



Optimal Approaches for Real-Time Machine Learning with Apache Spark on Kubernetes: Best Practices and Strategies

Hichem Kenniche

Open-Source Software Product Architect

instaclustr
Now part of **Spot by NetApp**

A few words about me!

Hichem Kenniche,

OSS Product Architect @Instaclustr (part of NetApp)

Previously at:

- Databricks
- Capgemini Invent
- Sony PlayStation

Disclaimer

- This is not a contribution to any OSS project !
- My vision of things is necessarily biased !
- Most of this work is based on the principles of OSS, open data, and a culture of knowledge sharing ❤️
- (Human) Learning is a Lifelong WIP ...

Our agenda for today

- The fundamentals of Real-Time ML.
- The biggest challenges facing Data teams.
- The motivation behind running Spark on Kubernetes
- Some challenges of Running Spark on Kubernetes and solutions
- Conclusion and takeaways

The fundamentals of Real-Time ML.

Finding the shortest, fastest Cycling Route

OpenStreetMap Modifier Historique Exporter

Traces GPS Journaux des utilisateurs Communautés Droits d'auteur Aide À propos Se connecter S'inscrire

88, Rue de Paris, Cité des Sentes, Les Lilas, I
Esplanade de La Défense, Tunnel de la Défer
À vélo (GraphHopper) Aller
Inverser les directions

Itinéraire

Distance: 14 km. Temps: 0:48.
Croissant: 164 m. Décroissant: 246 m.

- ↑ 1. Continuez sur Rue de Paris 200m
- ↑ 2. Restez sur la droite sur Rue de Paris 110m
- ↑ 3. Restez sur la droite sur Avenue de la Porte des Lilas 200m
- ↗ 4. Tournez légèrement à droite sur Avenue de la Porte des Lilas 100m
- ↘ 5. Tournez fort à droite 400m
- ↗ 6. Tournez légèrement à droite sur Avenue de la Porte du Pré Saint-Gervais 70m
- ↶ 7. Tournez à gauche 10m
- ↶ 8. Tournez fort à gauche sur Avenue de la Porte du Pré Saint-Gervais 50m

Couches de carte

- Standard
- CyclOSM
- Carte cyclable
- Carte de transport
- Topo de Tracestack
- Humanitaire

Autoriser les superpositions pour réparer la carte

- Notes de carte
- Données de carte
- Traces GPS publiques

© Les Contributeurs d'OpenStreetMap. Style de carreaux par CyclOSM hébergé par OpenStreetMap France. Conditions d'utilisation du site web et de l'API

Finding the shortest, fastest, least traffic?

OpenStreetMap Modifier Historique Exporter

Traces GPS Journaux des utilisateurs Communautés Droits d'auteur Aide À propos Se connecter S'inscrire

88, Rue de Paris, Cité des Sentes, Les Lilas, I
Esplanade de La Défense, Tunnel de la Défer
À vélo (GraphHopper) Aller
Inverser les directions

Itinéraire

Distance: 14 km. Temps: 0:48.
Croissant: 164 m. Décroissant: 246 m.

- ↑ 1. Continuez sur Rue de Paris 200m
- ↑ 2. Restez sur la droite sur Rue de Paris 110m
- ↑ 3. Restez sur la droite sur Avenue de la Porte des Lilas 200m
- ↗ 4. Tournez légèrement à droite sur Avenue de la Porte des Lilas 100m
- ↘ 5. Tournez fort à droite 400m
- ↗ 6. Tournez légèrement à droite sur Avenue de la Porte du Pré Saint-Gervais 70m
- ↖ 7. Tournez à gauche 10m
- ↙ 8. Tournez fort à gauche sur Avenue de la Porte du Pré Saint-Gervais 50m

Couches de carte

- Standard
- CyclOSM
- Carte cyclable
- Carte de transport
- Topo de Tracestack
- Humanitaire

Autoriser les superpositions pour réparer la carte

- Notes de carte
- Données de carte
- Traces GPS publiques

© Les Contributeurs d'OpenStreetMap. Style de carreaux par CyclOSM hébergé par OpenStreetMap France. Conditions d'utilisation du site web et de l'API

Finding the shortest, fastest, least traffic!

Adresses enregistrées...

Recherches récentes

NetApp 55 min

NetApp

88 Rue de Paris

55 min (14,8 km)

via Av. du Roule
↑ 62 m · ↓ 130 m

114 m —
28 m

⚠ Attention : Les itinéraires à vélo peuvent ne pas toujours tenir compte des conditions réelles

88 Rue de Paris
93260 Les Lilas

- ↑ Prendre la direction sud-ouest sur Rue de Paris/D117 vers Rue du Coq Français
150 m
- ➡ Prendre à droite sur Rue du Pré Saint-Gervais/D20
📍 Continuer de suivre D20
170 m
- ⬅ Prendre à gauche sur Av. du Belvédère
400 m
- ↑ Continuer sur Rue Alexander Fleming
400 m
- ↑ Continuer sur Rue Sigmund Freud
600 m
- ⬅ Prendre à gauche sur Av. de la Prte Chaumont/D35BIS
📍 Traverser le rond-point
290 m
- ➡ Au rond-point, prendre la 2e sortie sur Rue Manin


The collage consists of several elements:

- Top Left:** A map snippet showing a route through Paris, with a red location pin and a blue 'OUVRIR' button.
- Top Center:** A photo of a man in a dark suit standing in a heavy traffic jam on a multi-lane road.
- Top Right:** A photo of a black car that has crashed into a tree, with significant damage to the front and side.
- Bottom Left:** A larger map of Paris showing a route through the city, with various landmarks and districts labeled.
- Bottom Center:** A night-time photo of a man in a dark suit standing on a sidewalk next to a street with traffic lights and buildings.
- Bottom Right:** Another map snippet of Paris, showing a different route through the city, with a 'Zoomer' and 'Afficher le curseur' control overlay.

Finding the shortest, fastest, least traffic!

Adresses enregistrées

55 min (14,8 km)
via Av. du Roule
↑ 62 m · ↓ 130 m



Recherches récentes

NetApp 55 min

NetApp

88 Rue de Paris

93260 Les Lilas

- ↑ Prendre la direction sud-ouest sur Rue de Paris/D117 vers Rue du Coq Français
150 m
- ↪ Prendre à droite sur Rue du Pré Saint-Gervais/D20
Continuer de suivre D20
170 m
- ↶ Prendre à gauche sur Av. du Belvédère
400 m
- ↑ Continuer sur Rue Alexander Fleming
400 m
- ↑ Continuer sur Rue Sigmund Freud
600 m
- ↶ Prendre à gauche sur Av. de la Prte Chaumont/D35BIS
Traverser le rond-point
290 m
- ↪ Au rond-point, prendre la 2e sortie sur Rue Manin

⚠ Attention : Les itinéraires à vélo peuvent ne pas toujours tenir compte des conditions réelles

Johanne Original Poster

Aug 14, 2020

How does [redacted] calculate the shortest / preferred cycling route in the Netherlands?

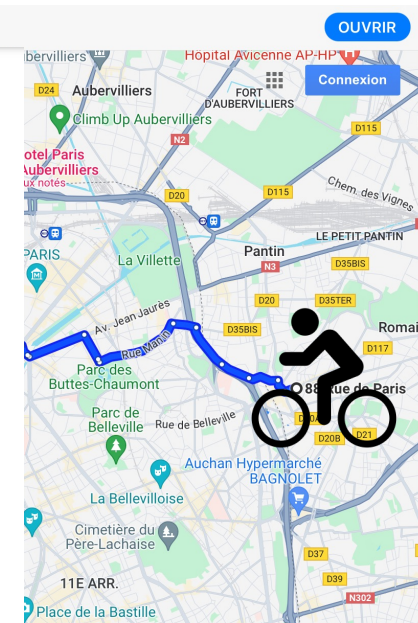
I live in the Netherlands and I often use [redacted] it to look up where I'm going and what route I need to take. It almost never gives me the shortest or best route; [redacted] is -at least in the Netherlands- extremely bad at showing cycling routes. I wonder why? Calculating the shortest route by taking all available bike paths/roads and getting the shortest within the network should be a very easy thing to do.

Often when having been shown a route, I will then proceed to -while on my computer- drag it around a bit and often decrease the distance by about 10% of the total distance. While doing this I also see that it actually recognises these roads as cycling paths, but just doesn't use or suggest them.

I also notice that it there are many roads [redacted] NEVER suggested by [redacted]. So I'm [redacted] roads, since they are actually the most su [redacted] been there for many years to have been n

I'm just really curious how [redacted] ge [redacted] that, how I can help to improve them? Per [redacted] behind, or maybe it has to do with the unc

Details
[Directions and Navigation, Desktop - Other](#)



S Saket Pofali

Sep 4, 2020

I've noticed this as well. I think [redacted] algorithms still based on recognizing roads and streets meant for motorized vehicular traffic which would mean that there would be minimum road width would be required to be recognized as legitimate street.

Finding the Least Air Pollution Exposure Cycling Routes

OpenStreetMap Modifier Historique Exporter

Traces GPS Journaux des utilisateurs Communautés Droits d'auteur Aide À propos Se connecter S'inscrire

88, Rue de Paris, Cité des Sentes, Les Lilas, I
Esplanade de La Défense, Tunnel de la Défer
À vélo (GraphHopper) Aller
Inverser les directions

Itinéraire

Distance: 14 km. Temps: 0:48.
Croissant: 164 m. Décroissant: 246 m.

