

Automated Debugging of Bad Deployments



SRECon17

Joe Gordon

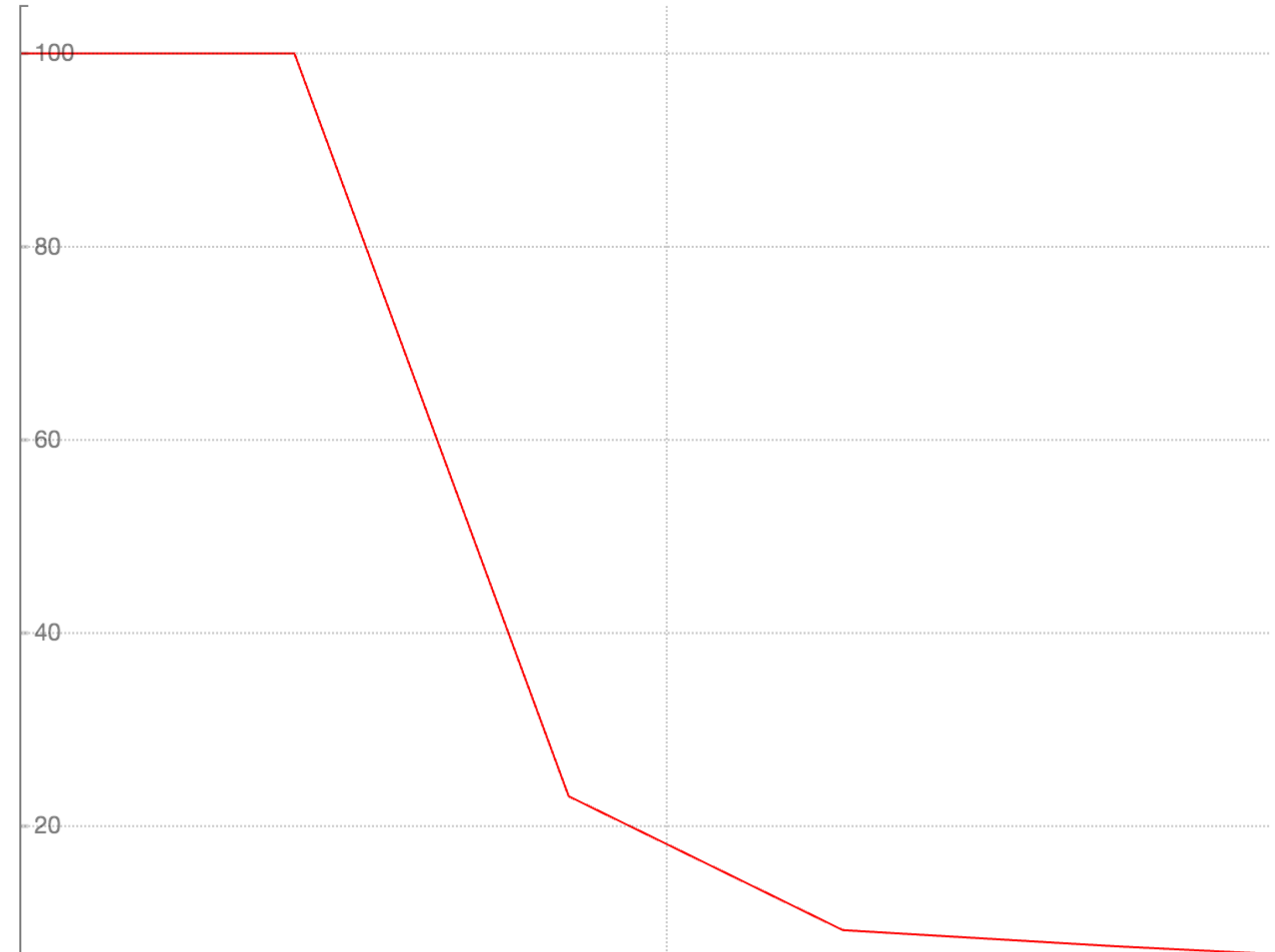
8/31/2017

Problem

Bad Deploy

Bad code is being deployed and the success rate drops. Where is the bad code?

