Speeding to resolution with human-in-the-loop automation

Tina Huang, Founder and CTO

Zoom:
August 24, 2020

Transposit

Slack:
May 12, 2020

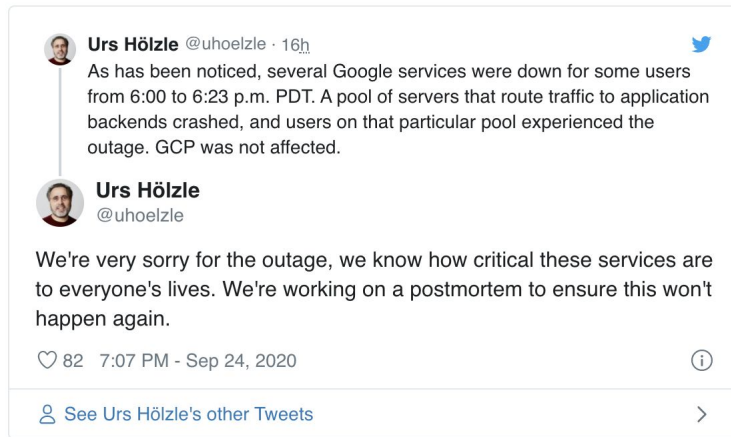Transposit

Gmail:
August 20, 2020
September 24, 2020

Microsoft 365:
September 28, 2020

Urs Hölzle @uhoelzle · 16h

As has been noticed, several Google services were down for some users from 6:00 to 6:23 p.m. PDT. A pool of servers that route traffic to application backends crashed, and users on that particular pool experienced the outage. GCP was not affected.

Urs Hölzle
@uhoelzle

We're very sorry for the outage, we know how critical these services are to everyone's lives. We're working on a postmortem to ensure this won't happen again.

♡ 82   7:07 PM - Sep 24, 2020

See Urs Hölzle's other Tweets

Microsoft 365 Status
@MSFT365Status

We're investigating an issue affecting access to multiple Microsoft 365 services. We're working to identify the full impact and will provide more information shortly.

3:44 PM · Sep 28, 2020

♡ 3.3K   2.9K people are Tweeting about this

Transposit

Digital dependency is at an all time high.

We are open! Sea Forager CSF will continue to deliver. Click here to sign up or manage deliveries. ✕

SEAFOOD SUBSCRIPTIONS   SIGN UP OR LOG IN

Sea Forager

In The Net   Shop   Tours   The Book   About   🛒 0

The Fishwives Band

FROM HOOK TO COOK

Seafood subscriptions for the passionate home cook.