- ↑ 1. Continuez sur Rue de Paris 200m
- ↑ 2. Restez sur la droite sur Rue de Paris 110m
- ↑ 3. Restez sur la droite sur Avenue de la Porte des Lilas 200m
- ↗ 4. Tournez légèrement à droite sur Avenue de la Porte des Lilas 100m
- ↘ 5. Tournez fort à droite 400m
- ↗ 6. Tournez légèrement à droite sur Avenue de la Porte du Pré Saint-Gervais 70m
- ↖ 7. Tournez à gauche 10m
- ↙ 8. Tournez fort à gauche sur Avenue de la Porte du Pré Saint-Gervais 50m

Couches de carte

- Standard
- CyclOSM
- Carte cyclable
- Carte de transport
- Topo de Tracstack
- Humanitaire

Autoriser les superpositions pour réparer la carte

- Notes de carte
- Données de carte
- Traces GPS publiques

© Les Contributeurs d'OpenStreetMap. Style de carreaux par CyclOSM hébergé par OpenStreetMap France. Conditions d'utilisation du site web et de l'API

Finding the Least Air Pollution Exposure Cycling Routes

OpenStreetMap [Modifier](#) [Historique](#) [Exporter](#)

Air quality adjusted routing for cyclists and pedestrians

Authors: [Sebastian Müller](#), [Agnès Voisard](#) [Authors Info & Claims](#)

EM-GIS '15: Proceedings of the 1st ACM SIGSPATIAL International Workshop on the Use of GIS in Emergency Management • November 2015 • Article No.: 19 • Pages 1–6 • <https://doi.org/10.1145/2835596.2835609>

Published: 03 November 2015 [Publication History](#) [Check for updates](#)

4 116 [Get Access](#)



Itinéraire

- Distance: 14 km. Temp Croissant: 164 m. Dé
1. Continuez sur Avenue de la Porte des Lilas
 2. Restez sur la Avenue de la Porte des Lilas
 3. Restez sur la Avenue de la Porte des Lilas
 4. Tournez légèrement à droite sur Avenue de la Porte des Lilas
 5. Tournez fort à droite sur Avenue de la Porte des Lilas
 6. Tournez légèrement à droite sur Avenue de la Porte du Pré Saint-Gervais
 7. Tournez à gauche sur Avenue de la Porte du Pré Saint-Gervais
 8. Tournez fort à gauche sur Avenue de la Porte du Pré Saint-Gervais

ABSTRACT

Air quality adjusted routing for cyclists and pedestrians provides the amount of air pollution. The data is available on Open Source Routing Machine at geofabrik.de. Our application is a transportation mode profile.

EAI Endorsed Transactions on Scalable Information Systems

Research Article **EAI.EU**

Predicting the least air polluted path using the neural network approach

K.Krishna Rani Samal¹*, Korra Sathya Babu¹, Santos Kumar Das¹

¹National Institute of Technology, Rourkela, India

Abstract

Air pollution exposure during daily transportation is becoming a critical issue worldwide due to its adverse effect on human health. Predicting the least air polluted healthier path is the best alternative way to mitigate personal air pollution exposure risk. Computing the least polluted path for the current time might not be helpful for real-time applications. Therefore, we develop a routing algorithm based on a neural network-based CNN-LSTM-EBK (CLE), a temporal-spatial interpolation model. The proposed model predicts pollution levels at high temporal granularity. This paper introduces a weight function to compute air pollution concentration at the road network. It also predicts the least air polluted path among all possible paths from a source to a destination at different time granularity. The results show that the predicted path may be longer than the shortest route but minimize pollution exposure risk all the time, which proves its effectiveness during daily transportation.

Received on 06 April 2021; accepted on 19 June 2021; published on 29 June 2021


Keywords: Air quality modelling, Routing, Deep learning, GIS, Kriging

Copyright © 2021 K.Krishna Rani Samal *et al.*, licensed to EAI. This is an open access article distributed under the terms of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>), which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

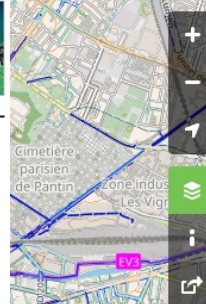
doi:10.4108/eai.29-6-2021.170250

UKIEAI - Direction des routes de France (UKIEAI) - www.sytadin.fr

Traces GPS Journaux des utilisateurs Communautés Droits d'auteur Aide À propos [Se connecter](#) [S'inscrire](#)



Journal of Transport & Health
Volume 26, September 2022, 101410



Health-oriented routes for active mobility

Paulo J.G. Ribeiro, Gabriel J.C. Dias, José F.G. Mendes

Show more

Real-time Route Planning using Mobile Air Pollution Detectors and Citizen Scientists

Publisher: IEEE [Cite This](#) [PDF](#)

Richard O. Sinnott; Yuan Wang; Yiqun Wang **All Authors**

79 Full Text Views

- Abstract
 - Document Sections
 - I. Introduction (Heading 1)
 - II. Related Work
 - III. Data Collection
 - IV. Implementation of the Routing Algorithm
 - V. Discussion
- Show Full Outline

Abstract:

The increasing urbanization of society is resulting in numerous challenges. One of these challenges is transport congestion and the associated increase in pollution that is widely accepted as driving global warming. For many individuals and especially those with respiratory issues, e.g., those with asthma or chronic obstructive pulmonary disease, high levels of pollution can cause direct health events. The ability to measure pollution accurately in real time at disaggregated levels and subsequently avoid pollution hotspots is thus highly desirable. This paper describes a Cloud-based infrastructure and associated mobile application that utilizes real time, mobile pollution measurement technology to help individuals avoid pollution hotspots through real-time pollution-aware routing algorithms.

Published in: 2021 17th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)

Date of Conference: 11-13 October 2021 **DOI:** 10.1109/WiMob52687.2021.9606267

Couches de carte

- Standard
- CyclOSM
- Itinéraire

les superpositions pour la carte

des de carte

nées de carte

es GPS publiques

Finding the Least Air Pollution Exposure Routes (in real time)

OpenStreetMap Modifier Historique Exporter

Traces GPS Journaux des utilisateurs Communautés Droits d'auteur Aide À propos Se connecter S'inscrire

Couches de carte

- Standard
- CyclOSM
- Carte cyclable
- Carte de transport
- Topo de Tracstack
- Humanitaire

Itinéraire

Distance: 14 km. Temps: 0:48.
Croissant: 164 m. Décroissant: 246 m.

1. Continuez sur Rue de Paris
2. Restez sur la droite sur Rue de Paris
3. Restez sur la droite sur Avenue de la Porte des Lilas
4. Tournez légèrement à droite sur Avenue de la Porte des Lilas
5. Tournez fort à droite
6. Tournez légèrement à droite sur Avenue de la Porte du Pré Saint-Gervais
7. Tournez à gauche
8. Tournez fort à gauche sur Avenue de la Porte du Pré Saint-Gervais

Carte temps réel

Les vitesses Les bouchons Viabilité hivernale

Etat des bouchons le mardi 05 mars 2024 à 17:20

La pollution en direct en Île-de-France

Saisissez une adresse ou cliquez sur la carte

05/03/2024 17:00:00

Qualité de l'air Dégradée

Concentration par polluants :

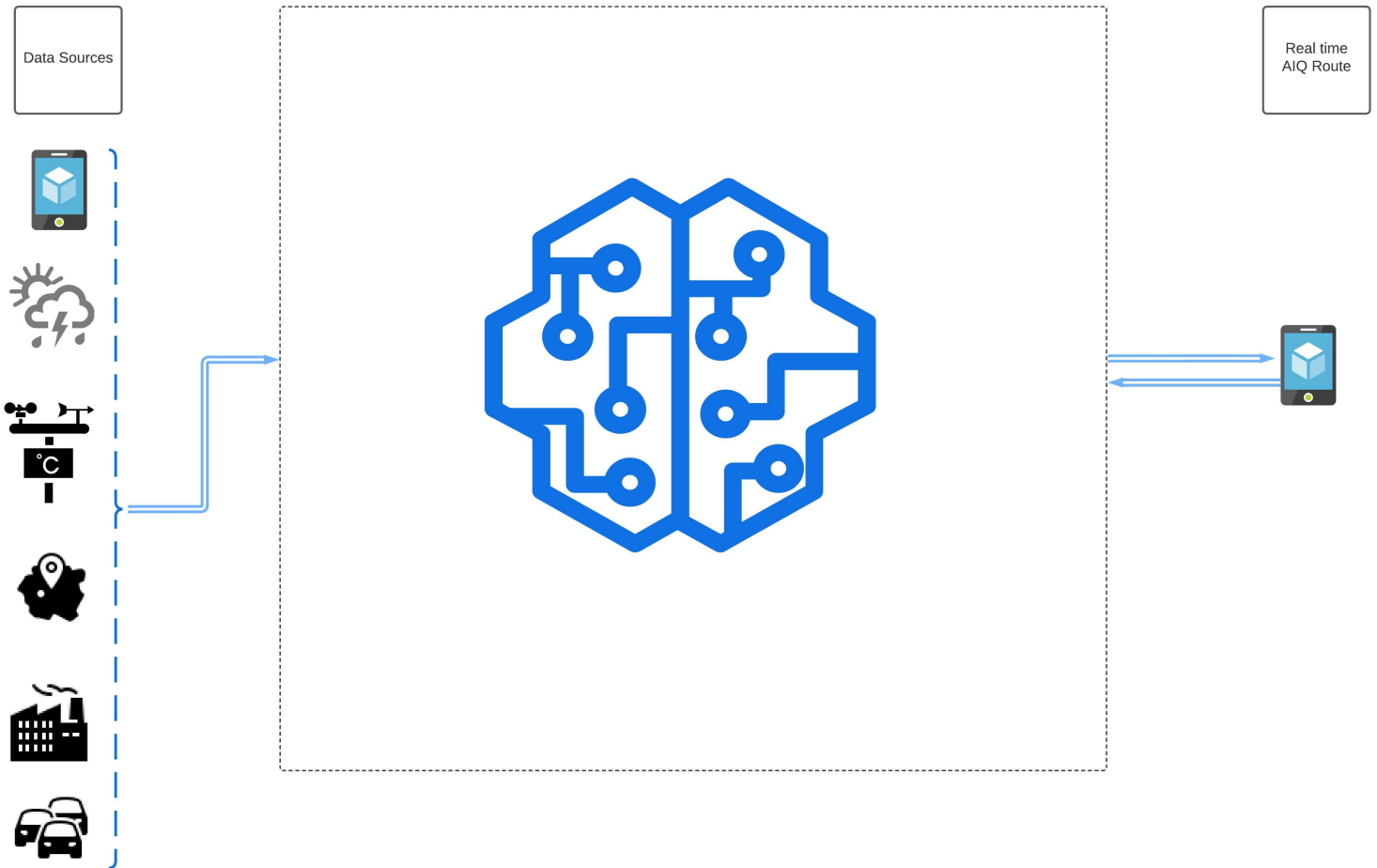
Ozone (O ₃)	45 µg/m ³
Dioxyde d'Azote (NO ₂)	92 µg/m ³
Particules (diamètre inférieur à 10µm) (PM ₁₀)	28 µg/m ³
Particules fines (diamètre inférieur à 2,5µm) (PM _{2,5})	18 µg/m ³