Success Rate



Bad Deploy

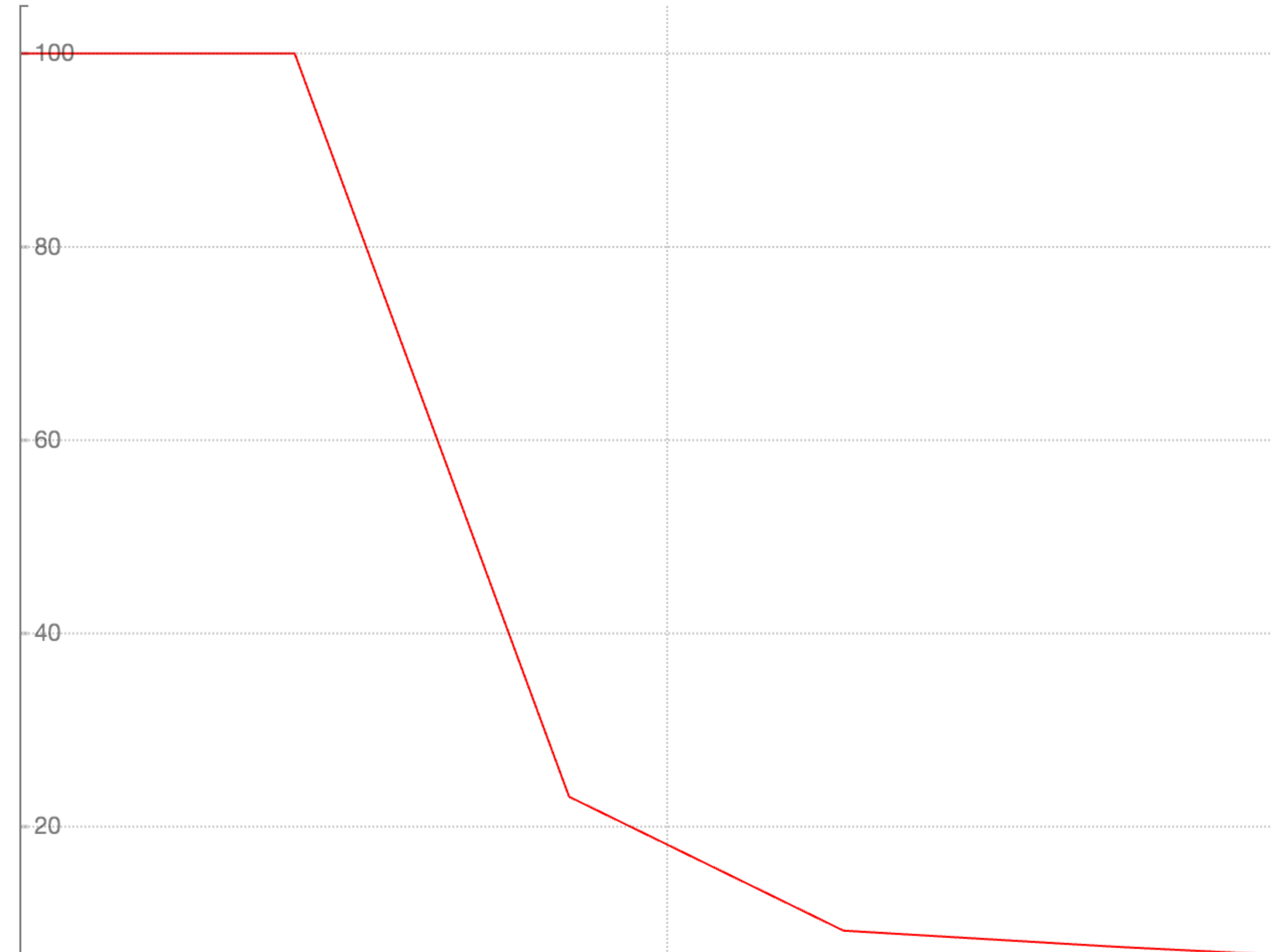
Bad code is being deployed and the success rate drops. Where is the bad code?

