Service Ownership

Learn Faster

Holly Allen
Service Engineering
@hollyjallen

Holly Allen

Software development and leadership for 18 years

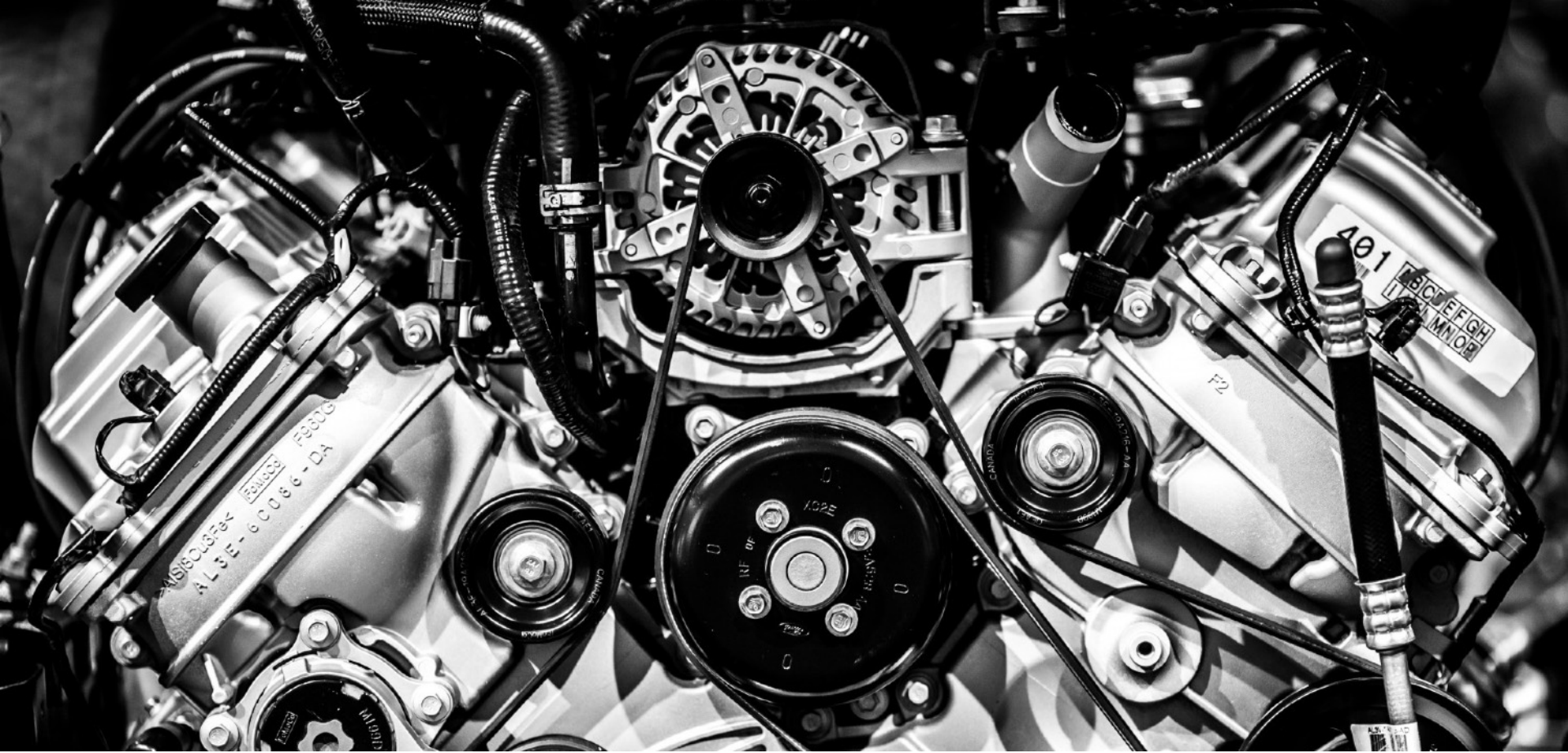
BioTrove

18F

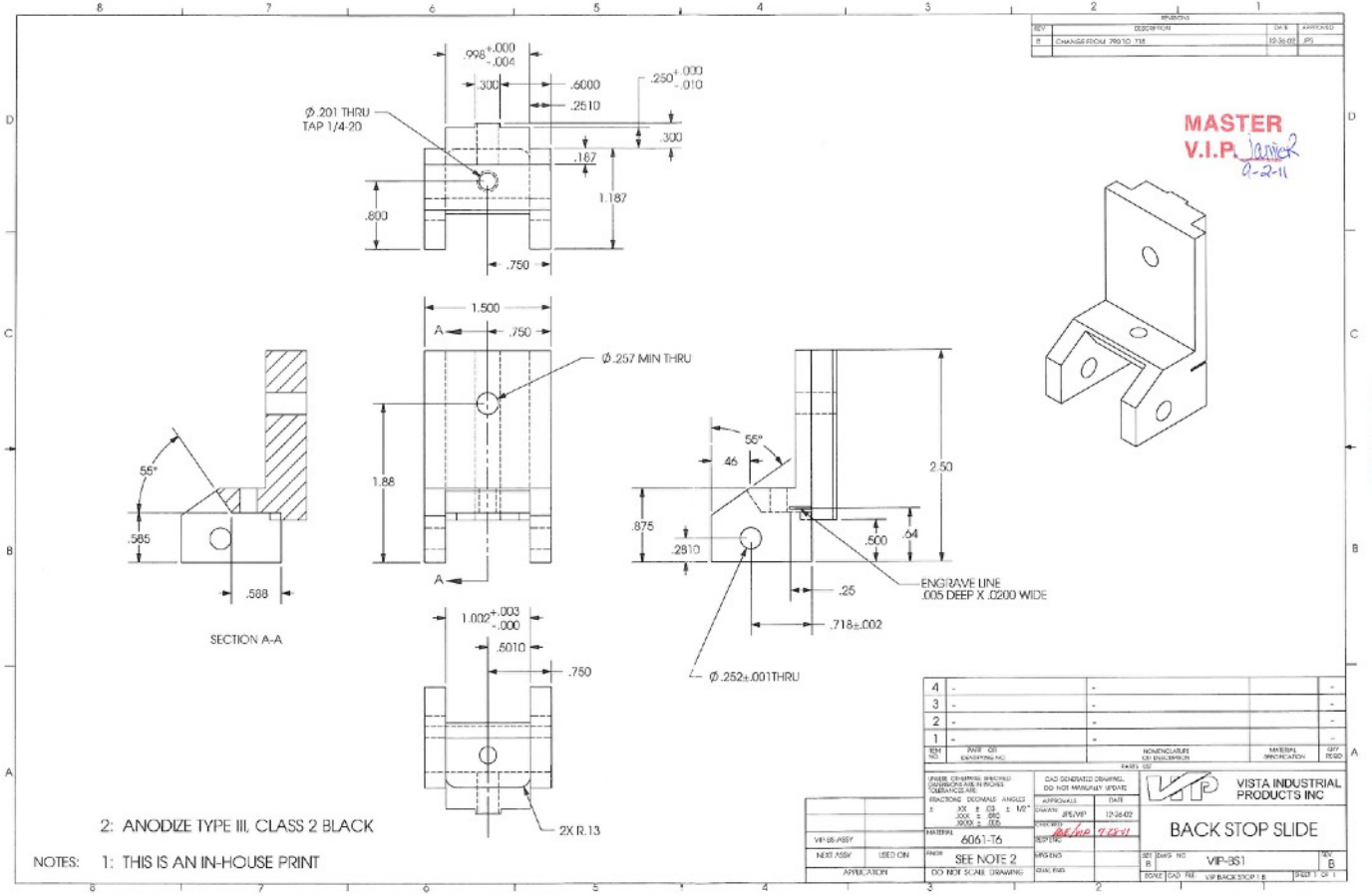
PLOS

slack

PDI DreamWorks Animation SKG



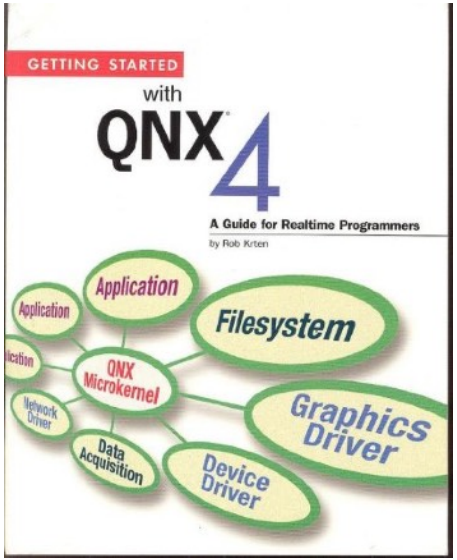
slack

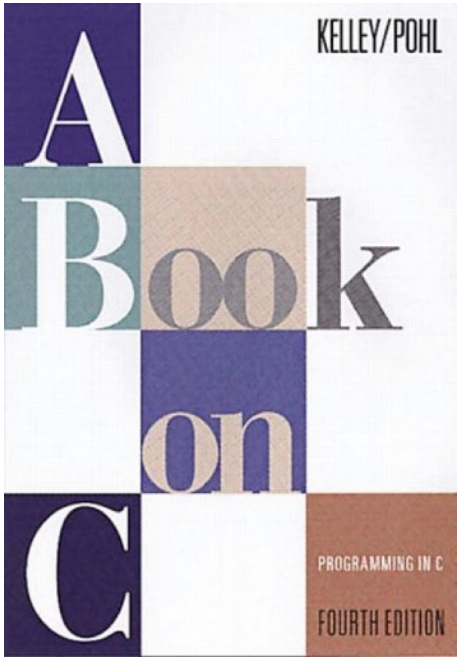

MASTER
V.I.P.
Ø.201 THRU
TAP 1/4-20
.998 +.000 -.004
.300
.6000
.2510
.250 +.000 -.010
.300
.167
1.187
.800
.750
1.500
A
.750
Ø.257 MIN THRU
1.88
A
.55°
.585
.588
SECTION A-A
1.002 +.003 -.000
.5010
.750
2X R.13
.55°
.46
2.50
.875
.2810
.500
.64
.25
.718±.002
ENGRAVE LINE
.005 DEEP X .0200 WIDE
Ø.252±.001THRU
2: ANODIZE TYPE III, CLASS 2 BLACK
NOTES: 1: THIS IS AN IN-HOUSE PRINT
VISTA INDUSTRIAL PRODUCTS INC
BACK STOP SLIDE
6061-T6
SEE NOTE 2
VIP-BS1
JUL 26 2011

slack

slack

Software! 😍


GETTING STARTED
with
QNX 4
A Guide for Realtime Programmers
by Rob Krten
Application
Application
Filesystem
QNX Microkernel
Graphics Driver
Network Driver
Data Acquisition
Device Driver


KELLEY/POHL
A Book on C
PROGRAMMING IN C
FOURTH EDITION

return (_map.ContainsKey(key));
public bool Contains(KeyValuePair
return (_map.ContainsKey(key));
public bool ContainsKey(TKey key)
return (_map.ContainsKey(key));
public bool ContainsValue(TValue item)
(item == null)
map.Any(kvp =>

