Deploy Containerized State-Of-The-Art Deep Learning Models in Your Apps

Jaffa Sztejnbok
jaffas@il.ibm.com
Developer Advocate, Israel

IBM **Developer**

IBM

Important Links for this event ...

1. PPT https://github.com/jaffaszt/MAX

2. Signup for IBM Cloud account : https://ibm.biz/BdfY4S

3. Go to Catalog

4. Search for "**Watson Studio"** and create a service

5. Please answer a short survey https://ibm.biz/BdfY4v

Sign up for an IBM Cloud account... and access a wide catalog of services

Service Type

IBM Service

Third Party Service

Status

New Service (2019)

Service Refresh (2019) *(Plan Add)*

AI

Watson Assistant
Watson Studio
Compare and Comply
Discovery
Knowledge Catalog
Knowledge Studio
Language Translator
Machine Learning
Natural Language Classifier
Natural Language Understanding
Personality Insights
Speech to Text
Text to Speech
Tone Analyzer
Visual Recognition
Voice Agent with Watson
Watson OpenScale
PowerAI

Analytics

Analytics Engine
DB2 Warehouse on Cloud
IBM Cognos Dashboard
Master Data Management
SQL Query
Streaming Analytics
Weather Company Data
AccountScore

Security & Identity

App ID
Certificate Manager
Cloud HSM
FortiGate Security Appliance
Hardware Firewall
Hardware Firewall (Dedicated)
Hyper Protect Crypto Services
Hyper Protect Virtual Servers
Internet Services
Key Protect
Security Advisor
SSL Certificates
Contrast Security
FusionAuth
Twilio Authy
Twilio Verify

Dev Tools

Availability Monitoring
Continuous Delivery
DevOps Insights
Cloud Event Management
Globalization Pipeline
Monitoring
Schematics
Toolchain
IBM Cloud Activity Tracker with LogDNA
IBM Cloud Monitoring with Sysdlg
IBM Log Analysis with LogDNA
Pager Duty

VPC

Block Storage for VPC
Load Balancer for VPC
Virtual Private Cloud
Virtual Server for VPC
VPN for VPC

Web & Mobile

Mobile Foundation
Push Notifications
Email Delivery, powered by Sendgrid
Bitbar Testing Cloud
Esri ArcGIS for Developers
Nexmo
Phunware Location Based Services
Phunware Mobile Marketing Automation
SPLICE Pre-CAT Insurance Notifications
Telstra Messaging API
Twilio Programmable SMS
Twilio Programmable Video
Twilio Programmable Voice

Web & App

Apache Spark
Consult with IBM Garage
Information Server
Managed Financial Data API
Simulated Instruments Analytics API
Accern-API
Network ACL for VPC
Subnet for VPC
Alloy
Bondevalue-API
CloudAMQP
Difitek
Dwolla
ElephantSQL
Envestnet | Yodlee
FundingShield – Wire Accounts
Verification Service (WAVS)
HazardHub Property Risk Data API
Health Score
Hydrogen
Morningstar
Natural Language Generation APIs
Payeezy
Plaid
Powerlytics Behavior/Propensity Model API
Powerlytics Consumer Income API
Powerlytics Investable Assets & Wealth API
Quovo
Rainbow
RelSci
Risk Engine
SizeUp Small Business Intelligence
Strands Business Financial Management
Totum Risk
TradeIt
uCloud Multitenant Core Platform for VMWare
WealthEngine API
Xignite Market Data APIs
Ylabs

IoT

IoT Platform
Weather Company Data
AT&T Flow Designer
AT&T IoT Data Plans
Bosch IoT Rollouts
Car Diagnostic API
Precision Location
Unification Engine

Starter Kits

Custom Vision Model for Core ML with Watson
Internet of Things Platform Starter
Java Microservice with MicroProfile and Java EE
Java Microservice with Spring
Java Web App with Spring
Mobile Basic
Node.js Microservice with Express.js
Node.js Web App with Express.js
Python Microservice Flask
Python Web App with Django
Python Web App with Flask
Watson Assistant Basic
Watson Natural Language Understanding Basic
Watson Visual Recognition Basic
Node-RED Starter

Databases

Blockchain Platform
Cloudant
Databases for PostgreSQL
Databases for Redis
Databases for Elasticsearch
Databases for MongoDB
Messages for RabbitMQ
Databases for etcd
Compose Enterprise
Compose for RethinkDB
Db2
Db2 Hosted
Db2 Warehouse
Hyper Protect DBaaS for Mongo DB
Hyper Protect DBaaS for PostgreSQL
Informix
SQL Query
GEO Web Services
InfluxCloud

Storage

Block Storage
Block Storage for VPC
File Storage
IBM Cloud Backup
Object Storage
OSNEXUS Software Defined Storage for Bare Metal
Actifio GO
box

Network

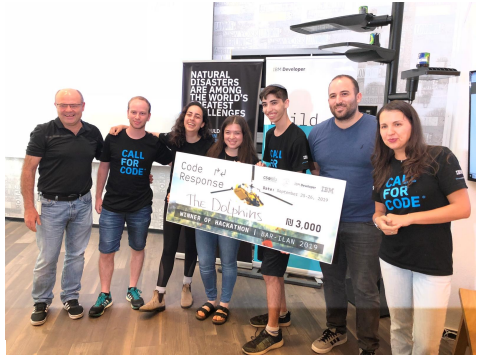
Load Balancers
Content Delivery Network
Direct Link Connect
Direct Link Dedicated
Direct Link Dedicated Hosting
Direct Link Exchange
Domain Name Service
FortiGate Security Appliance
Gateway Appliance
Internet Services
IPSec VPN
Load Balancer for VPC
Subnets/IPs
Virtual Private Cloud
VLAN
VPN for VPC

Integration

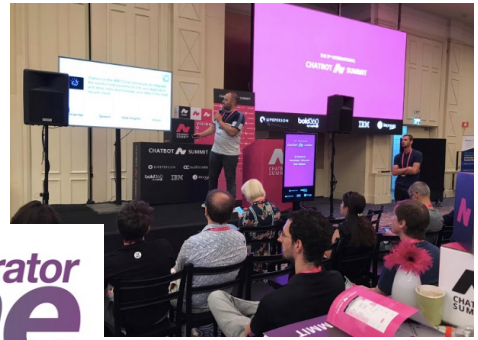
API Connect
APP Connect
Lift CLI
MQ
Secure Gateway
Rocket Mainframe Data
SPLICE PRE-Cat Insurance
Notifications

Compute

Bare Metal Server
Cloud Foundry Enterprise Environment
HPCaaS from Rescale
Power System Virtual Server
Virtual Server
Virtual Server for VPC
WebSphere Application Sever
VMware vCenter Server on IBM Cloud
VMware vSphere on IBM Cloud
NetApp ONTAP Select
Single-node Trial for Migration and App Modernization
Single-node Trial for Data Protection & Disaster Recovery
Caveon RiskForesight on IBM Cloud
F5 on IBM Cloud
FortiGate Security Appliance on IBM Cloud
FortiGate Virtual Appliance on IBM Cloud
HCX on IBM Cloud
HyTrust CloudControl on IBM Cloud
HyTrust CloudControl on IBM Cloud
HyTrust DataControl on IBM Cloud
HyTrust KeyControl on IBM Cloud
IBM Cloud Private Hosted
IBM Cloud Secure Virtualization
IBM SPECTRUM Protect Plus on IBM Cloud
KMIP for VMware on IBM Cloud
Veeam Availability Suite for Virtual Server
Veeam on IBM Cloud
VMware vRealize Operations and vRealize Log Insight on IBM Cloud
Mission Critical VMWare on IBM Cloud
Managed Services from IMI
Managed Services for Veeam on IBM Cloud
Managed Services for Zerto on IBM Cloud
IBM Cloud Backup
OSNEXUS Software Defined Storage for Bare Metal
Plesk Onyx Linux
Unlimited for Virtual Server
Red Hat Enterprise Linux OS for Virtual Server
Liberty for Java
SDK for Node.js
ASP.NET Core
Runtime for Swift
Go
PHP
Python
Ruby
Tomcat
Red Hat OpenShift Cluster
Kubernetes Service
Container Registry



IBM **Developer**



IBM Alpha Zone accelerator

Startup & Developer – IBM Israel

https://www.meetup.com/IBM-Startup-Developer-Tel-Aviv

https://www.crowdcast.io/ibmdevelopereurope

https://www.facebook.com/groups/941786926573388

Call for Code®

—

ibm.biz/callforcode

What is Call for Code

Inaugurated in 2018, Call for Code was created by **David Clark Cause** and supported by:

- **IBM**, Founding Partner
- **United Nations Human Rights**, Charitable Partner
- **the Linux Foundation**, Program Affiliate

Invites developers and problem solvers around the world to build and contribute to sustainable open source software solutions, that address social and humanitarian issues, while ensuring solutions are deployed to make a **real difference**.

