OPEN SOURCE WITHIN CORPORATE WALLS

openexpo europe

20-06-2019

HONG PHUC DANG

@hpdang

About @hpdang

InnerSource manager **@Zalandotech**

Founder **@FOSSASIA** - FOSSASIA.ORG

Board director of **Open Source Initiative (OSI)**

Zalando at a glance

~ 5.4 billion EUR
revenue 2018

> 214 million
visits per month

> 15.000
employees in Europe

> 70%
of visits via mobile devices

> 26 million
active customers

> 300.000
product choices

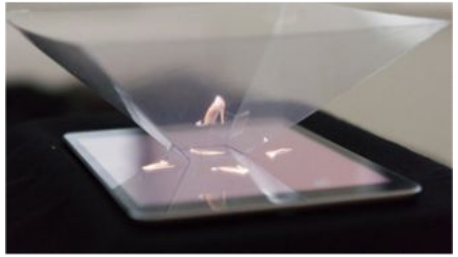
~ 2.000
brands

17
countries

zalando

We are constantly innovating technology

HOME-BREWED, CUTTING-EDGE & SCALABLE
technology solutions



help our brand to
WIN ONLINE

~ 2,000
employees from

77
nations

7 tech locations
(HQs in Berlin)

zalando

OPEN SOURCE CONTRIBUTION

199 active Zalando projects written in **16** different languages,with **883** contributors

Zalando GitHub Orgs

zalando

Zalando SE

The org page for Zalando, Europe's leading online fashion platform. Visit opensource.zalando.com for project stats.

Berlin, Dublin, Helsinki, D… https://tech.zalando.com opensource@zalando.de

Pinned repositories

Customize pinned repositories

patroni
A template for PostgreSQL High Availability with ZooKeeper, etcd, or Consul
Python ★ 2.4k 256

connexion
Swagger/OpenAPI First framework for Python on top of Flask with automatic endpoint validation & OAuth2 support
Python ★ 2.2k 371

skipper
An HTTP router and reverse proxy for service composition, including use cases like Kubernetes Ingress
Go ★ 1.7k 157

zalenium
A flexible and scalable container based Selenium Grid with video recording, live preview, basic auth & dashboard.
Java ★ 1.6k 328

restful-api-guidelines
A model set of guidelines for RESTful APIs and Events, created by Zalando
CSS ★ 959 178

nakadi
A distributed event bus that implements a RESTful API abstraction on top of Kafka-like queues
Java ★ 476 103

