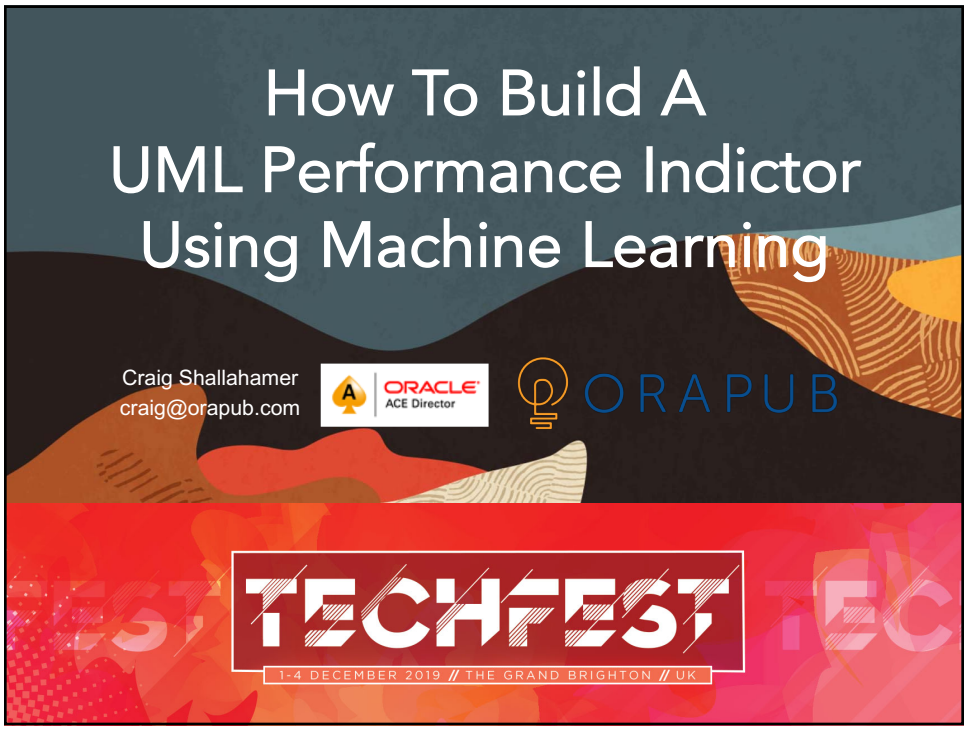
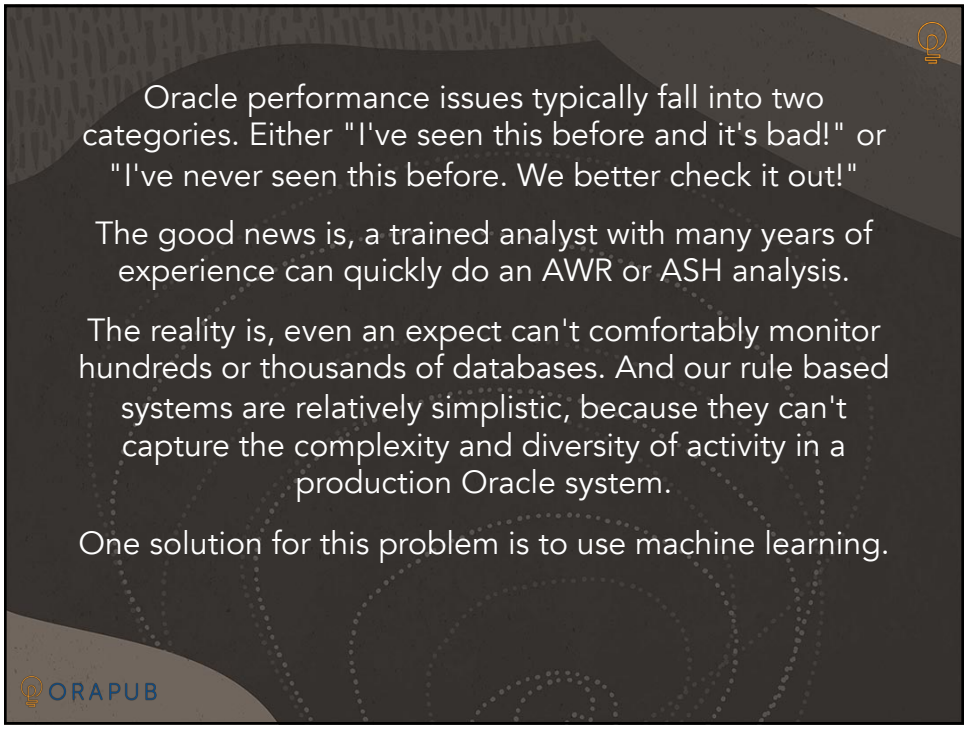


This presentation was given by Craig Shallahamer (craig@orapub.com) at the December 2019 UKOUG (UK Oracle User Group). There is likely a more recent version at www.orapub.com



One solution for this overwhelming monitoring and analysis problem is to use machine learning.