Sponsors:

ARROW

intuit. turbotax quickbooks mint

INGRAM MICRO®

ibm.biz/callforcode

Call for Code has become the only global, **always-on** tech for good open source platform to deploy & scale top projects through a host of offerings:

- Annual Call for Code Global Challenge
- University Challenge
- Spot Challenges to address urgent & unexpected societal issues facing global citizens
- Call for Code for Racial Justice
- Open-Source Projects & Deployments

A community of:

400k+ Developers	**179** Nations	**15k+** Applications

Call for Code through the years

Natural Disasters

In 2018, this challenge looked to build solutions to improve preparedness for natural disasters and accelerate relief. In 2019, an emphasis was added on the health and well-being of individuals and communities.

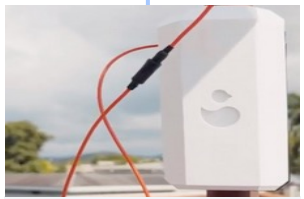
Climate Change
COVID-19
Racial Justice

2020 revealed the need for many solutions. Call for Code pivoted to provide problem solvers the opportunity to address them all.

Climate Change

"The impacts of climate change are global in scope and unprecedented in scale. Without drastic action today, adapting to these impacts in the future will be more difficult and costly."

2018



Deployed Solution:

Project OWL

Keeping first responders and victims connected in natural disasters.

Deployment underway in Barcelona, Spain and Australia, and Puerto Rico.

2019

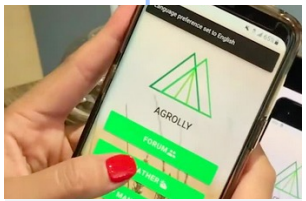
Deployed Solution:

Prometeo

Real-time & predictive analysis of firefighter toxicity exposure.

Deployments in Puerto Rico, Australia, and more.

2020

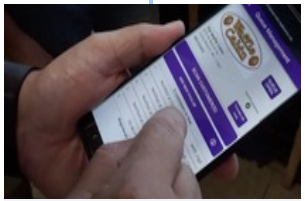


Deployed Solution:

Agrolly

Helping farmers grasp the climate change impacts on crops to help manage their resources more effectively.

Currently utilized in Mongolia



Deployed Solution:

Safe Queue

An app letting you queue at a safe distance by notifying you when it's your turn to enter.

Deployed in NY.

Call for Code®
for Racial Justice

Open source solution starters:

CFCFRJ

– Fair Change
– Five Fifths Voter
– IARS
– Legit-Info
– Open Sentencing
– TakeTwo
– Truth Loop

2021

Call for Code®

Answer the Call

See your solution get deployed in 2021!

Why join Call for Code

Skill-building

You can learn about industry-leading open hybrid cloud technologies and develop skills that can enhance your professional career while making a positive impact on the world.

Social Good

By fighting back against the most pressing humanitarian and societal issues of our time, you are building and/or contributing to projects that have the potential to be deployed into communities in need.

Community

Are you ready to be a part of a community of societal problem solvers? Call for Code is your opportunity to meet and network with people from around the world who share similar passions for using #TechforGood.

Ideas to Action

This isn't your average hackathon. If you want the chance to build a solution that can make true impact in the field, Call for Code can turn your idea into action through deployment support from the Linux Foundation.

$200K Grand Prize

The winner of the Call for Code Global Challenge wins $200k and support from IBM and its partners who help set up the winning team from concept to deployment.

ibm.biz/callforcode

Call for Code Starter Kits

Clean Water and Sanitation

from intelligent solutions for small farmers to recycling showers, technology can make a significant impact on the availability of water and its consumption.

Get started →

Zero Hunger

communities can harness data insights from their environment for better crop resilience and overall yield for sustainable food production systems. More crops mean better access to food for the community.

Get started →

Responsible Production and Green Consumption

from recommendations on energy efficiency to highlighting the carbon footprint of online purchases. The goal is to take into account all phases of resource use to do more and better with less.

Get started →

Answer
the call.
ibm.biz/callforcode

Agenda

- Data Science and Deep Learning
- Model Asset Exchange (MAX)
- Containers and deployment options :

  - Docker Desktop (locally)

  - Openshift based on Kubernetes

  - Serverless : IBM Code Engine
- Hands-on

What is Data Science?

Goal?

Find a solution to the business problem

How?

Transform into well posed questions

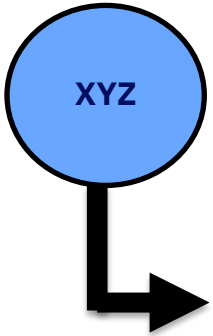
Using?

Mathematics, programming, and scientific method

Finally

Communicate results and its business impact

Solving business problems with data science



Employee Attrition

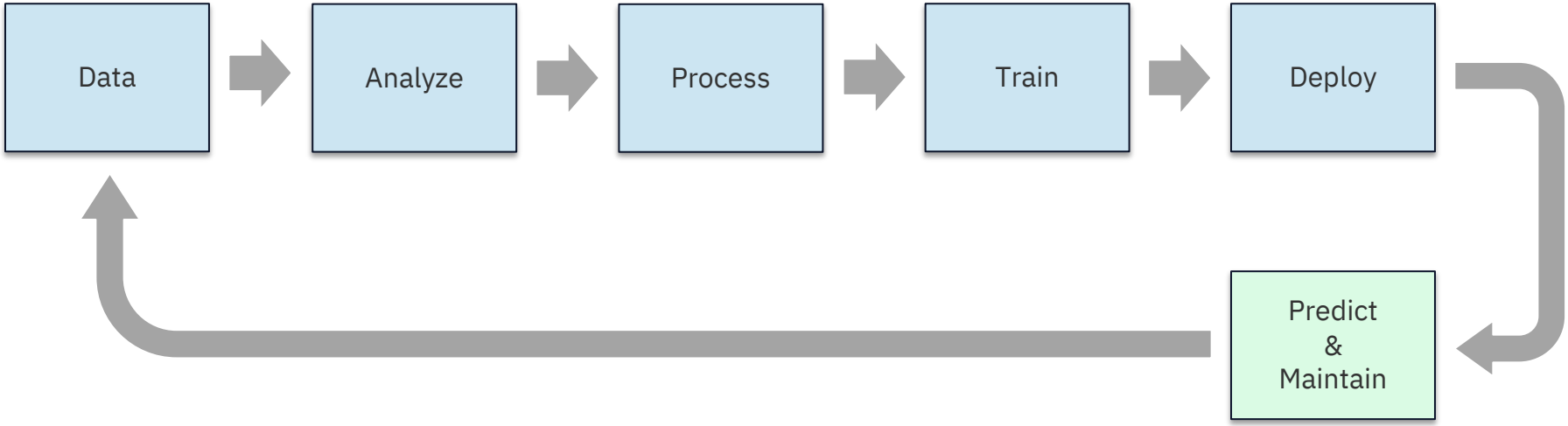
Data Science Team

Data Science Questions

1. What are the factors causing employee attrition?
2. Can we predict if an employee will quit or not based on current status?

