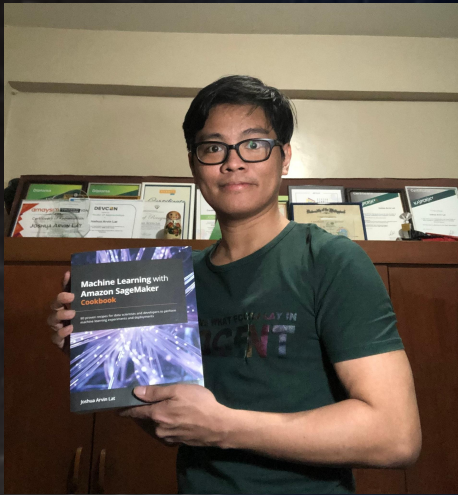


# Hacking and Securing Machine Learning Environments and Systems

Joshua Arvin Lat



— aws —  
machine learning  
**HERO**



- **Chief Technology Officer** of NuWorks Interactive Labs
- AWS Machine Learning Hero
- Author of 📖 **Machine Learning with Amazon SageMaker Cookbook**



# Machine Learning with Amazon SageMaker Cookbook

80 proven recipes for data scientists and developers to perform machine learning experiments and deployments

Joshua Arvin Lat

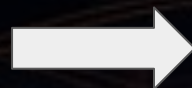
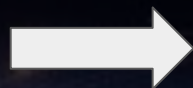
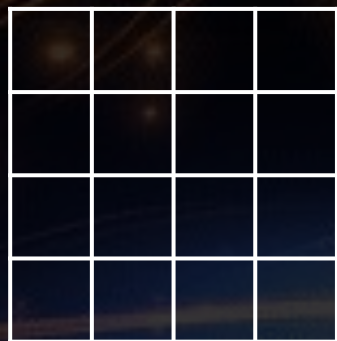


Author of 

## Machine Learning with Amazon SageMaker Cookbook

80 proven recipes for data scientists and developers to perform machine learning experiments and deployments

- 
- Data Collection
  - Data Preparation and Cleaning
  - Data Visualization and Analysis
  - Feature Engineering
  - Model Training and Parameter Tuning
  - Model Evaluation
  - Model Deployment



**CLASS IMBALANCE**

**DPPL**

**TREATMENT EQUALITY**



# **CYBERSECURITY ATTACK CHAIN**

SHORT-TERM FINANCIAL OBJECTIVES	VERY HIGH
LONG-TERM FINANCIAL OBJECTIVES	HIGH
CLIENT AND CUSTOMER HAPPINESS	HIGH
COMPLIANCE	LOW

**CUSTOM IMPLEMENTATION**

**CONFIGURATION**

**ML SERVICE**







Virtual Private Cloud

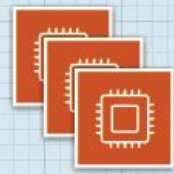
Availability Zone



Public Subnet



Private Subnet



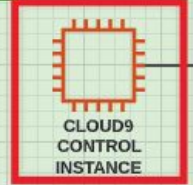
Availability Zone



Public Subnet



Private Subnet



CLOUD9  
CONTROL  
INSTANCE

**HIGH  
RISK?**



Elastic  
Kubernetes  
Service

**LIBRARIES,  
FRAMEWORKS,  
PACKAGES, AND  
DEPENDENCIES**

**CUSTOM CODE**

**DATASET**

**JUPYTER NOTEBOOK**

**INFRASTRUCTURE**

**ML SERVICE**



## Navigation

- [Why joblib: project goals](#)
- [Installing joblib](#)
- [On demand recomputing: the \*Memory\* class](#)
- [Embarrassingly parallel for loops](#)
- [Persistence](#)
  - [Use case](#)
  - [A simple example](#)
  - [Persistence in file objects](#)
  - [Compressed joblib pickles](#)

## Examples

### Development

[joblib.Memory](#)  
[joblib.Parallel](#)  
[joblib.dump](#)  
[joblib.load](#)

# Persistence

## Use case

`joblib.dump()` and `joblib.load()` provide a replacement for pickle to work efficiently on arbitrary Python objects containing large data, in particular large numpy arrays.