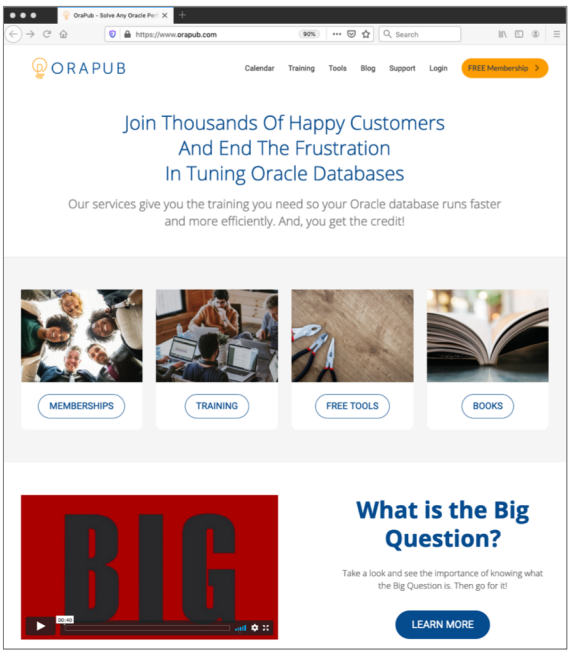
How to use machine learning to quickly and automatically detect an anomalous performance situation (before the phone rings), using an unsupervised algorithm, is what this presentation is all about.

ORAPUB

About Me...

- Long time Oracle DBA
- Specialize in Oracle Database performance and **machine learning**
- Performance researcher
- Blogger: A Wider View About Oracle Performance Tuning
- Author: Oracle Performance Firefighting and Forecasting Oracle Performance.
- Conference speaker
- Teacher and mentor
- Oracle ACE Director
- Quest/IOUG DBA Track Manager

ORAPUB



OraPub works with IT to deploy machine learning into their monitoring and alerting processes.

OraPub works with Oracle DBAs empowering them to beat bots, AI, machine learning and autonomous anything.

ORAPUB

Want my updated slides, Python code, data set and view the demo recording?

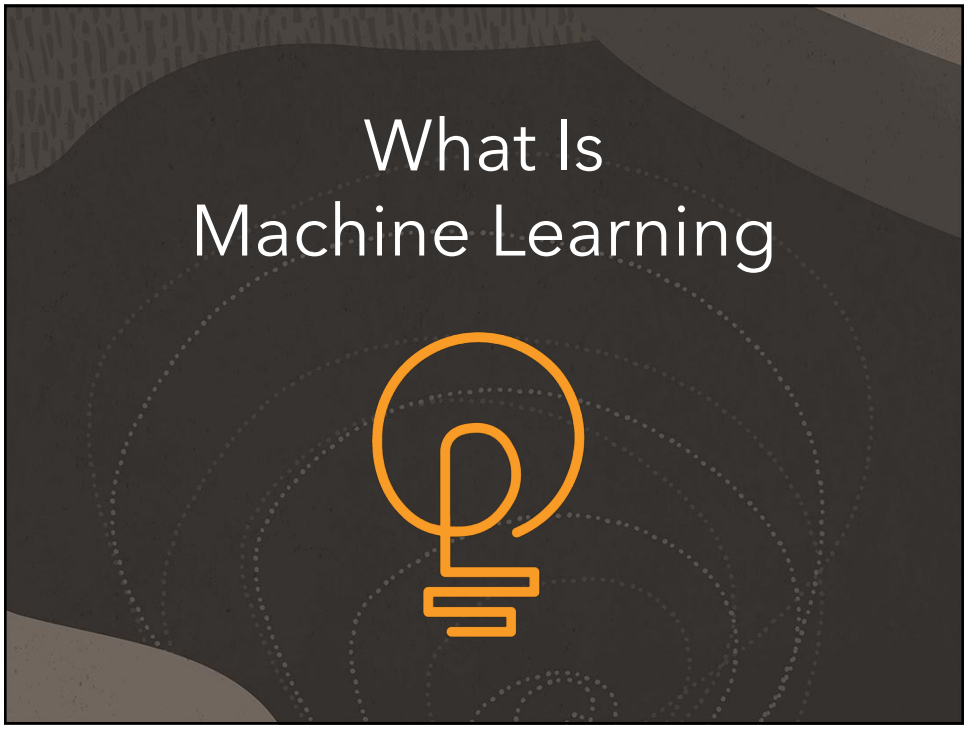
`www.orapub.com > (become a free member) > Login > Tools & Presentations`