Qualité de l'air

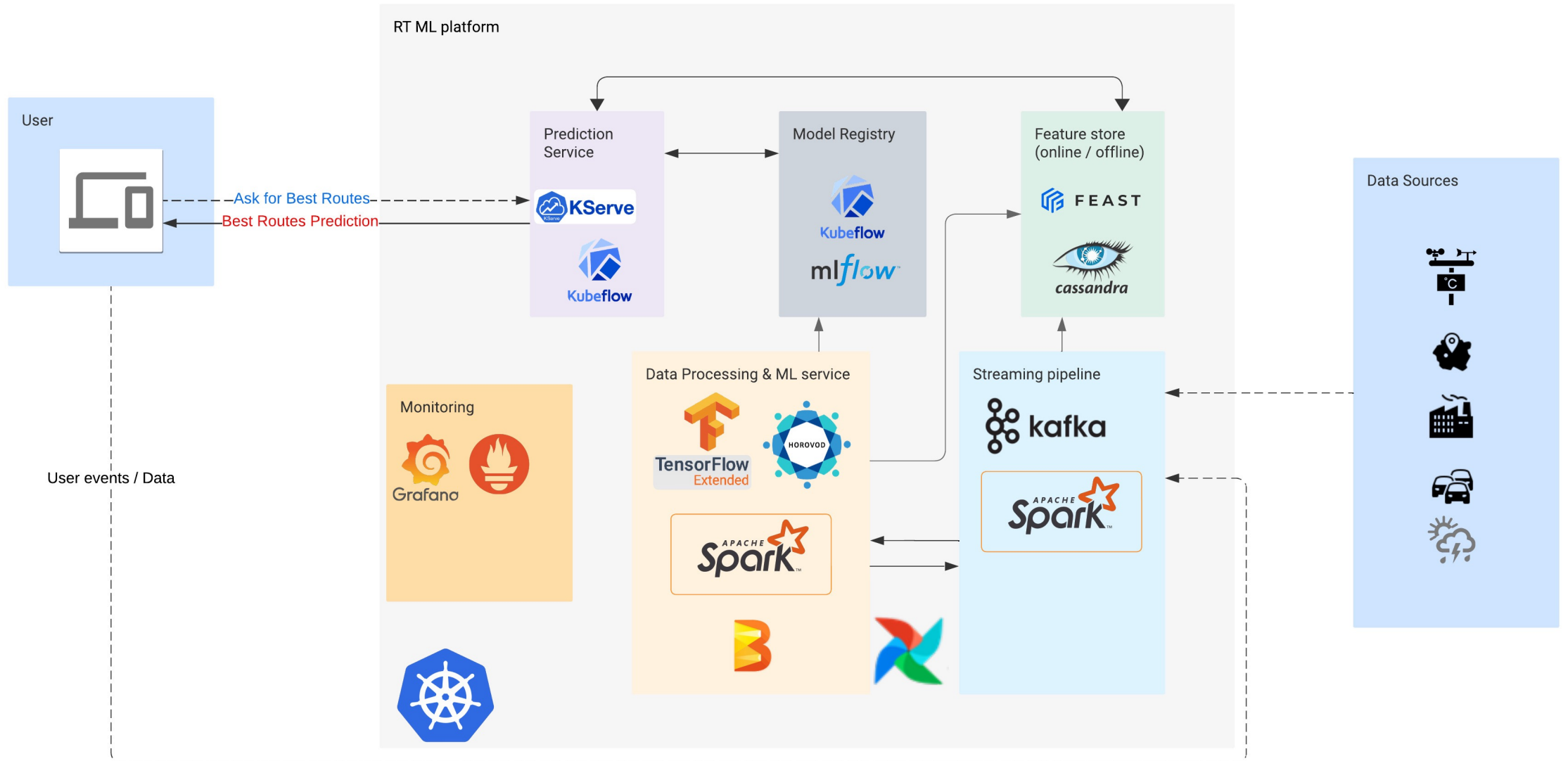
Bonne Moyenne Dégradée Mauvaise Très mauvaise Extrêmement mauvaise

Finding the Least Air Pollution Exposure Routes (in real time)

Real-time machine learning: the application of machine learning models to generate predictions or decisions in real-time and adapt to the changing environment.



Real-time Machine Learning Platform



The biggest challenges
facing Data teams.

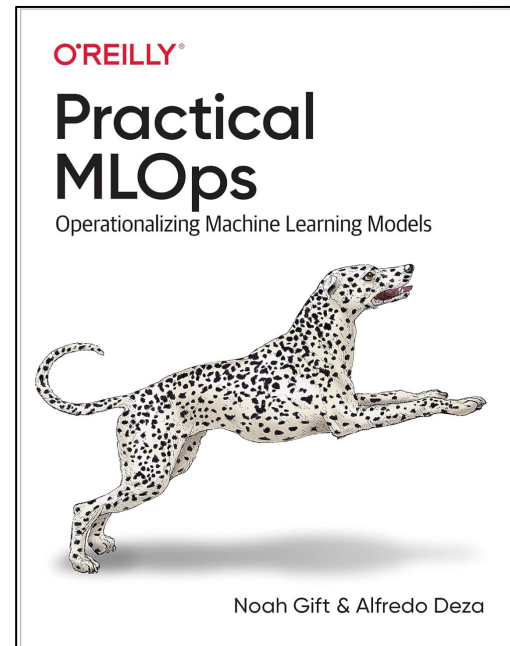
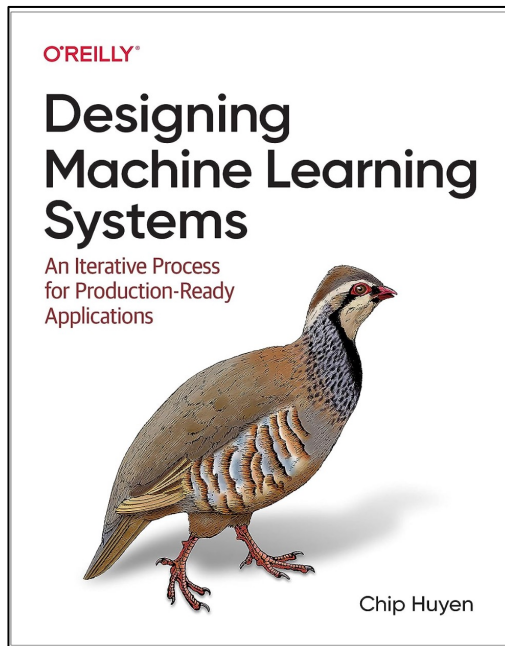
Real-time Machine Learning Challenges

Chip Huyen

[Blog](#) [Books](#) [List 100](#) [MLOps Guide](#) [Tiếng Việt](#)

Real-time machine learning: challenges and solutions

Jan 2, 2022 • Chip Huyen



Feature Engineering

Incremental Learning
(online learning)

Model Updating

Model / Data Drift

Performance Evaluation

MLOps

Stream Processing

Scalability

Latency

Monitoring

Distributed Training &
Inference

Resource Management
/ Cost

Real-time Machine Learning Challenges

Real-time machine learning challenges (our experience) are largely an infrastructure problem.

Stream Processing

Scalability

Latency

Monitoring

**Distributed Training &
Inference**

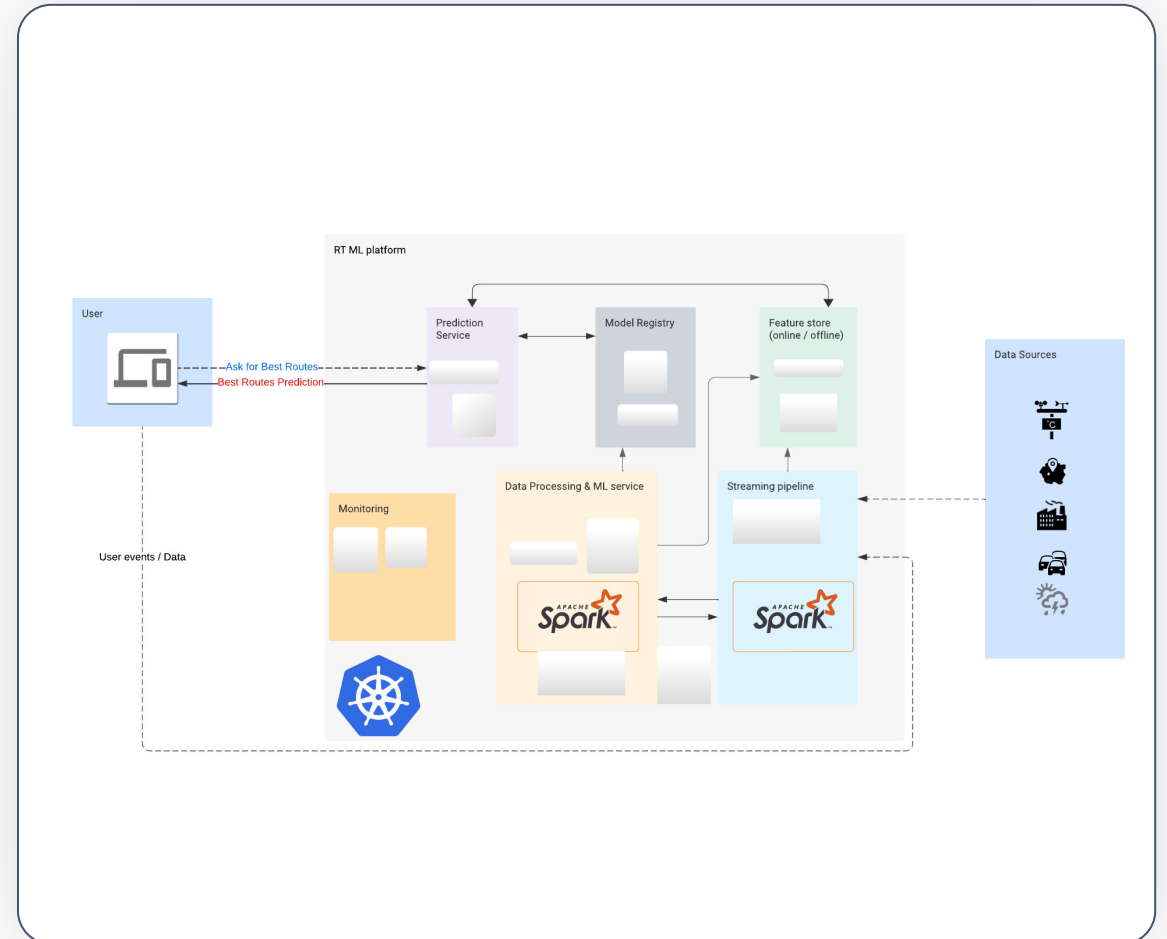
**Resource Management
/ Cost**

Solving some Real-time Machine Learning Challenges

Addressing these challenges requires a significant investment in advanced (OSS) technologies.