Sign up today!
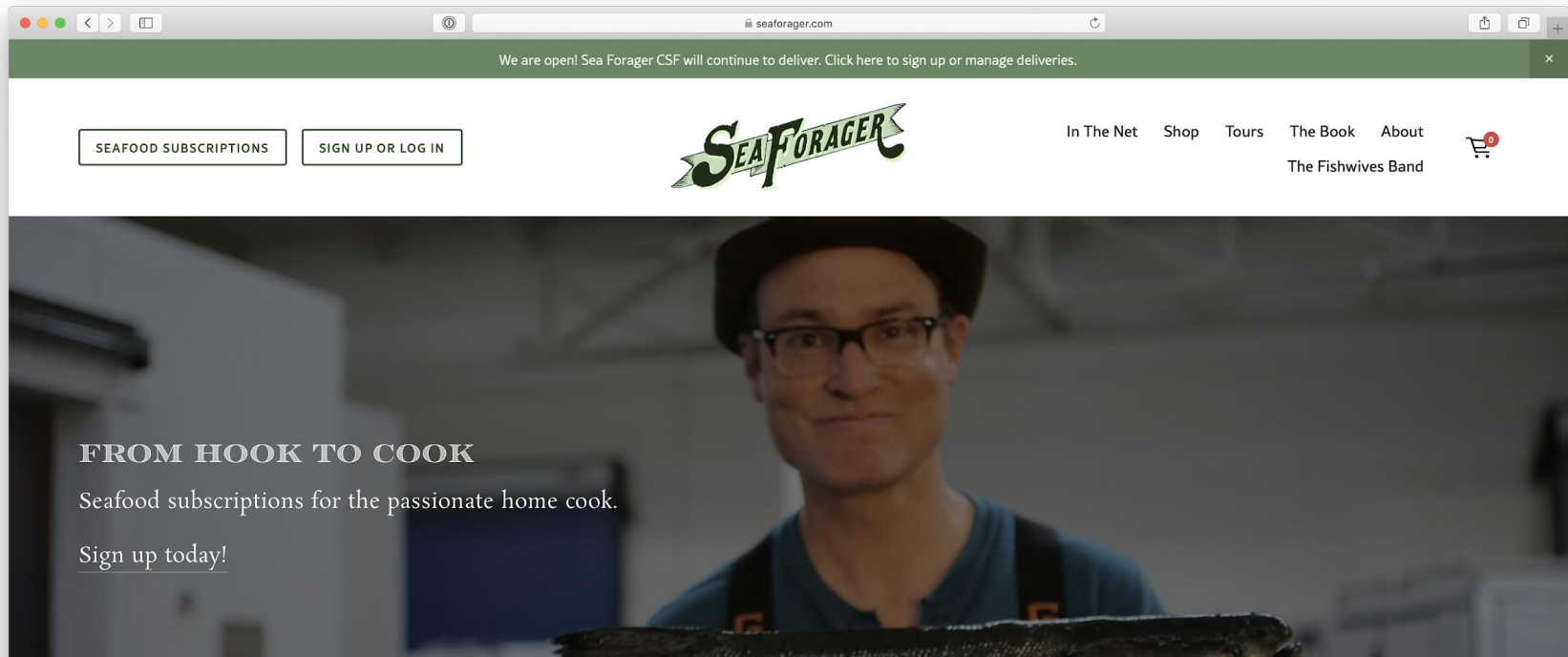
Agile requires us to rethink operations and how to make reliability sustainable.

Agile makes customer support an integral part of reliability.

Transposit

Sustainable agile requires faster resolution and continuous learning

Transposit
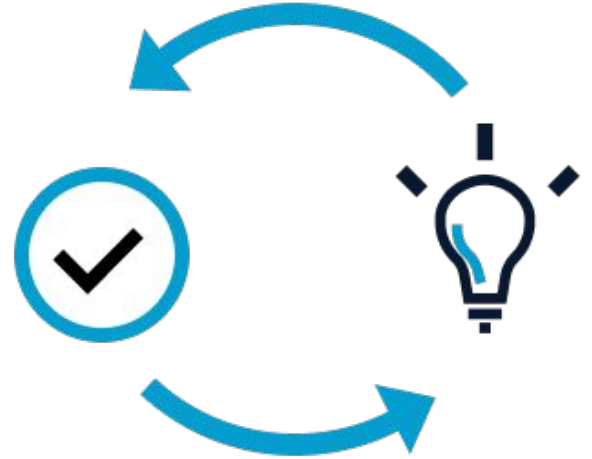
Automation and post-mortems are not the panacea for incident management.

Resolving faster is more than the technical, it's about helping people communicate better.
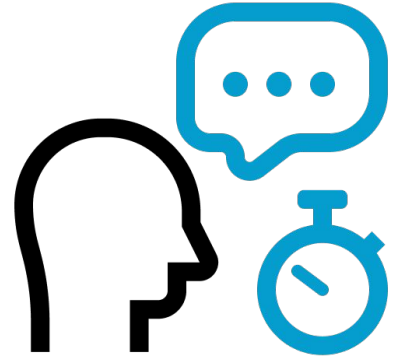
Why is automation hard?