ORAPUB

Our Journey For Today

- What is machine learning (ML)
- How does this relate to me?
- Let's do some machine learning
  - Supervised learning
  - Unsupervised learning anomaly detection
- Demonstration using Python, unsupervised K-means algorithm (and more) to detect an anomalous performance event

ORAPUB

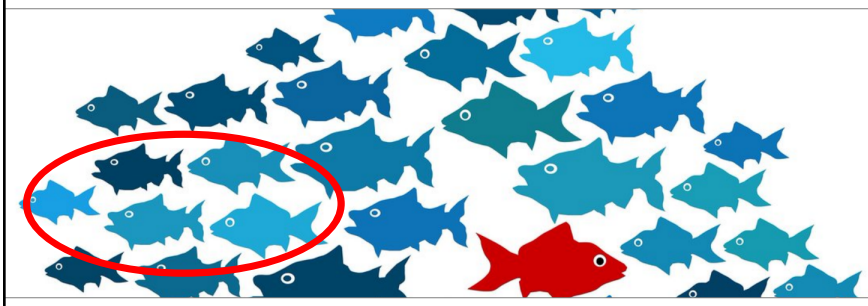


What Is Machine Learning?

- ML fits under the umbrella of AI.
- At it's core, ML is about understanding data; extracting interesting and useful patterns. But this is done **methodically** and using a wide variety of **algorithms**.
- ML contains a growing **set of algorithms** to analyze data. Here is a short list: Support Vector Machines (SVM), Decision Tree Learning, Instance-Based Learning, Generalized Linear Models, Artificial Neural Network, Centroid-Based Clustering, Hierarchical Clustering, Density-Based Clustering.
- ML involves using a variety of advanced statistical and computing **techniques** to process data to find patterns; feature selection, feature engineering, imputation, stratification, principle component analysis, cross fold validation, residual analysis, data transformation, centering and scaling, etc.

ORAPUB

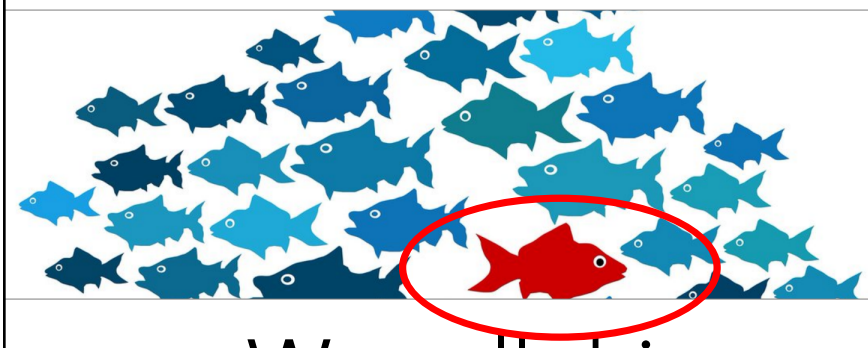
"I've seen this before!"



We call this,
"classification."