slack

S L O W 😩

slack

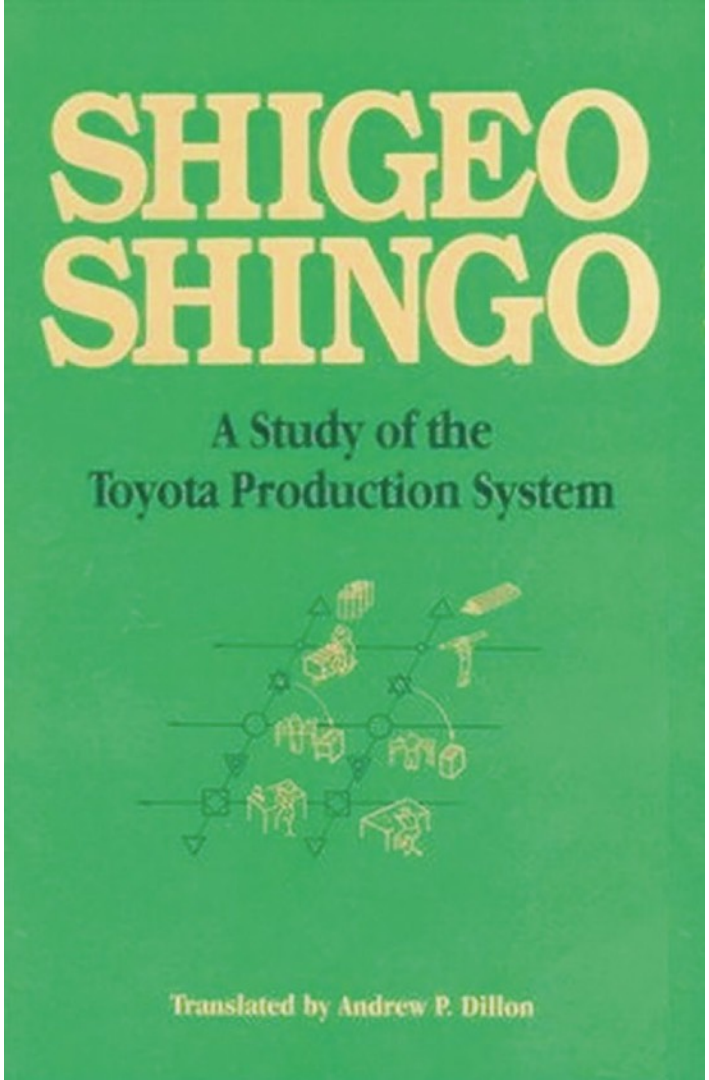
Design

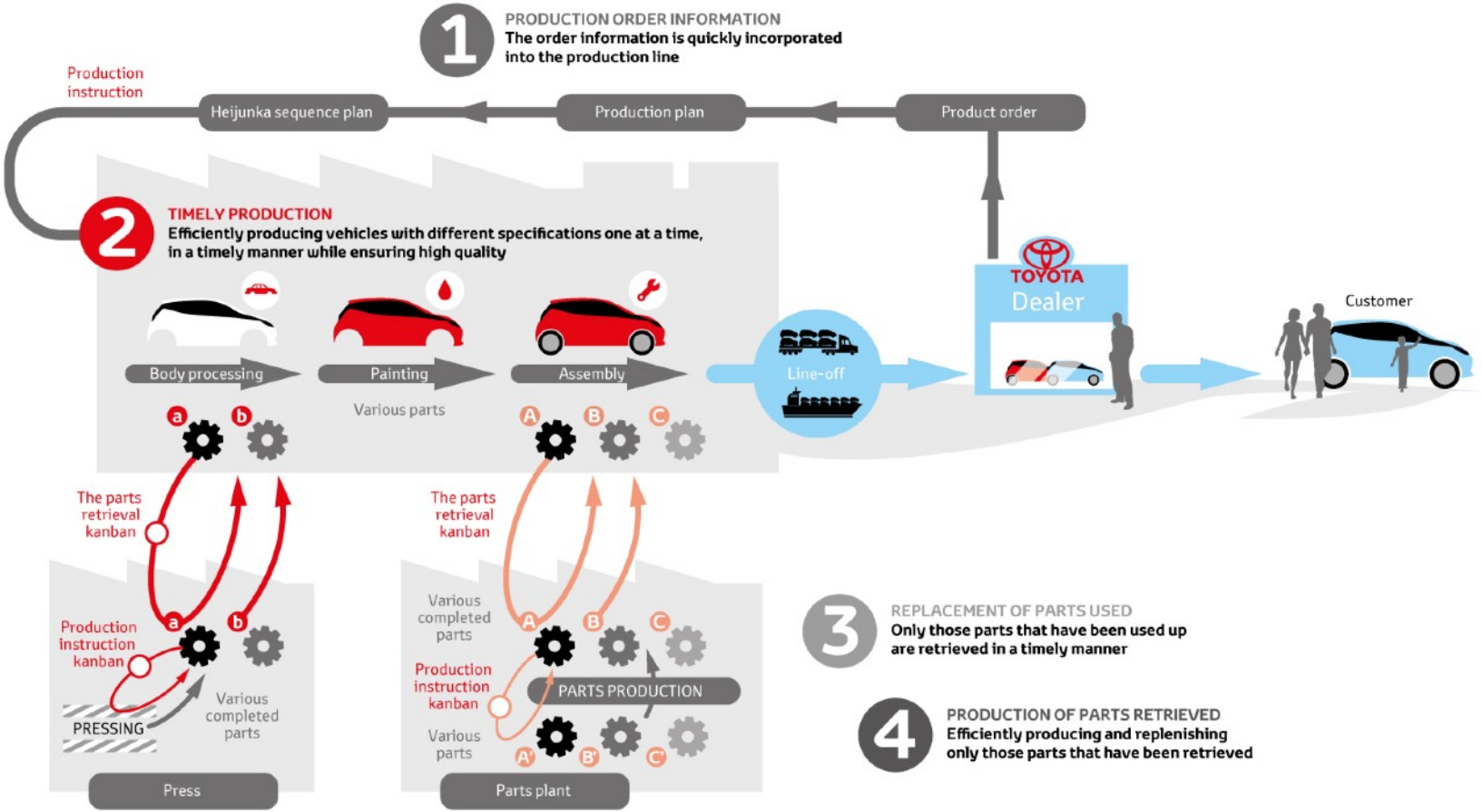
Measure

Learn

slack

Toyota Production System



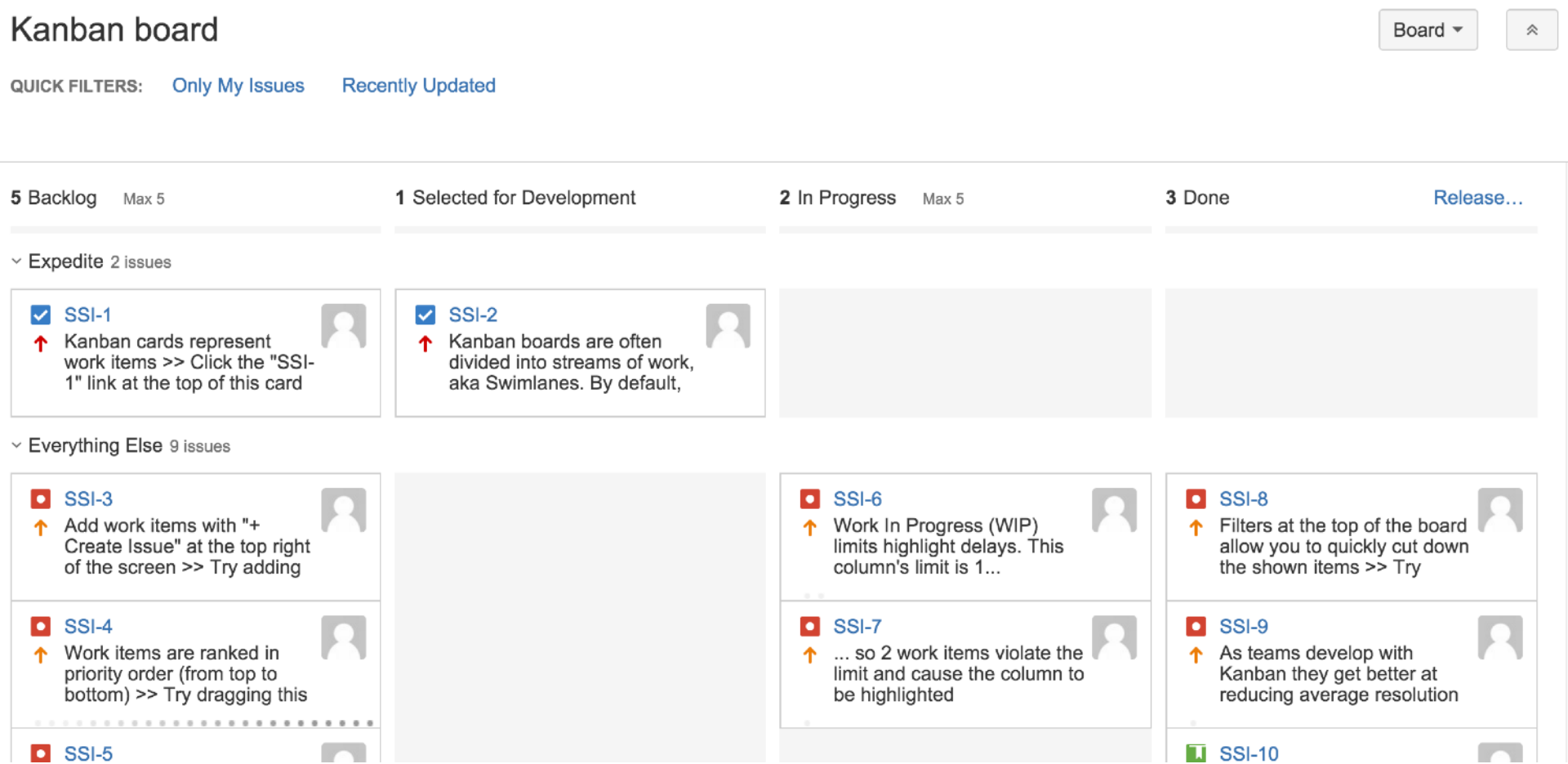

1
PRODUCTION ORDER INFORMATION
The order information is quickly incorporated into the production line
Production instruction
Heijunka sequence plan
Production plan
Product order
2
TIMELY PRODUCTION
Efficiently producing vehicles with different specifications one at a time, in a timely manner while ensuring high quality
Body processing
Painting
Assembly
Various parts
Line-off
TOYOTA
Dealer
Customer
The parts retrieval kanban
Production instruction kanban
PRESSING
Various completed parts
Press
The parts retrieval kanban
Various completed parts
Production instruction kanban
PARTS PRODUCTION
Various parts
Parts plant
3
REPLACEMENT OF PARTS USED
Only those parts that have been used up are retrieved in a timely manner
4
PRODUCTION OF PARTS RETRIEVED
Efficiently producing and replenishing only those parts that have been retrieved

slack

slack

PARTS RUNNING
GM LINE #2
KANBAN BOARD
PRIORITY
NEEDS TO RUN
RUN ONLY IF NO RED OR YELLOW CARDS
KANBAN
MONTHLY SHIPMENTS
AAO-5143
AAO-5172
AAO-5338
AAO-5348
WEEK #1
WEEK #2