Find a repository… Type: All Language: All New

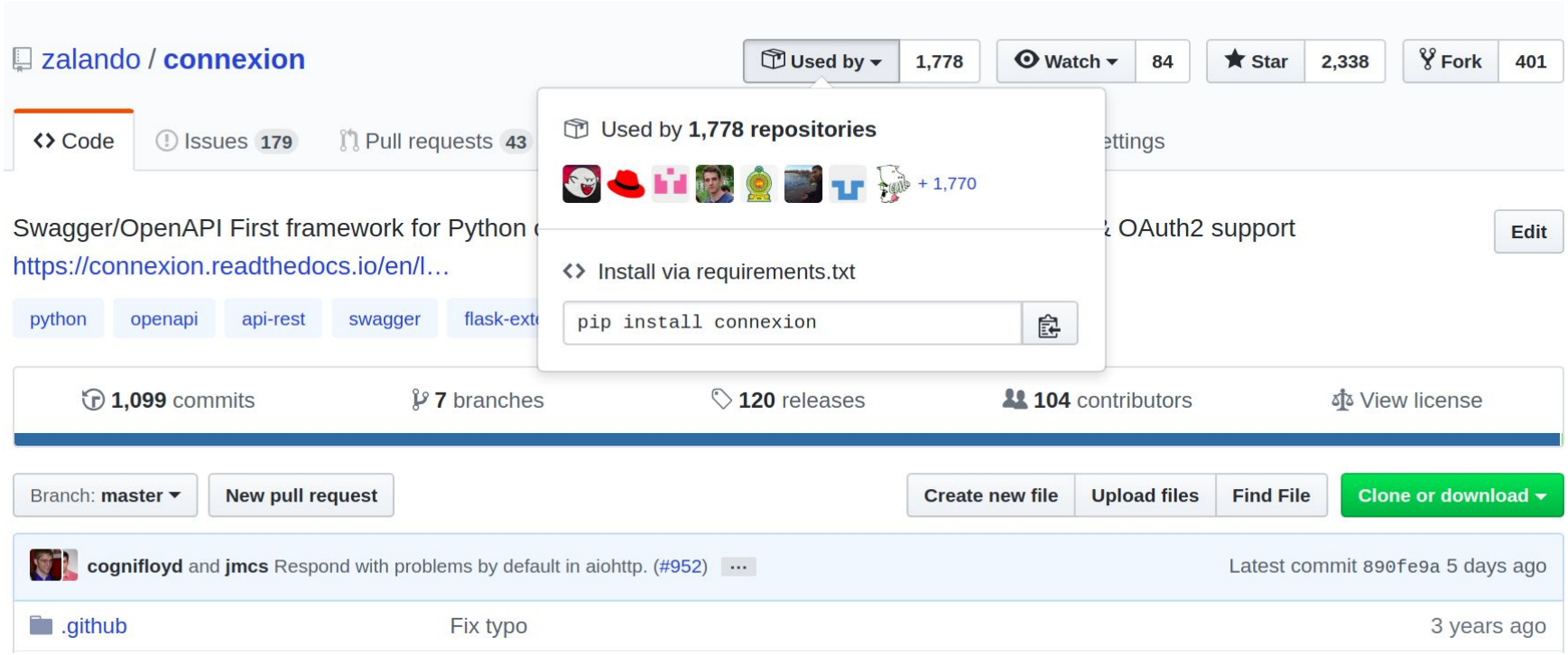
zally

A minimalistic, simple-to-use API linter

kotlin api spring-boot api-management linter swagger

Top languages

Java Python Go Scala CSS


zalando / connexion
Used by 1,778
Watch 84
Star 2,338
Fork 401
Used by 1,778 repositories
+ 1,770
Code
Issues 179
Pull requests 43
ettings
Swagger/OpenAPI First framework for Python
OAuth2 support
Edit
https://connexion.readthedocs.io/en/l…
python
openapi
api-rest
swagger
Install via requirements.txt
pip install connexion
1,099 commits
7 branches
120 releases
104 contributors
View license
Branch: master
New pull request
Create new file
Upload files
Find File
Clone or download
cognifloyd and jmcs Respond with problems by default in aiohttp. (#952)
Latest commit 890fe9a 5 days ago
.github
Fix typo
3 years ago

25	11,239	4,966	357
New Projects	New Commits	Pull Requests Created	New Releases

Work done on our open source projects in 2018

Community and contribution

49%

Of our org members are contributing

Out of the current 510 people in Zalando organisations on GitHub, half of them have made an accepted code contribution to a Zalando open source project - this is a 4% increase since June 2018.

31%

Community contributions

For all current projects, 31% of all pull requests are from non-employees.

521

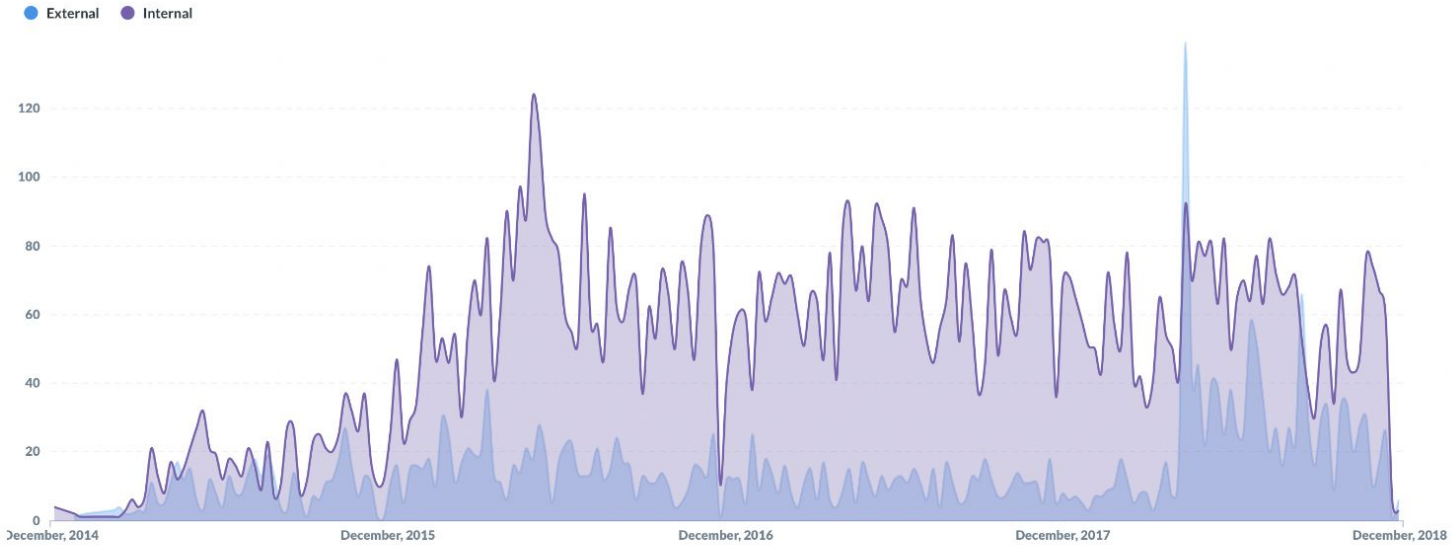
Contributions made to techradar projects

Zalandos contributed to 18 projects on the Techradar "Adopt" tier, which are part of our tech stack and have a high strategic value to Zalando