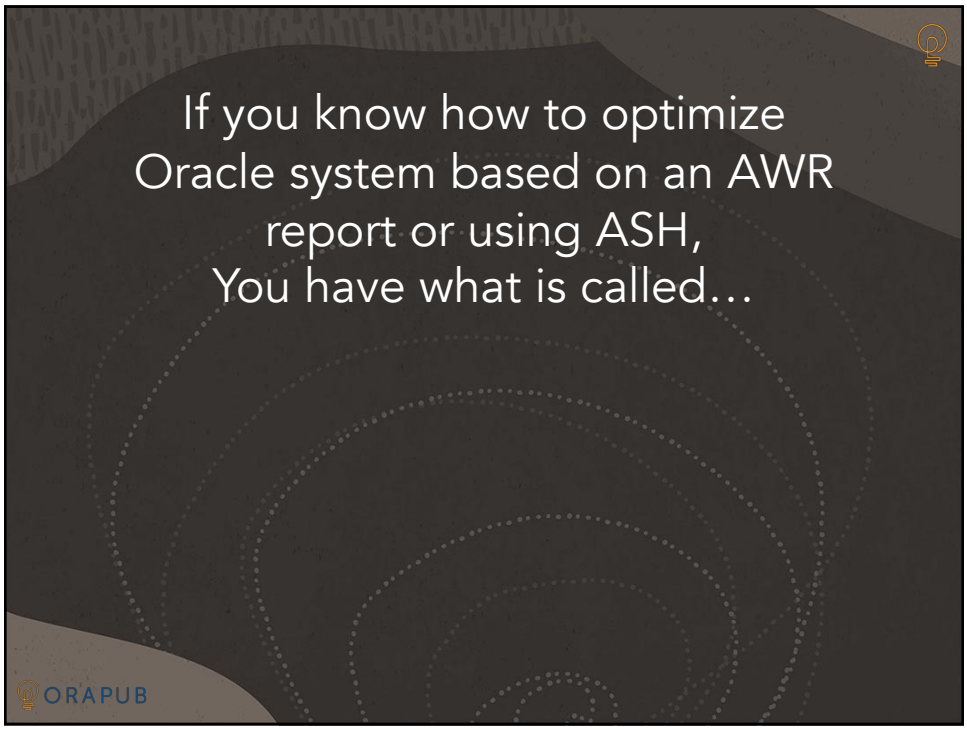
ORAPUB

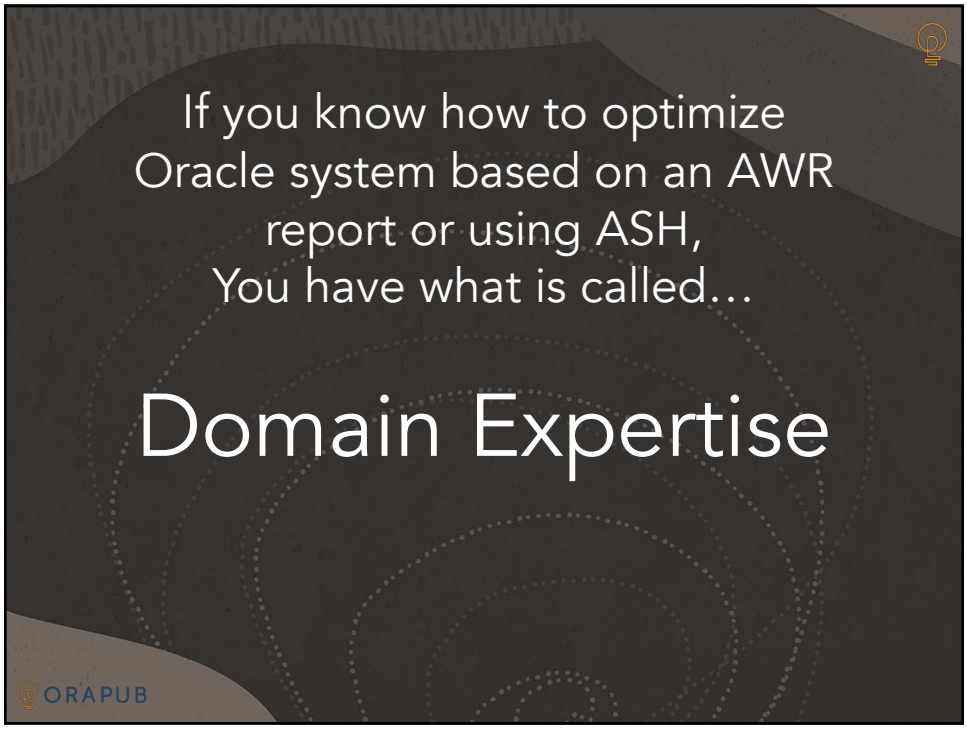
"I've never seen this before!"



We call this,
"anomaly detection."