Spark on k8s:

- Stream processing
- Training
- Scalability & Latency
- Resource Efficiency



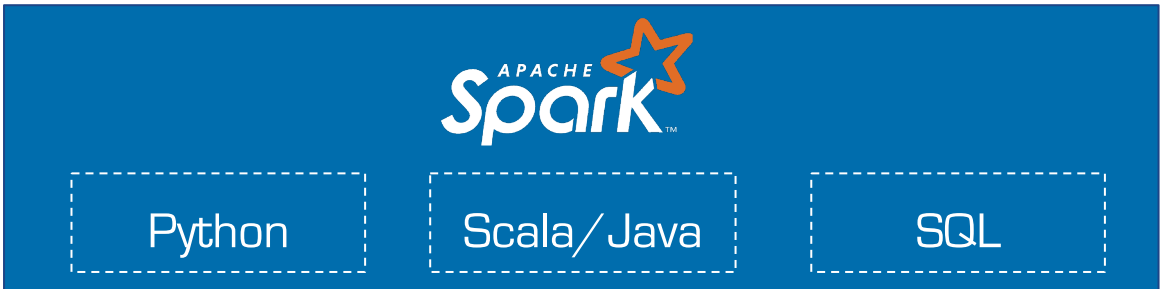
The motivation behind running Spark on kubernetes

Apache Spark is the #1 analytics engine for Big Data & AI

Fast*
Massively parallelizable, efficient read and write



Easy
Interfaces with well-known programming languages

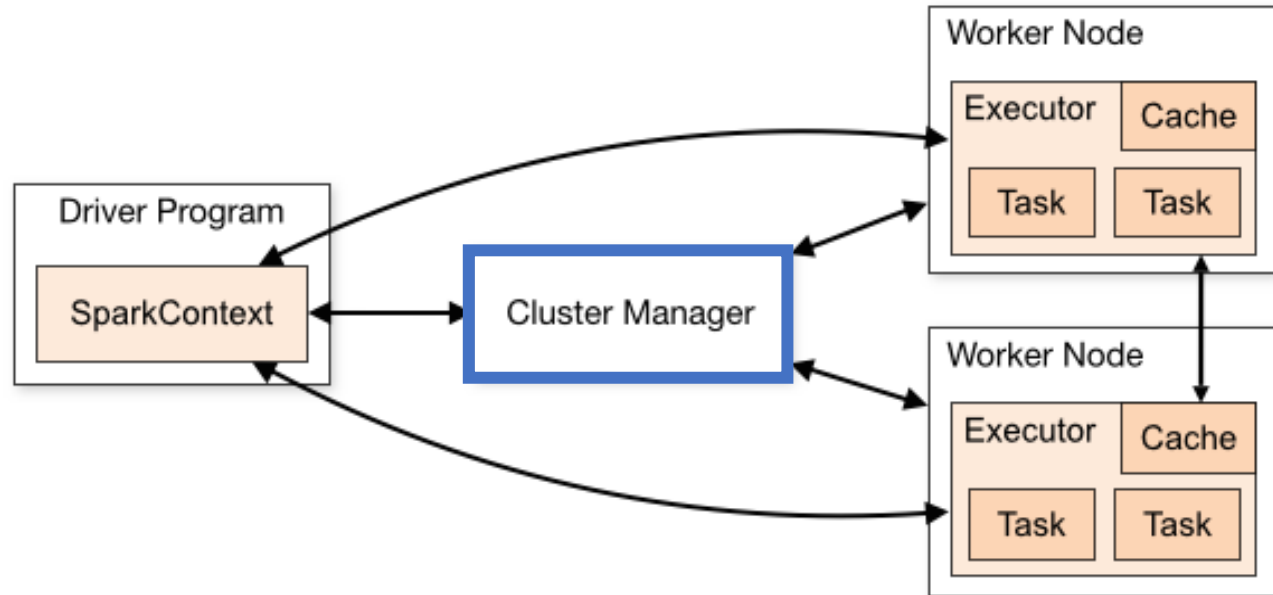


Versatile
Across multiple use cases



The role of resource manager in a Spark cluster

Spark depends on cluster manager for orchestration of a job on a cluster



Kubernetes is the latest cluster manager for Spark

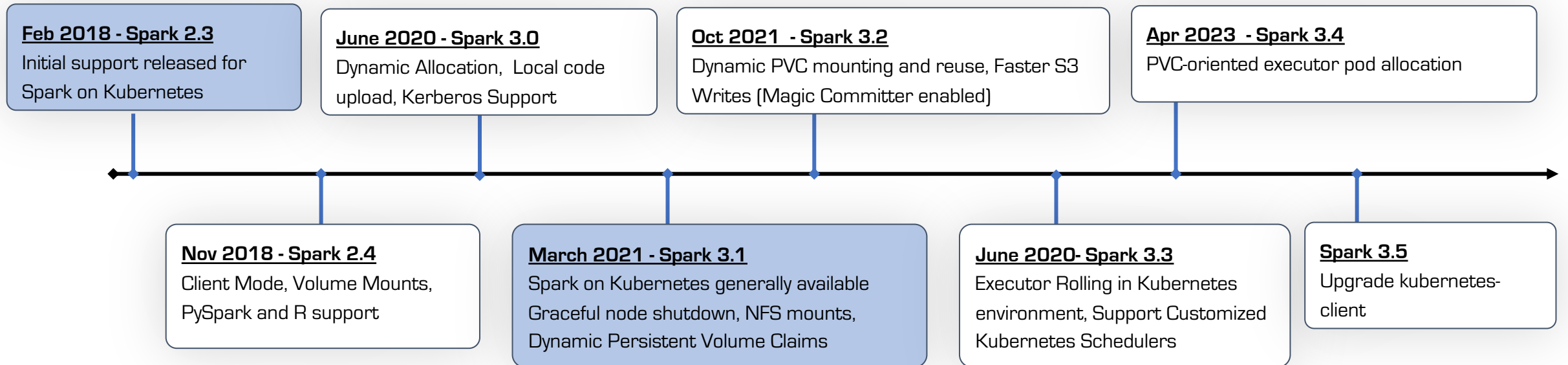
Standalone: built-in, limited functionalities

Apache Mesos: deprecated as of Spark 3.2.0

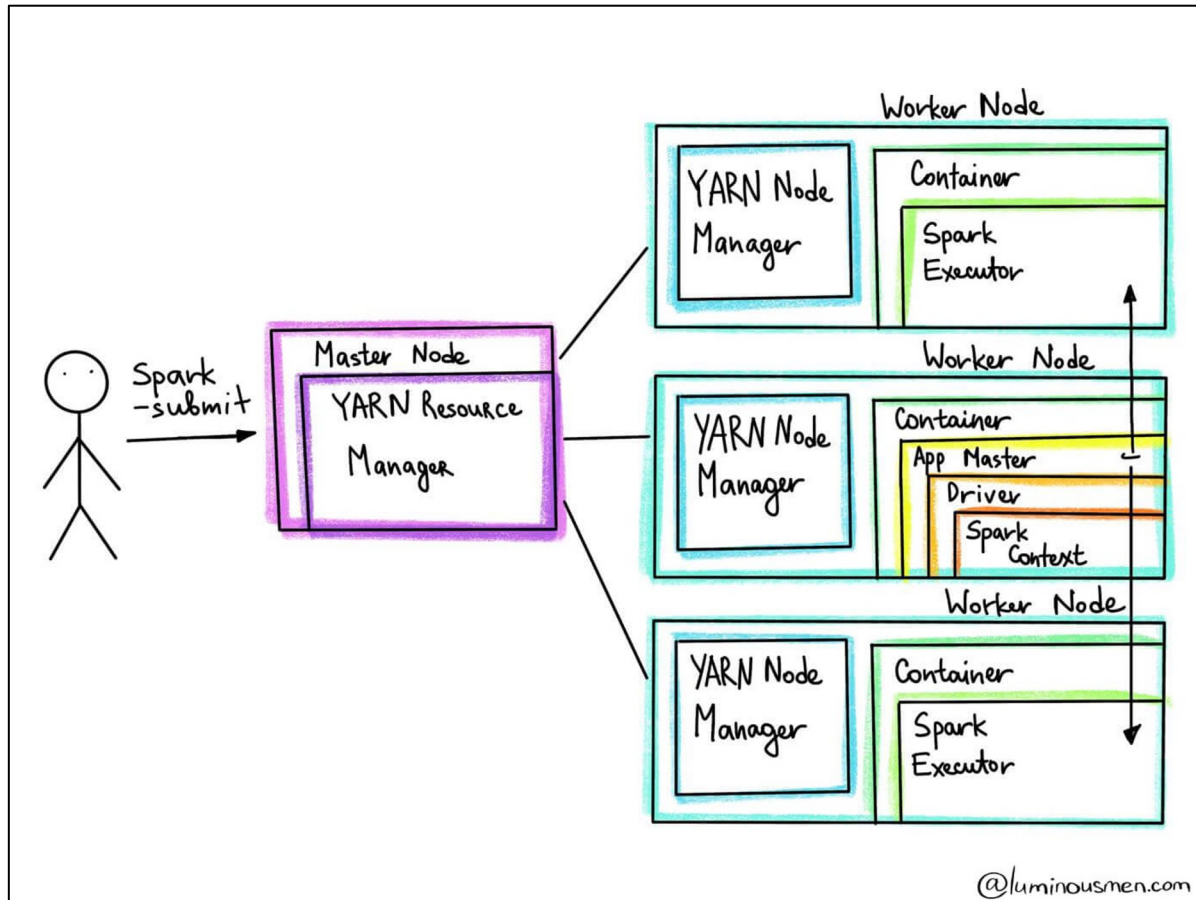
Hadoop YARN: most widely used today

Kubernetes: most popular among new deployments

The Spark on Kubernetes Journey



Spark on YARN: architecture & pain points



Global Spark version and shared libraries

- You'll have a Spark 2.4 cluster, a Spark 3.0 cluster, a Spark 3.1 cluster.
- Transient clusters are recommended for stability, but increase costs.