Tedious to deal with because:

- Many commits per deploy
- Many deploys per day over many services

Can we do better?

Success Rate



Bad Deploy: Resolution

1. Roll back deploy

2. Debug

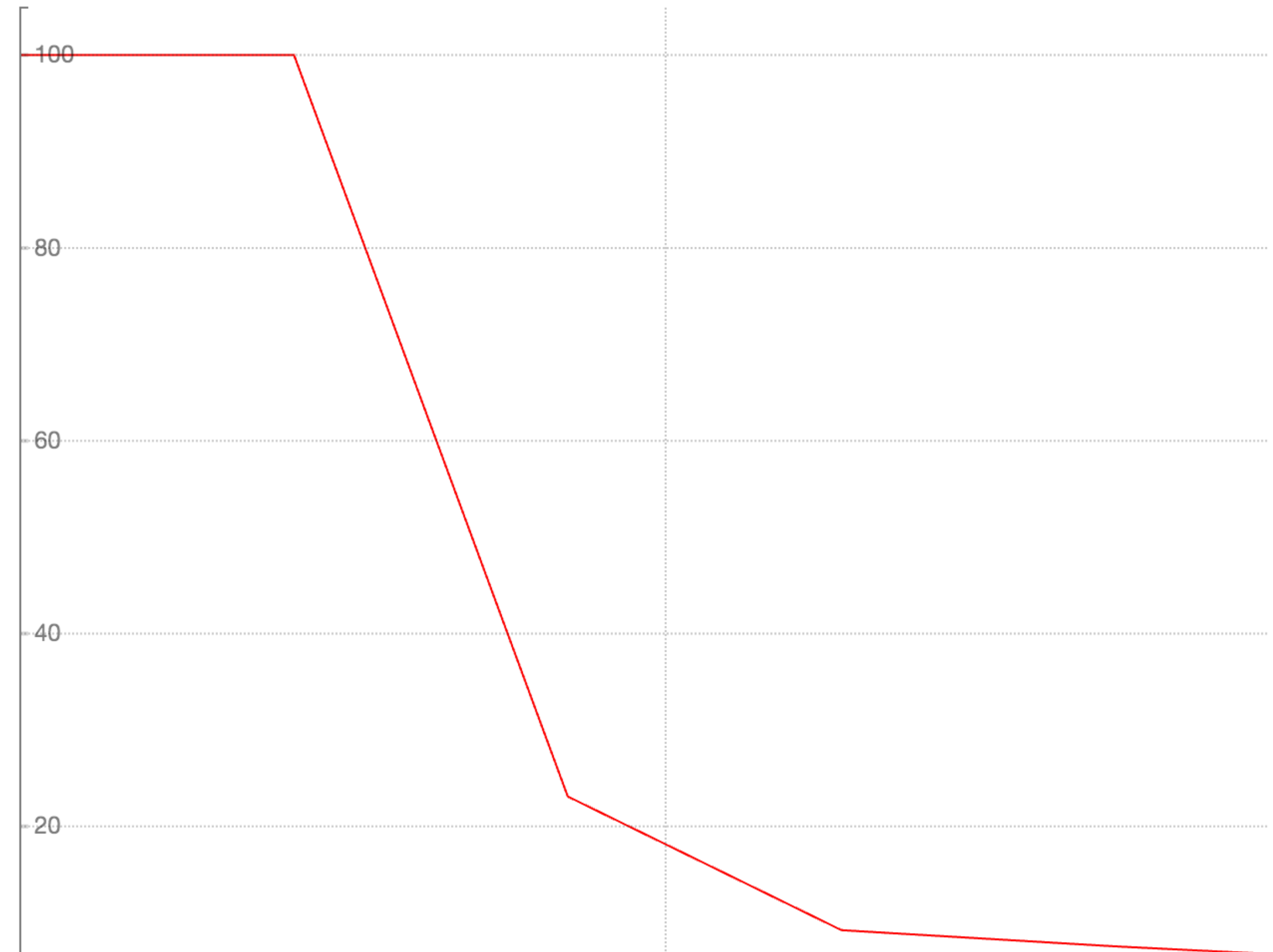
2.1. Find the error

2.2. What commit caused it?

3. Fix problem

4. Restart deploy

Success Rate



Types of Errors

- Performance regressions
- broken webpages
- Empty results
- unexpected status codes (404 etc.)
- ...