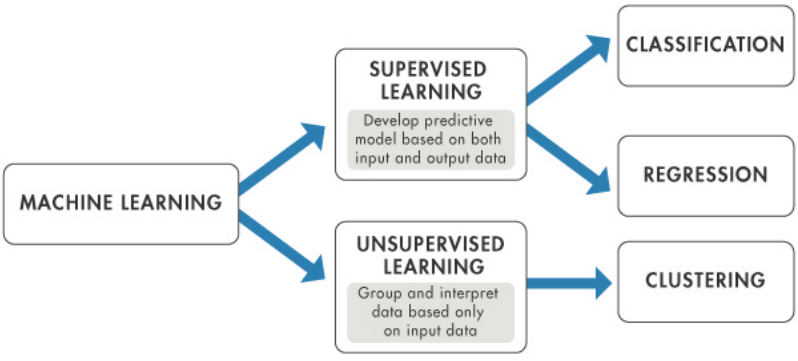
ORAPUB


If you know how to optimize
Oracle system based on an AWR
report or using ASH,
You have what is called…
ORAPUB


If you know how to optimize
Oracle system based on an AWR
report or using ASH,
You have what is called…
Domain Expertise
ORAPUB

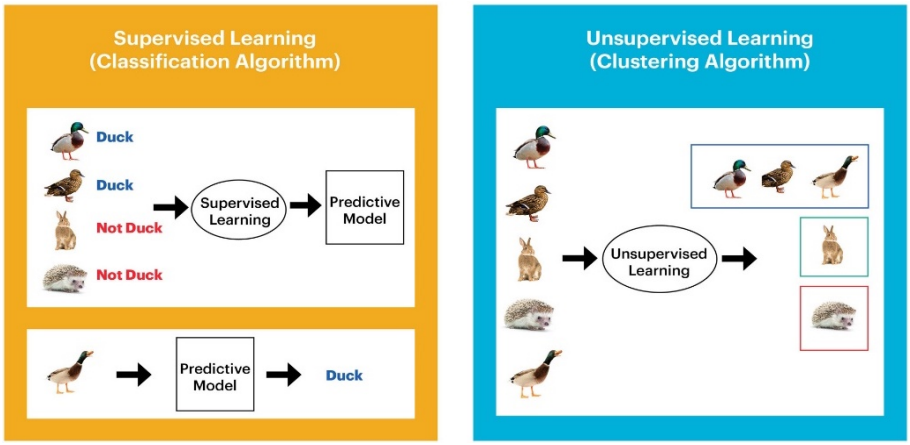
Let's do some Machine learning

Supervised and Unsupervised Learning



ORAPUB

Notice the difference?

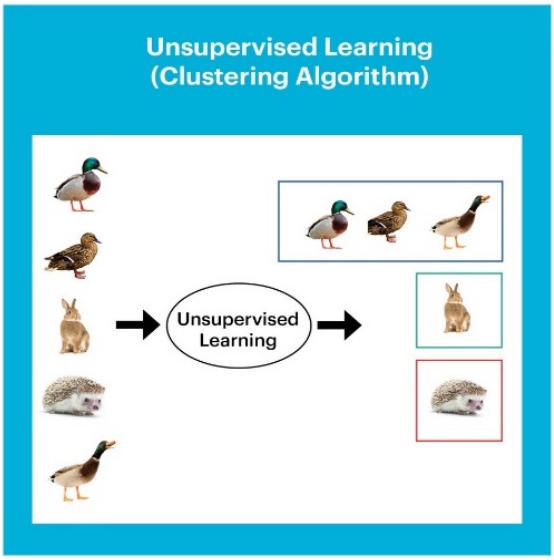


We may know a duck means "bad performance" or perhaps you'll say, "I've never seen a hedge hog before!"

In **unsupervised learning** (UML), no labels are provided, and the learning algorithm focuses solely on **detecting structure**, in unlabeled input data.

ORAPUB

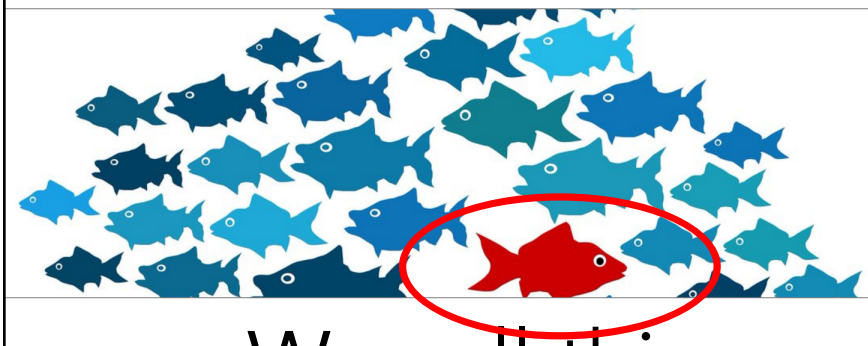
Unsupervised Learning Example



Perhaps "rabbits" are very rare, so if we see one, we should take a closer look.

