

Data lake: how Red Hat maintains data quality across multiple Drupal sites

Florida Drupalcamp 2023

Melissa Bent

Senior Software Engineer

April Sides

Senior Software Engineer



- Problem
- Discovery
- Solution
- Integration

Problem

Redhat.com is built on
**Drupal and single-page
applications.**

Organization-level data
should be **consistent** and
accurate across
redhat.com.

Organization-level data is
 duplicated and managed
by **different teams** .

Managing **data quality**
across redhat.com is a
manual process.

There has to be a **better way to share data** across the redhat.com ecosystem.

Discovery

Requirements

- ▶ Share data in a scalable and maintainable way
- ▶ Connect with different tech stacks
- ▶ Serve as a single source of truth
- ▶ Provide flexible data model/schema

Data Repository Types

- ▶ Relational database
- ▶ Data warehouse
 - Data mart
 - Operational data store
- ▶ Data lake



✘ Relational database

- Structured transactional data
- Main features:
 - Data normalization
 - Compliancy

Example: Drupal database



Data warehouse

- Business intelligence focus
- Structured data
- Schema-on-write



Data Lake

- Flexible structure
- Scalable
- Schema-on-read
- Low maintenance

Data Lake Challenges

- ▶ Data rot
- ▶ Data governance
- ▶ Data compliance
- ▶ Data security
- ▶ Data availability

Data Lake Advantages

- ▶ Multiple data sources—Consistent access
- ▶ Protection against downtime
- ▶ Provides an additional, query-level, caching layer for production

Solution

Tech Stack

- ▶ Database Backend: MongoDB
- ▶ Indexing: Search API + Custom module
- ▶ Retrieval
 - GraphQL
 - Direct query

Indexing

- ▶ Custom Search API Backend
- ▶ Search API's index management
- ▶ Flexible schema for each data source
- ▶ Drupal provides access control, the data model, and the editorial experience

Retrieval

- ▶ Single-page applications: GraphQL
- ▶ Drupal: PHP MongoDB Driver

Integration

Integration: Products



Red Hat Customer Portal

Product Experience

access.redhat.com/products

Customer Portal Products

- ▶ Customer Portal helps our customers get the most out of their subscriptions
- ▶ Product information is core to our data organization
- ▶ Multiple teams and sites combine to make what is the Customer Portal

The screenshot shows a web browser window with the URL `access.redhat.com/products/?product-tab=glossary`. The page title is "Red Hat Products and Documentation". Below the title, there are two tabs: "A-Z" (which is selected and underlined with a red line) and "Categories". A search bar contains the text "# A C D E F G H I J M N O P Q R S U V". Below the search bar, there is a list of products, each with a "Product Info" and "Documentation" link. The visible products are:

- Red Hat 3scale API Management
- Red Hat Advanced Cluster Management for Kubernetes
- Red Hat Advanced Cluster Security for Kubernetes
- Red Hat AMQ

The screenshot shows a web browser window with the URL `access.redhat.com/products/?product-tab=category`. The page title is "Red Hat Products and Documentation". Below the title, there are two tabs: "A-Z" and "Categories", with "Categories" being the active tab. A list of product categories is displayed, each with a right-pointing chevron icon. An "Expand All" button is located to the right of the list.

Category	Expand
Red Hat Developer	>
Partner Certification	>
Infrastructure and Management	>
Integration and Automation	>
Runtimes	>
Subscription Tools	>
Cloud Computing	>

The screenshot shows a web browser window with the URL `access.redhat.com/products/?product-tab=category`. The page title is "모든 Red Hat 제품 및 문서" (All Red Hat products and documents). Below the title, there are two tabs: "A-Z" and "카테고리" (Category), with "카테고리" being the active tab. On the right side, there is a button labeled "모두 펼치기" (Expand all). Below this, a list of categories is displayed, each with a right-pointing chevron icon:

- Red Hat 개발자
- 파트너 인증
- 인프라 및 관리
- 통합 및 자동화
- 런타임
- 서브스크립션 툴
- 클라우드 컴퓨팅

Strategy

- ▶ Product data managed in Drupal
- ▶ Indexed to the Data Lake
- ▶ Queried via GraphQL
- ▶ Page built via GitLab pipeline (statically generated)
- ▶ Refreshed every 30 minutes