WEEK #3
WEEK #4

slack


Kanban board
Board
QUICK FILTERS:
Only My Issues
Recently Updated
5 Backlog
Max 5
1 Selected for Development
2 In Progress
Max 5
3 Done
Release...
Expedite 2 issues
SSI-1
Kanban cards represent work items >> Click the "SSI-1" link at the top of this card
SSI-2
Kanban boards are often divided into streams of work, aka Swimlanes. By default,
Everything Else 9 issues
SSI-3
Add work items with "+ Create Issue" at the top right of the screen >> Try adding
SSI-6
Work In Progress (WIP) limits highlight delays. This column's limit is 1...
SSI-8
Filters at the top of the board allow you to quickly cut down the shown items >> Try
SSI-4
Work items are ranked in priority order (from top to bottom) >> Try dragging this
SSI-7
... so 2 work items violate the limit and cause the column to be highlighted
SSI-9
As teams develop with Kanban they get better at reducing average resolution
SSI-5
SSI-10

slack

To Do
Doing
Done
USE KANBAN
TRY KANBAN TOOL
LEARN ABOUT KANBAN
GET SOME STICKY NOTES!
GET A WHITE-BOARD

slack

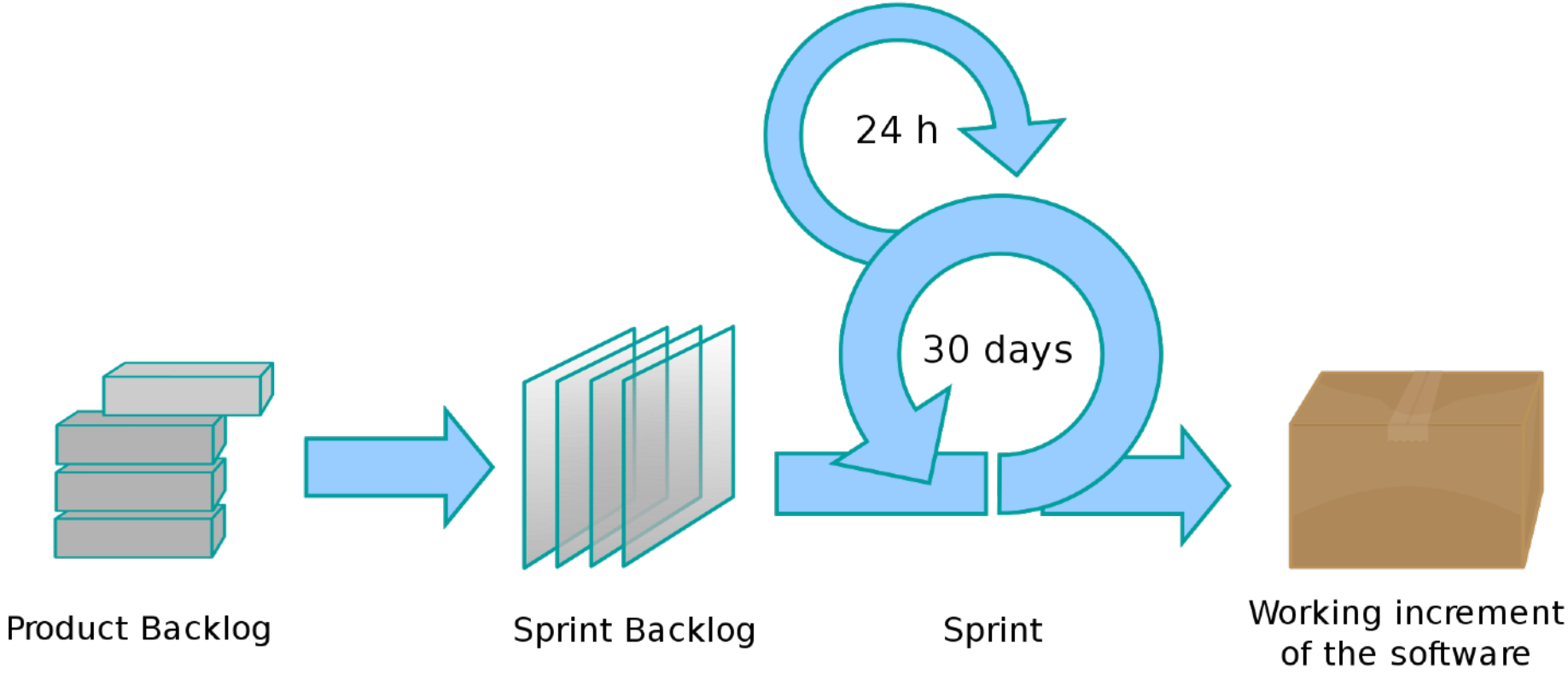
Kaizen
Continuous Improvement

Design

Measure

Learn

slack


24 h
30 days
Product Backlog
Sprint Backlog
Sprint
Working increment
of the software