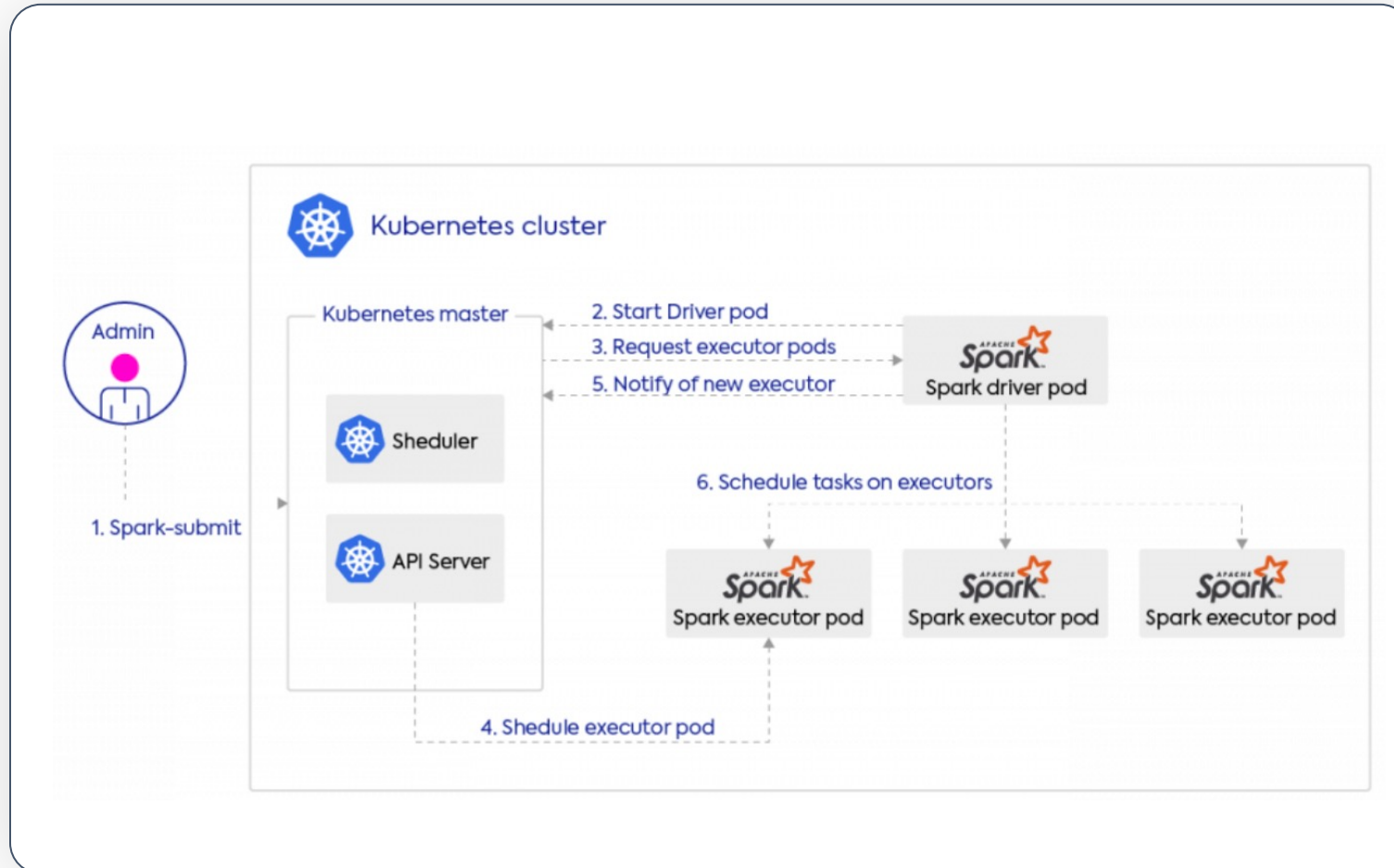
Limited Docker image support*

- Environment is built from AMIs and bash scripts, flaky runtime library installation
- Debugging is painful - there's no way to run Spark locally, environment is subtle

Resource Overhead

- Slow startup time
- YARN master node, YARN Node Mgr are JVM processes using a lot of resources.

Spark on Kubernetes: architecture & benefits



Native Dockerization

- Simpler dependency management
- Reliable executions across environments (locally during development, staging, production)
- Faster startup time

A single long-running cluster

- Quick to scale up (and down) based on load
- Mix different Spark versions
- Mix Spark and non-Spark apps
- Mix use cases (notebooks, batch/streaming jobs)

A standard, agnostic infrastructure layer

- Reduce lock in
- Simplify your operations
- Leverage the open-source tools from the cloud-native ecosystem

Two ways to run Spark apps on k8s

Spark-submit

- “Vanilla” way from Spark main open source repo
- Requires Spark distribution on client
- Configs spread between Spark config (mostly) and k8s manifests
- Less pod customization support (improving)
- App management is more manual

spark-on-k8s operator

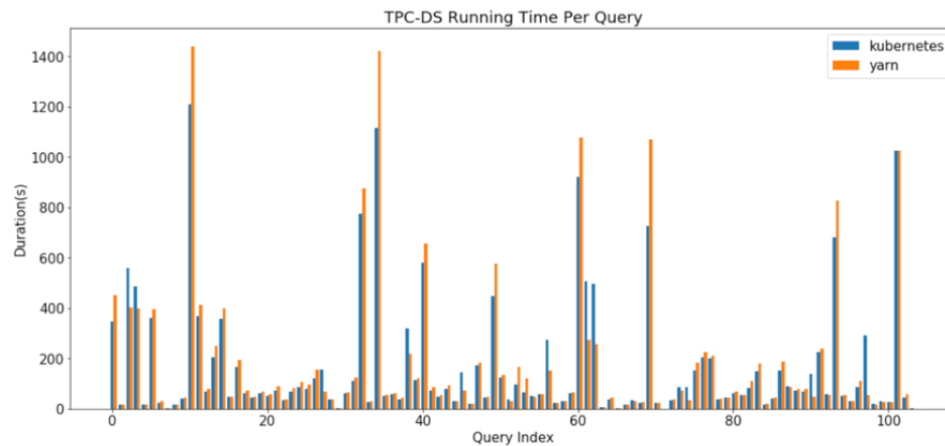
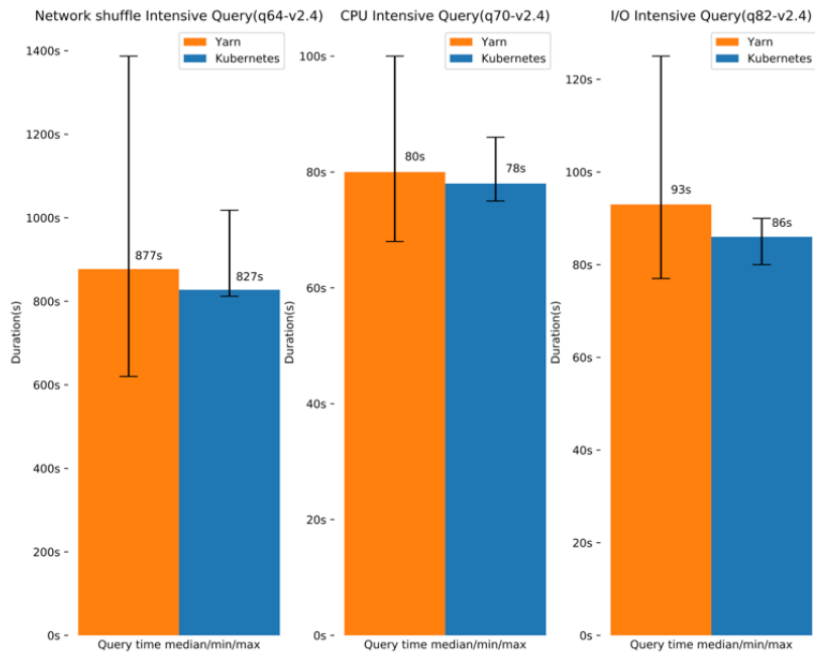
Overview

The Kubernetes Operator for Apache Spark aims to make specifying and running [Spark](#) applications as easy and idiomatic as running other workloads on Kubernetes. It uses [Kubernetes custom resources](#) for specifying, running, and surfacing status of Spark applications. For a complete reference of the custom resource definitions, please refer to the [API Definition](#). For details on its design, please refer to the [design doc](#). It requires Spark 2.3 and above that supports Kubernetes as a native scheduler backend.

The Kubernetes Operator for Apache Spark currently supports the following list of features:

- Supports Spark 2.3 and up.
- Enables declarative application specification and management of applications through custom resources.
- Automatically runs `spark-submit` on behalf of users for each `SparkApplication` eligible for submission.
- Provides native [cron](#) support for running scheduled applications.
- Supports customization of Spark pods beyond what Spark natively is able to do through the mutating admission webhook, e.g., mounting ConfigMaps and volumes, and setting pod affinity/anti-affinity.
- Supports automatic application re-submission for updated `SparkApplication` objects with updated specification.
- Supports automatic application restart with a configurable restart policy.
- Supports automatic retries of failed submissions with optional linear back-off.
- Supports mounting local Hadoop configuration as a Kubernetes ConfigMap automatically via `sparkctl`.
- Supports automatically staging local application dependencies to Google Cloud Storage (GCS) via `sparkctl`.
- Supports collecting and exporting application-level metrics and driver/executor metrics to Prometheus.