Types of Errors

- Performance regressions
- broken webpages
- Empty results
- Unexpected status codes (404 etc.)

- ...

- **Stack traces**

```
Traceback (most recent call last):
  File "../core/managers/user_manager.py", line 2835, in add_typed_search_query
    query)
  File "/mnt/virtualenv_A/local/lib/python2.7/site-packages/thrift_utils/thrift_client_mixin.py", line 210, in wrap_with_pool
    return method(client, *args, **kwargs)
  File "/mnt/virtualenv_A/local/lib/python2.7/site-packages/thrift_utils/thrift_client_mixin.py", line 484, in method_wrapper
    result = method(self._client, *args, **kwargs)
  File "../services/dataservices/thrift_libs/DataServices.py", line 1772, in createOrUpdateTypedSearchQuery
    self.send_createOrUpdateTypedSearchQuery(context, userId, queryType, query)
  File "../services/dataservices/thrift_libs/DataServices.py", line 1782, in send_createOrUpdateTypedSearchQuery
    args.write(self._oprot)
  File "../services/dataservices/thrift_libs/DataServices.py", line 8975, in write
    oprot.writeString(self.query.encode('utf-8'))
UnicodeDecodeError
```

Focus on stack traces: common and lots of information

Debugging Steps

1. Finding New Stack Traces

Challenges:

- Finding stack traces
- Finding *new* stack traces
- Deduplicating similar stack traces

2. Finding the Bad Commit

- **In: Stack trace and a range of commits**

```
Traceback (most recent call last):
  File "../core/managers/user_manager.py", line 2835, in add_typed_search_query
    query)
  File "/mnt/virtualenv_A/local/lib/python2.7/site-packages/thrift_utils/thrift_client_mixin.py", line 210, in wrap_with_pool
    return method(client, *args, **kwargs)
  File "/mnt/virtualenv_A/local/lib/python2.7/site-packages/thrift_utils/thrift_client_mixin.py", line 484, in method_wrapper
    result = method(self._client, *args, **kwargs)
  File "../services/dataservices/thrift_libs/DataServices.py", line 1772, in createOrUpdateTypedSearchQuery
    self.send_createOrUpdateTypedSearchQuery(context, userId, queryType, query)
  File "../services/dataservices/thrift_libs/DataServices.py", line 1782, in send_createOrUpdateTypedSearchQuery
    args.write(self._oprot)
  File "../services/dataservices/thrift_libs/DataServices.py", line 8975, in write
    oprot.writeString(self.query.encode('utf-8'))
UnicodeDecodeError
```

- **Out: Commit that caused the stack trace**

```
commit 101d6d7b71ed2d7c9bf570d826ad7e620e8d07b9
```