slack

slack

slack

Executive dedication to learning

High Trust Teams

Design

Measure

Learn

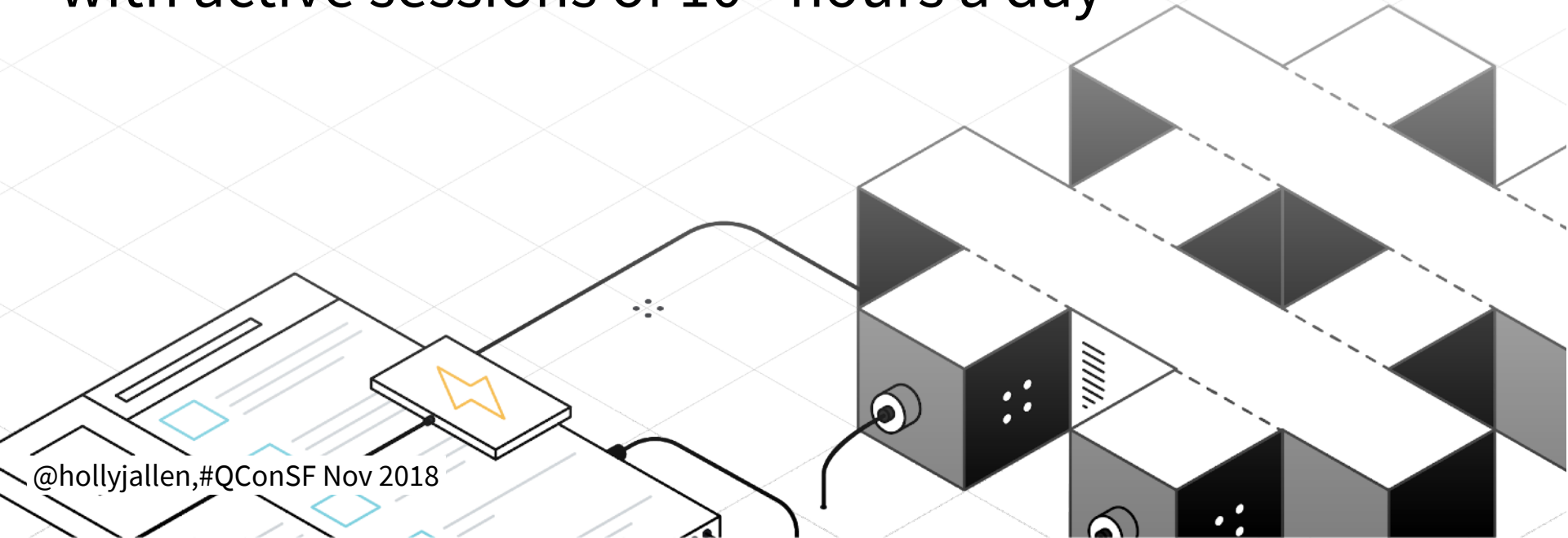
slack

slack

Slack launched
February 2014

5 Years

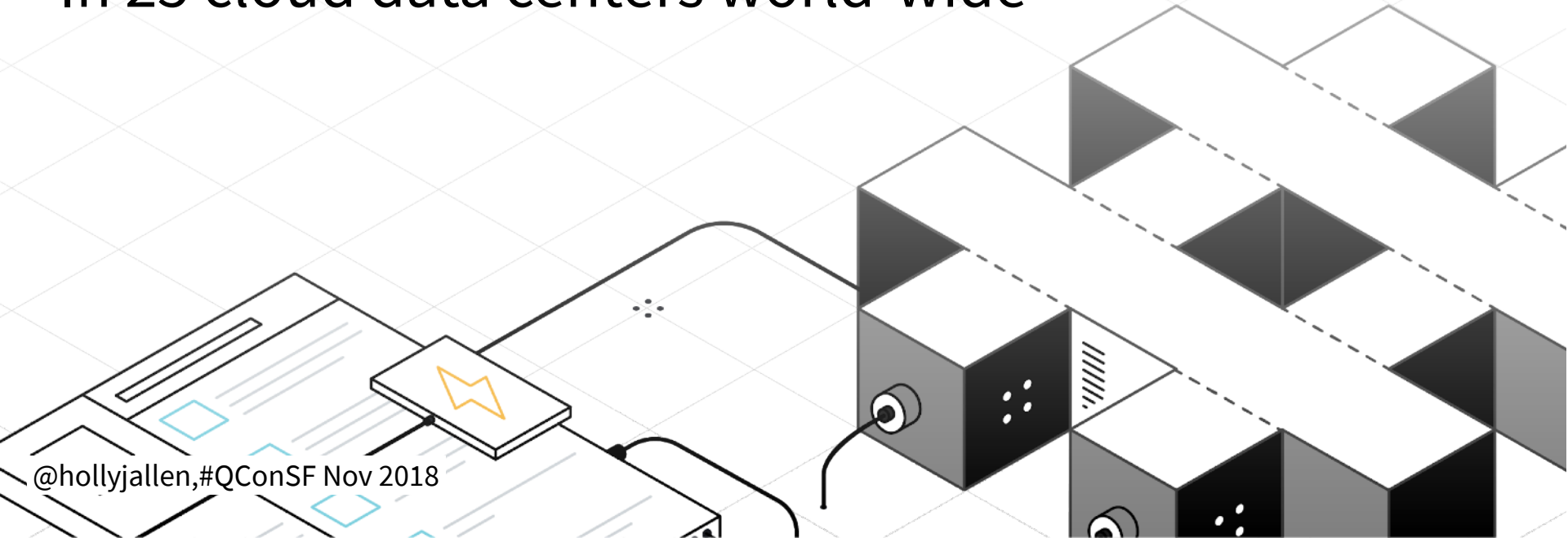
Grew to 13+ million weekly active users,
with active sessions of 10+ hours a day



