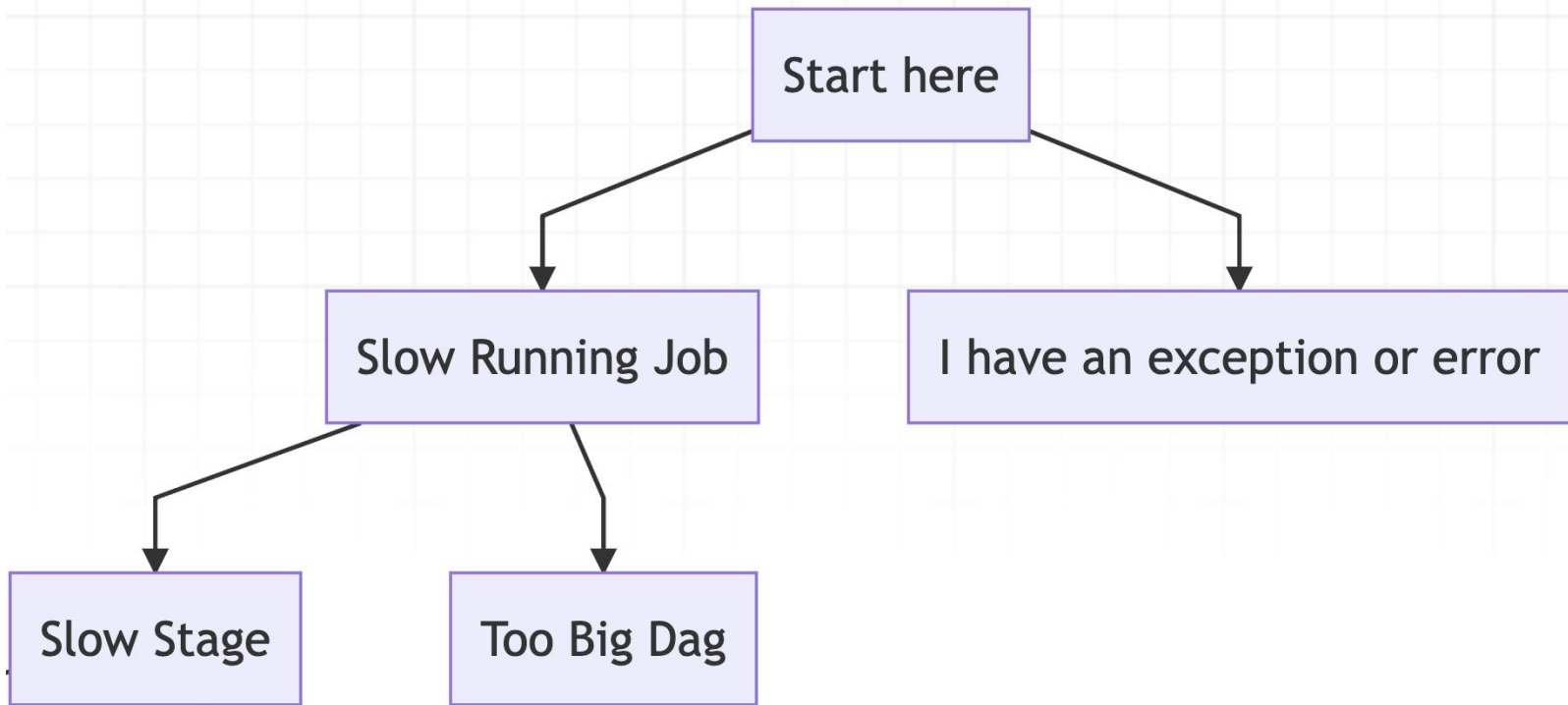
A flow chart would be nice here.