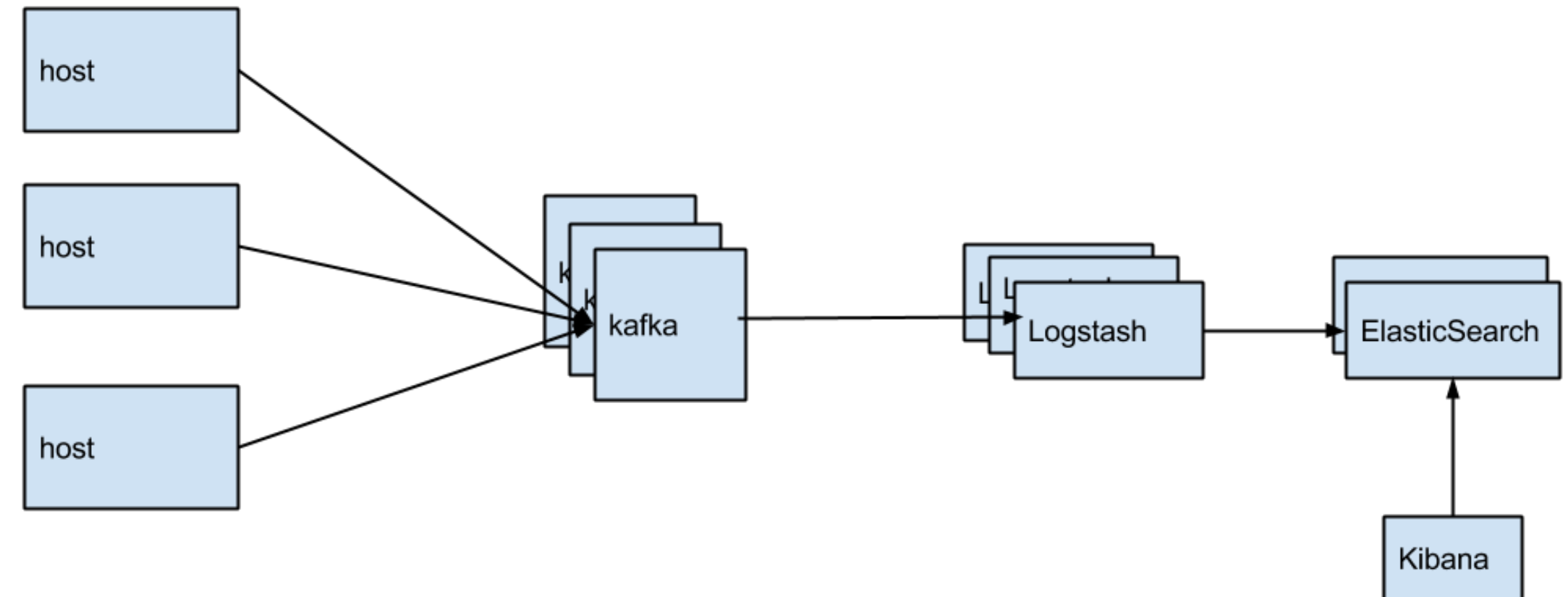
Finding New Stack Traces

Finding Stack Traces

Log Aggregation

- Collect logs with Kafka and Elasticsearch+Logstash+Kibana (ELK Stack)
- Prioritize real time versus completeness
 - Incomplete data is better than none
- Easy to search for known stack traces

Log source



"Line in a stacktrace"



[pinlogger-]YYYY.MM.DD

2,153,244 hits

Selected Fields

? _source

Available Fields

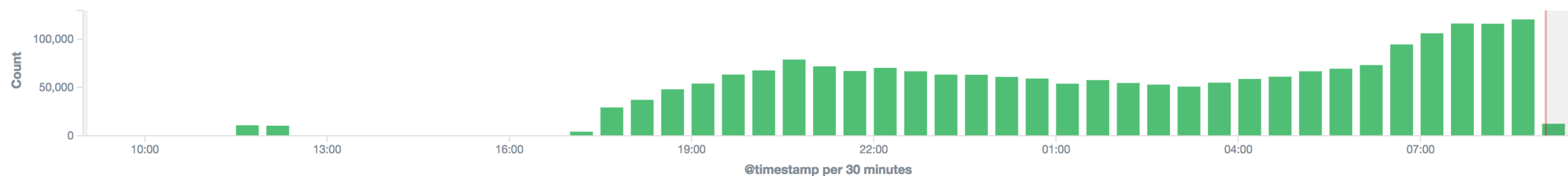
Popular

† env_name

† host

† payload.exception_name

March 10th 2017, 09:03:35.921 - March 11th 2017, 09:03:35.921 — [by 30 minutes](#)