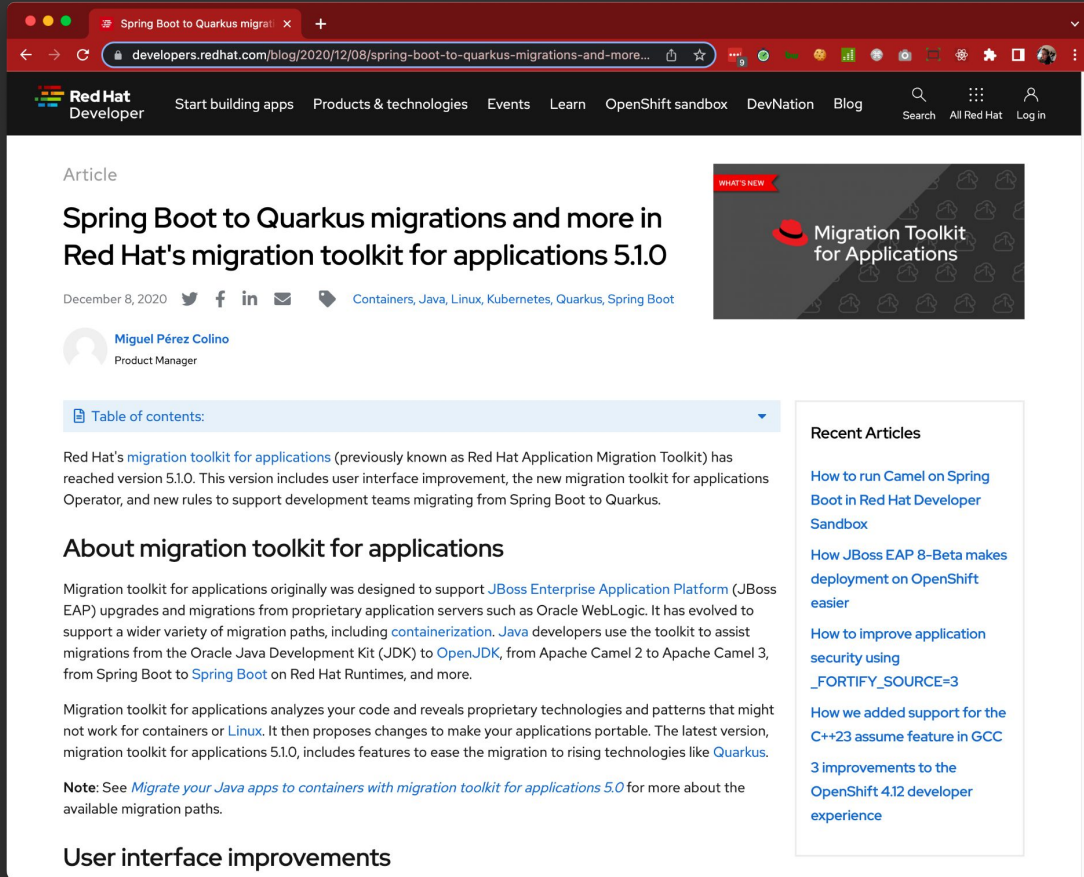
Integration: Learning Paths

Learning Path

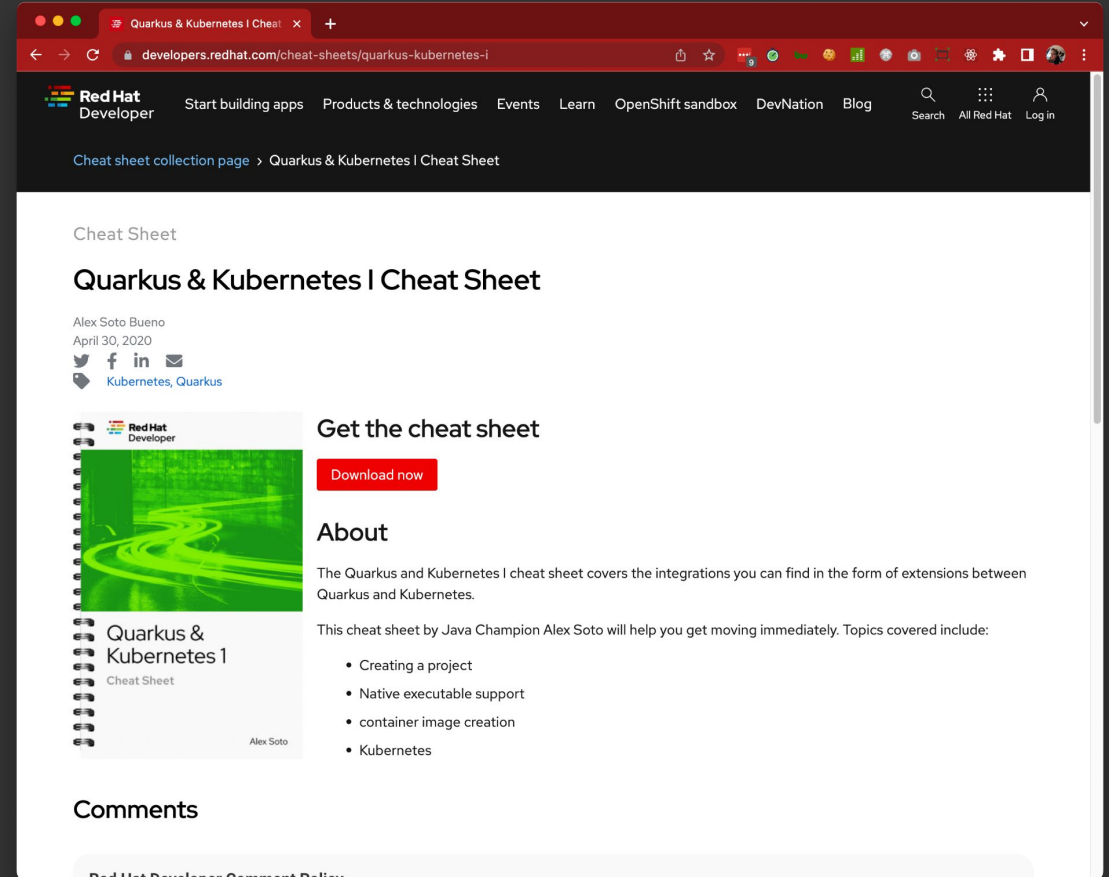
A curated collection of content, directing users to learn more about a particular topic or product.



developers.redhat.com



Article



Cheat Sheet

developers.redhat.com

The screenshot shows a browser window with the URL `developers.redhat.com/learning/devnation:kubernetes-and-friends/resource/resources...`. The page title is "Kubernetes and friends" and the breadcrumb is "DevNation > Kubernetes and friends > Spring Boot to Quarkus migrations and more in Red Hat's migration toolkit for applications 5.1.0". The main content area features an article titled "Spring Boot to Quarkus migrations and more in Red Hat's migration toolkit for applications 5.1.0" by Miguel Pérez Colino, published on December 8, 2020. The article text discusses the Red Hat migration toolkit for applications (formerly Red Hat Application Migration Toolkit) reaching version 5.1.0, which includes user interface improvements and new migration rules for Spring Boot to Quarkus. A sidebar on the left lists a learning path with six items: Overview: Kubernetes and friends, Beginner: Kubernetes I Tutorial, Kubernetes in the sandbox, Quarkus & Kubernetes I Cheat Sheet, Video: Kubernetes Fundamentals, and Elementary: Kubernetes II Tutorial. The current article is the third item in the path.

Article in Learning Path

The screenshot shows a browser window with the URL `developers.redhat.com/learning/devnation:kubernetes-and-friends/resource/resources...`. The page title is "Kubernetes and friends" and the breadcrumb is "DevNation > Kubernetes and friends > Quarkus & Kubernetes I Cheat Sheet". The main content area features a cheat sheet titled "Quarkus & Kubernetes I Cheat Sheet" by Alex Soto Bueno, published on April 30, 2020. The cheat sheet covers integrations between Quarkus and Kubernetes. A "Download now" button is visible. The sidebar on the left lists a learning path with six items: Overview: Kubernetes and friends, Beginner: Kubernetes I Tutorial, Kubernetes in the sandbox, Quarkus & Kubernetes I Cheat Sheet, Video: Kubernetes Fundamentals, and Elementary: Kubernetes II Tutorial. The current cheat sheet is the third item in the path.