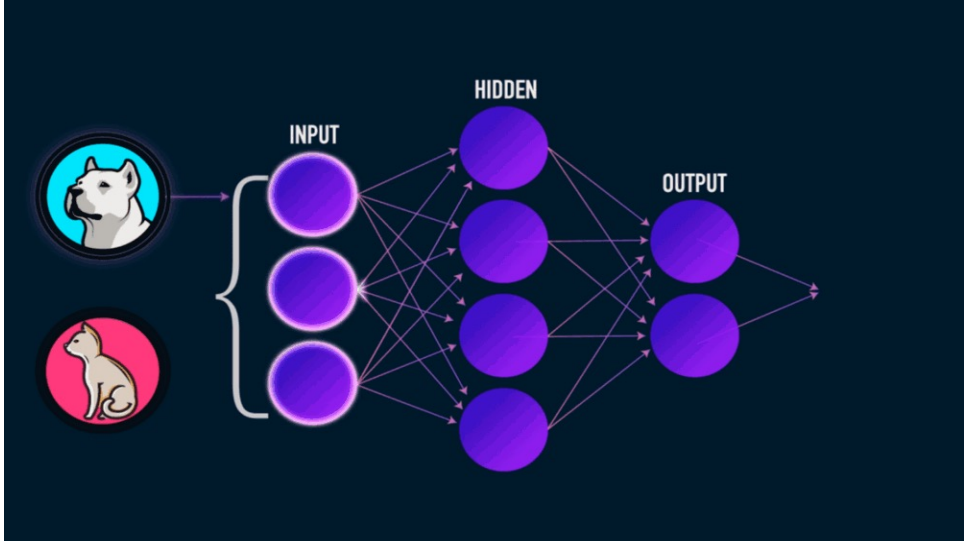
Explore, analyze, use tools, create models and interpret the results

Machine Learning Workflow



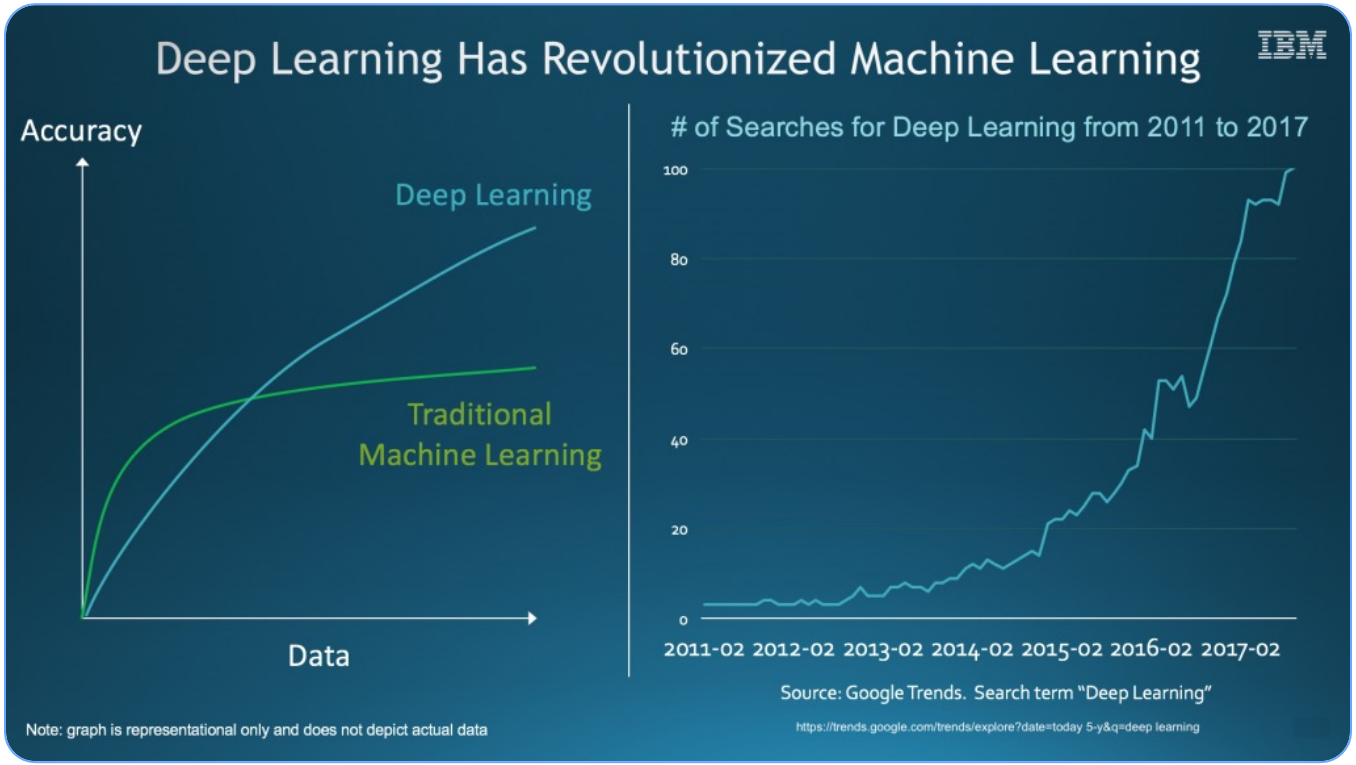
Deep Learning

Deep learning is a specialized type of machine learning. It refers to a general set of models and techniques that tries to loosely emulate the way the human brain solves a wide range of problems. Deep learning has recently been very successful and hence is becoming an increasingly popular and important tool for data science.



Why is Deep Learning so popular?

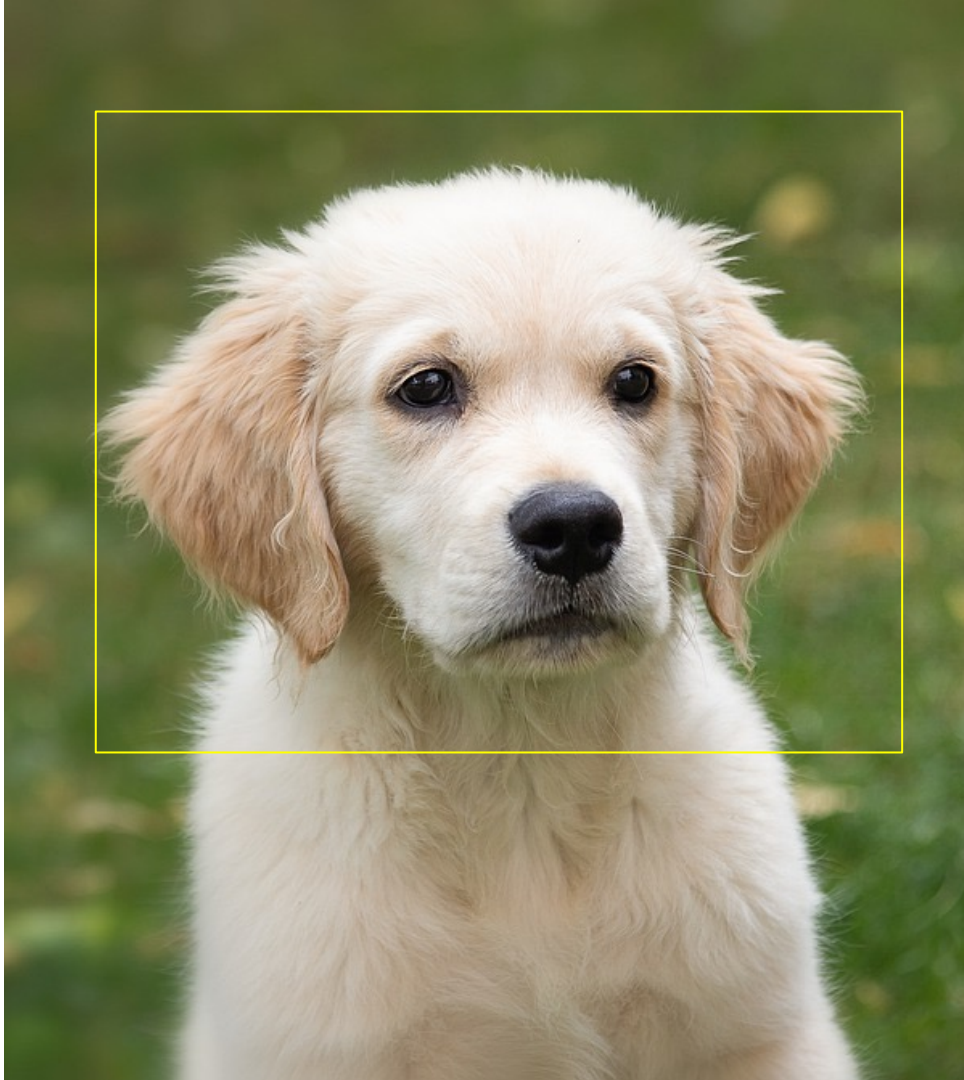
- Access to huge amounts of data
- Increases in computational power
- Great at supervised learning
- Automatically learns important features



Object detection

What's in
the picture?

