
DATA
SCIENCE

Who

Daniel Machado
Machine Learning Engineering
& Solution Architect @ dtidigital

Why

	Individual Development Environment	Team Development Environment	Test/UAT Environment	Production Environment
Who	Data Scientist	Data Scientist, Engineer & Analyst	Production and UAT Team	Production Team
Development Tool	Jupyter Notebook	Jupyter Notebook, SQL Editor & …	None	None
Code	Python	Python, SQL, & others ….	Python, SQL, & others ….	Python, SQL, & others ….
Code Version	Version 1	Version 2 & 3	Version 3	Version 3
Run, Test, & Monitor	Ad hoc run: check unit tests	Ad hoc run: check unit, functional tests	Ad hoc run: check functional & monitor tests	Scheduled run: check monitor tests
Hardware	Laptop & test data	Development Servers & DB	Test Servers & DB	Production Servers & DB
Operating System & Libraries	Windows + libraries	Unix + Various libraries	Unix + Various libraries	Unix + Various libraries
Data	Small Development Data	Development Data	Copy Production Data	Production Data

Infraestrutura "as a code"

Conceito para facilitar o processo de gerenciamento e provisionamento de diferentes recursos, e possibilita utilização de ferramentas de versionamento, teste validação e CI/CD.

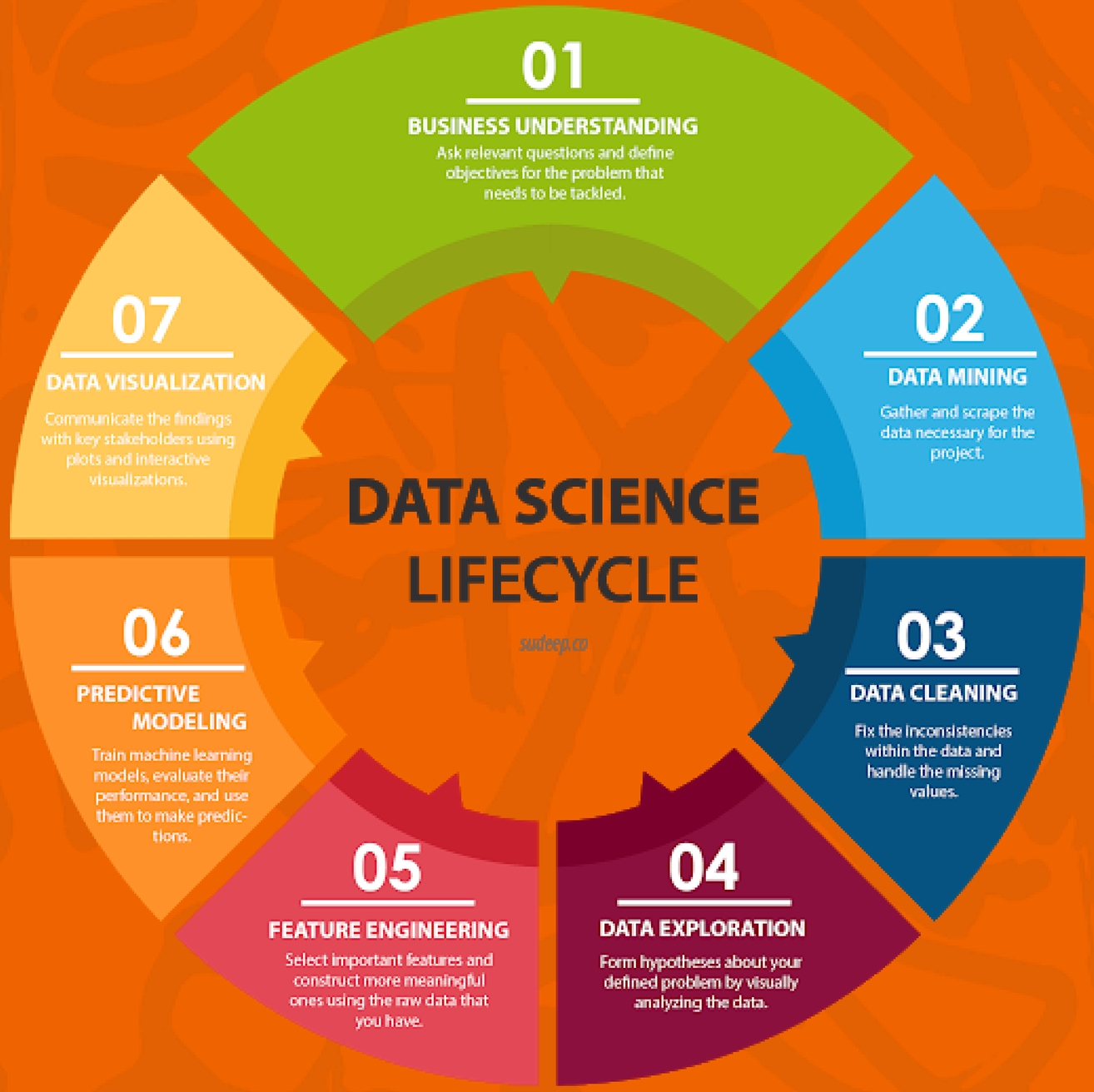
Ansible, Terraform, Kubernets, Docker

IaC

Terraform

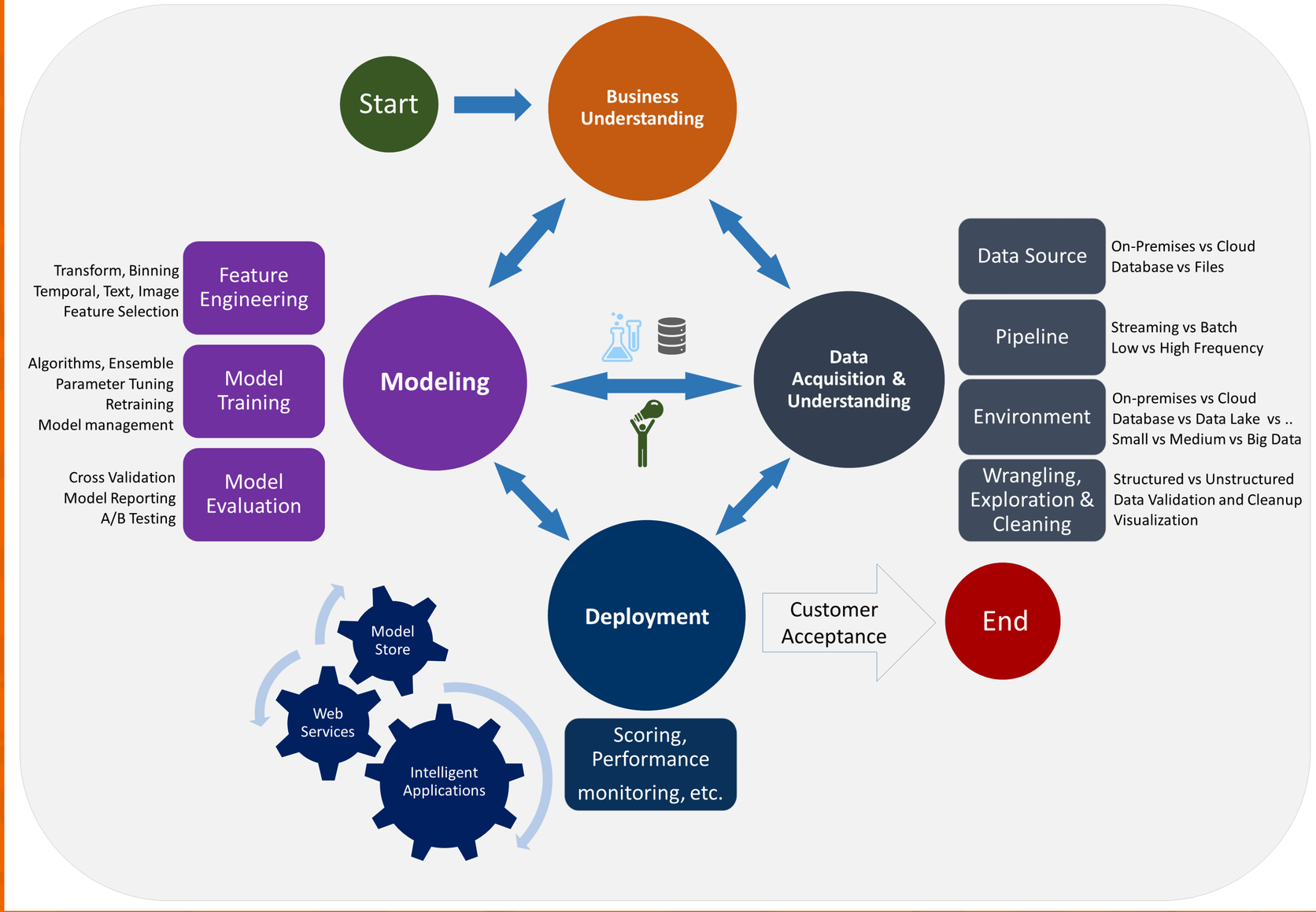
- Terraform vs Ansible
- Terraform vs Kubernets