5 Years

From 10 to 15,000 servers

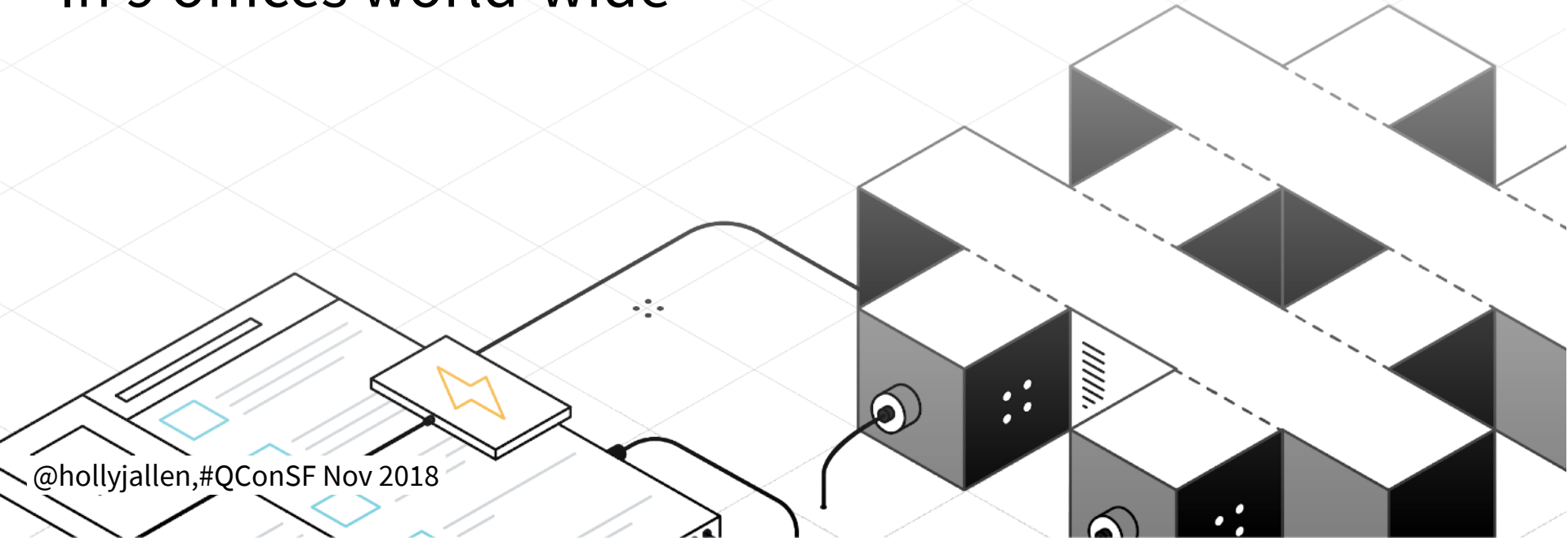
In 25 cloud data centers world-wide



5 Years

From 8 to 1,200 people

In 9 offices world-wide



Design

Measure

Learn

slack

slack

✅ Continuous Deployment

✅ Experiment Frameworks

✅ User Research

Something didn't scale...

Centralized Operations

Who should be responsible for the management, monitoring and operation of a production application?

Centralized Operations Division of Labor

Devs

Features

Scale

Architecture

Ops

Cloud Infra

Deployment

Monitoring

Ops is getting the pages

Product Development grew faster
than Operations,
A lot faster

20 Product Developers

1 Ops Engineer

How can operations reliably reach the developers when there's a problem?

"Call Maude, she knows how this works"

Devs

I've never been on-call before, this is scary!

Ops

Now I know I can find a developer when I need to.

Ops is getting the ~~pages~~ first pages
Ultra-senior devs on-call

Design

Measure

Learn

slack

How can operations reliably reach the developers when there's a problem?

Most devs
go on-call
Fall 2017

Kaizen
Continuous Improvement

"Wait, I'm on-call now?"

Devs

I'm glad I'm only on call a few times a year

Ops

I'll be able to reach a search engineer if I need to.

Learn by Doing

On-call 3 times a year 🤔

Ops is getting the ~~pages~~ first pages

~~Ultra~~-senior devs on-call

Seven ~~One~~ dev rotations

Continuous Deployment
100+ prod deploys a day

What Changed?

Deploy Chicken APP 4:02 PM

duretti deployed d0F90HB6BES

> PR #114458 from button_button_button_must_be_up_to_somethin_whoo - duretti
> PR #114467 from user_groups_lfs_ac - allisoncraig
> PR #114507 from date_picker_bug_fixes - wkimeria

incoming-webhook APP 5:47 PM

🍎 **job-queue-relay-04d256538630bc4dc-6586 logged an error**

> E 1027 17:47:10.844089 6586 topic:157 {} Consul error while reading redis host: Consul error while trying to get key value for job-queue-relay/redis_hosts/indexing, Unexpected response code: 500

👢 Relay process starting: job-queue-relay-04d256538630bc4dc-17582 v97

Grafana - PDE APP 6:36 PM

> **[Alerting] Total Percent Job Dequeues to Enqueues alert**
> We're dequeuing jobs at a lower rate than normal. Runbook: https://slack-github.com/slack/docs/blob/master/appeng/alert_runbooks/job_queue.md#total-percent-job-dequeues-to-enqueues-alert
> **rate of dequeue to enqueue**
> 71.190912049222
>
> Grafana v5.3.0 | Oct 27th

Page the dev

Devs

I don't understand this part of the code

Ops

These are the machine alerts I'm seeing

Human Routers

"Call Andy, he knows how this works"

Postmortems weren't a great place for learning

Can we catch problems earlier?

sieve APP 7:52 AM

🔵 **PHP Errors in api (at least 30 from 06:52 - 07:52)**

30 occurrences of at least 2 Errors:

21 Error(s):

```
E_RECOVERABLE_ERROR: Argument 2 passed to
chat_delete_ephemeral_message() must be an instance of id_t, null
given in /var/www/html/slack1/include/lib_unfurl.php on line 2166
```

9 Error(s):

```
E_RECOVERABLE_ERROR: Argument 2 passed to
chat_delete_ephemeral_message() must be an instance of id_t, null
given in /var/www/html/slack2/include/lib_unfurl.php on line 2166
```

Stack Trace:

```
route.php:5#route_main() -> route.php:201#include() ->
api.php:19#controller_run() ->
lib_controller.php:10#call_user_func() -> #local_api_main() ->
api.php:58#local_api_main_inner() -> api.php:461#api_dispatch() -
> lib_api.php:133#util_call_user_func() ->…
```

Show more

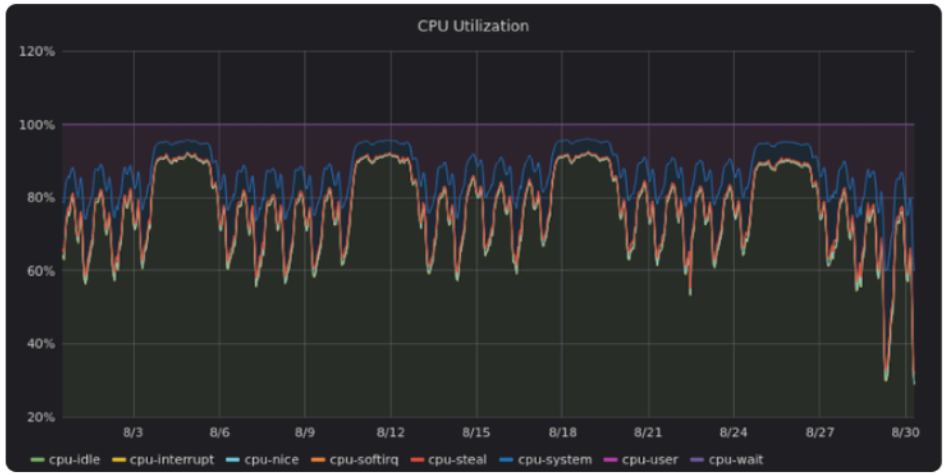
Filter **Snooze 24h**

Kevin Marshall 7:33 AM

https://grafana.tinyspeck.com/d/000000440/hhvm?panelId=8&fullscreen&orgId=1&from=1533063239053&to=1535639557117&var-region=All&var-host=slack-www-hhvm-0547650a48afc01cc

HHVM

CPU Utilization (0 bytes)