### Warning:

`joblib.dump()` and `joblib.load()` are based on the Python pickle serialization model, which means that arbitrary Python code can be executed when loading a serialized object with `joblib.load()`.

`joblib.load()` should therefore never be used to load objects from an untrusted source or otherwise you will introduce a security vulnerability in your program.

### Note:

As of Python 3.8 and numpy 1.16, pickle protocol 5 introduced in [PEP 574](#) supports efficient serialization and de-serialization for large data buffers natively using the standard library:

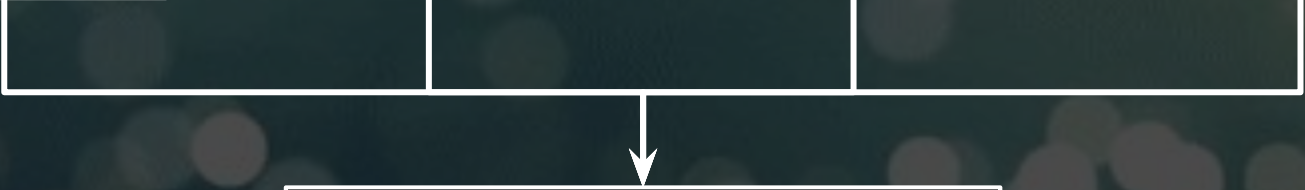
```
pickle.dump(large_object, fileobj, protocol=5)
```