ORAPUB

"I've never seen this before!"

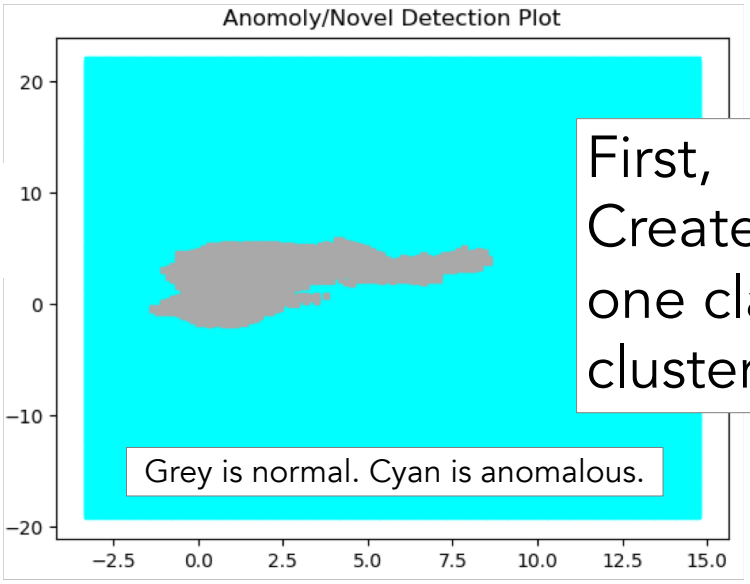


We call this,
"anomaly detection."

ORAPUB

This presentation was given by Craig Shallahamer (craig@orapub.com) at the December 2019 UKOUG (UK Oracle User Group). There is likely a more recent version at www.orapub.com

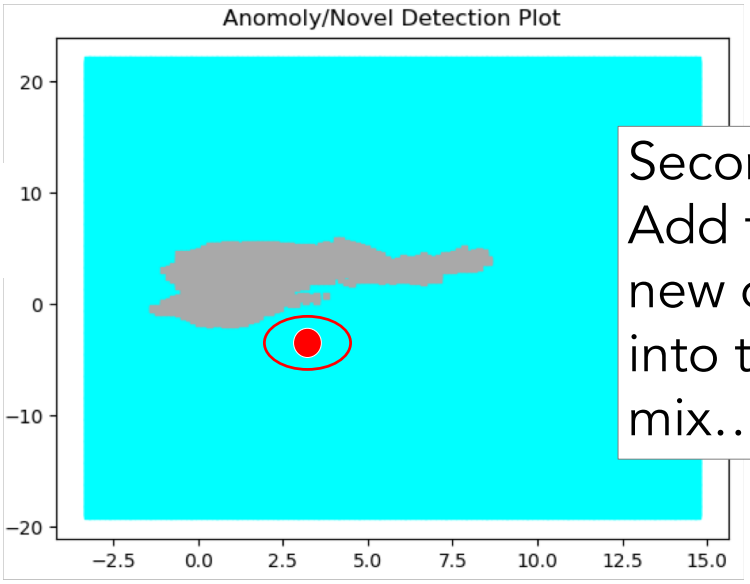
Anomaly Detection



Local Outlier Factor (LOF) is used in this graphic.

ORAPUB

Anomaly Detection

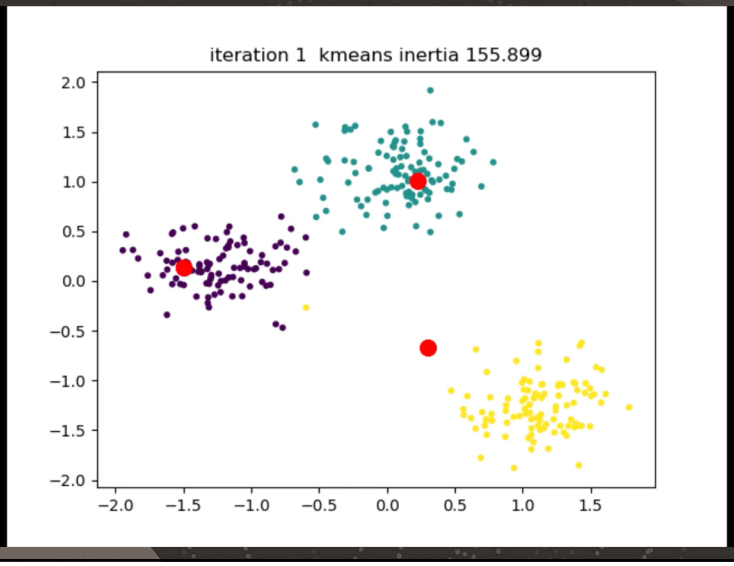


Second, Add the new data into the mix…

Local Outlier Factor (LOF) is used in this graphic.