Kevin Marshall 8:15 PM

I was curious which memcached keys / operations we do the most per API - https://analytics.tinyspeck.com/v2/explore/428600845589

kevin

Slack Analytics: Explore

```
select
    replace(http_uri, '?_x_id=...') as http_uri,
    prefix,
    op,
    round(approx_percentile(elapsed / 1000.0, .50), 2) as p50,
```

Show more

Sep 3rd (78 kB)

http_uri	prefix	op	p50	p95	hit
/api/dnd.teamInfo	USERS_PREFS	GET	0.99	2.02	1
/api/conversations.history	USER-TEAM	GET	1.03	2.32	0.8
/api/conversations.history	USER-TEAM	MGET	1.58	4.17	1
/api/conversations.history	USERS_PREFS_ALL	GET	1.2	3.05	0.9
/api/chat.postMessage	USER-TEAM	GET	1.03	2.3	0.9
/api/conversations.history	TEAM	GET	1.06	1.94	1
/api/conversations.history	SESSIONS	GET	1.02	2.2	0.9
/api/conversations.history	PINS_COUNT	GET	1	2.15	0.9
/api/conversations.mark	USERS_PREFS_ALL	GET	1.21	3.09	0.9
/api/conversations.mark	USER-TEAM	GET	1.04	2.45	0.9
and 90 more rows...					

Investing in tech to make detection and remediation faster

Reorg!

Fall 2017

Operations is out

Service Engineering is in

How can Slack ensure that developers know when there's a problem?

~~Centralized Operations~~

Service Ownership

Design

Measure

Learn

slack

"We are the toolsmith and specialists. We empower Service Ownership"

Devs

Features

Reliability

Performance

Postmortems

Service

Cloud Platform

Observability tools

Service Discovery

Define best practice

I joined Slack in February 2018

How to empower development teams to improve service reliability?

Define service health and operational maturity

- At least one alerting health metric, like latency or throughput

Send metrics to Prometheus

Observability team is here to help! 🔮

Define service health and operational maturity

- Team should be on-call ready
- At least 4, preferably 6 engineers participating to make it sustainable
- 24/7 or during the weekday, depending on the service

Define service health and operational maturity

- Runbooks for standard actions and troubleshooting
- Central location in our code repository
- Up to date and useable by any engineer

Define service health and operational maturity

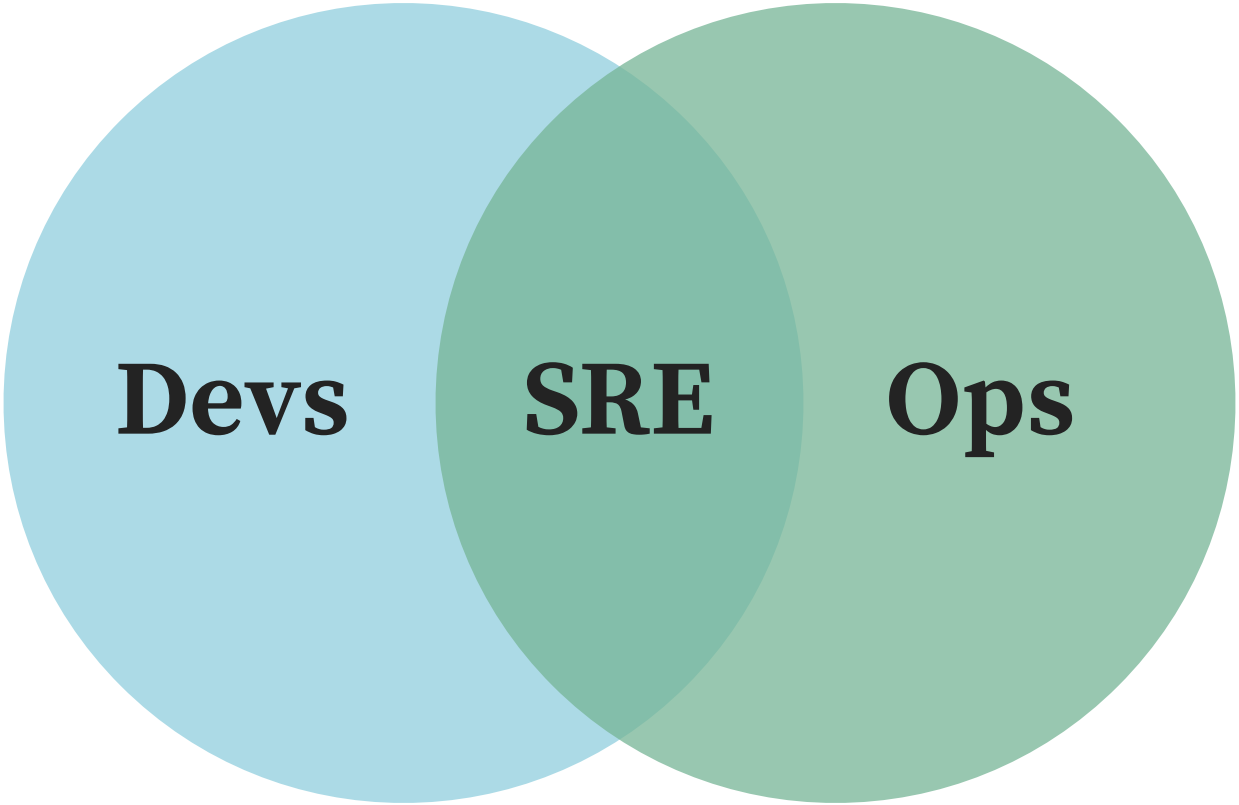
- Paging alerts should link to the runbook
- Make responding to an page easy
- Practice incident response

Incident Lunch

Site Reliability Engineers

- Devops generalists
- Emotional intelligence
- Mentoring
- Ambassadors
- Operational maturity

SRE embedded in dev teams


Devs
SRE
Ops

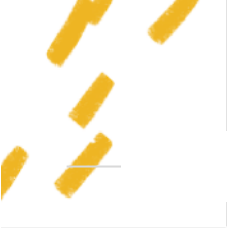
Devs

Um, where are the SREs?

SREs

I'm over here doing operational tasks

SRE ~~Ops~~ is still getting the first pages



How do we lower operational burden on the SREs?