Cheat Sheet in Learning Path

Learning Path content type

The screenshot displays the Workbench interface for managing learning paths. The main view shows a list of resources under the heading "Resources". A modal window titled "Add resource" is open, allowing the user to add a new resource to the path. The modal includes a "Library" tab and a "Create resource" link. It features dropdown menus for "Product" (set to "Red Hat OpenShift API Management") and "Topics" (set to "- Any -"). There is also a "Title" input field and "Apply" and "Reset" buttons. Below the form, three resource cards are visible, each with a thumbnail, a title, and a duration. The first card is titled "Getting started with the workshop" and has a duration of 15 minutes. The second card is titled "Set up your instance of Red Hat's single sign-on technology on the Red Hat OpenShift API Management Developer Sandbox" and has a duration of 15 minutes. The third card is titled "What is Red Hat OpenShift API Management?" and has a duration of 15 minutes.

Resources

The resources within this learning path. They will be listed in the order in which they are added.

Create new revision
Revisions are required.

Add resource

Title

+

 Red Hat OpenShift architecture

+

 Explore Red Hat OpenShift

+

 Create your first project

+

 Deploy an application using a pre-existing container image

+

 Work with application labels to identify component parts

Add resource

Title

+

 Red Hat Op

+

 Explore Red

+

 Create you

+

 Deploy an a

+

 Work with a

Add resource

Library **Create resource**

Product Red Hat OpenShift API Management

Topics - Any -

Title

Apply **Reset**

Page 1

Getting started with the workshop

15 minutes

Learn how to to consume and produce Kafka messages.

Page 2

Set up your instance of Red Hat's single sign-on technology on the Red Hat OpenShift API Management Developer Sandbox

Set up your instance of Follow these steps only if you are using the OpenShift API Management Developer Sandbox. Red Hat's single sign-on technology on the Red Hat OpenShift API Management Developer Sandbox.

Page 1

What is Red Hat OpenShift API Management?

Red Hat OpenShift API Management is a cloud service for creating, securing, and publishing your APIs. Get an introduction to how it works.

Unlock

Resource content type

Add resource

Text format: Red Hat Developers Approved HTML [About text formats](#)

Red Hat Learning Paths: Resource Item *

OpenShift

- [RHDP Cheat Sheet (en)] Debezium on OpenShift Cheat Sheet
- [RHDP Cheat Sheet (en)] 10 tips for running Node.js applications on OpenShift
- [RHDP Cheat Sheet (en)] Red Hat OpenShift Container Platform Cheat Sheet
- [RHDP E-books (en)] OpenShift 3 for Developers: A Guide for Impatient Beginners
- [RHDP E-books (en)] Devops with OpenShift: Cloud Deployments Made Easy
- [RHDP E-books (en)] Deploying to OpenShift
- [RHDP E-books (en)] OpenShift for Developers, Second Edition
- [RHDP E-books (en)] DevOps Culture and Practice with OpenShift
- [RHDP E-books (en)] Getting GitOps: A practical platform with OpenShift, Argo CD,

Add resource

Text format: Red Hat Developers Approved HTML

Red Hat Learning Paths: Resource Item *

drhc001:b740e1db-0f86-43fd-9ca1-2ad480e00c39

Select a resource item from the Learning Paths Data Lake by title.

Duration

The screenshot shows a browser window with the URL `developers.redhat.com/learning/devnation:kubernetes-and-friends/resource/resources...`. The page title is "Kubernetes and friends" and the breadcrumb is "DevNation > Kubernetes and friends > Spring Boot to Quarkus migrations and more in Red Hat's migration toolkit for applications 5.1.0". The article title is "Spring Boot to Quarkus migrations and more in Red Hat's migration toolkit for applications 5.1.0" by Miguel Pérez Colino, published on December 8, 2020. A sidebar on the left lists the learning path items: Overview: Kubernetes and friends, Beginner: Kubernetes I Tutorial, Kubernetes in the sandbox, Quarkus & Kubernetes I Cheat Sheet, Video: Kubernetes Fundamentals, Elementary: Kubernetes II Tutorial, and OpenShift in the sandbox II. The article content discusses the migration toolkit for applications 5.1.0 and its features.