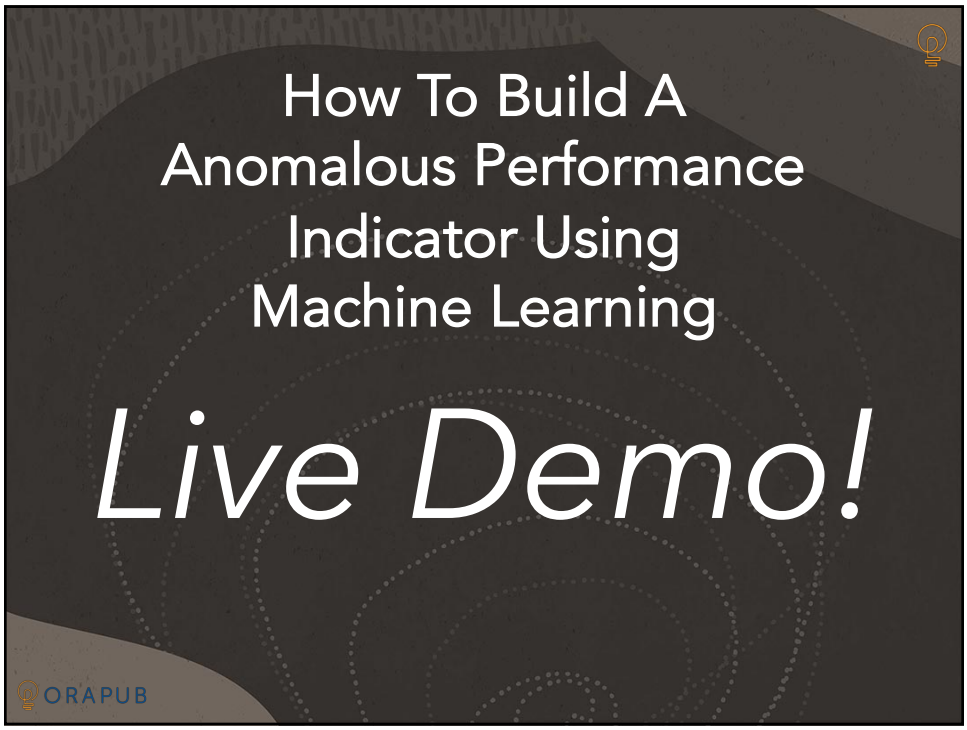
ORAPUB

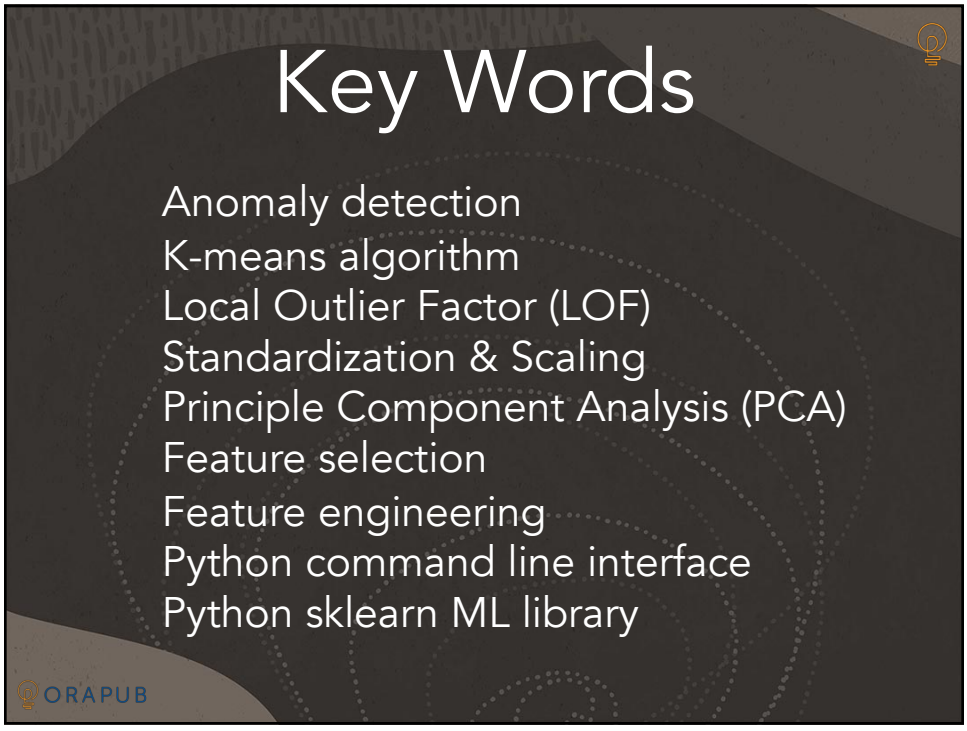
Watch the K-means algorithm finds the best three cluster centers.