Data science

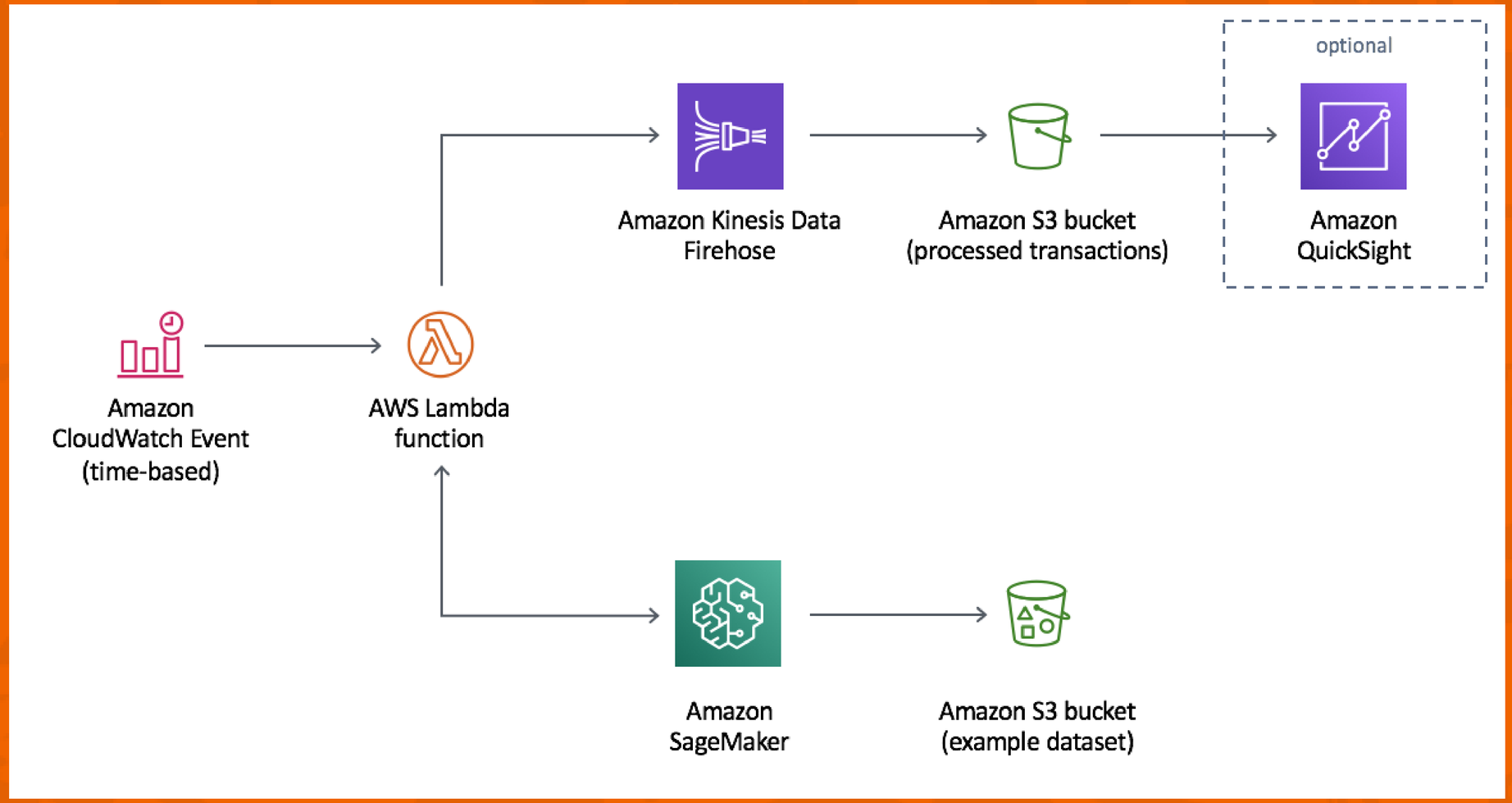


Data Science

Data Science Lifecycle



Data Science



code

code

1. Criação de roles (IAM)

```
resource "aws_iam_role"
"fraud_detection_firehose_role"

resource "aws_iam_role"
"sm_notebook_instance_role"

resource "aws_iam_role"
"fraud_detection_lambda_role"
```

code

2. Armazenamento do nosso notebook (s3_lambda)

resource "aws_s3_bucket" "fraud_detection_function_bucket"

3. Armazenamento de dados de teste (s3 sagemaker)

resource "aws_s3_bucket" "s3_bucket_1"