TPC-DS 1T Benchmark

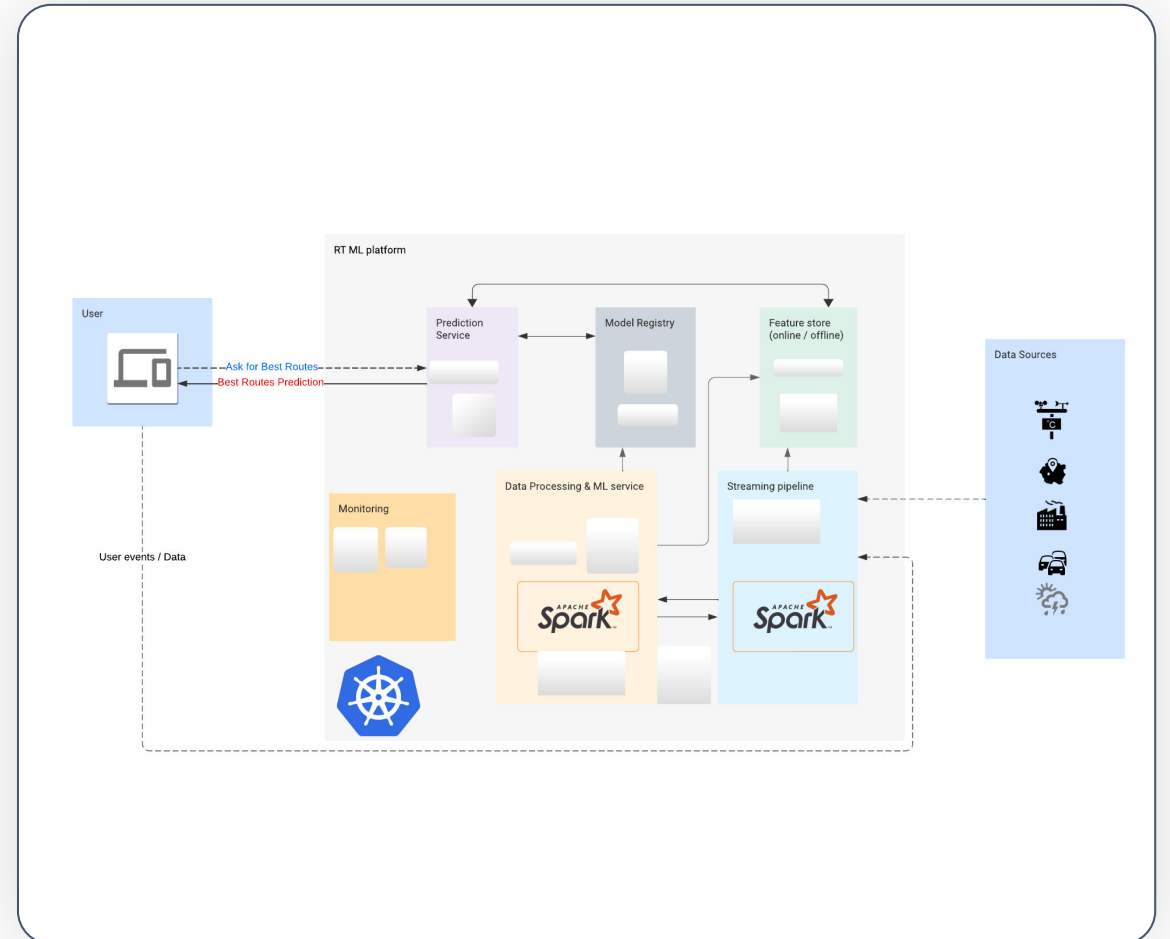


<https://aws.amazon.com/fr/blogs/containers/optimizing-spark-performance-on-kubernetes/>

Some challenges of running Spark on kubernetes and solutions

Challenges in the context of R-L M-L

- Monitoring
- Scalability
- Latency
- Models Training



Monitoring: logs, logs and more logs

Key information is buried under a lot of noisy one.

- Spark event/driver/executor logs.
- kubernetes logs
- Hard to reconcile with Spark jobs/stages/tasks

Solution

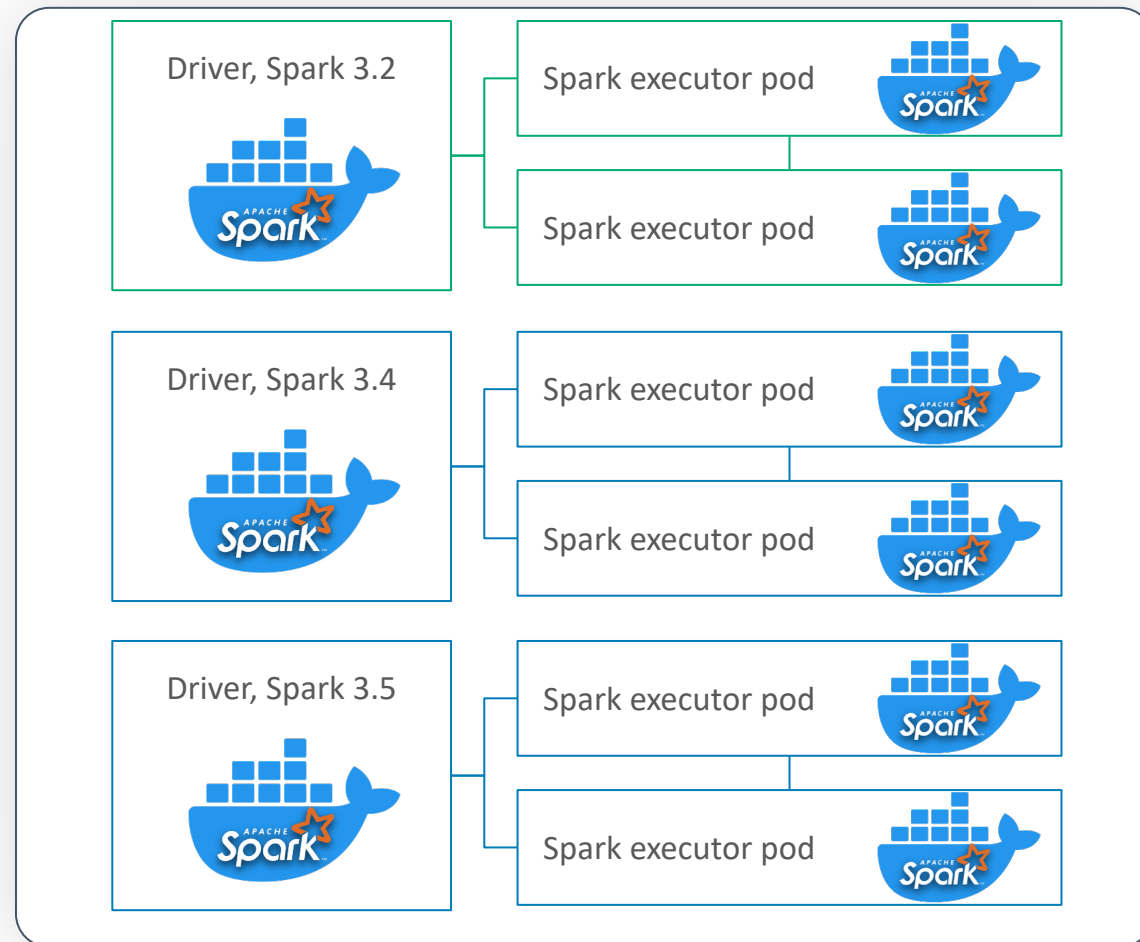
- Logs shipping tools : fluentbit & logstash
- Prometheus: Spark has a built-in Prometheus



Scalability

Key factors to consider

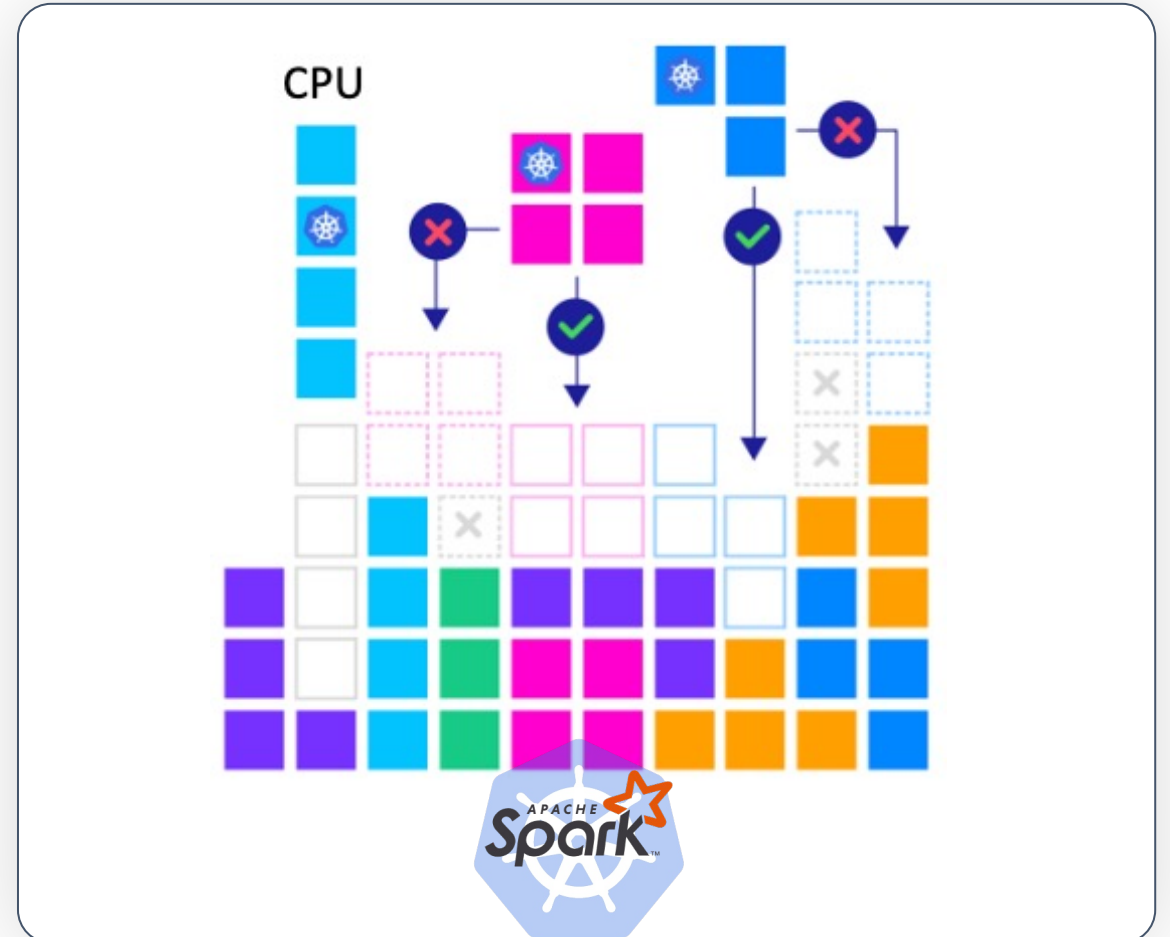
- Cluster sizing, infra choice/ specs.
- Dynamic Allocation
- Shuffle data (NO external shuffle service YET)



Scalability: the right sizing

For the sizing:

- Continuous and repeated exercise: know your data sources
- When selecting the cluster focus on enhancing parallelism in relation to the source.
- Streaming is CPU-bounded, State matters too (avoid spills)
- Deep Learning models with relatively long training and inference time: mix CPU with GPU (when required).



Scalability: Dynamic Allocation (A two-sided problem)

Dynamic Allocation in Spark Structured Streaming

- Designed for batch jobs, it is compatible with batch and Spark structured streaming. Works poorly for certain applications !

Dynamic Allocation (within k8s)

- This feature **may** cause issues with Spark Scalability on k8s !