Where is it in
the picture?



Find model

...that does what you *need*

...that is *free* to use

...that is *performant* enough

Get Code and Cleanup

Many open source repositories

Research vs production code

Code licenses

Train

Multiple frameworks

Data licenses

Deploy and Consume

Adjust inference code

Package inference and model code, and pre-trained weights

Deploy your package

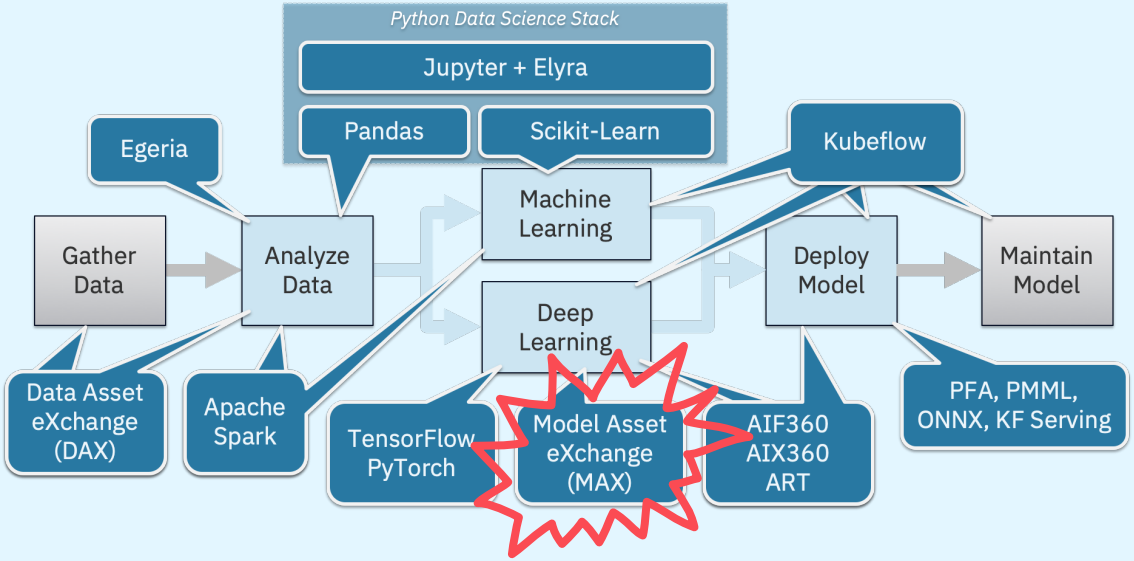
With the
Model Asset
eXchange,
you can!

Center for Open Source Data & AI Technologies

codait.org

- CODAIT aims to make AI solutions dramatically easier to create, deploy, and manage in the enterprise.
- We contribute to and advocate for the open-source technologies that are foundational to IBM's AI offerings.
- 30+ open-source developers!

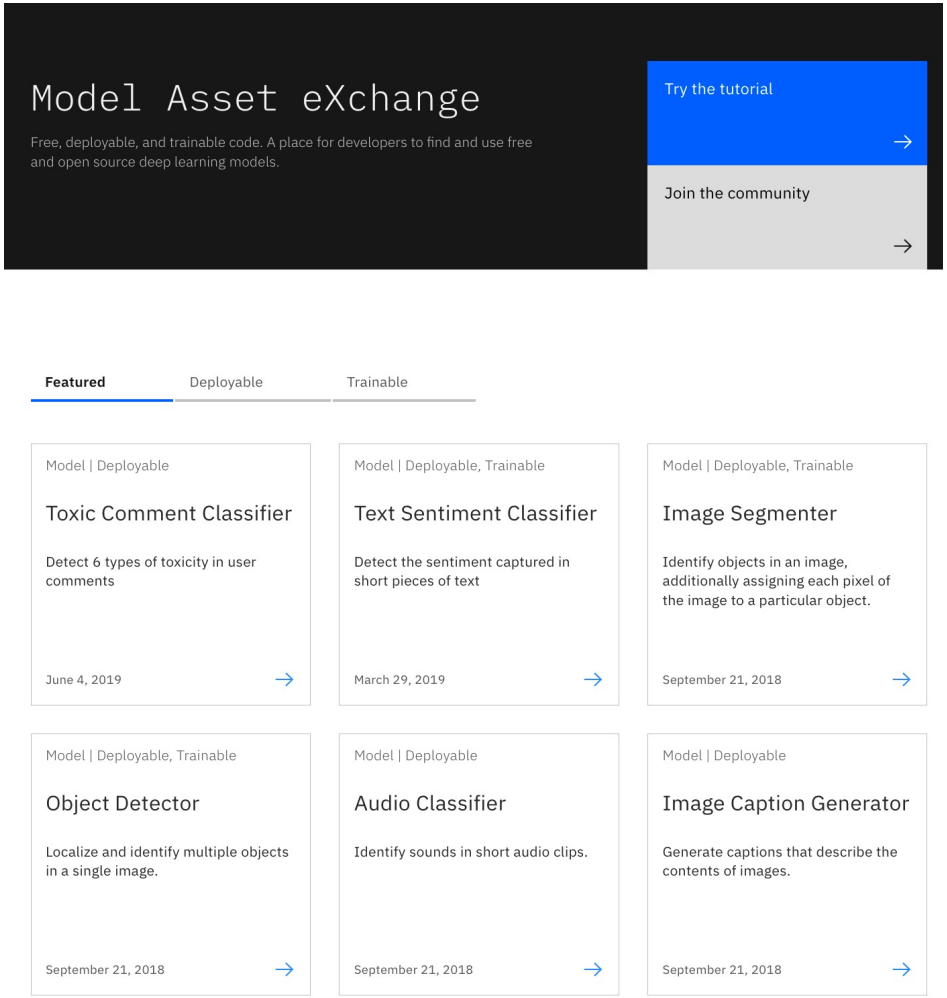
Improving the Enterprise AI Lifecycle in Open Source