Article in Learning Path

The screenshot shows a browser window with the URL `developers.redhat.com/learning/devnation:kubernetes-and-friends/resource/resources...`. The page title is "Kubernetes and friends" and the breadcrumb is "DevNation > Kubernetes and friends > Quarkus & Kubernetes I Cheat Sheet". The article title is "Quarkus & Kubernetes I Cheat Sheet" by Alex Soto Bueno, published on April 30, 2020. A sidebar on the left lists the learning path items: Overview: Kubernetes and friends, Beginner: Kubernetes I Tutorial, Kubernetes in the sandbox, Quarkus & Kubernetes I Cheat Sheet, Video: Kubernetes Fundamentals, Elementary: Kubernetes II Tutorial, and OpenShift in the sandbox II. The article content includes a "Download now" button and a list of topics covered: Creating a project, Native executable support, container image creation, and Kubernetes.

Cheat Sheet in Learning Path

Spring Boot to Quarkus migrati

developers.redhat.com/learning/devnation:kubernetes-and-friends/resource/resources...

Red Hat Developer Start building apps Products & technologies Events Learn OpenShift sandbox DevNation Blog Search All Red Hat Log in

DevNation > Kubernetes and friends > Spring Boot to Quarkus migrations and more in Red Hat's migration toolkit for applications 5.1.0

Kubernetes and friends

Learning path | 21 resources | 12 hrs and 45 mins | Published on August 10, 2021

Spend time getting to know your new friend - Kubernetes. Learn more about this open source container orchestration system and make notes on commands, tips, and tricks to bring it to life.

- Overview: Kubernetes and friends
- 1 Beginner: Kubernetes I Tutorial Website | 30 mins
- 2 Kubernetes in the sandbox Website | 1 hr
- 3 Quarkus & Kubernetes I Cheat Sheet Cheat Sheet | 1 hr
- 4 Video: Kubernetes Fundamentals Website | 1 hr
- 5 Elementary: Kubernetes II Tutorial Website | 30 mins
- 6 OpenShift in the sandbox II Website | 30 mins

Article

Spring Boot to Quarkus migrations and more in Red Hat's migration toolkit for applications 5.1.0

Containers, Java, Linux, Kubernetes, Quarkus, Spring Boot

Miguel Pérez Colino December 8, 2020

Red Hat's [migration toolkit for applications](#) (previously known as Red Hat Application Migration Toolkit) has reached version 5.1.0. This version includes user interface improvement, the new migration toolkit for applications Operator, and new rules to support development teams migrating from Spring Boot to Quarkus.

About migration toolkit for applications

Migration toolkit for applications originally was designed to support [JBoss Enterprise Application Platform](#) (JBoss EAP) upgrades and migrations from proprietary application servers such as Oracle WebLogic. It has evolved to support a wider variety of migration paths, including [containerization](#). Java developers use the toolkit to assist migrations from the Oracle Java Development Kit (JDK) to [OpenJDK](#), from Apache Camel 2 to Apache Camel 3, from [Spring Boot](#) to [Spring Boot](#) on Red Hat Runtimes, and more.

Article in Learning Path

Quarkus & Kubernetes I Cheat

developers.redhat.com/learning/devnation:kubernetes-and-friends/resource/resources...

Red Hat Developer Start building apps Products & technologies Events Learn OpenShift sandbox DevNation Blog Search All Red Hat Log in

DevNation > Kubernetes and friends > Quarkus & Kubernetes I Cheat Sheet

Kubernetes and friends

Learning path | 21 resources | 12 hrs and 45 mins | Published on August 10, 2021

Spend time getting to know your new friend - Kubernetes. Learn more about this open source container orchestration system and make notes on commands, tips, and tricks to bring it to life.