The screenshot shows a JIRA issue page for Spark. The title is "[SPARK-24815] Structured Streaming should support dynamic allocation". The issue is categorized as an "Improvement" with a "Minor" priority, affecting Spark version 2.3.1. It is currently in an "OPEN" status and is "Unresolved". The issue is reported by Karthik Palaniappan and has 11 votes. It was created on 16/Jul/18 01:22 and last updated on 01/Mar/24 22:30. The description explains that dynamic allocation is useful for matching resources to workload in multi-tenant clusters but can cause issues with structured streaming. It lists three quick thoughts: 1) Dynamic allocation should be pluggable, 2) Structured streaming should have its own algorithm, and 3) Spark should warn when dynamic allocation is enabled during structured streaming.

[[SPARK-24815](#)] Structured Streaming should support dynamic allocation

Spark / SPARK-24815
Structured Streaming should support dynamic allocation

Details

Type:	Improvement	Status:	OPEN
Priority:	Minor	Resolution:	Unresolved
Affects Version/s:	2.3.1	Fix Version/s:	None
Component/s:	Scheduler, Spark Core, Structured Streaming		
Labels:	pull-request-available		

People

Assignee:	Unassigned
Reporter:	Karthik Palaniappan
Votes:	11 Vote for this issue
Watchers:	26 Start watching this issue

Dates

Created:	16/Jul/18 01:22
Updated:	01/Mar/24 22:30

Description

For batch jobs, dynamic allocation is very useful for adding and removing containers to match the actual workload. On multi-tenant clusters, it ensures that a Spark job is taking no more resources than necessary. In cloud environments, it enables autoscaling. However, if you set `spark.dynamicAllocation.enabled=true` and run a structured streaming job, the batch dynamic allocation algorithm kicks in. It requests more executors if the task backlog is a certain size, and removes executors if they idle for a certain period of time.

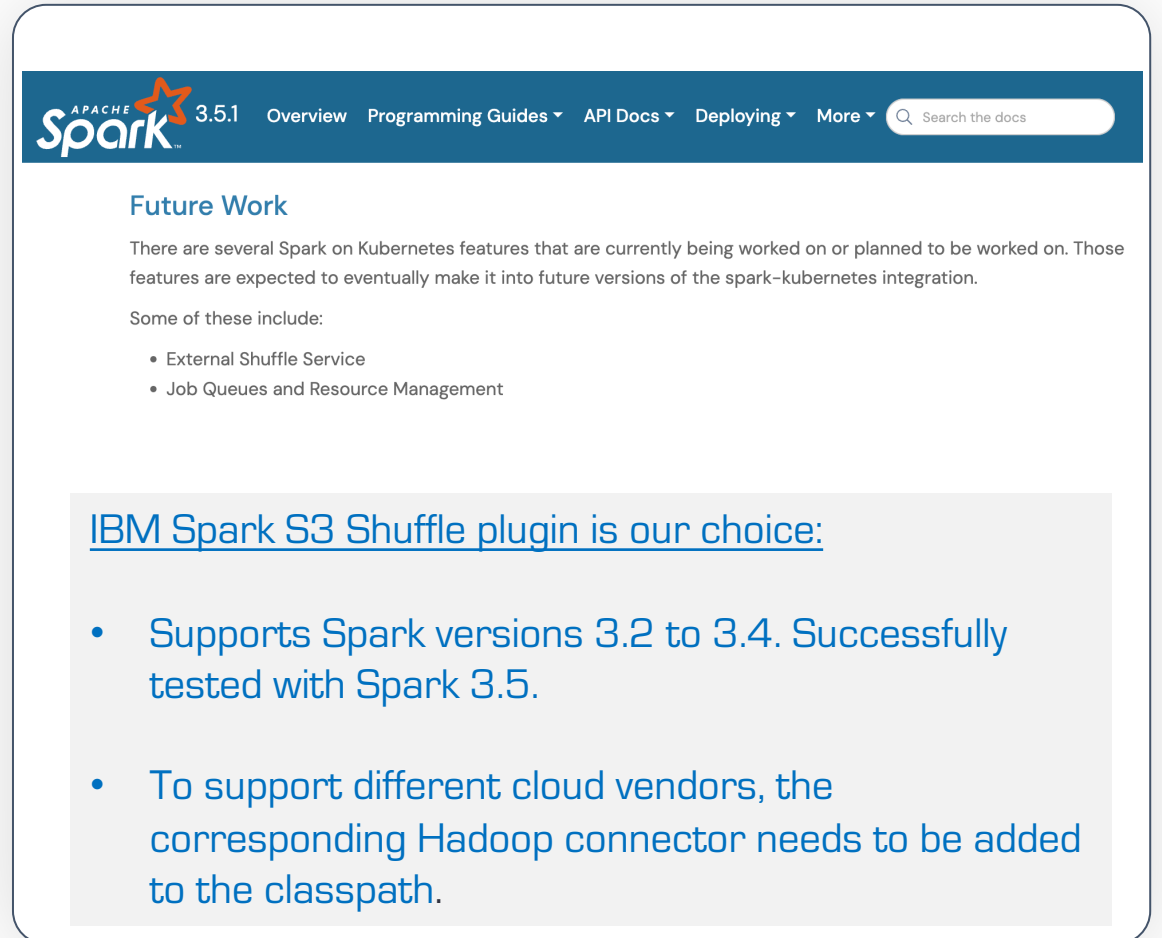
Quick thoughts:

- 1) Dynamic allocation should be pluggable, rather than hardcoded to a particular implementation in `SparkContext.scala` (this should be a separate JIRA).
- 2) We should make a structured streaming algorithm that's separate from the batch algorithm. Eventually, continuous processing might need its own algorithm.
- 3) Spark should print a warning if you run a structured streaming job when Core's dynamic allocation is enabled

Scalability: Shuffle data

External shuffle service for Spark on kubernetes is not supported yet. There are 4 options (+1):

- [Cloud Shuffle Storage Plugin for Apache Spark - AWS Glue](#)
- [IBM/spark-s3-shuffle: Shuffle plugin for Apache Spark and S3 compatible service](#)
- [GitHub - oap-project/remote-shuffle: Spark shuffle plugin for support shuffling data through a remote Hadoop-compatible file system](#) (Intel)
- [Apache Spark on Kubernetes - Local Storage](#) (main project)
- [AWS S3 CSI driver](#) and [High-Performance Storage – S3 Express One Zone, AWS](#)



The screenshot shows the Apache Spark 3.5.1 documentation page. The navigation bar includes links for Overview, Programming Guides, API Docs, Deploying, and More, along with a search box. The main content area is titled 'Future Work' and discusses several Spark on Kubernetes features currently being worked on or planned. It lists 'External Shuffle Service' and 'Job Queues and Resource Management' as examples. A highlighted section states that the IBM Spark S3 Shuffle plugin is the chosen option, supporting Spark versions 3.2 to 3.4 and tested with Spark 3.5. It also notes that to support different cloud vendors, the corresponding Hadoop connector needs to be added to the classpath.

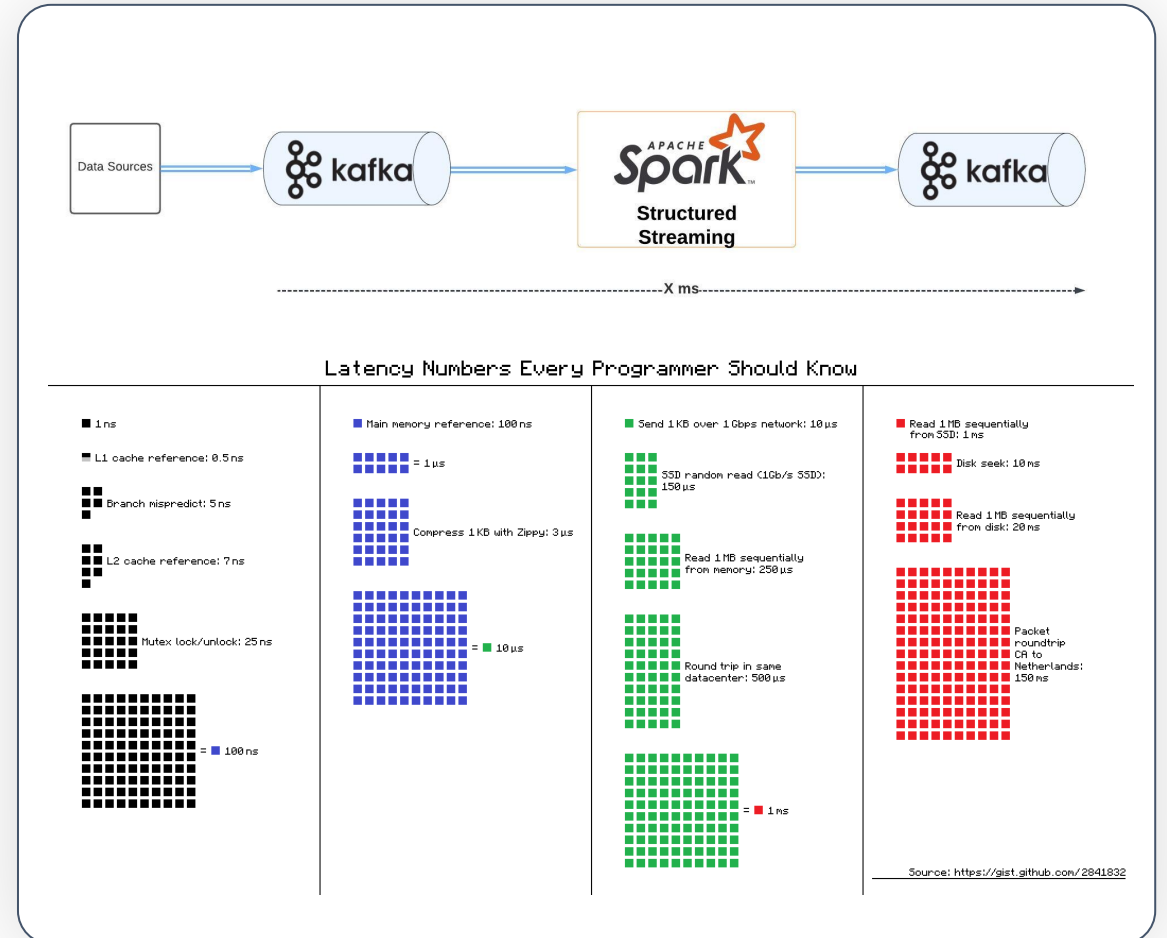
Latency

Key factors to consider

- Spark configurations.
- Sub-second latency expectations are a challenge
- Stateful vs Stateless Pipelines

Consider the following:

- Use only simple computations involving data transformation or enrichment !
- Always use a message bus (e.g., Apache Kafka or Apache Pulsar) and fast key-value stores (e.g., Apache Cassandra or Redis)
- RocksDB state store provider



M-L Training

Key factors to consider

- Most ML frameworks were designed for single-node environments
- Spark MLlib is lagging behind !