Transposit
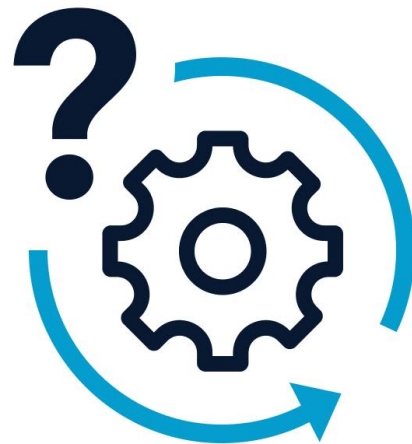
Automation is scary

Nuance is hard to automate

Innovation breeds incidents

APIs are hard

Let's rethink automation

Transposit

Human in the loop automation is checklists on steroids
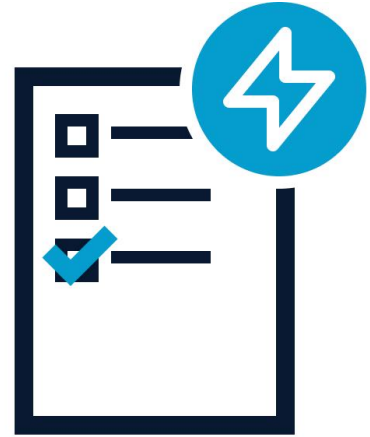
What makes a great SRE?

- Having lots of operational expertise
- Lots of domain knowledge
- Great operational and organizational intuition

SRE

Transposit

Turning human intuition into human judgement

Transposit
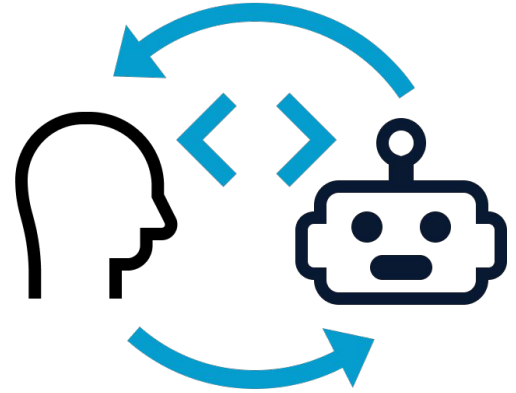
With so much data, why is incident management not more data driven?

Incident management is fundamentally broken.

Build deploy

DB upgrade

Bug fix

Feature launch

Build deploy

Alert fired

Ack alert

Create incident

?

Resolve incident

Postmortem

JIRA issues filed

Bug fix

Bug fix

Build deploy
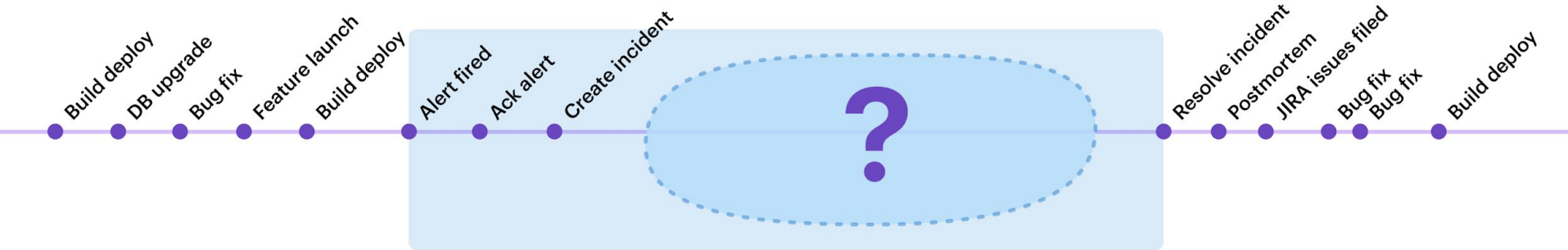
At 3am no one is thinking about capturing data

Transposit
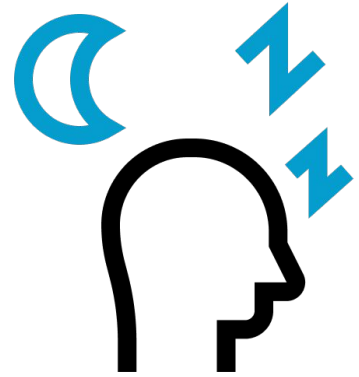
You need an incident management platform to capture data