91 Hours

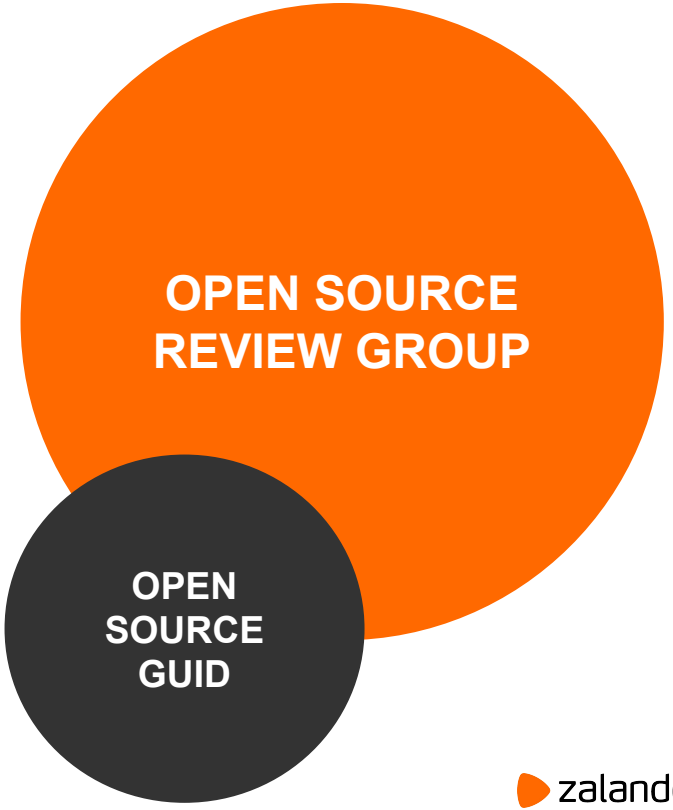
Average time to get a pull request merged

The average response time to pull requests have increased from 85 hours to 91, the number is within an acceptable response time.



Open source activity by Zalando employees (purple) is at a overall stable level with seasonal variations (summer holidays and the run up to black friday are typical low-activity periods). We are however seeing a positive increase in External contribution (blue) to our projects. The single big spike in the chart above is caused by a GitHub bot doing over 100 automated pull requests.

Organisational Structure Of Zalando Open Source



zalando

Open Source Release Process

1. Get sign-off
2. Prepare your repository
3. Submit your project for review
4. Incubation period
5. Graduation
6. Offboarding a Project

zalando

INNERSOURCE

= OPEN SOURCE WITHIN THE WALLS

“InnerSource is the use of open source software development best practices and the establishment of an open source-like culture within organizations.”

- Tim O’Reilly 2000

Tools and Practices

- Correctness
- Compliance
- GDPR
- Security
- Cost Efficiency
- 24x7 On Call
- Governance
- Resilience
- Capacity
- ...

zalando

APPLICATIONS
Pipelines

TOOLS
Repositories
Clusters
Kube Resource Report
Lightstep Manager

TV Link

Welcome to the Developer Console

SOFTWARE DEVELOPMENT | DATA SCIENCE

You can now find consolidated documentation covering the whole Developer Journey in one place:

The Developer's Manual

Idea

- How to write a Narrative
- Template for Narratives

Design

- API Portal
- API Guidelines
- API Linter
- API Review
- Fashion Content Platform (AI & Data APIs)
- Service Guidelines

Test

Plan

- Tech Jira
- GitHub Enterprise
- Rules of Play
- Tech Radar

Code

- Confluence
- Nexus
- Database as a Service
- Data Processing Platform (Databricks)
- Open Source Guidelines
- Nakadi