# Given a symptom, how would my lead engineer troubleshoot?





# Introducing spark-flowchart



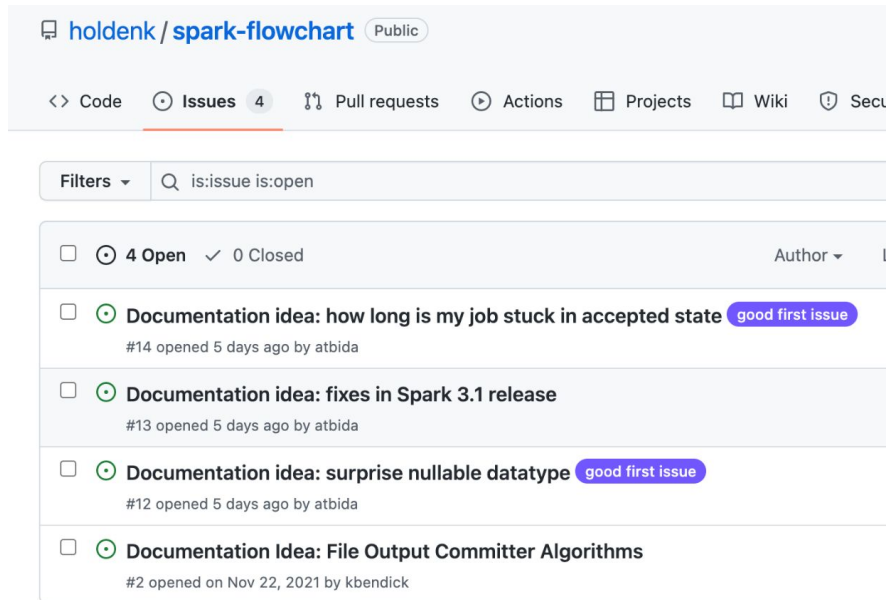
# How spark-flowchart works

- Hosted mkdocs server
- Read access for everyone

## How to contribute

```
git pull /spark-flowchart  
pip install -r requirements.txt  
mkdocs serve
```

- Add pro-tip text. View local changes instantly.
- Create a PR



holdenk / spark-flowchart Public

<> Code Issues 4 Pull requests Actions Projects Wiki Secu

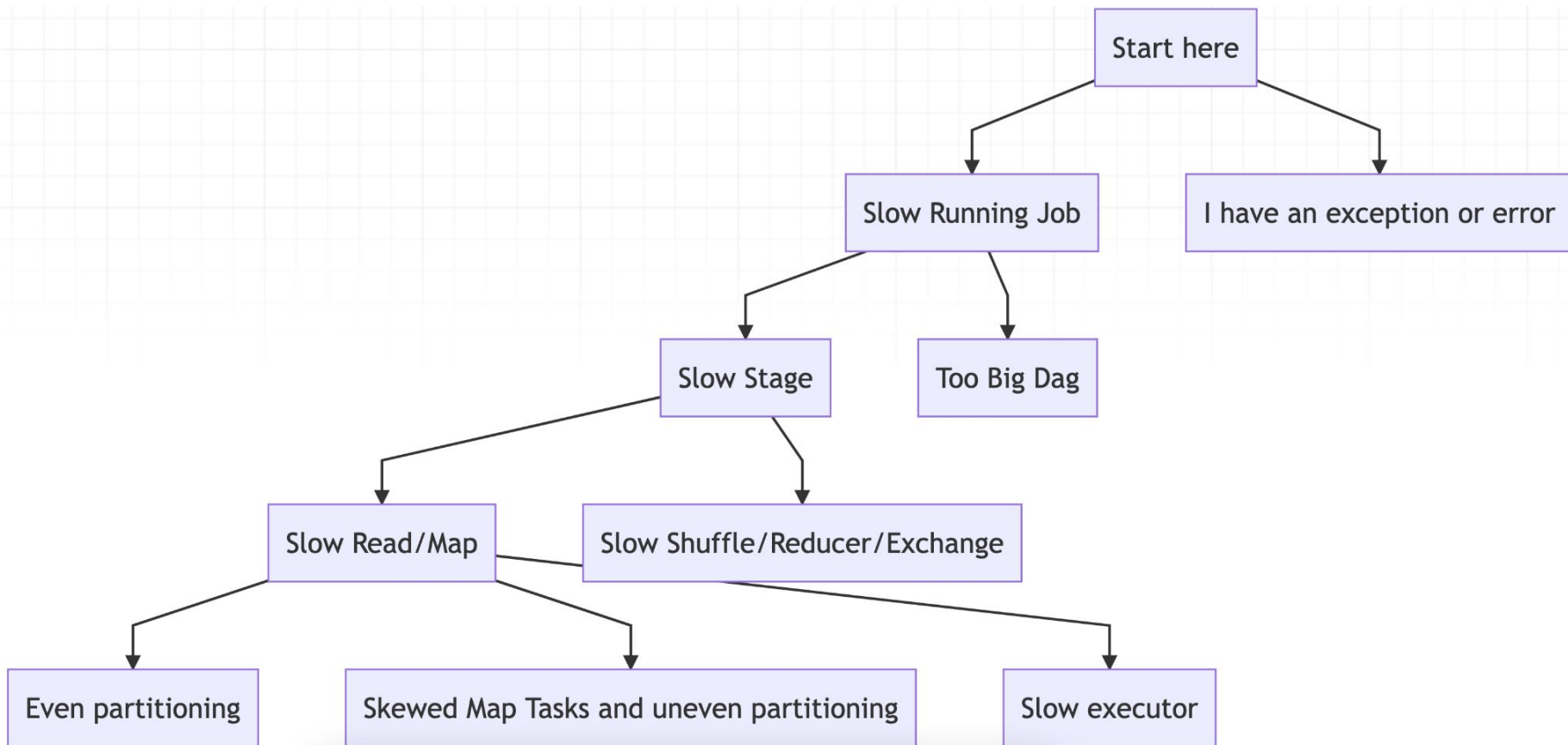
Filters Q is:issue is:open

4 Open ✓ 0 Closed Author

- Documentation idea: how long is my job stuck in accepted state** good first issue  
#14 opened 5 days ago by atbida
- Documentation idea: fixes in Spark 3.1 release**  
#13 opened 5 days ago by atbida
- Documentation idea: surprise nullable datatype** good first issue  
#12 opened 5 days ago by atbida
- Documentation Idea: File Output Committer Algorithms**  
#2 opened on Nov 22, 2021 by kbendick