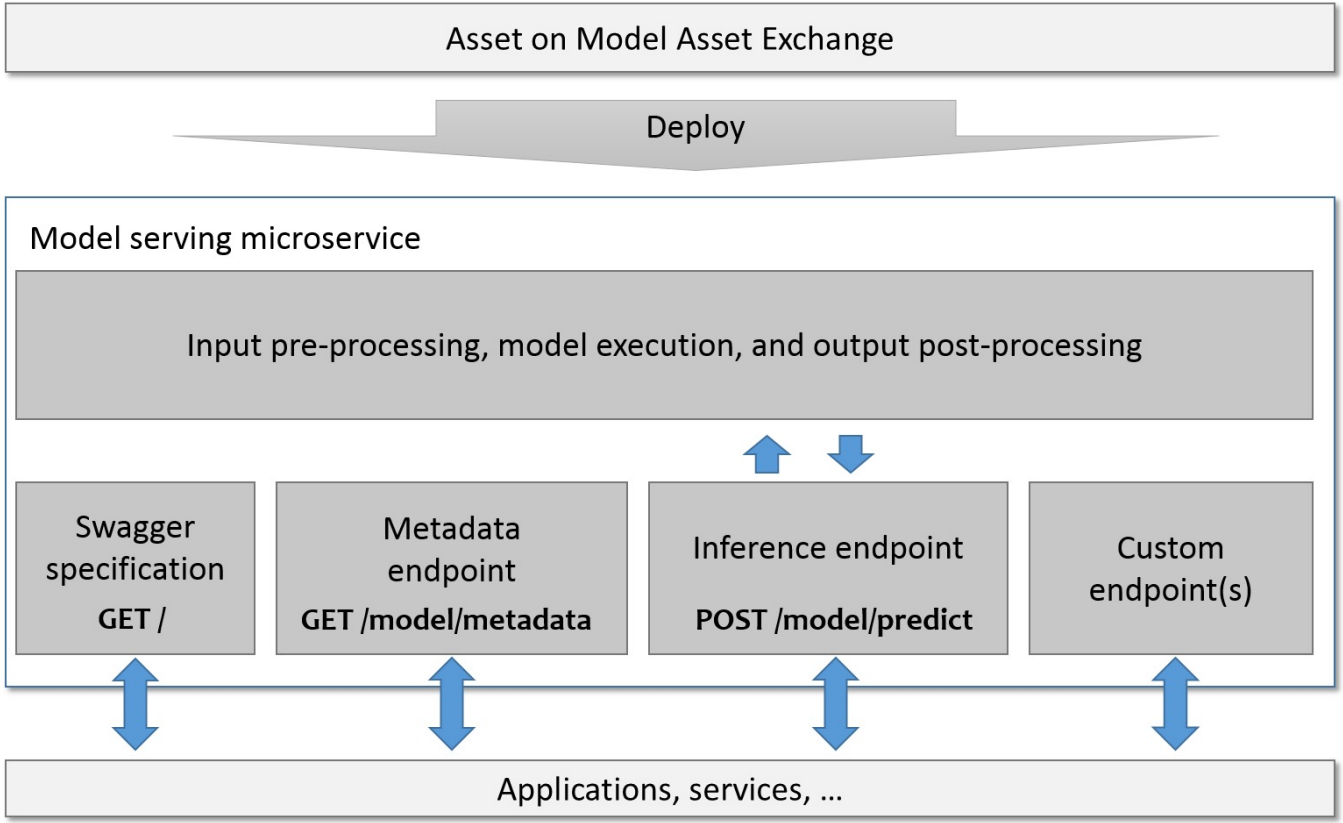
Model Asset eXchange (MAX)

- State-of-the-art deep learning models
- Free and open source
- Wide variety of domains
- Tested code and IP
- Pre-trained & trainable models
- Standardized REST API & containerized microservices

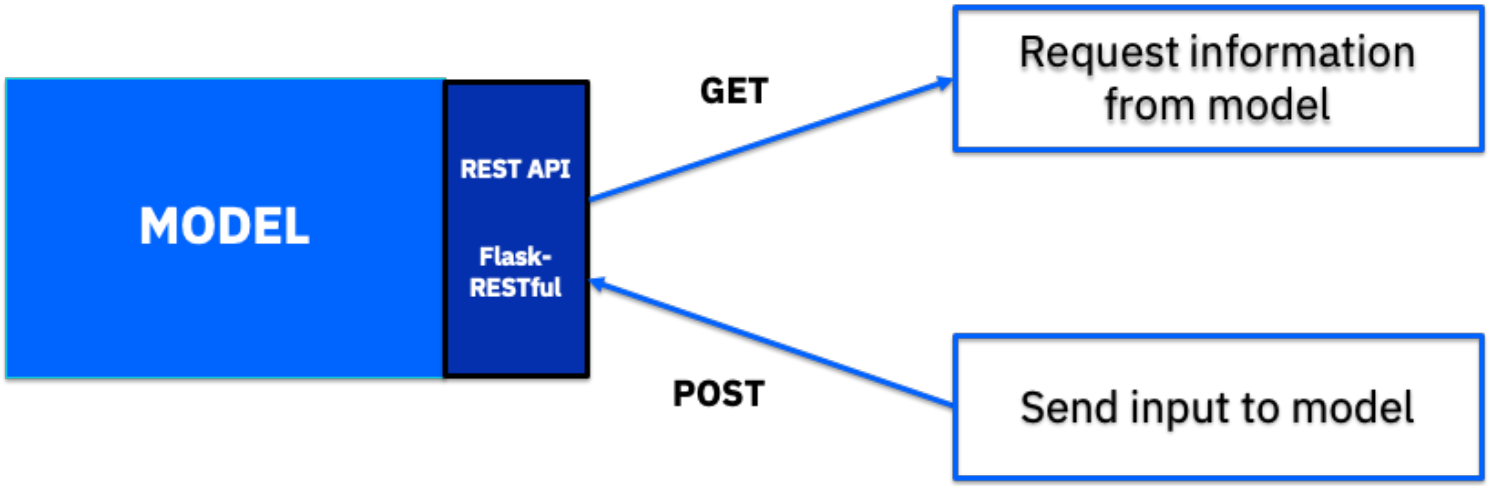
https://developer.ibm.com/exchanges/models/



Deployable Asset on MAX



MAX model consumption



Highlights

- Image, audio, text, healthcare, time-series and more
- Pre- / post-processing & inference wrapped up in Docker container
- Generic API framework code - Flask RESTPlus
- Swagger specification for API
- One-line deployment locally and on a Kubernetes cluster

Model Asset Exchange Server 0.1

[Base URL: /]
http://localhost:5000/swagger.json

An API for serving models

model Model information and inference operations

GET `/model/labels` Return the list of labels that can be predicted by the model

GET `/model/metadata` Return the metadata associated with the model

POST `/model/predict` Make a prediction given input data

Parameters | Cancel

Name	Description
image * required file *(formData)*	An image file (encoded as PNG or JPG/JPEG) Choose File No fil...hosen
threshold number *(query)*	Probability threshold for including a detected object in the response in the range [0, 1] (default: 0.7). Lowering the threshold includes objects the model is less certain about. 0.8

Execute | Clear

Containers and deployment options :

- Docker Desktop (locally)
- Openshift based on Kubernetes
- Serverless : IBM Code Engine

What is a container?

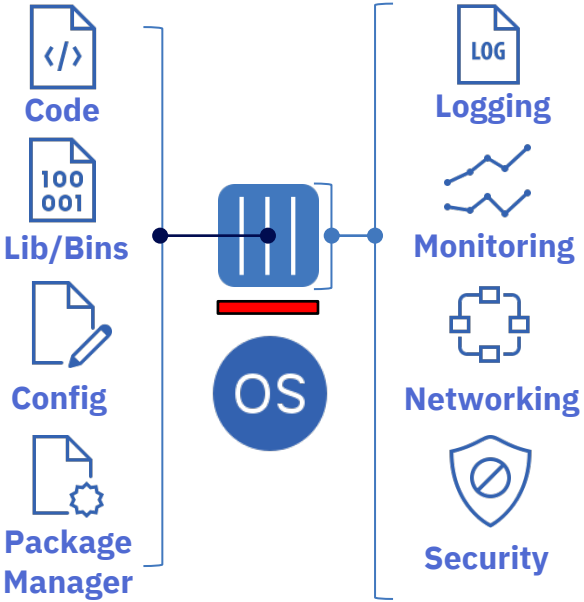
A standard way to **package** an application and all its dependencies so that it can be moved between environments and **run** without changes.

Containers work by **isolating** the differences between applications **inside** the container so that everything **outside** the container can be standardized.