Plan: Send paging alerts to the development teams

Devs

We need training

SREs

We're going to plan this out perfectly

slack

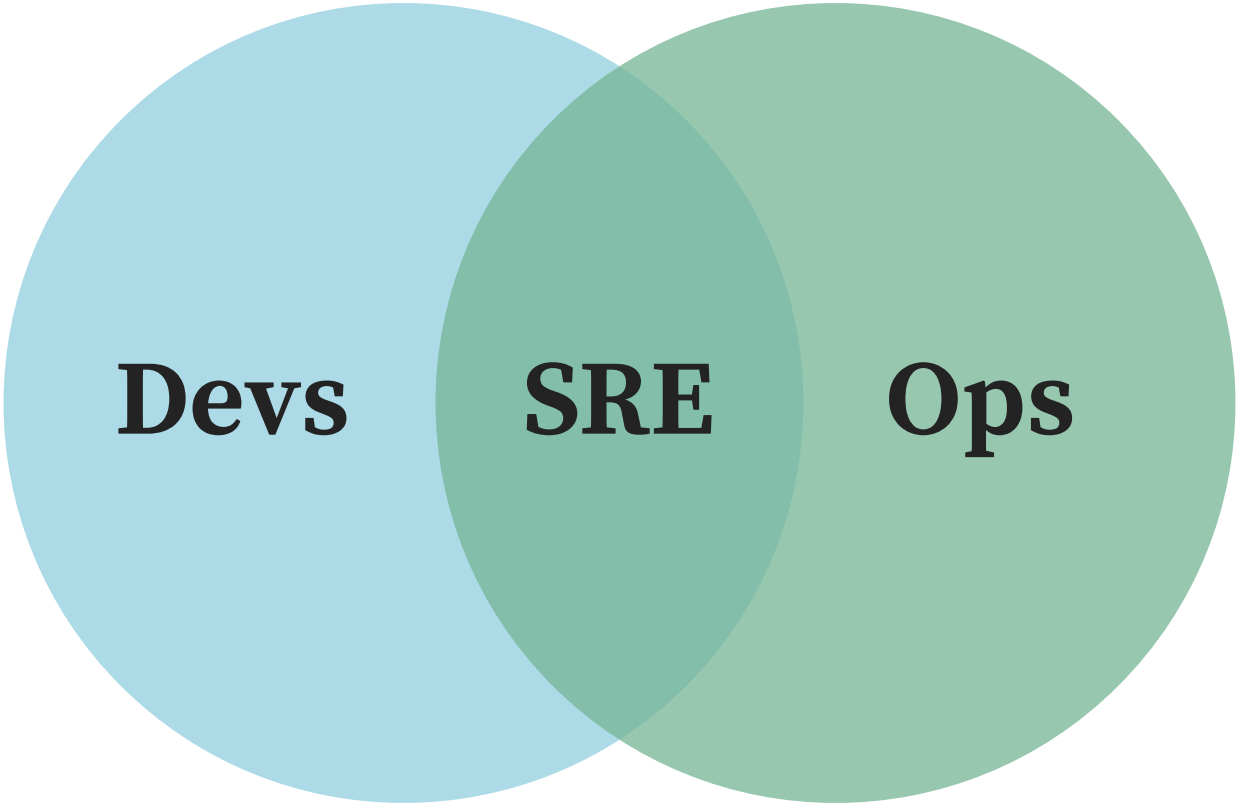
Host level alerts
Hundreds of them

Test with the users

slack

Everything
was fine!

Empowered
Continuous Improvement


Devs
SRE
Ops

How do we test our understanding of how Slack will fail?

"Disasterpiece Theater is an ongoing series of exercises in which we will purposely cause a part of Slack to fail."

Design

Measure

Learn

slack

Success Metrics

- Increased engineer confidence
- Validate reliability improvements
- Learn something new
- Practice incident response

~~Centralized Operations~~

Service Ownership

How do we ensure the teams are being alerted, instead of skillsets?

How do we make postmortems a place for learning again?

How do we make sure a capable incident commander is available for all incidents?

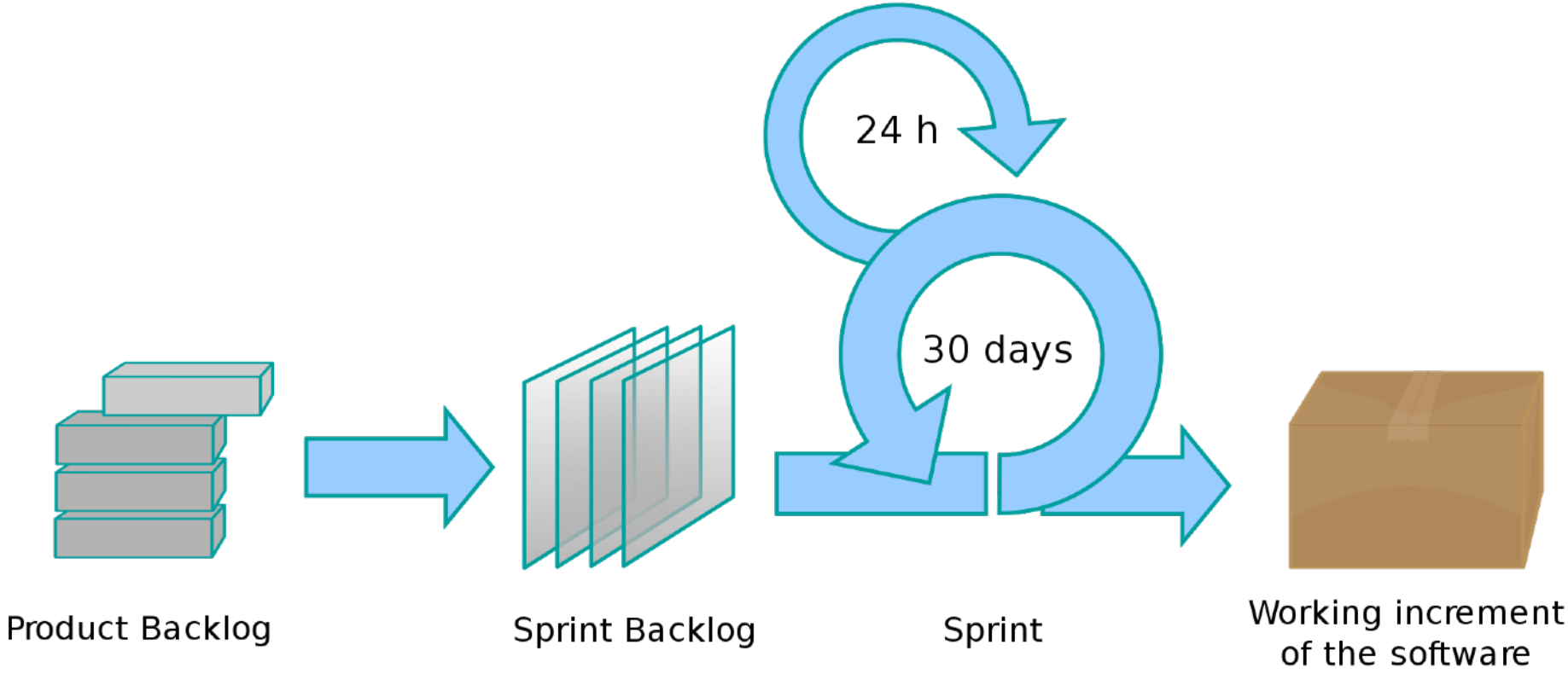
Design

Measure

Learn

slack

slack


24 h
30 days
Product Backlog
Sprint Backlog
Sprint
Working increment
of the software

slack

Copy the questions
Not the answers

Change is possible
You don't have to be ready

Empowered
Continuous Improvement

Design

Measure

Learn

slack

Learn Faster

Thank You

Holly Allen
Service Engineering
@hollyjallen