Transposit

#BeachOps

ON CALL
FROM THE BEACH
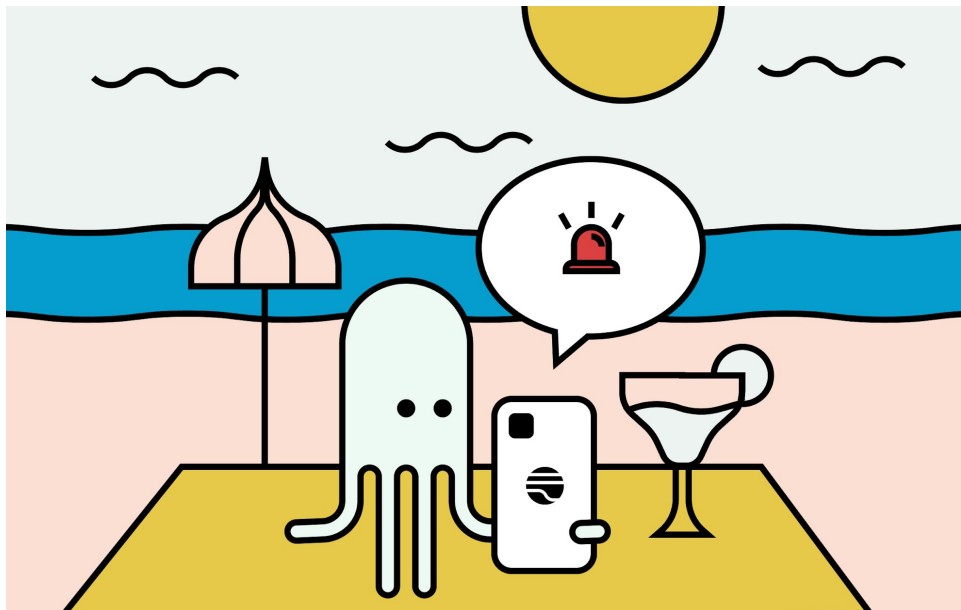
Having data isn't enough. How can we make data accessible and useful during incidents?

Make data access friendly with search and recommendations

Transposit

Automating the learning process

Iggy's Surfshop ˅
● Tina Huang

🔍 Threads
@ Mentions & reactions
🖊 Drafts
🔖 Saved items
👥 People
▦ Apps
▤ Files
↑ Show less

˅ Channels ＋
＃ general
＃ alerts
＃ oncall
＃ random
＃ TR-1234
＋ Add a channel

˅ Direct messages ＋
💚 Slackbot
● Tina Huang (you)
＋ Invite people

˅ Apps ＋

🔒 **#oncall**
☆ | 👥 4 | ⚡ 1 | 🖊 Add a topic

🔍 Search

Web 500s Are Over Threshold

Jump to...   Capacity Problems ˅

Capacity Problems

• Check the ELB logs to see if there are a lot of 500s there. **503s** indicate that the underlying service is unavailable, **504s** indicate service timeout. In either case, adding capacity probably helps.
• Check is the **ECS metrics** . If CPU or memory are high, we probably need more capacity. Before you add capacity, check that we are actually running the desired number of tasks. If not, sometimes that means a task is wedged and restarting ECS tasks can help. Otherwise, add some capacity.

[Restart ECS]   [ECS Metrics]   [Scale ECS]

[Trigger Terraform using Jenkins]

[Close]

Transpo...
joined #

Transpo...
JIRA iss...
IS-963
The pay...   ng assessment in TR-1234.

Tasks   Current Runbook   Browse Runbooks

Message #oncall

⚡ B I S </> 🔗 ≔ ⚬ ⌐ Aa @ 😊 📎 ➤

Web 500s Are Over Threshold

Jump to...   Capacity Problems ▾

Scale ECS

Service

Execution ▾

Instances

6 ▾

Cancel   Run

Iggy's Surfshop ⌄
● Tina Huang

Threads
Mentions & reactions
Drafts
Saved items
People
Apps
Files
↑ Show less

⌄ Channels +
general
alerts
oncall
random
TR-1234
Add a channel

⌄ Direct messages +
💚 Slackbot
● Tina Huang (you)
Invite people

⌄ Apps +

🔒 **#oncall**
⭐ | 👥 4 | 📌 1 | 📎 Add a topic

Search

Transposit APP 3:10 PM
joined #tr-1234 along with @tina

Transposit APP 3:10 PM
JIRA issue created by @tina
IS-963 Payments server is down - **Priority P1**
The payments server was erroring and was offline at 3:06 PM today, as reported. Ongoing assessment in TR-1234.

[Tasks] [Current Runbook] [Browse Runbooks]

@Tina Huang ran Scale ECS

Restarting tasks for ecs service excution on cluster production-cluster. Updates will be posted in this channel.