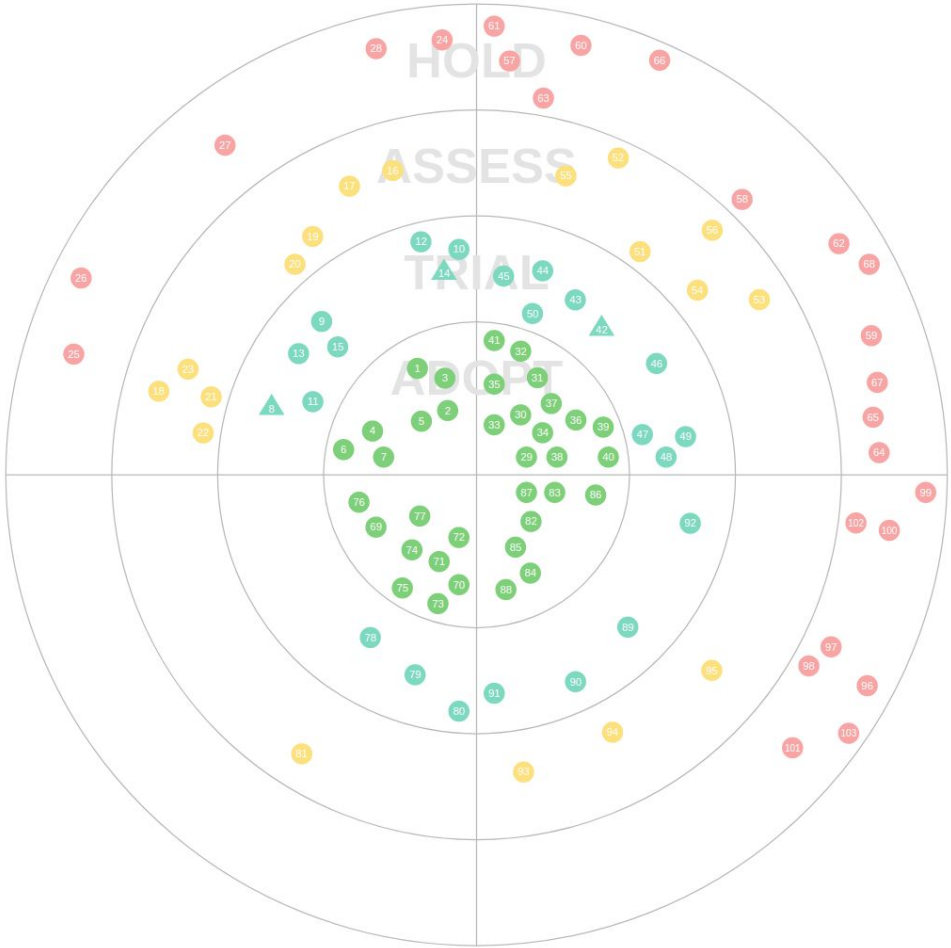
Setup

- ZACK Role Management
- GitHub Enterprise
- Create Git Repository
- YOUR TURN
- Zappr
- Create database cluster
- AWS Account Chooser

Build

- Continuous Delivery Platform (CDP)
- CloudLobster Jenkins

Zalando Tech Radar — 2018.10



Frameworks

ADOPT
1. Akka (Scala)
2. Node.js
3. Play (Scala)
4. ReactJS
5. RxJava (Android)
6. scikit-learn
7. Spring

TRIAL
8. Akka-Http
9. Angular
10. AspectJ
11. Camel
12. Camunda
13. OpenNLP
14. TensorFlow
15. Thymeleaf

ASSESS
16. Aurelia
17. Ember.js
18. gRPC
19. Http4s
20. jOOQ
21. Redux
22. Vert.x
23. Vue.js

HOLD
24. Activiti
25. AngularJS 1.x
26. BackboneJS
27. Drools
28. Spray

Infrastructure

ADOPT
69. Docker
70. HAProxy
71. Hystrix
72. Jetty
73. Kubernetes
74. Nginx
75. STUPS
76. Tomcat
77. ZMON

TRIAL
78. Failsafe
79. OpenTracing
80. Undertow

ASSESS
81. AWS Lambda

HOLD

Data Management

ADOPT
29. AWS EMR
30. AWS S3
31. AWS SNS
32. AWS SQS
33. Cassandra
34. Elasticsearch
35. etcd
36. Kafka
37. Nakadi
38. PostgreSQL
39. Redis
40. Solr
41. Spark

TRIAL
42. Airflow
43. AWS Data Pipeline
44. AWS DynamoDB
45. Flink
46. Google BigQuery
47. HDFS
48. KairosDB
49. Presto
50. RabbitMQ

ASSESS
51. AWS Kinesis
52. Consul
53. Google Bigtable
54. Hadoop
55. RocksDB
56. YARN

HOLD
57. ActiveMQ
58. Aerospike
59. CouchBase
60. Esper
61. HBase
62. HornetQ
63. Memcached
64. MongoDB
65. MySQL
66. Oracle DB
67. Riak
68. ZooKeeper

Languages

ADOPT
82. Go
83. Java
84. JavaScript
85. OpenAPI (Swagger)
86. Python
87. Scala
88. Swift

TRIAL
89. Clojure
90. Haskell
91. Kotlin
92. TypeScript

ASSESS
93. Elm
94. R
95. Rust

HOLD
96. .NET languages
97. C languages
98. CoffeeScript
99. Erlang
100. Groovy
101. Perl
102. PHP
103. Ruby

Zalando Tech Radar — 2018.10

Frameworks

ADOPT
1. Akka (Scala)
2. Node.js
3. Play (Scala)
4. ReactJS
5. RxJava (Android)
6. scikit-learn
7. Spring

TRIAL
8. Akka-Http
9. Angular
10. AspectJ
11. Camel
12. Camunda
13. OpenNLP
14. TensorFlow
15. Thymeleaf

ASSESS
16. Aurelia
17. Ember.js
18. gRPC
19. Http4s
20. jOOQ
21. Redux
22. Vert.x
23. Vue.js