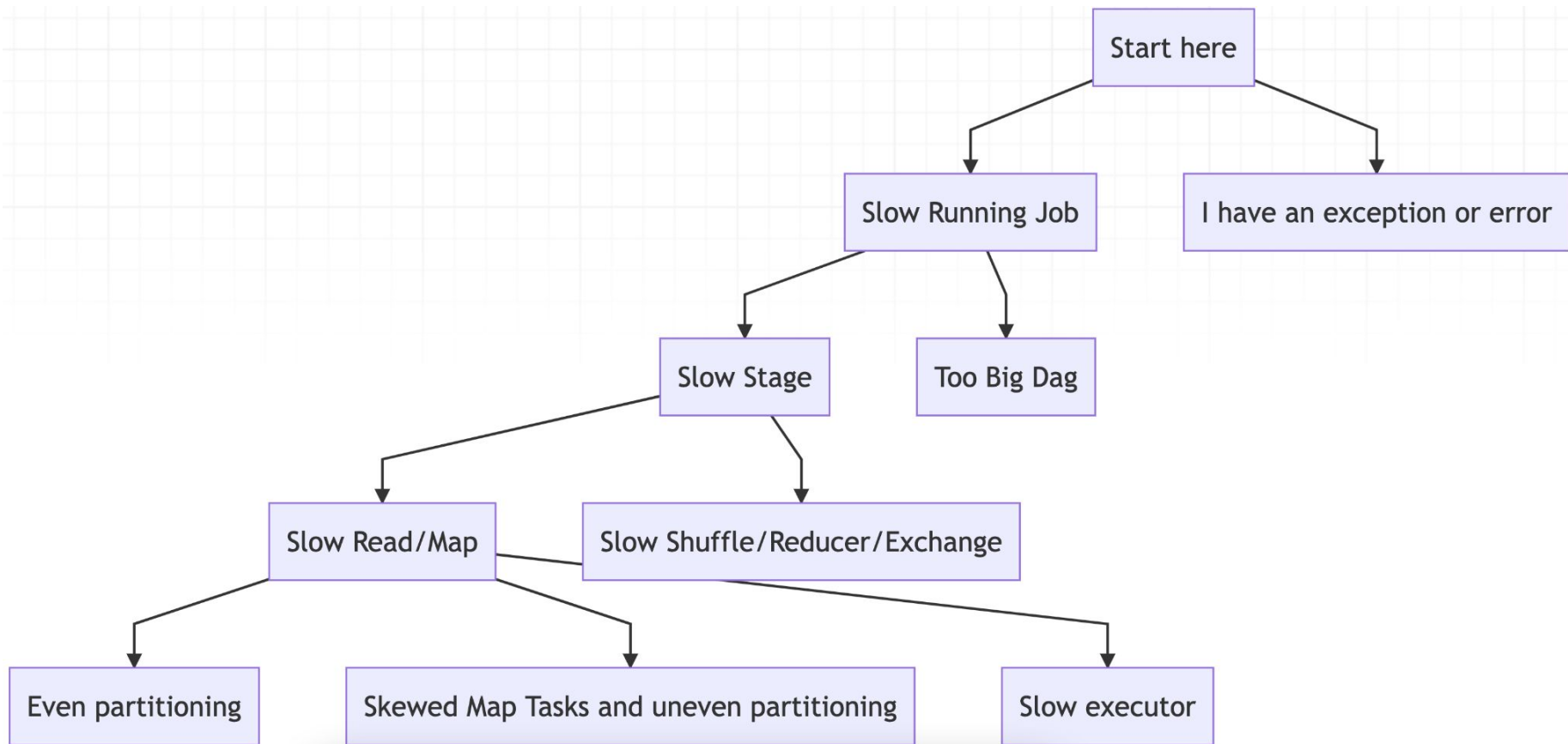


# Introducing [spark-flowchart](#)



# Let's find some patterns