12:39:54 PM (PDT): (service excution) has started 6 tasks: (task 5154eabb-f324-4c38-88f9-00fe57991692) (task 747a8acc-f076-407e-8177-a3cf39247b50)

12:39:55 PM (PDT): (service excution) registered 6 targets in (target-group arn:aws:elasticloadbalancing:us-west-2:668521361712:targetgroup/production-cluster-excution/6c7c5b650e48c849)

12:39:56 PM (PDT): (service excution) deregistered 6 targets in (target-group arn:aws:elasticloadbalancing:us-west-2:668521361712:targetgroup/production-cluster-excution/6c7c5b650e48c849)

12:39:57 PM (PDT): (service excution) has begun draining connections on 1 tasks

12:39:57 PM (PDT): (service excution) has stopped 6 running tasks: (task ef7aa378-ed05-4cdf-bf10-bb10b7a49f1e) (task ebd20d4e-1be5-4862-9d2c-9985dc46906b)

12:39:59 PM (PDT): (service excution) has reached a steady state. :white_check_mark:

Message #oncall

⚡ **B** *I* ~~S~~ </> 🔗 1. ≣ ≣ ⌨️   Aa @ 😊 📎 ➤

Show All Events

- T-132
- T-133
- T-134
- T-135
- T-136
- T-137
- Alerts
- PagerDuty
- OpsGenie
- Zendesk

Time ↓   Event

🔍   Event Type   People   Date   Time

12:43 PM   **Leslie Bauman** @leslie   · · ·
"Looks like someone on the analytics team is running queries against prod. I'm going to kill that one.."

12:43 PM   **Running DB Queries**   · · ·
run by Leslie Bauman

```
SELECT * FROM users AS U JOIN teams AS T WHERE T.id = U.teamId AND T.id IN
(SELECT * FROM teams WHERE org = "Engineering") (6009ms)

SELECT * FROM teams WHERE id = '529' (29ms)
```

12:43 PM   **Running DB Queries**   · · ·
run by Leslie Bauman

12:43 PM   **Nina Yang** @nina   · · ·
"Not sure why an alert didn't fire for the DB, but CPU load seems high"

12:43 PM   **Nina Yang** @nina   · · ·
"Looks like the website is lagging when customers add items to the shopping cart"

12:43 PM   **Leslie Bauman** @leslie   · · ·
@nina I'm not really sure what's going on but can you help poke around?

12:43 PM   **Nina Yang** @nina   · · ·
Started runbook Latency Exceeds Maximum

12:43 PM   **Create a new JIRA issue** (Done)   · · ·
Nina Yang created a new Jira issue created
Jira T-539

12:43 PM   **Alert**   · · ·
Pagerduty: Web 500s are over threshold

Runbooks   —

Show All Events

T-132
T-133
T-134
T-135
T-136
T-137
Alerts
PagerDuty
OpsGenie
Zendesk

Time ↓   Event

🔍   Event Type   People   Date   Time

12:43 PM   **Nina Yang** @nina
Started runbook Latency Exceeds Maximum

12:43 PM   **Alert**
Pagerduty: Web 500s are over threshold

Latency Exceeds Maximum   ⌄   ⤢   ✕

Jump to...   ⌄

Capacity Problems

• Check the ELB logs to see if there are a lot of 500s there. **503s** indicate that the underlying service is unavailable, **504s** indicate service timeout. In either case, adding capacity probably helps.

• Check the **ECS metrics**. If CPU or memory are high, we probably need more capacity. Before you add capacity, check that we are actually running the desired number of tasks. If not, sometimes that means a task is wedged and restarting ECS tasks can help. Otherwise, add some capacity.