HOLD
24. Activiti
25. AngularJS 1.x
26. BackboneJS
27. Drools
28. Spray

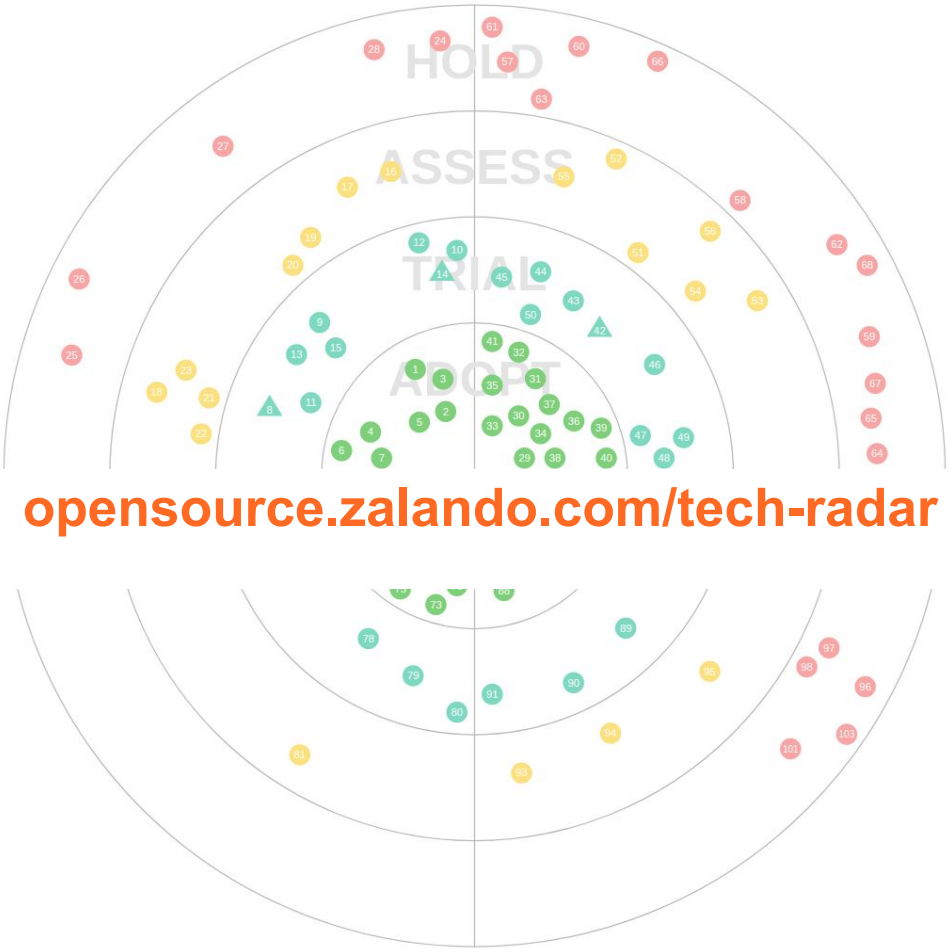
Infrastructure

ADOPT
69. Docker
70. HAProxy
71. Hystrix
72. Jetty
73. Kubernetes
74. Nginx
75. STUPS
76. Tomcat
77. ZMON

TRIAL
78. Failsafe
79. OpenTracing
80. Undertow

ASSESS
81. AWS Lambda

HOLD



opensource.zalando.com/tech-radar

Data Management

ADOPT
29. AWS EMR
30. AWS S3
31. AWS SNS
32. AWS SQS
33. Cassandra
34. Elasticsearch
35. etcd
36. Kafka
37. Nakadi
38. PostgreSQL
39. Redis
40. Solr
41. Spark

TRIAL
42. Airflow
43. AWS Data Pipeline
44. AWS DynamoDB
45. Flink
46. Google BigQuery
47. HDFS
48. KairosDB
49. Presto
50. RabbitMQ

ASSESS
51. AWS Kinesis
52. Consul
53. Google Bigtable
54. Hadoop
55. RocksDB
56. YARN

HOLD
57. ActiveMQ
58. Aerospike
59. CouchBase
60. Esper
61. HBase
62. HornetQ
63. Memcached
64. MongoDB
65. MySQL
66. Oracle DB
67. Riak
68. ZooKeeper

Languages

ADOPT
82. Go
83. Java
84. JavaScript
85. OpenAPI (Swagger)
86. Python
87. Scala
88. Swift

TRIAL
89. Clojure
90. Haskell
91. Kotlin
92. TypeScript

ASSESS
93. Elm
94. R
95. Rust

HOLD
96. .NET languages
97. C languages
98. CoffeeScript
99. Erlang
100. Groovy
101. Perl
102. PHP
103. Ruby

Zalando Tech Radar — 2018.10

Frameworks

ADOPT
1. Akka (Scala)
2. Node.js
3. Play (Scala)
4. ReactJS
5. RxJava (Android)
6. scikit-learn
7. Spring

TRIAL
8. Akka-Http
9. Angular
10. AspectJ
11. Camel
12. Camunda
13. OpenNLP
14. TensorFlow
15. Thymeleaf

ASSESS
16. Aurelia
17. Ember.js
18. gRPC
19. Http4s
20. jOOQ
21. Redux
22. Vert.x
23. Vue.js

HOLD
24. Activiti
25. AngularJS 1.x
26. BackboneJS
27. Drools
28. Spray

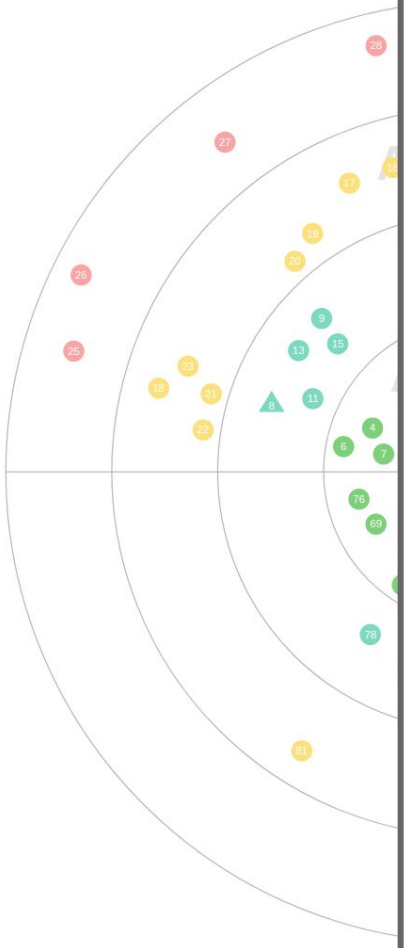
Infrastructure

ADOPT
69. Docker
70. HAProxy
71. Hystrix
72. Jetty
73. Kubernetes
74. Nginx
75. STUPS
76. Tomcat
77. ZMON