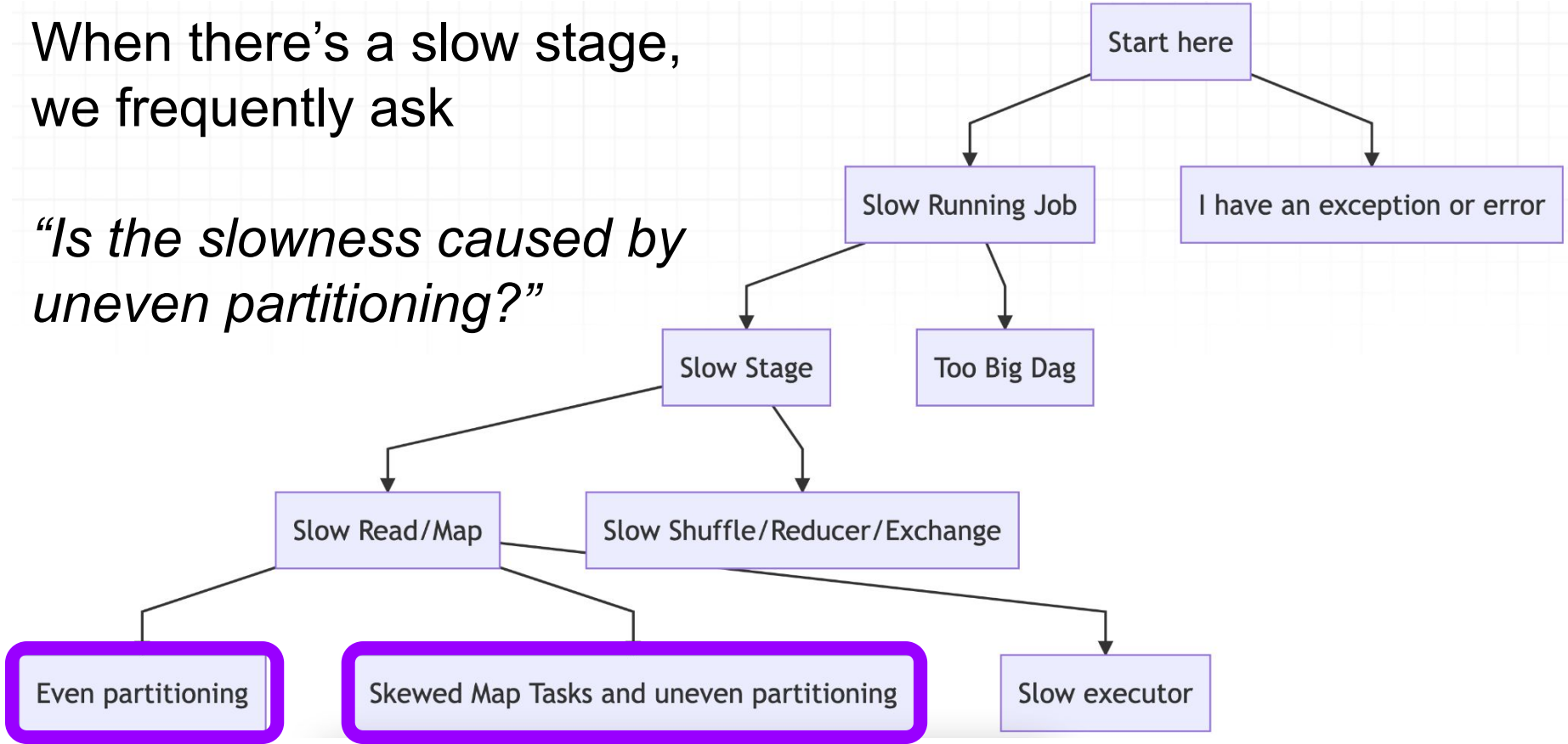
- 1. Slow stage → Partitioning
- 2. Exception → OOM



Pattern:

When there's a slow stage,  
we frequently ask

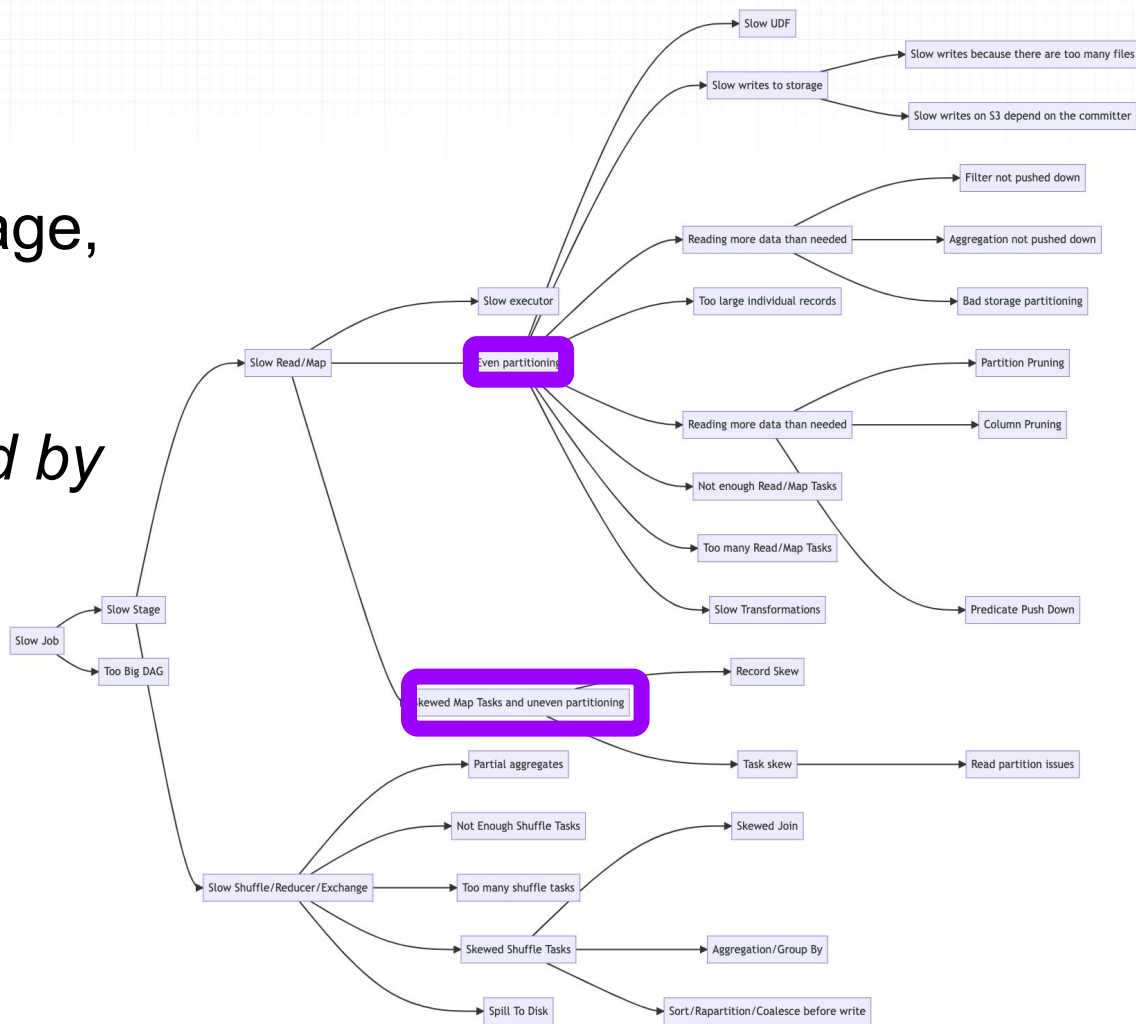
*“Is the slowness caused by  
uneven partitioning?”*