The screenshot shows the Apache Spark MLlib website and its documentation. The top section, titled "MLlib is Apache Spark's scalable machine learning library," includes a "Latest News" sidebar with release dates for Spark 3.5.1 (Feb 23, 2024), Spark 3.3.4 (Dec 16, 2023), Spark 3.4.2 (Nov 30, 2023), and Spark 3.5.0 (Sep 13, 2023). The main content is divided into four sections: "Ease of use" (usable in Java, Scala, Python, and R), "Performance" (100x faster than MapReduce), "Runs everywhere" (Hadoop, Mesos, Kubernetes), and "Built-in Libraries" (SQL, DataFrames, Streaming, MLlib, GraphX). A bar chart compares "Running time (s)" for Hadoop (110s) and Spark (0.9s) in a logistic regression task. The Hadoop logo is also visible. Below the website content is a snippet of the PySpark documentation for the `TorchDistributor` class, showing its parameters and version changes.

```
class pyspark.ml.torch.distributor.TorchDistributor(num_processes: int = 1, local_mode: bool = True, use_gpu: bool = True, ssl_conf: str = "pytorch.spark.distributor.ignoreSsl")
```

Parameters: `num_processes` : *int, optional*
An integer that determines how many different concurrent tasks are allowed. We expect `spark.task.gpus = 1` for GPU-enabled training. Default should be 1; we don't want to invoke multiple cores/gpus without explicit mention.

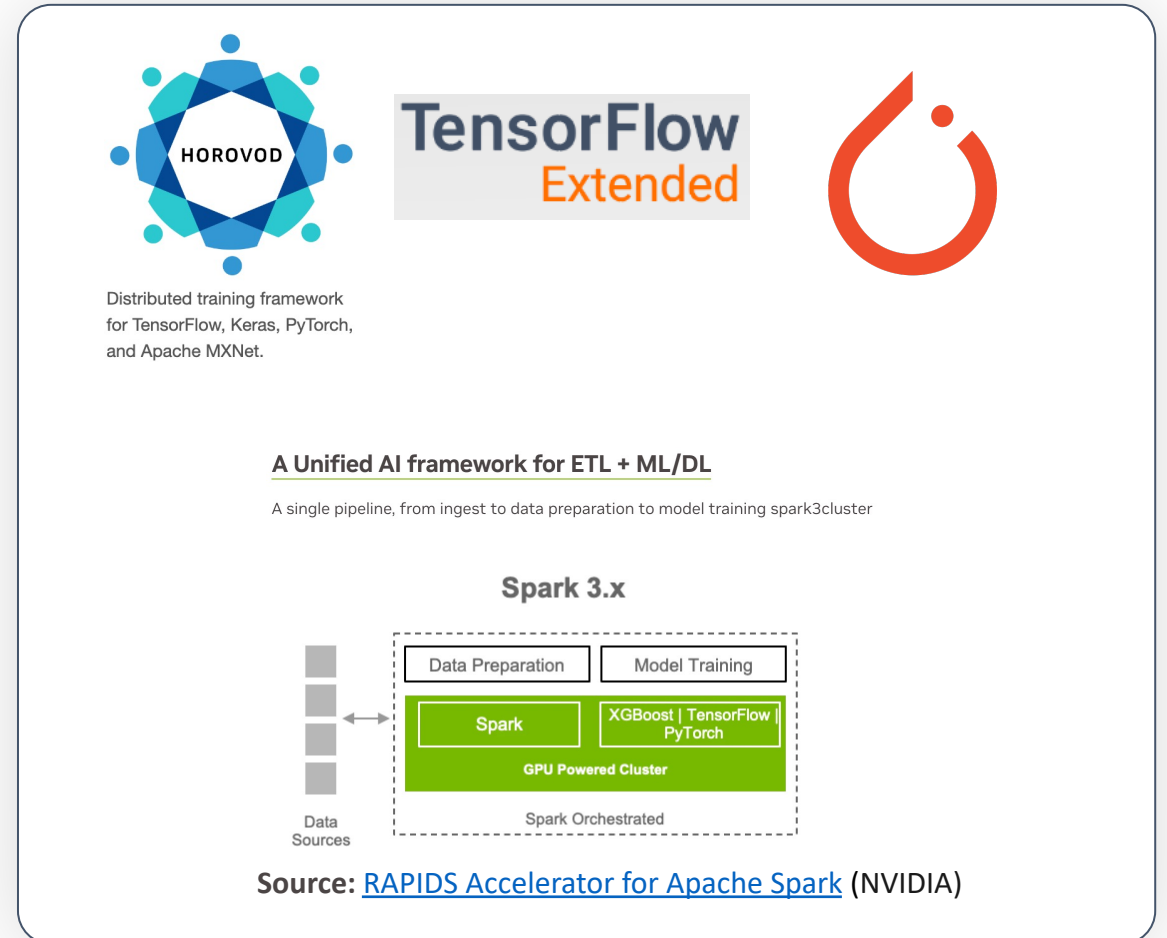
`local_mode` : *bool, optional*
A boolean that determines whether we are using the driver node for training. Default should be false; we don't want to invoke executors without explicit mention.

`use_gpu` : *bool, optional*
A boolean that indicates whether or not we are doing training on the GPU. Note that there are differences in how GPU-enabled code looks like and how CPU-specific code looks like.

M-L Training

Consider the following:

- Use TensorFlow, Keras, and PyTorch
- Accelerator, Distributed ML & GPU :
 - Horovod
 - NVIDIA RAPIDS Accelerator for Spark



M-L Training (Batch)

With the default scheduler: workloads experience higher rates of resource starvation, leading to performance degradation or failure !

Solutions for the default scheduler:

- Use custom k8s Scheduler support. Enabling YARN-like capabilities such as queue, gang scheduling, etc



VOLCANO
Kubernetes Native Batch System

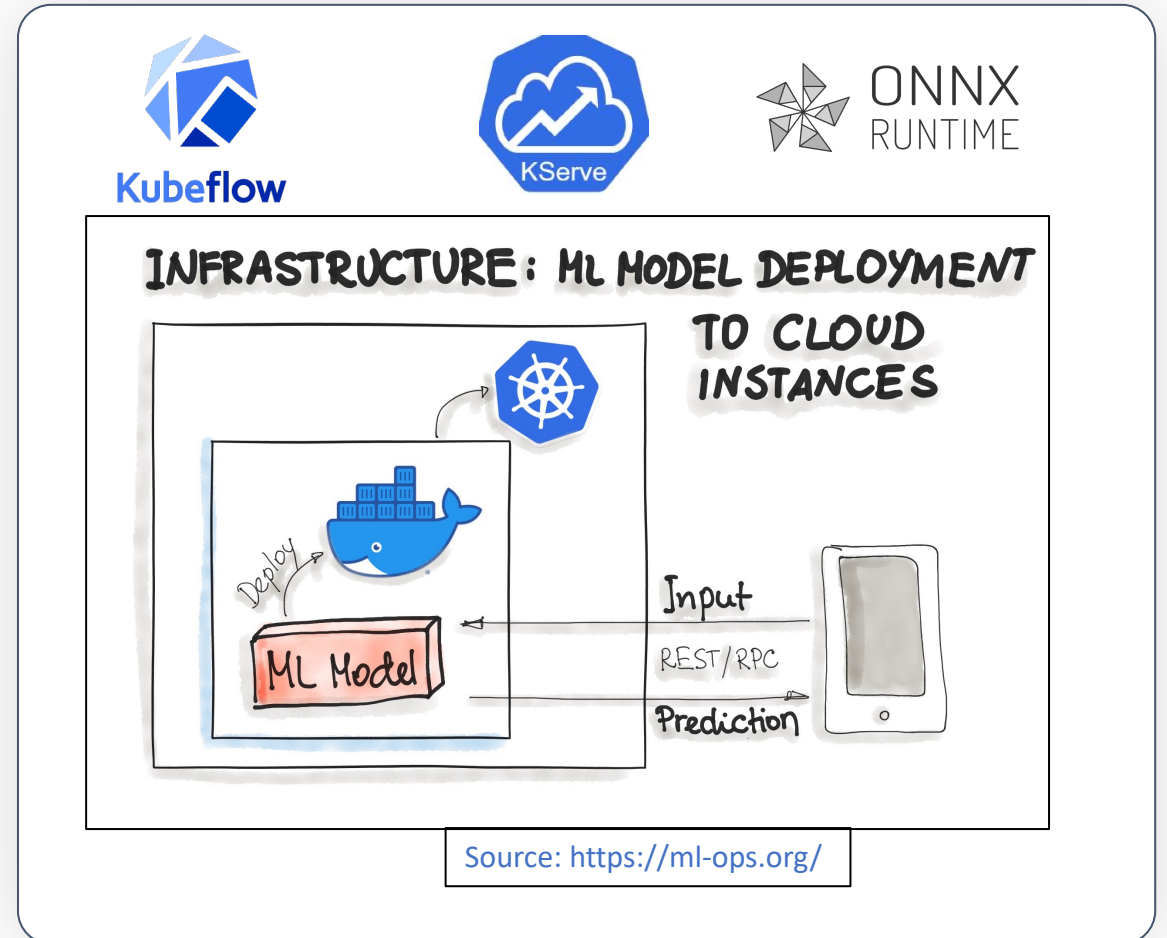


Apache YuniKorn

M-L Deployment & Serving (WIP)

Key factors to consider

- Package the whole ML tech stack (dependencies) and the code for ML model prediction into a Docker container.
- Model optimization and Model compression



Conclusion and takeaways

Why Spark on k8s integration is Important for R-L M-L?

- Native Integration
- k8s best practices apply to Spark on k8s for free!
- Scalability, Latency, Fault-tolerance
- Models Training and Serving
- Integration with a rich ecosystems



Key takeaways

- Use the k8s Spark operator
- Design and build your logging and monitoring stack
- Keep adhering to Spark best practices compatible with k8s
- Use the rich k8s (Monitoring, Mlops, etc) ecosystem
- Contribute to the OSS (share your experiences, code, ideas, challenges)
- Keep Cycling ...



Thank you



Blogs



Ocean for Apache Spark