TRIAL
78. Failsafe
79. OpenTracing
80. Undertow

ASSESS
81. AWS Lambda

HOLD



ZALANDO TECH RADAR

9fd968d / master-459

edit this page

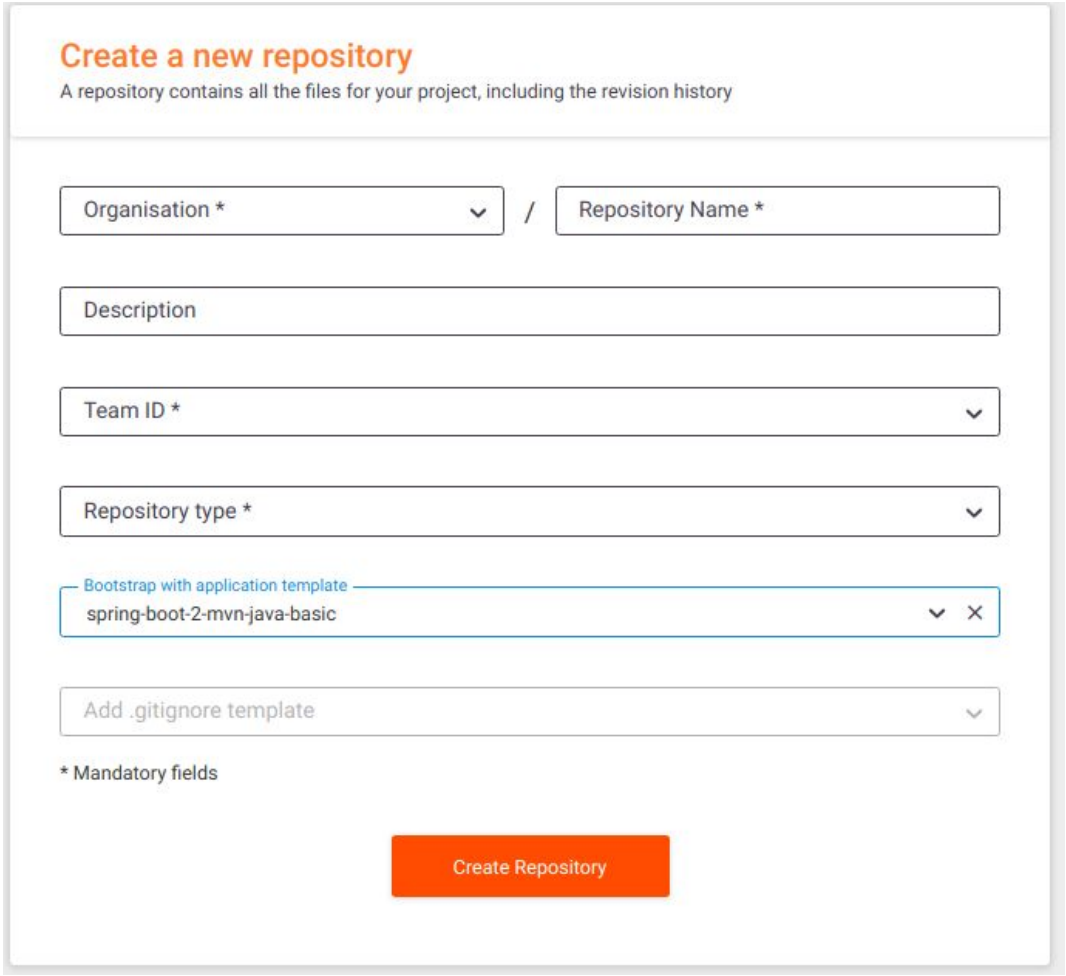
Spark ADOPT

[redacted]

Apache Spark is a fast and general engine for large-scale data processing. It uses an advanced DAG execution engine that supports acyclic data flow and in-memory computing.

Existing experience

- Team [redacted] — We use Apache Spark to create machine learning features from different sources and labelled data sets. These jobs run on temporary AWS EMR clusters and consume data from S3 in various formats such as Apache Parquet, JSON or CSV. We also use the DataFrame based MLlib to train models which perform real time predictions in production.
- Team [redacted] — We use Apache Spark on a daily basis to generate on-demand and scheduled reports from Google AdWords (AdWords data source), to upload campaigns using Spark Streaming, to upload campaigns back to AdWords on our standalone cluster and to process various feeds on EMR cluster.
- Team [redacted] — We use Apache Spark usually when we need to do parallel processing over a big amount of data. We use it in either scala or java and we run it on EMR. We have many projects that use it. Some of them are:
  - Spank (java): It does feature extraction and training of one of the models used in Home, catalog and PDP.
  - Offline-Data-Extraction (scala): data extraction and sequence generation for training. In this project we use spark to calculate word2vec features for each item.
  - Offline-Evaluation (java): We use it for creating test cases and evaluating offline KPIs.
  - Kpi-Calculation (java): We use it for aggregating our logs and calculating our KPIs.
  - Entity-Reco (scala-java): We use it for calculating our baseline for many entities like items, brands, etc.
  - snitch (java): We use it for aggregating the logs based on some rules that can be establish by configuration and then we do anomaly detection over the