**DATASET**

**CUSTOM CODE**

**DOCKER  
CONTAINER IMAGE**

**CONFIGURATION**



**ML SERVICE**

**MODEL  
ARTIFACTS**



MODEL DEPLOYED IN AN EC2 INSTANCE



MODEL DEPLOYED IN A CONTAINER IN AN EC2 INSTANCE



BUILT-IN ALGORITHM + SAGEMAKER ENDPOINT



CUSTOM CONTAINER + SAGEMAKER ENDPOINT



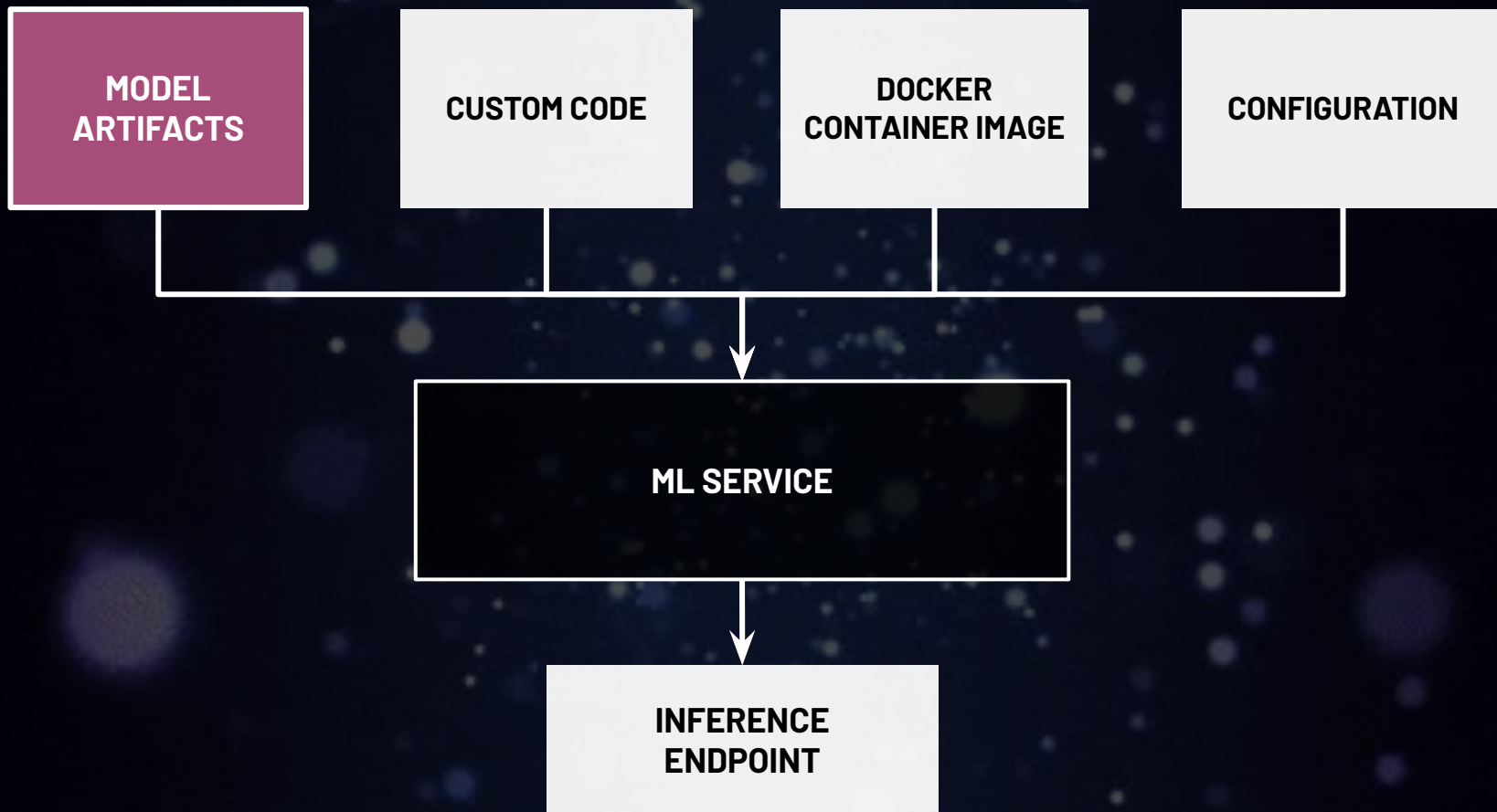
MODEL DEPLOYED INSIDE A LAMBDA FUNCTION



LAMBDA TRIGGERING A SAGEMAKER ENDPOINT



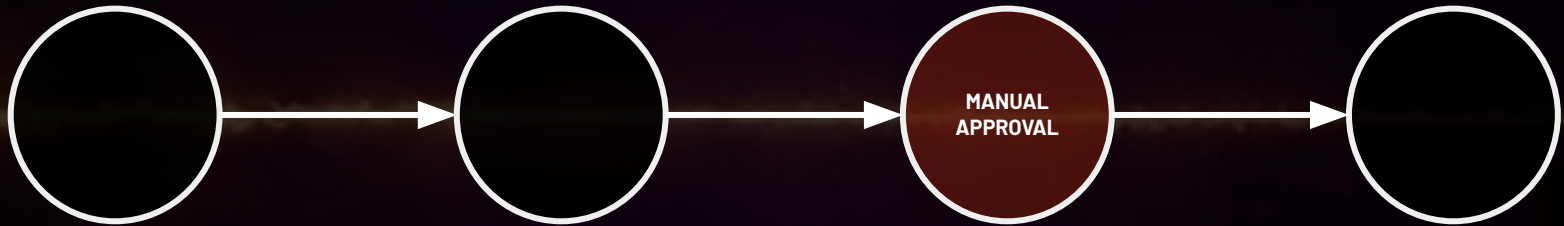
MODEL DEPLOYED IN FARGATE






# **NETWORK ISOLATION**







# **AUTOMATED VULNERABILITY MANAGEMENT**

# VULNERABILITY ASSESSMENT TOOL



Finding summary  
Package findings  
0 Critical 43 High 71 Medium

**Findings (100+)** 🔄  
Choose a row to view the finding details. All findings are related to this instance.