Deb
The Developer

- Package apps with all dependencies
- Deploy to any environment in seconds
- Easily accessed and shared

Worries about what's « **inside** » the container

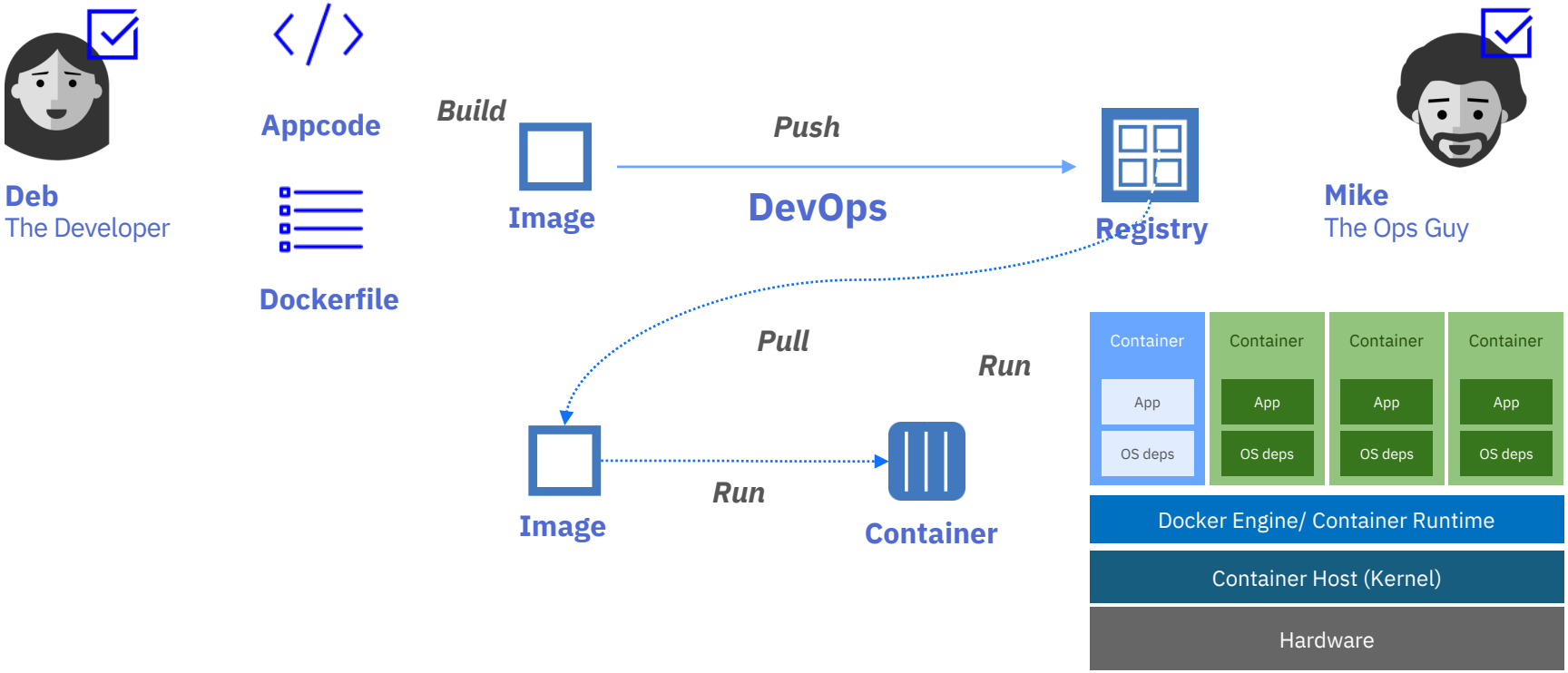


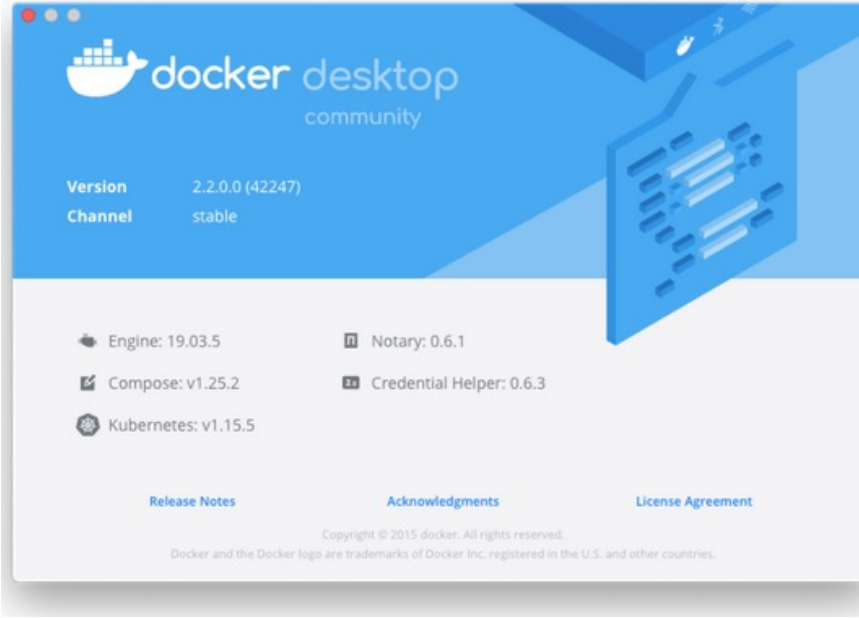
Mike
The Ops Guy

- Application processes on a shared kernel
- Simpler, lighter, and denser than VMs
- Portable across different environments