# Pattern:

When there's a slow stage, we frequently ask

*“Is the slowness caused by uneven partitioning?”*

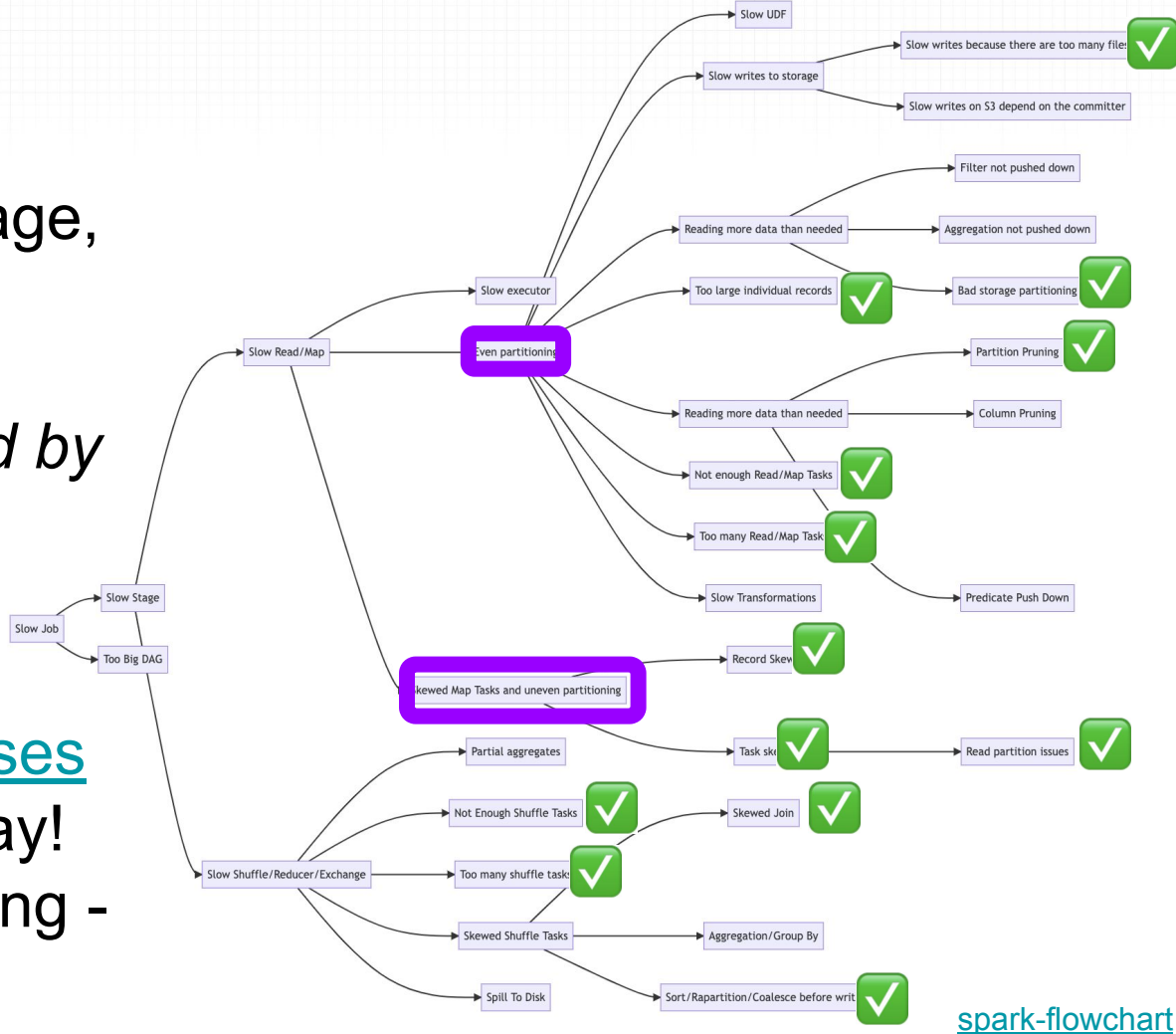


Pattern:

When there's a slow stage, we frequently ask

*"Is the slowness caused by uneven partitioning?"*

Many different root causes can be identified this way!  
Related (✅) to partitioning - or not!