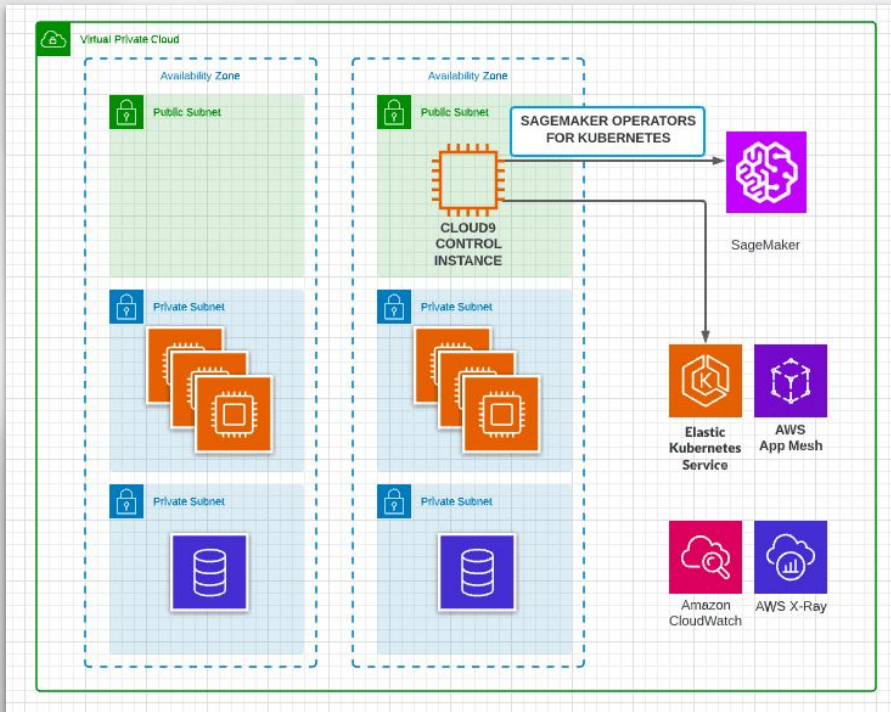
Active   Resource ID EQUALS

< 1 2 3 4 5 6 7 8 ... >

Severity	Title	Impacted resource	Type	Age
High	CVE-2018-20669 - kernel		Package Vulnerability	an hour
High	CVE-2019-19074 - kernel		Package Vulnerability	an hour
High	CVE-2021-3347 - kernel		Package Vulnerability	an hour
High	CVE-2020-8648 - kernel		Package Vulnerability	an hour
High	CVE-2019-19319 - kernel		Package Vulnerability	an hour
High	CVE-2020-25670 - kernel		Package Vulnerability	an hour
High	CVE-2021-3656 - kernel		Package Vulnerability	an hour
High	CVE-2021-3656 - kernel		Package Vulnerability	an hour
High	CVE-2021-3656 - kernel		Package Vulnerability	an hour
High	CVE-2021-3656 - kernel		Package Vulnerability	an hour