Worries about what's « **outside** » the container

Basic deployment process

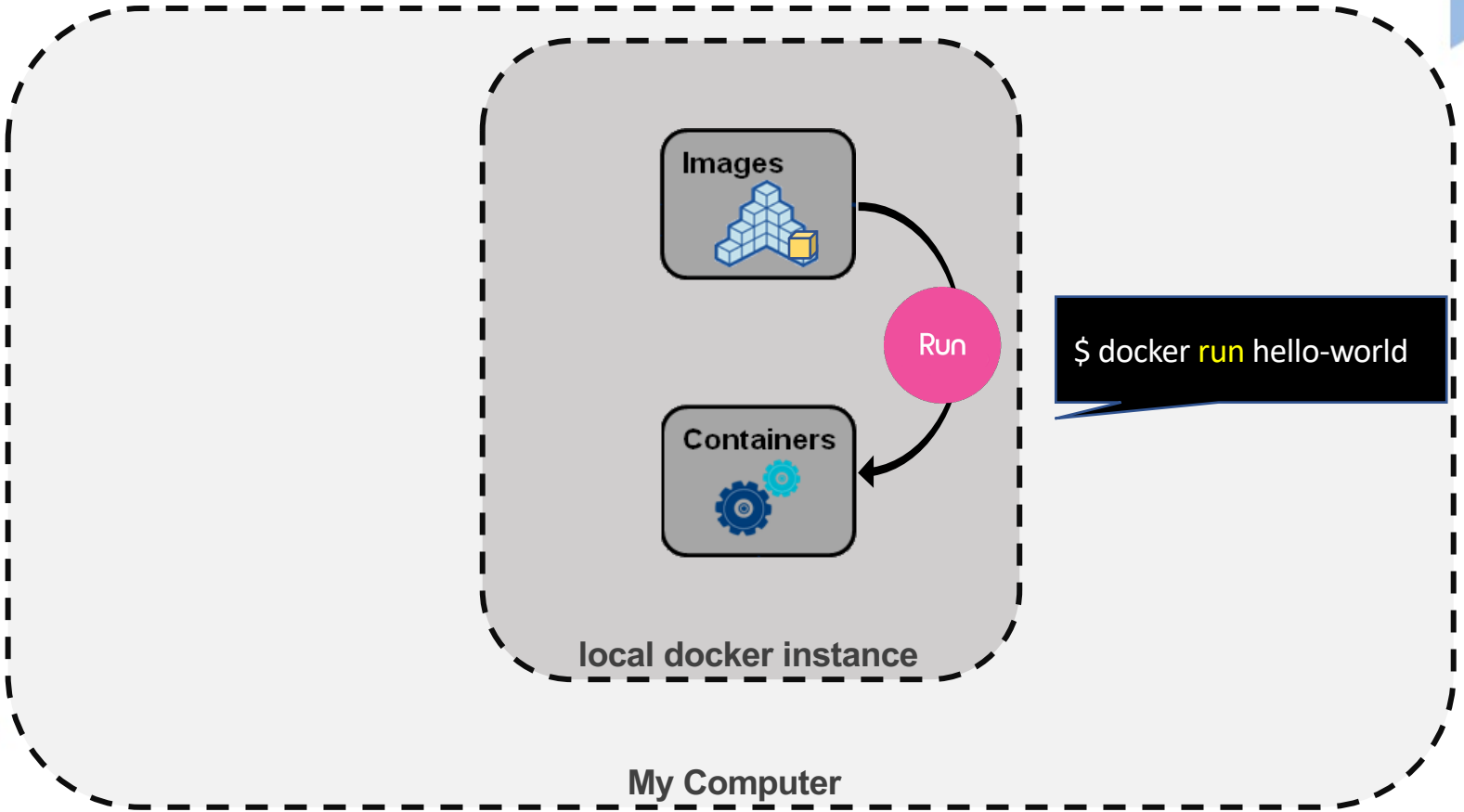




Docker Desktop

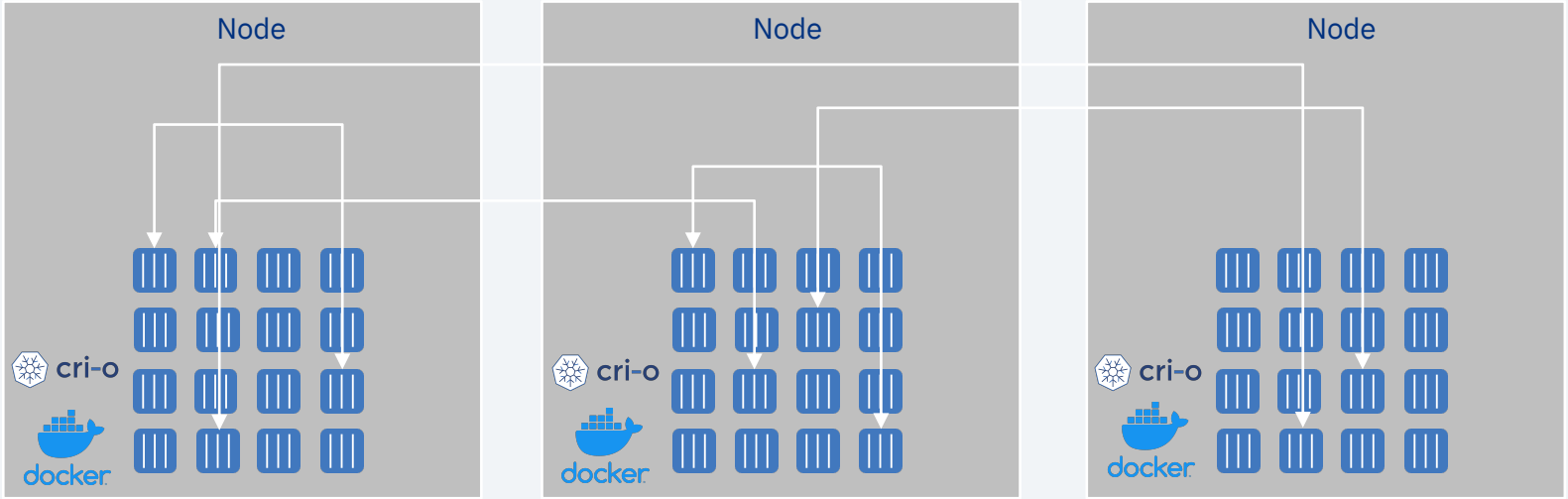
Docker Desktop is a native application that delivers all of the Docker tools to your Mac or Windows Computer.

1. Open Docker Desktop. (Download here if you don't have it).
2. Type the following command in your terminal: `docker run -dp 80:80 docker/getting-started`
3. Open your browser to http://localhost
4. Have fun!


Images
Run
Containers
local docker instance
$ docker run hello-world
My Computer

Docker registry

Welcome to Kubernetes



Intelligent Deployment

Auto repair (*Self-healing*)

Horizontal scaling

Service discovery and load balancing

Automation deployments and rollback

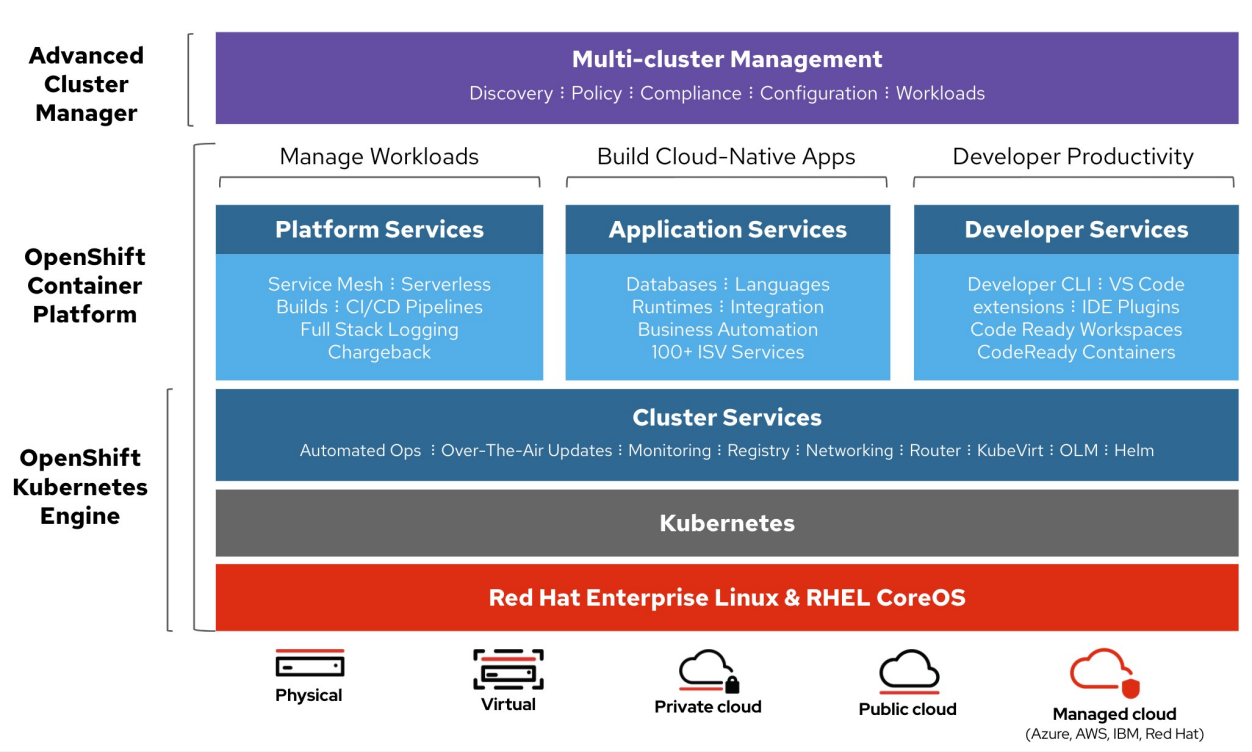
Key Management (Secret) and configuration

What is Red Hat OpenShift?

Red Hat OpenShift is Kubernetes for business, in Hybrid Cloud mode.

OpenShift value

- Certified platform
- Support & SLA
- Security
- Full Stack Installation
- Automation
- Product life cycle
- Collaboration
- Catalogue
- Hybrid, Multi-Cluster Management
- Operator Framework
- OpenShift Service Mesh (Istio)
- OpenShift Serverless (Knative)
- CodeReady Workspaces (Che)



Code Engine is a Kubernetes-based Container Platform that:

allows me to deploy my container-based workloads **easily**, **quickly** and **securely**.

dynamically scales my containers according to load.