Identifying New Stack Traces

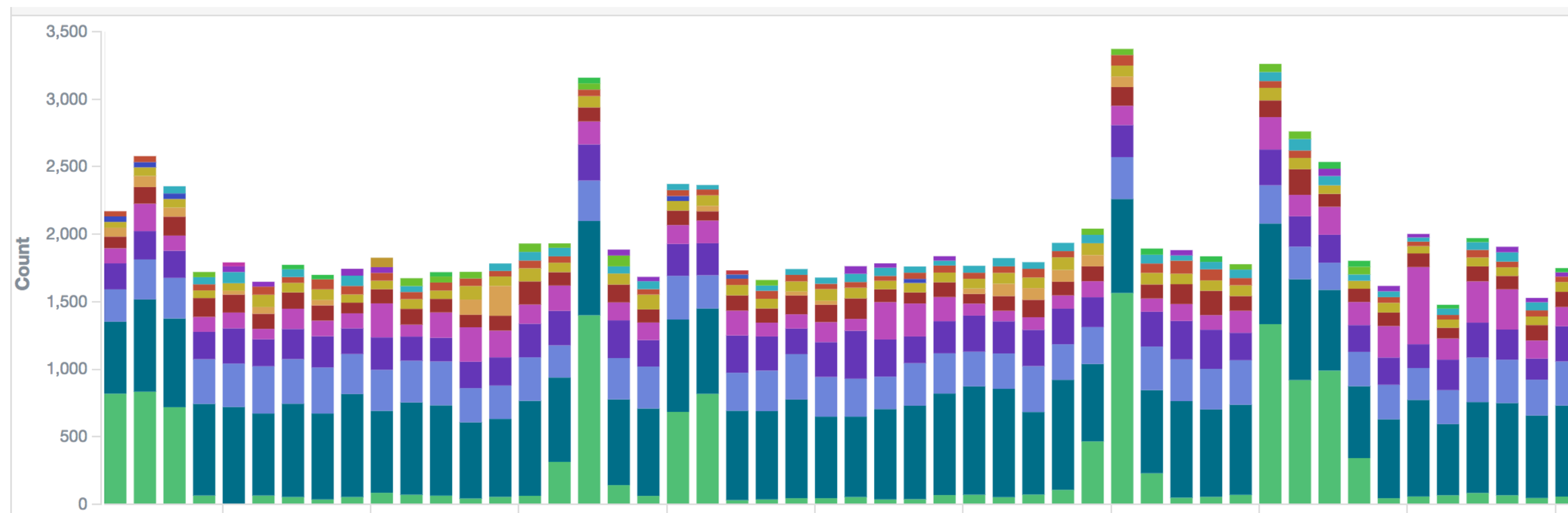
- **old vs. new stack traces**

- Too many stack traces to look at manually
- Use Elasticsearch REST API

```
for trace in canary_traces:  
    if trace not in production_traces:  
        new_traces.append(trace)
```

- **compare 2 Elasticsearch queries**

- Canary vs production
- Before and after deploy



Deduplicating Similar Stack Traces

Simple string matching doesn't work

- Line numbers change during a deploy
- Optional code paths (sampled metrics collection)

```
def bad_code():  
    5 + None
```

```
bad_code()
```

Elasticsearch filter and aggregation query

- Filter: logs containing stack traces on set of hosts X, during time window Y
- Aggregate by:
 1. Exception name (*TypeError*)
 2. File name (*test.py*)
 3. Exception value (*unsupported operand type(s) for +: 'int' and 'NoneType'*)

Exception values may contain variables

```
Traceback (most recent call last):
```

```
File "test.py", line 5, in <module>  
    bad_code()
```

```
File "test.py", line 3, in bad_code  
    5 + None
```

```
TypeError: unsupported operand type(s) for +: 'int' and 'NoneType'
```