Restart ECS   ECS Metrics   Trigger Terraform

Scale ECS   Commonly used following alert
Pagerduty: Web 500s are over threshold

Next step: Server Errors   Previous   Next

≔ **Show All Events**

⌕   Event Type   People   Date   Time

Time ↓   **Event**

〜 **T-132**

〜 T-133

〜 T-134

〜 T-135

〜 T-136

〜 T-137

〜 Alerts

〜 PagerDuty

〜 OpsGenie

〜 Zendesk

12:43 PM   **Leslie Bauman** @leslie
"Looks like someone on the analytics team is running queries against pro

12:43 PM   **Running DB Queries**
run by Leslie Bauman

```
SELECT * FROM users AS U JOIN teams AS T WHERE T.id = U.teamId A
(SELECT * FROM teams WHERE org = "Engineering") (6009ms)

SELECT * FROM teams WHERE id = '529' (29ms)
```

12:43 PM   **Running DB Queries**
run by Leslie Bauman

12:43 PM   **Nina Yang** @nina
"Not sure why an alert didn't fire for the DB, but CPU load seems high"

12:43 PM   **Nina Yang** @nina
"Looks like the website is lagging when customers add items to the shopping cart"

12:43 PM   **Leslie Bauman** @leslie
@nina I'm not really sure what's going on but can you help poke around?

12:43 PM   **Nina Yang** @nina
Started runbook Latency Exceeds Maximum

12:43 PM   ▶ **Create a new JIRA issue** (Done)
Nina Yang created a new Jira issue created
Jira T-539

12:43 PM   🔔 **Alert**
Pagerduty: Web 500s are over threshold

Event type

Workflow ▾

Within alert type

Web 500s are over threshold ▾

Clear   Apply

Runbooks   —

☰ Show All Events

∿ T-132

∿ T-133

∿ T-134

∿ T-135

∿ T-136

∿ T-137

∿ Alerts

∿ PagerDuty

∿ OpsGenie

∿ Zendesk

Time ↓   Event   🔍   **Event Type**   People   Date   Time

▶ 12:43 PM **Running DB Queries**   ⋯
            run by Leslie Bauman

```
SELECT * FROM users AS U JOIN teams AS T WHERE T.id = U.teamId AND T.id IN
(SELECT * FROM teams WHERE org = "Engineering") (6009ms)

SELECT * FROM teams WHERE id = '529' (29ms)
```

T-82

▶ 8:33 AM   **Create a new JIRA issue** (Done)   ⋯
            Nina Yang created a new Jira issue created
            Jira T-539
            T-123

June 8, 2020

▶ 12:43 PM **Scale ECS**   ⋯
            run by Leslie Bauman

```
12:39:54 PM (PDT): (service excution) has started 6 tasks: (task
5154eabb-f324-4c38-88f9-00fe57991692) (task
747a8acc-f076-407e-8177-a3cf39247b50)
12:39:55 PM (PDT): (service excution) registered 6 targets in (target-group
arn:aws:elasticloadbalancing:us-west-2:668521361712:targetgroup/production-cl
uster-excution/6c7c5b650e48c849)
12:39:56 PM (PDT): (service excution) deregistered 6 targets in (target-group
arn:aws:elasticloadbalancing:us-west-2:668521361712:targetgroup/production-cl
uster-excution/6c7c5b650e48c849)
```

T-142

Runbooks   —

Development

Run   Discard code   Commit code

Auths & settings
Authorize data connections

Stash

Dynamic Tasks

Operations   +

list_commits
SQL

input_prompt
PY

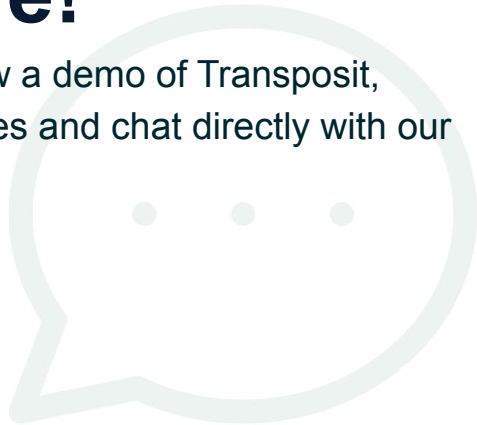Data connections   +

GitHub
transposit/github

```sql
1  SELECT * FROM github.list_commits
2    WHERE owner='tina@transposit.com'
3    AND repo='transposit/terraform'
5    AND author=*
6    AND since='09/01/2020'
7    LIMIT 100
8
9
10
11
```

The Future of Operations is Now

Join me in our Slack Q&A to chat more!

Visit our booth to view a demo of Transposit, enter to win cool prizes and chat directly with our team!

Join us for

Happy Hour