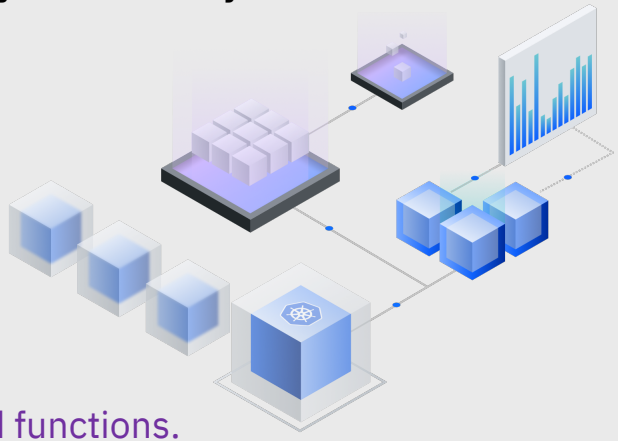
has **monitoring**, **logging** and a **service mesh** “built in”.

gives me the **ability to run serverless** workloads.

does not “lock me in” and is **built on open-source** projects.

... and ...

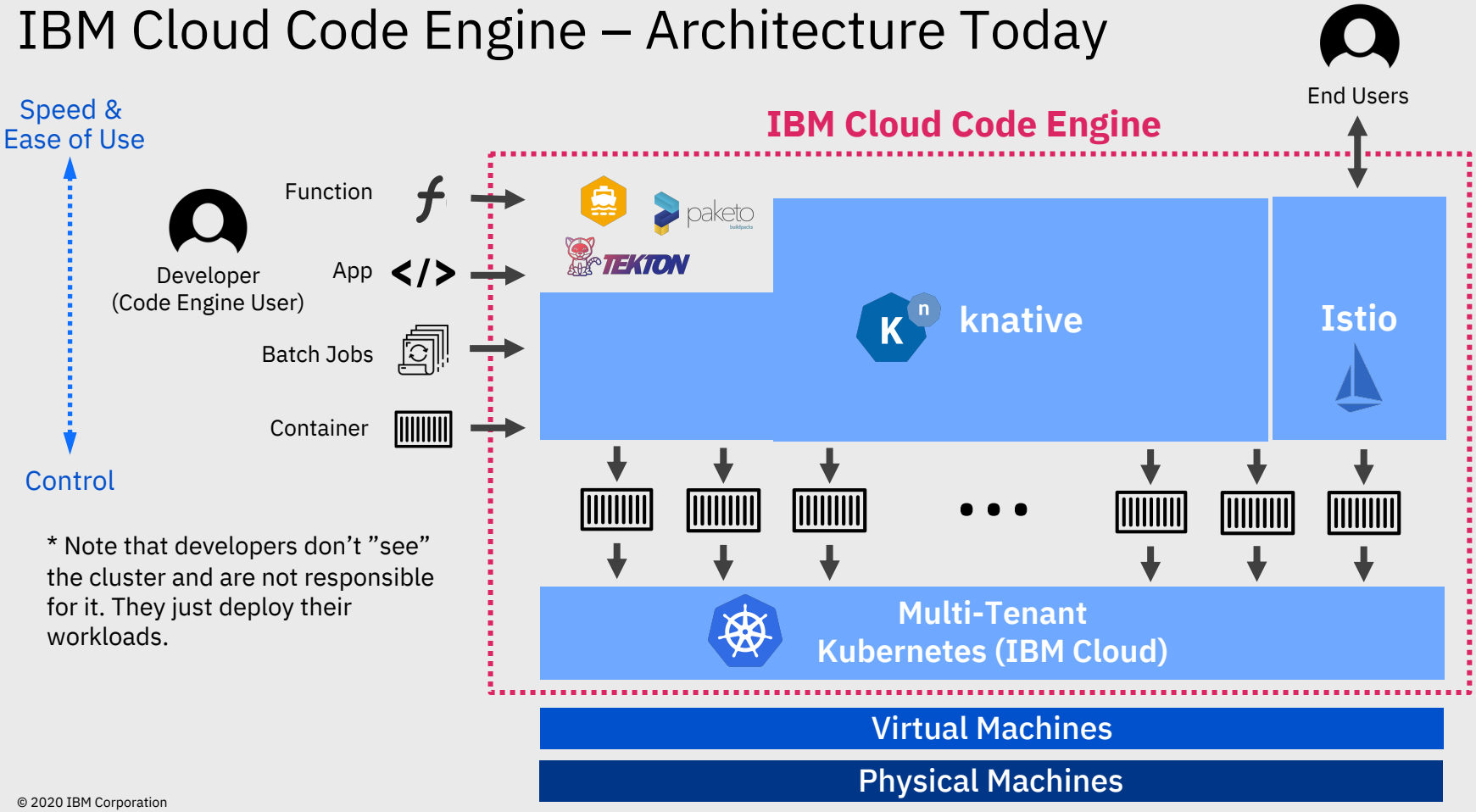
unifies the deployment of containers, applications, batch jobs and functions.



starts as a **production-grade, multi-tenant shared container service**,
but will be extendable to other locations (via IBM Cloud Satellite) for higher isolation.

© 2020 IBM Corporation

IBM Cloud Code Engine – Architecture Today



* Note that developers don't "see" the cluster and are not responsible for it. They just deploy their workloads.

© 2020 IBM Corporation

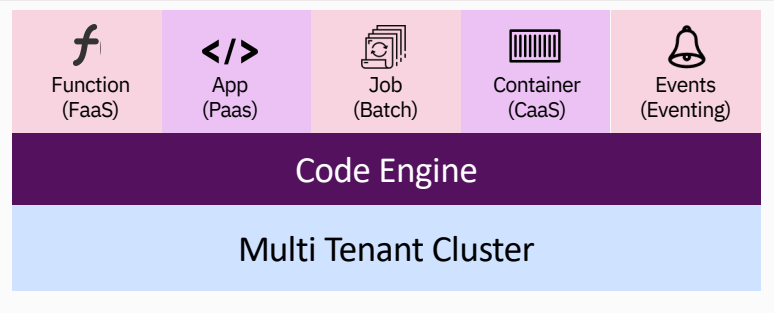
"I just want to write & run my applications"

"I don't want to learn Kubernetes"

"I want smart infrastructure that does what it needs to do, not only what/when I tell it to do it"

Code Engine:

- **Managed, simplified, intelligent hosting experience**
- **All your container-based workloads in one environment**
- **Pay only for what you use**
- **Designed to let your developers get back to coding**
- **The power of Kubernetes without the complexity**



Try it today for free: https://cloud.ibm.com/codeengine

Why?

- Kubernetes, while feature rich, is too complex
- Kubernetes is missing some newer features - e.g. serverless
- Forcing a choice of *aaS is artificial and unnecessary
- Developers should code, not manage infrastructure

What?

- Public managed container hosting platform
- Supports containerized workloads (web, APIs, batch, ...)
- Key features
  - Request-based auto-scaling - including to zero
  - Built-in security - e.g. networking, runtime isolation
  - App versioning
  - Build images from Source - Dockerfiles or Buildpacks
  - Integration with IBM Cloud Services (service bindings)
  - No artificial limits - e.g. execution time
  - Runtime choice - e.g. cpu, memory
  - Event driven integrations (Cloud Object Storage, Ping)
  - Private Registry Support

How?

- Built on open-source technologies - Kube, Istio, Knative, ...
- Infrastructure still available if necessary, goal is to hide it
- Integrates with existing Kubernetes workloads

Important Links for this event ...

1. PPT https://github.com/jaffaszt/MAX
2. Signup for IBM Cloud account : https://ibm.biz/BdfY4S
3. Go to Catalog
4. Search for “**Watson Studio”** and create a service
5. Please answer a short survey https://ibm.biz/BdfY4v

IBM