Create a new repository
A repository contains all the files for your project, including the revision history
Organisation *
/
Repository Name *
Description
Team ID *
Repository type *
Bootstrap with application template
spring-boot-2-mvn-java-basic
Add .gitignore template
* Mandatory fields
Create Repository

zalando

Branch: **master** | **New pull request** | **Create new file** | **Upload files** | **Find file** | **Clone or download**

machinery-cd-robot add zapprfile — Latest commit 79136e7 7 days ago

.mvn	Initializing with template	7 days ago
config	Initializing with template	7 days ago
deploy/apply	Initializing with template	7 days ago
meta/credentials	Initializing with template	7 days ago
spring-config	Initializing with template	7 days ago
src	Initializing with template	7 days ago
.gitignore	Initializing with template	7 days ago
.zappr.yaml	add zapprfile	7 days ago
README.md	Initializing with template	7 days ago
configure.sh	Initializing with template	7 days ago
delivery.yaml	Initializing with template	7 days ago
lombok.config	Initializing with template	7 days ago
mvnw	Initializing with template	7 days ago
mvnw.cmd	Initializing with template	7 days ago
pom.xml	Initializing with template	7 days ago

zalando

Developer's Manual

This is the very first release of the Developer's Manual, feel free to send us your feedback to mostafa.nageeb@zalando.de. By using the "Edit on GitHub" link on the top, you can directly contribute changes via pull requests.

Tip

Check out the Zalando Developer Console on dev.zalando.net which links to all relevant documentations, tools and support channels.

Contents:

Welcome to Documentation at Zalando

> **Tip**
>
> Use the Zalando Developer Console on dev.zalando.net to search existing documentation.

Looking to create a documentation site? You've come to the right place! Continue reading to learn how to get a documentation site up and running in under two minutes!

Overview

Documentation is a key component for collaboration and the sharing of ideas and Developer Productivity's goal is to make documentation and documentation discovery as easy as possible. Using our Cloud Native Application Runtime functionality, one can publish static html which can be uploaded and hosted on Zalando Docs. This allows for the creation of a documentation site with its own sub domain under `*.docs.zalando.net`, authenticated with Platform IAM and integrated with a Git / CDP workflow. Wow, that's a lot. But fear not, this site covers how to get a documentation site up and running in under two minutes.

Move onto the next step by clicking the link below.

Discover great APIs.

The central hub for all APIs in the Zalando universe!

Search for an API

Get started
Short introduction to API Portal.

Publish API
How to publish API specifications.

Improve API design
API Guidelines - rules for high quality APIs.

Lint API online
Use Zally Web UI for checking compliance with guidelines.

Get feedback for API
Feedback quality assurance process for API design.

Register functional component
Central registry for functional names.

Lint API in IntelliJ
Use IntelliJ Plugin for checking compliance with guidelines.

Getting started

In API Portal you can find current and historical information on API deployments, including details on service applications and OpenAPI specifications.

How to search

API Portal offers a powerful **full-text search** that enables you to find exactly what you are looking for. The easiest way to get started is to **simply type terms** into the search box. If a term consists of multiple words, just use **double quotes**, e.g. `"api portal"`. Things you can search for include the title, endpoints and schemas of an API as well as the URL, application and application owner of an API deployment. To see **all** existing APIs you can search for *.

Advanced queries

In order to further narrow down the search results, it is possible to write a more advanced query using fields and operators.

Knowledge Sharing

Internal Guilds

Zalando Tech Community ~2000 people

Scala	API	OPEN SOURCE	DATA SCIENCE
Tech Interviewing	WEB GUILD	TECHNOLOGISTS	Agility
Producer	Databases	Python	Clojure
Cassandra	SRE	Diversity Guild	General Gaming