# Let's find some patterns

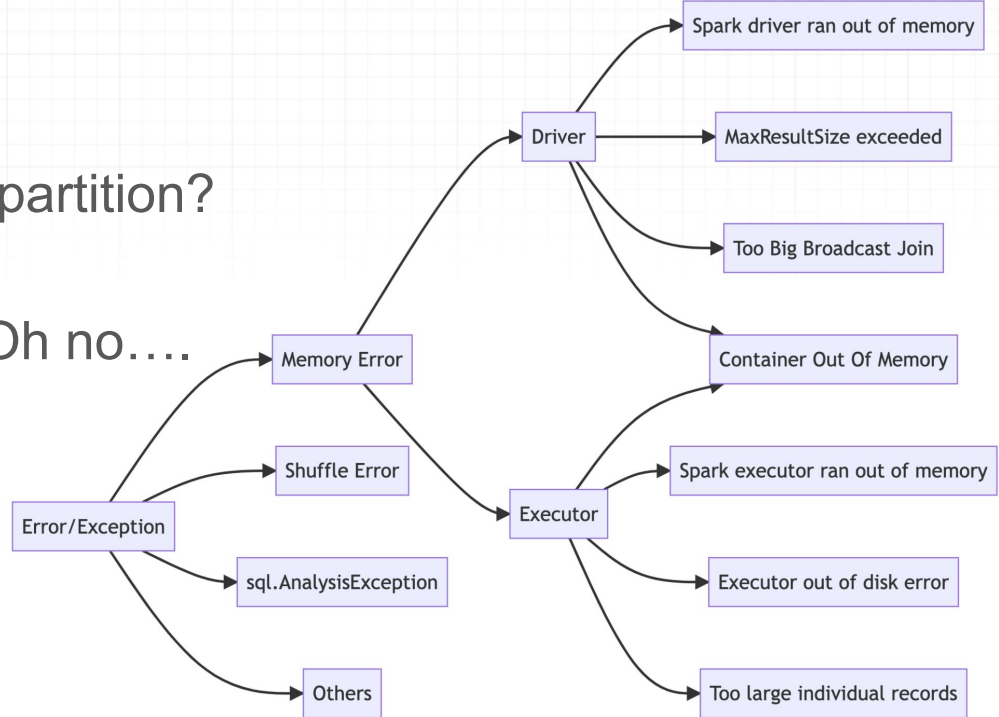
1. Slow stage → Partitioning

 2. Exception → OOM

# Pattern: The many different OOMs of Spark



- Driver OOM
  - No collect for you
- Executor OOM
  - Ooops is all my data in one partition?
- Container OOM
  - Wait am I running Python? Oh no....
- .... mystery OOM



```
org.apache.spark.scheduler.TaskSetManager: Starting task 121.0 in stage 2.0 (TID 428, 24-8a-07-d6-7f-30.pa4.hpc.i
org.apache.spark.scheduler.TaskSetManager: Finished task 110.0 in stage 2.0 (TID 416) in 63410 ms on 24-8a-07-d6-
org.apache.spark.scheduler.TaskSetManager: Starting task 122.0 in stage 2.0 (TID 429, 24-8a-07-d6-7f-30.pa4.hpc.i
org.apache.spark.scheduler.TaskSetManager: Finished task 111.0 in stage 2.0 (TID 417) in 62711 ms on 24-8a-07-d6-
org.apache.spark.scheduler.TaskSetManager: Starting task 124.0 in stage 2.0 (TID 430, 24-8a-07-d6-7f-30.pa4.hpc.i
org.apache.spark.scheduler.TaskSetManager: Finished task 112.0 in stage 2.0 (TID 418) in 63410 ms on 24-8a-07-d6-
org.apache.spark.scheduler.TaskSetManager: Starting task 125.0 in stage 2.0 (TID 431, 24-8a-07-d6-7f-30.pa4.hpc.i
org.apache.spark.scheduler.TaskSetManager: Finished task 113.0 in stage 2.0 (TID 419) in 63410 ms on 24-8a-07-d6-
org.apache.spark.scheduler.TaskSetManager: Starting task 126.0 in stage 2.0 (TID 432, 24-8a-07-d6-7f-30.pa4.hpc.i
org.apache.spark.scheduler.TaskSetManager: Finished task 114.0 in stage 2.0 (TID 420) in 63410 ms on 24-8a-07-d6-
org.apa
org.apa
org.apa
org.apa
```

I nailed the root cause!

Like a boss!

```
org.apache.spark.scheduler.TaskSetManager: Starting task 127.0 in stage 2.0 (TID 433, 24-8a-07-d6-7f-30.pa4.hpc.i
org.apache.spark.scheduler.TaskSetManager: Finished task 115.0 in stage 2.0 (TID 421) in 63410 ms on e4-1d-2d-19-
-7f-d0.pa4.hpc.i
org.apache.spark.scheduler.TaskSetManager: Starting task 128.0 in stage 2.0 (TID 434, 24-8a-07-d6-7f-30.pa4.hpc.i
org.apache.spark.scheduler.TaskSetManager: Finished task 116.0 in stage 2.0 (TID 422) in 63410 ms on e4-1d-2d-19-
-7f-d0.pa4.hpc.i
org.apache.spark.scheduler.TaskSetManager: Starting task 129.0 in stage 2.0 (TID 435, 24-8a-07-d6-7f-30.pa4.hpc.i
org.apache.spark.scheduler.TaskSetManager: Finished task 117.0 in stage 2.0 (TID 423) in 63410 ms on e4-1d-2d-19-
-7f-d0.pa4.hpc.i
org.apache.spark.scheduler.TaskSetManager: Starting task 130.0 in stage 2.0 (TID 436, 24-8a-07-d6-7f-30.pa4.hpc.i
org.apache.spark.scheduler.TaskSetManager: Finished task 118.0 in stage 2.0 (TID 424) in 63410 ms on e4-1d-2d-19-
-7f-d0.pa4.hpc.i
org.apache.spark.scheduler.TaskSetManager: Starting task 131.0 in stage 2.0 (TID 437, 24-8a-07-d6-7f-30.pa4.hpc.i
org.apache.spark.scheduler.TaskSetManager: Finished task 119.0 in stage 2.0 (TID 425) in 63410 ms on e4-1d-2d-19-
-7f-d0.pa4.hpc.i
org.apache.spark.scheduler.TaskSetManager: Starting task 132.0 in stage 2.0 (TID 438, 24-8a-07-d6-7f-30.pa4.hpc.i
org.apache.spark.scheduler.TaskSetManager: Finished task 120.0 in stage 2.0 (TID 426) in 63410 ms on e4-1d-2d-19-
-7f-d0.pa4.hpc.i
org.apache.spark.scheduler.TaskSetManager: Starting task 133.0 in stage 2.0 (TID 439, 24-8a-07-d6-7f-30.pa4.hpc.i
org.apache.spark.scheduler.TaskSetManager: Finished task 121.0 in stage 2.0 (TID 428) in 63311 ms on 24-8a-07-d6-
org.apache.spark.scheduler.TaskSetManager: Starting task 142.0 in stage 2.0 (TID 441, 24-8a-07-d6-7f-30.pa4.hpc.i
org.apache.spark.scheduler.TaskSetManager: Finished task 122.0 in stage 2.0 (TID 429) in 63487 ms on 24-8a-07-d6-
org.apache.spark.scheduler.TaskSetManager: Starting task 146.0 in stage 2.0 (TID 442, 24-8a-07-d6-7f-30.pa4.hpc.i
```