Finding the Bad Commit

Finding the Bad Commit

Which of 71 commits caused this stack trace?

```
Traceback (most recent call last):
  File "../core/managers/user_manager.py", line 2835, in add_typed_search_query
    query)
  File "/mnt/virtualenv_A/local/lib/python2.7/site-packages/thrift_utils/thrift_client_mixin.py", line 210, in
wrap_with_pool
    return method(client, *args, **kwargs)
  File "/mnt/virtualenv_A/local/lib/python2.7/site-packages/thrift_utils/thrift_client_mixin.py", line 484, in
method_wrapper
    result = method(self._client, *args, **kwargs)
  File "../services/dataservices/thrift_libs/DataServices.py", line 1772, in createOrUpdateTypedSearchQuery
    self.send_createOrUpdateTypedSearchQuery(context, userId, queryType, query)
  File "../services/dataservices/thrift_libs/DataServices.py", line 1782, in send_createOrUpdateTypedSearchQuery
    args.write(self._oprot)
  File "../services/dataservices/thrift_libs/DataServices.py", line 8975, in write
    oprot.writeString(self.query.encode('utf-8'))
UnicodeDecodeError
```


Finding the Bad Commit

Which of 71 commits caused this stack trace?

```
Traceback (most recent call last):
  File "../core/managers/user_manager.py", line 2835, in add_typed_search_query
    query)
  File "/mnt/virtualenv_A/local/lib/python2.7/site-packages/thrift_utils/thrift_client_mixin.py", line 210, in
wrap_with_pool
    return method(client, *args, **kwargs)
  File "/mnt/virtualenv_A/local/lib/python2.7/site-packages/thrift_utils/thrift_client_mixin.py", line 484, in
method_wrapper
    result = method(self._client, *args, **kwargs)
  File "../services/dataservices/thrift_libs/DataServices.py", line 1772, in createOrUpdateTypedSearchQuery
    self.send_createOrUpdateTypedSearchQuery(context, userId, queryType, query)
  File "../services/dataservices/thrift_libs/DataServices.py", line 1782, in send_createOrUpdateTypedSearchQuery
    args.write(self._oprot)
  File "../services/dataservices/thrift_libs/DataServices.py", line 8975, in write
    oprot.writeString(self.query.encode('utf-8'))
UnicodeDecodeError
```

- git blame different files
- git whatchanged
- ...
- Slow, tedious and repetitive

git-stacktrace

[https://github.com/pinterest/
git-stacktrace](https://github.com/pinterest/git-stacktrace)

github.com / pinterest / git-stacktrace

Unwatch 9 Unstar

<> Code Issues 0 Pull requests 0 Projects 0 Wiki Insights

Easily figure out which git commit caused a given stacktrace <https://pypi.python.org/pypi/git-stac...>

git

37 commits 2 branches 9 releases 2 contributors

Branch: master New pull request Create new file Upload files Find file

jogo committed on GitHub Merge pull request #11 from jogo/fixup

git_stacktrace	Handle case where python traceback is missing lines of code
.gitignore	Get ready for pypi
.testr.conf	Initial open source commit
.travis.yml	Update travis to test python36
CHANGES.rst	Update CHANGES.rst in preparation for releasing 0.7.0
LICENSE	Initial open source commit
README.rst	Add line matching support
requirements.txt	Add python36 support
setup.cfg	Add python36 support
setup.py	Initial open source commit
test-requirements.txt	Initial open source commit
tox.ini	Update travis to test python36

README.rst

git-stacktrace

git-stacktrace is designed to make it easy to figure out which commit caused a given stacktrace.

git-stacktrace looks for:

- commits in given range that touched files in the stacktrace
- commits in given range that added/removed code present the stacktrace

git-stacktrace: Goals

Make it easy to figure out which commit caused a given stack trace

How a commit can trigger a stack trace:

- 1. An actual line of code from from the commit shows up in the stack trace**
- 2. One of the files touched in the commit shows up in the stack trace**
3. The commit caused a different code path to be executed which in turn caused the stack trace