- Overview: Kubernetes and friends
- 1 Beginner: Kubernetes I Tutorial Website | 30 mins
- 2 Kubernetes in the sandbox Website | 1 hr
- 3 Quarkus & Kubernetes I Cheat Sheet Cheat Sheet | 1 hr
- 4 Video: Kubernetes Fundamentals Website | 1 hr
- 5 Elementary: Kubernetes II Tutorial Website | 30 mins
- 6 OpenShift in the sandbox II Website | 30 mins

Cheat Sheet

Quarkus & Kubernetes I Cheat Sheet

Kubernetes, Quarkus

Alex Soto Bueno April 30, 2020

Get the cheatSheet

Download now

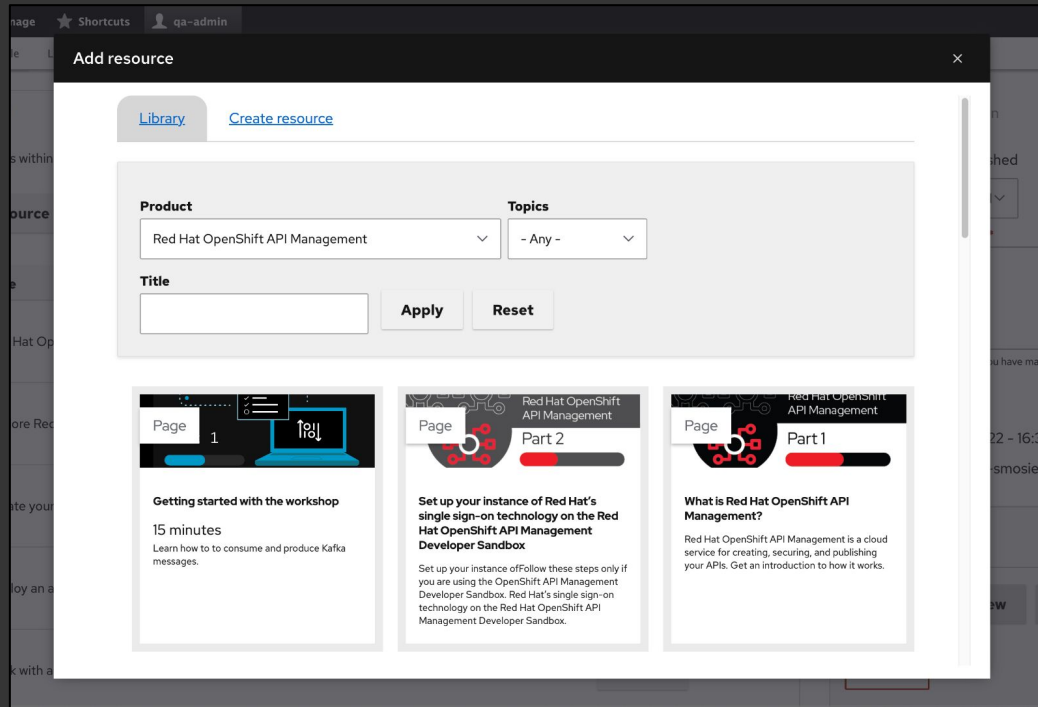
The Quarkus and Kubernetes I cheat sheet covers the integrations you can find in the form of extensions between Quarkus and Kubernetes.

This cheat sheet by Java Champion Alex Soto will help you get moving immediately. Topics covered include:

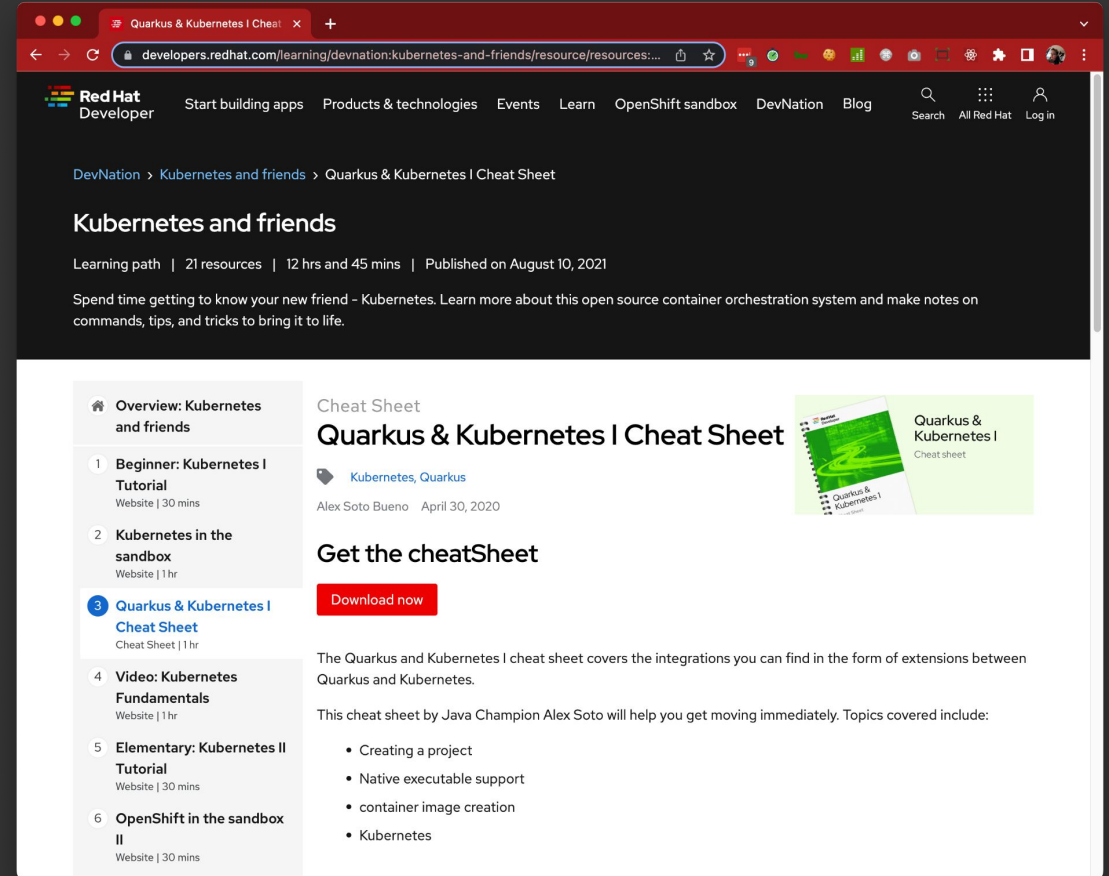
- Creating a project
- Native executable support
- container image creation
- Kubernetes

Cheat Sheet in Learning Path

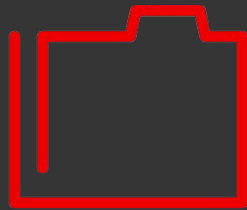
Resource displays



`hook_preprocess_node()`

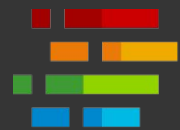


Shared module



Learning Paths shared module

- Data Lake Learning Paths schema
- Data Lake query service
- Reusable code:
 - Services
 - Controller
 - EventSubscriber
 - Blocks
 - Constraints/Validators



**Red Hat
Developer**



**Red Hat
Hybrid Cloud**

Integration: Future Plans

Content Syndicated Patterns

- ▶ “Shared patterns” with embedded content for banners, footers, marketing content, etc.

Customer Portal

- ▶ Autobuilding product pages from the Data Lake (currently are Drupal nodes)
- ▶ Standardizing Product taxonomy across Customer Portal microsites
- ▶ Integrating with external systems (product life cycle, case management, developers.redhat.com, etc.)

Thank you

Questions?

Melissa Bent



[linkedin.com/in/melissabent](https://www.linkedin.com/in/melissabent)



twitter.com/merauluka

April Sides



[linkedin.com/in/aprilsides](https://www.linkedin.com/in/aprilsides)



twitter.com/weekbeforenext

Red Hat



[linkedin.com/company/red-hat](https://www.linkedin.com/company/red-hat)



[youtube.com/user/RedHatVideos](https://www.youtube.com/user/RedHatVideos)



[facebook.com/redhatinc](https://www.facebook.com/redhatinc)



twitter.com/RedHat