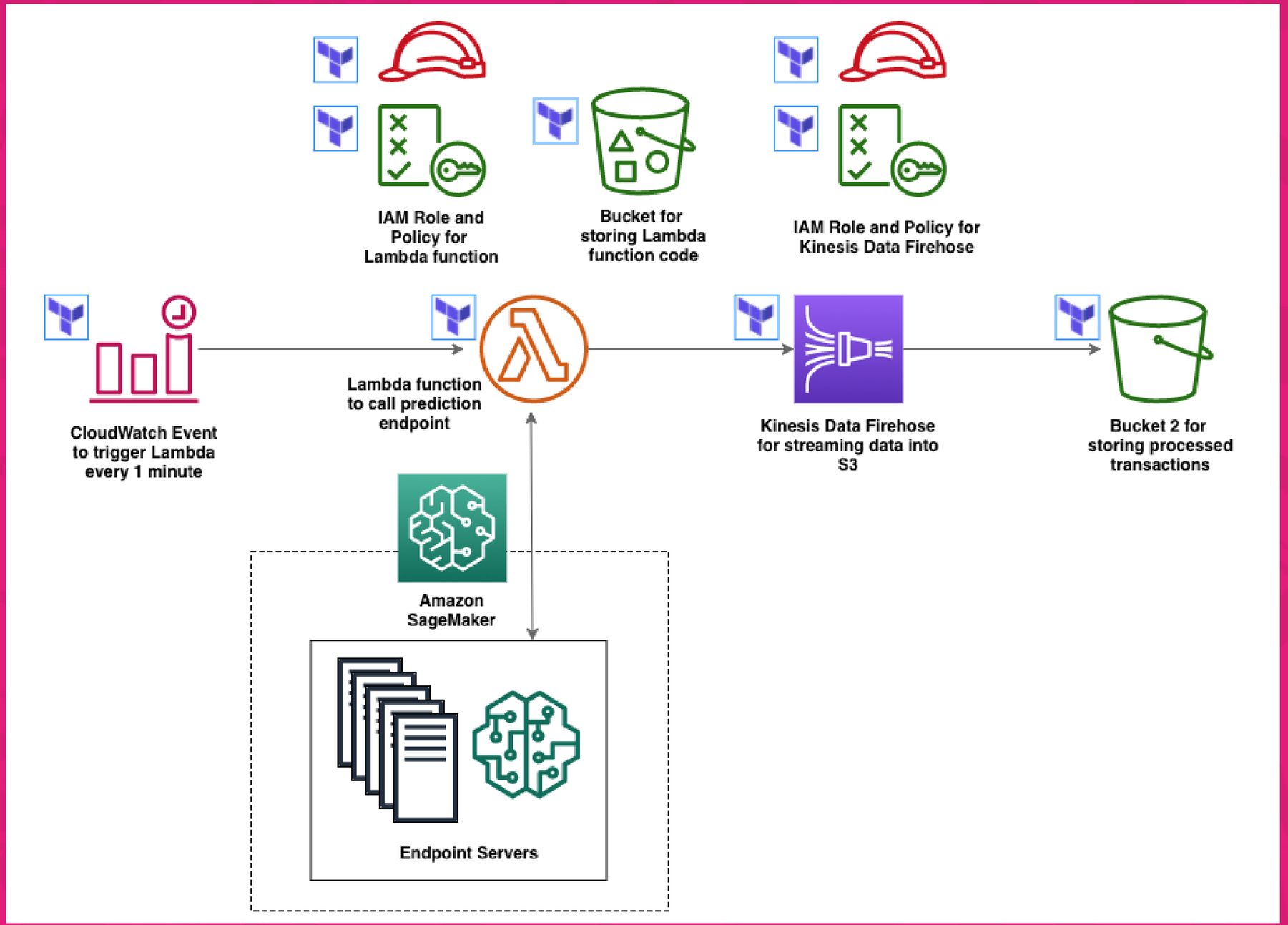
code

4. Notebook (sagemaker)

```
resource "aws_sagemaker_notebook_instance"
"basic"
```

```
resource
"aws_sagemaker_notebook_instance_lifecycle_config
uration" "basic_lifecycle"
```

code


IAM Role and Policy for Lambda function
Bucket for storing Lambda function code
IAM Role and Policy for Kinesis Data Firehose
CloudWatch Event to trigger Lambda every 1 minute
Lambda function to call prediction endpoint
Kinesis Data Firehose for streaming data into S3
Bucket 2 for storing processed transactions
Amazon SageMaker
Endpoint Servers

code

5. Lambda.tf

resource "aws_lambda_function"
"fraud_detection_event_processor"

6. Kinesis

resource "aws_kinesis_firehose_delivery_stream"
"fraud_detection_firehose_stream"

Deploy

```
terraform init
terraform validate
terraform plan -out=tfplan
terraform apply --auto-approve tfplan
```

```
terraform plan -destroy -out=tfplan
terraform apply tfplan
```

Tnks

https://aws.amazon.com/solutions/fraud-detection-using-machine-learning/

terraform.io/docs/providers/aws/r/sagemaker_notebook_instance.html