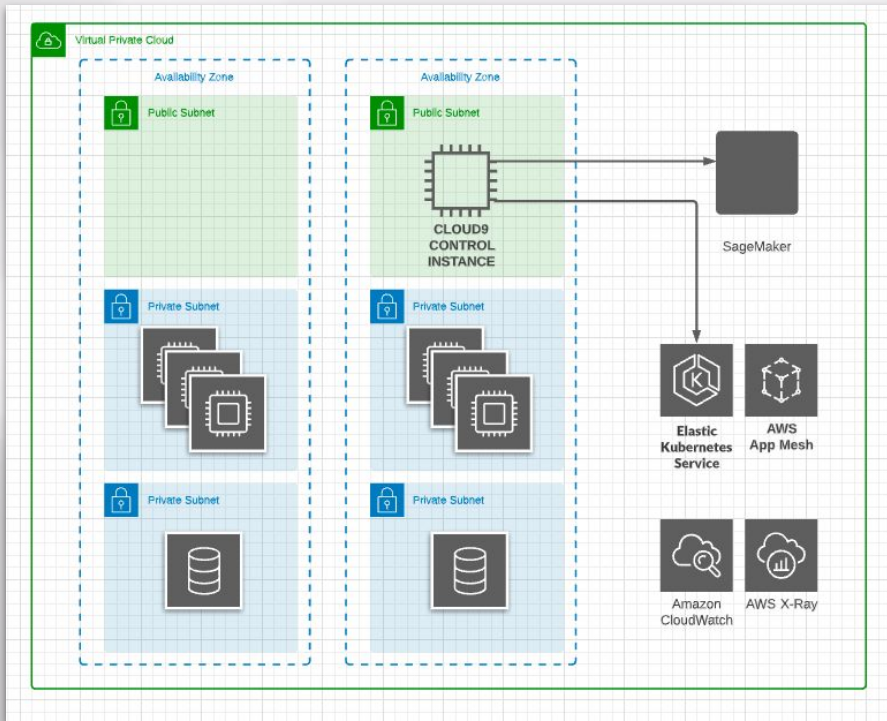


**(SECURE)  
INFRASTRUCTURE AS CODE**



```
{
  'key': 'value'
}
```

```
}
  ,key, : ,value,
{
```

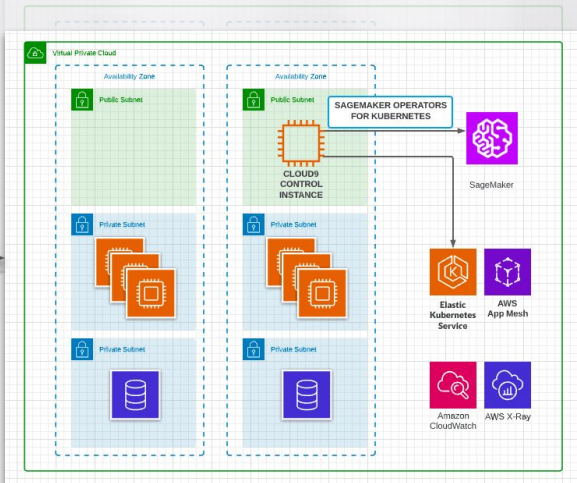
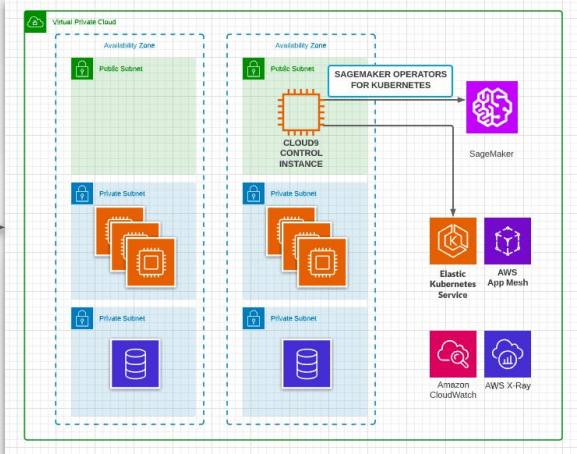


```
{
  'key': 'value'
}
```

```
}
  ,key, : ,value,
{
```

```
{  
  'key': 'value'  
}
```

```
{  
  'key': 'value'  
}
```



# **ACCOUNT ACTION MONITORING**



PREDICTION ENDPOINT



API GATEWAY



LAMBDA + SCIKIT-LEARN

PREDICTION ENDPOINT



API GATEWAY



LAMBDA + TENSORFLOW

PREDICTION ENDPOINT



API GATEWAY



LAMBDA + FB PROPHET



# **RESTRICTIVE IAM PERMISSIONS**

# **PRINCIPLE OF LEAST PRIVILEGE**



TOOL  
A

TOOL  
B

TOOL  
C

CONCEPT  
X

CONCEPT  
Y



**IMPACT ON COST?**

# **Hacking and Securing Machine Learning Environments and Systems**