# How did we “decide” what to put in this thing?

- Grouping the most common support questions together
- Google Sheets document

But honestly it's not like we really made decisions:

- Small volunteer working group (only you can prevent forest... wait OOMs)

e.g. whoever got annoyed enough answering the same question.

# What's this “export\_external.sh”?

- People told me perl was not cool anymore
- You can customize the flowchart to your company's environment
- At Netflix we have a bunch of... “special”... settings that live under “private/”
- export\_external lets you have special internal company notes that you don't share
- So please don't just fork this and not contribute upstream. (looking at you Joey...)
- Speaking of....

# Ok so lets say I did want to do that – how?

- Fork funtimes
- `{% include %}`
- `./private/`
- Ehhh lets just take a look at Netflix's fork that's less effort :p
- Then to contribute back upstream: `export_external` + make a PR :D



Contribute!

holdenk / spark-flowchart Public

Watch 5 Fork 12 Star 58

<> Code Issues 4 Pull requests Actions Projects Wiki Security Insights

Filters is:issue is:open Labels 10 Milestones 0 New issue

4 Open 0 Closed

- Documentation idea: how long is my job stuck in accepted state  
#14 opened 5 days ago by atbida
- Documentation idea: fixes in Spark 3.1 release  
#13 opened 5 days ago by atbida
- Documentation idea: surprise nullable datatype  
#12 opened 5 days ago by atbida
- Documentation Idea: File Output Committer Algorithm...  
#2 opened on Nov 22, 2021 by kbendick

good first issue

good first issue

# Let's do it!

- Let's all add something together!
- What are some common problems you all face?



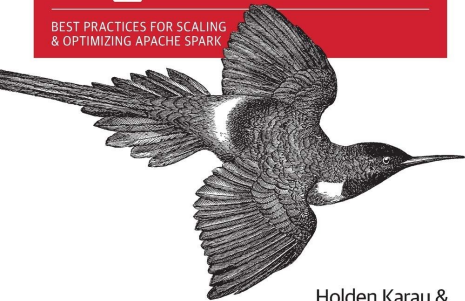
# You too can win a free book!

Ooor you can buy several copies with your corporate credit card. If you don't have a corporate credit card now is a great time to get one. And then buy.

O'REILLY

## High Performance Spark

BEST PRACTICES FOR SCALING  
& OPTIMIZING APACHE SPARK



Holden Karau &  
Rachel Warren

O'REILLY

## Scaling Python with Dask

From Data Science to Machine Learning



Early  
Release  
RAW &  
UNEDITED

Holden Karau

O'REILLY

## Scaling Python with Ray

Exploring Actors, Distributed Data, and Friends  
in Serverless and Cloud Environments



Early  
Release  
RAW &  
UNEDITED

Holden Karau  
& Boris Lublinsky

# TL;DR spark-flowchart

- Tackles your team's FAQs
- Does not replace documentation
- Provides pointers for faster recovery
- Give it to your users for fun and decreased questions

Link: [spark-flowchart](#) & [source](#)

<https://holdenk.github.io/spark-flowchart/flowchart/>

We'll curate your PRs.  
Thank you for your contributions!

P.S.

Come back at ~9 tomorrow to learn about  
upgrading your Spark jobs for "fun."

Holden



OSS Engineer, queer AF,  
co-author of some books

Anya



Tech Evangelist

Jacek\*



Freelance Awesomeness

Author: [Internals of Apache Spark](#) & other books

....reach out with your ideas!

Extra slides