Tradeoff: Too many matches vs. no matches

Idea: 5 commits are easier to review then 71

Return a ranked list of all possibly related commits

git-stacktrace

```
$ git stacktrace 81a25be..b3864b2 < stacktrace
```

```
Traceback (most recent call last):
```

```
  File "../core/managers/user_manager.py", line 2835, in add_typed_search_query
    query)
  File "/mnt/virtualenv_A/local/lib/python2.7/site-packages/thrift_utils/thrift_client_mixin.py", line 210, in
wrap_with_pool
    return method(client, *args, **kwargs)
  File "/mnt/virtualenv_A/local/lib/python2.7/site-packages/thrift_utils/thrift_client_mixin.py", line 484, in
method_wrapper
    result = method(self._client, *args, **kwargs)
  File "../services/dataservices/thrift_libs/DataServices.py", line 1772, in createOrUpdateTypedSearchQuery
    self.send_createOrUpdateTypedSearchQuery(context, userId, queryType, query)
  File "../services/dataservices/thrift_libs/DataServices.py", line 1782, in send_createOrUpdateTypedSearchQuery
    args.write(self._oprot)
  File "../services/dataservices/thrift_libs/DataServices.py", line 8975, in write
    oprot.writeString(self.query.encode('utf-8'))
UnicodeDecodeError
```

```
commit 101d6d7b71ed2d7c9bf570d826ad7e620e8d07b9
```

```
Commit Date: Thu, 9 Mar 2017 19:24:47 -0800
```

```
Author: John Doe <johndoe@pinterest.com>
```

```
Subject: A bad commit
```

```
Link: https://example.com/D1000
```

```
Files Modified:
```

- services/dataservices/thrift_libs/DataServices.py
- services/dataservices/thrift_libs/DataServices.py:8975

```
Lines Added:
```

- "args.write(self._oprot)"
- "oprot.writeString(self.query.encode('utf-8'))"

git-stacktrace: How It Works

One of the files touched in the commit shows up in the stack trace

```
git log --raw
:100644 100644 c99e4d0... da4d504... M git_stacktrace/git.py
:100644 100644 2bb13df... 78c7f4e... M git_stacktrace/tests/test_git.py
```

```
../services/dataservices/thrift_libs/DataServices.py == services/dataservices/thrift_libs/
DataServices.py
```

Match stack trace line numbers with lines changed in a commit

```
git log --patch
```

Breaks if a later commit adds/removes lines in the file

Use git pickaxe to determine if a commit added a line of code in the stack trace

```
git log -S
```

'Look for differences that change the number of occurrences of the specified string (i.e. addition/deletion) in a file'

Strongest signal

git-stacktrace

```
$ git stacktrace 81a25be..b3864b2 <stacktrace
```

```
Traceback (most recent call last):
```

```
File "../core/managers/user_manager.py", line 2835, in add_typed_search_query
    query)
File "/mnt/virtualenv_A/local/lib/python2.7/site-packages/thrift_utils/thrift_client_mixin.py", line 210, in
wrap_with_pool
    return method(client, *args, **kwargs)
File "/mnt/virtualenv_A/local/lib/python2.7/site-packages/thrift_utils/thrift_client_mixin.py", line 484, in
method_wrapper
    result = method(self._client, *args, **kwargs)
File "../services/dataservices/thrift_libs/DataServices.py", line 1772, in createOrUpdateTypedSearchQuery
    self.send_createOrUpdateTypedSearchQuery(context, userId, queryType, query)
File "../services/dataservices/thrift_libs/DataServices.py", line 1782, in send_createOrUpdateTypedSearchQuery
    args.write(self._oprot)
File "../services/dataservices/thrift_libs/DataServices.py", line 8975, in write
    oprot.writeString(self.query.encode('utf-8'))
UnicodeDecodeError
```

```
commit 101d6d7b71ed2d7c9bf570d826ad7e620e8d07b9
```

```
Commit Date: Thu, 9 Mar 2017 19:24:47 -0800
```

```
Author: John Doe <johndoe@pinterest.com>
```

```
Subject: A bad commit
```

```
Link: https://example.com/D1000
```

```
Files Modified:
```

- services/dataservices/thrift_libs/DataServices.py
- services/dataservices/thrift_libs/DataServices.py:8975

```
Lines Added:
```

- "args.write(self._oprot)"
- "oprot.writeString(self.query.encode('utf-8'))"

Not Author Date

Uses at
Pinterest

Elasticsearch Queries + git-stacktrace at Pinterest

- Run on every API and web deploy
- Allows us to quickly detect new rare errors
- Significantly reduced debugging time