Communications

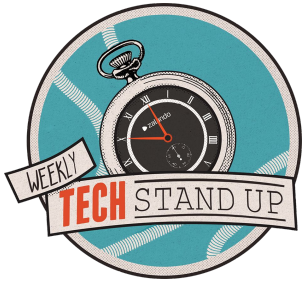
zLife

Sharing Platforms

TECH Talks

Meetup

DEMOS



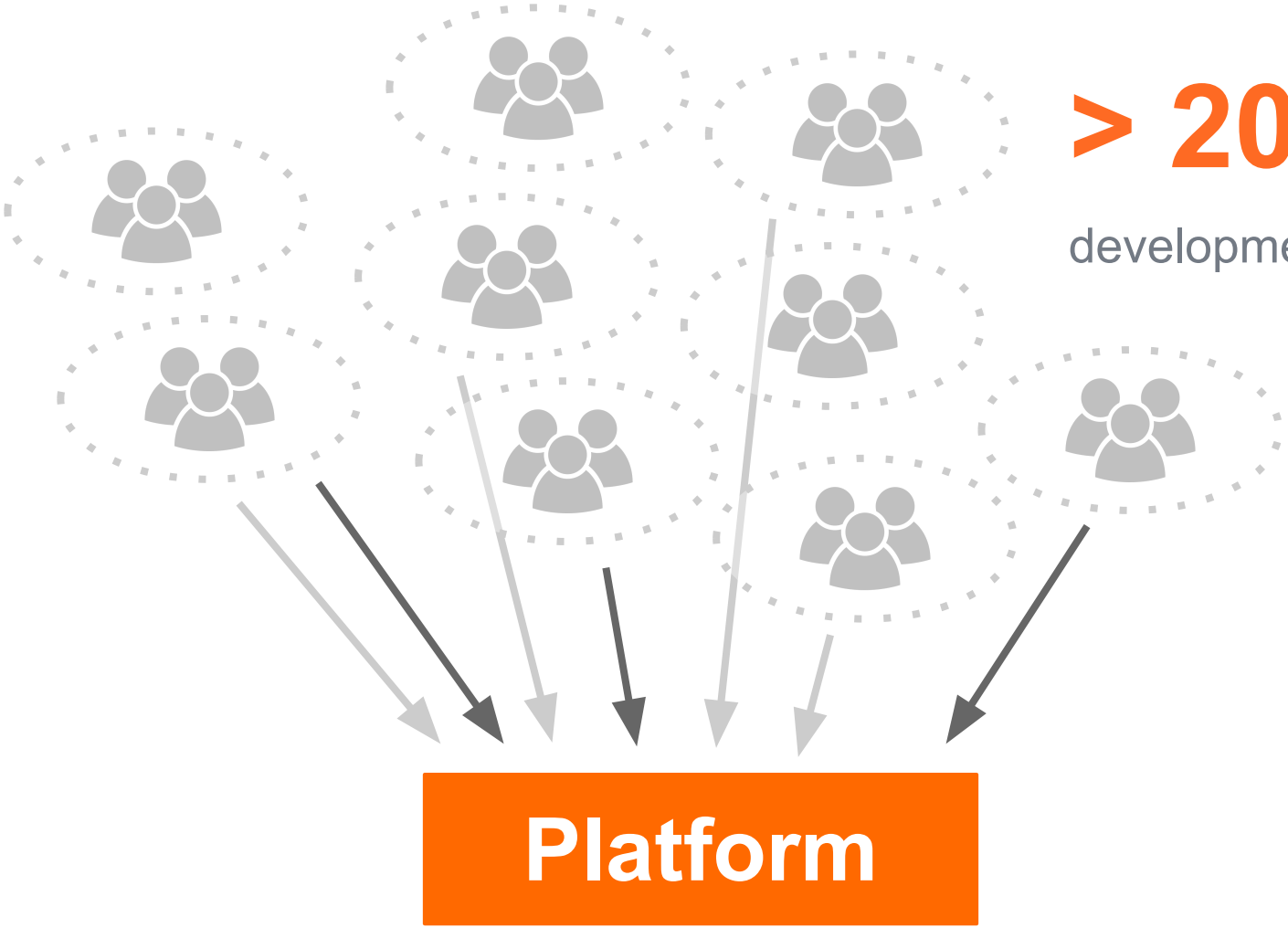
Team and Cross-Team Collaboration

The Way of Working

Agile
Daily Scrum
Sprint Planning
Backlog Refinement
Retrospective Meeting
Cross-team alignment with facilitators
All hands

“postits” by Peer 2 Peer University (https://flic.kr/p/bt51RX)

zalando

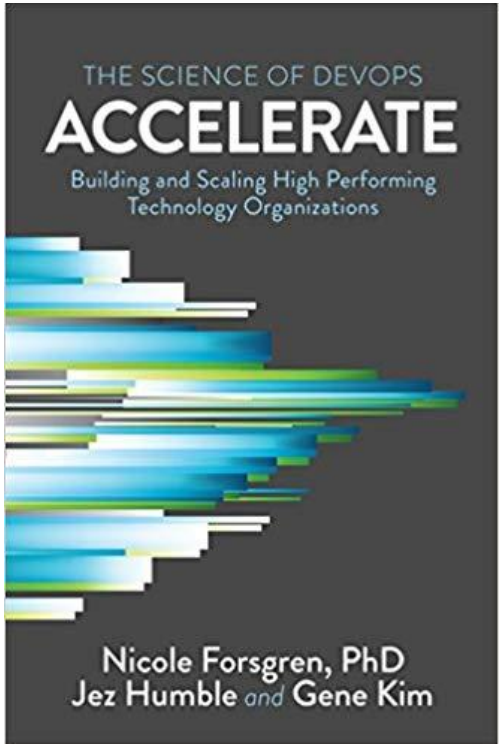


zalando

Measurement

Delivery Performance Metrics

- Lead Time
- Release Frequency
- Time to Restore Service
- Change Fail Rate



InnerSource: Lessons Learned

Understand the problems

Listen to your engineers

Explore existing tools

Make a business case

Define measurable impacts

zalando

QUESTIONS?

HONG PHUC DANG

@hpdang