ORAPUB

ORAPUB


How To Build A
Anomalous Performance
Indicator Using
Machine Learning
Live Demo!
ORAPUB


Key Words
Anomaly detection
K-means algorithm
Local Outlier Factor (LOF)
Standardization & Scaling
Principle Component Analysis (PCA)
Feature selection
Feature engineering
Python command line interface
Python sklearn ML library
ORAPUB

Demonstration #1

Determine if recent database activity is unusual enough to require a closer look.

ORAPUB

Many times Oracle performance data is in a relationally normalized format:

```
create table perf_stats as (
   snap_id number,
   metric_name varchar2(100),
   metric_value number
)
```

For ML work, we usually need to massively de-normalize the data:

```
create table perf_stats as (
   snap_id number,
   us_psec number,
   aas number,
   trx_psec number
)
```

ORAPUB

We can de-normalize the data by "pivoting."
Here is an example, before and
after we "pivot" the data.

Before Pivot:

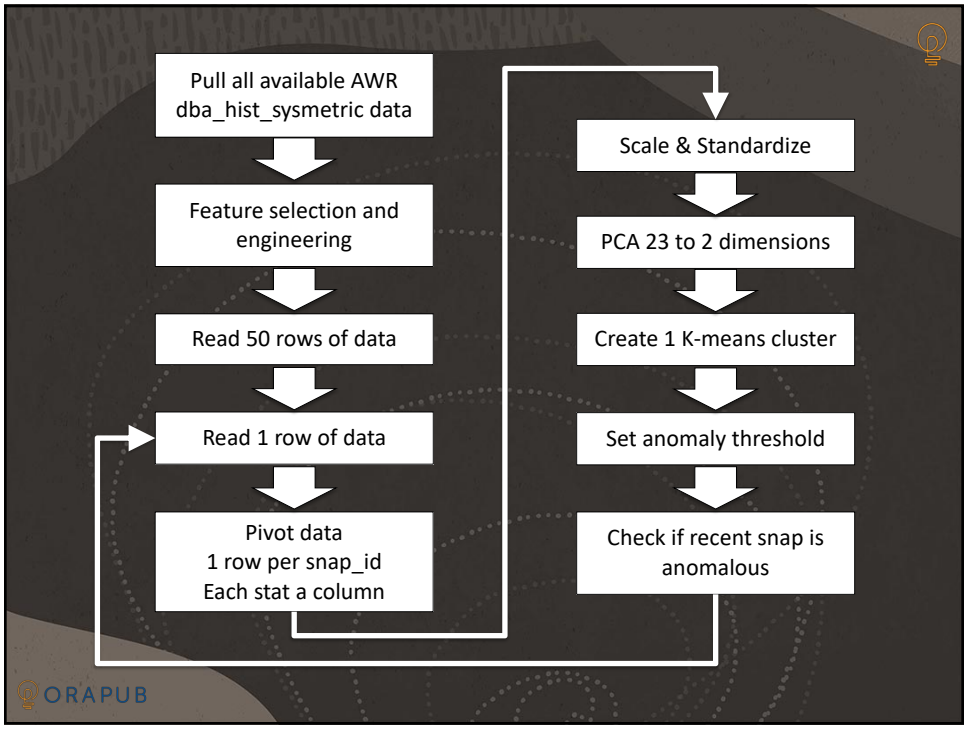
snap_id	metric_name	metric_value
1001	uc_psec	2500
1001	aas	34.25
1001	trx_psec	9.45
1002	uc_psec	1200
1002	aas	14.50
1002	trx_psec	6.50

After de-normalization, i.e., Pivot:

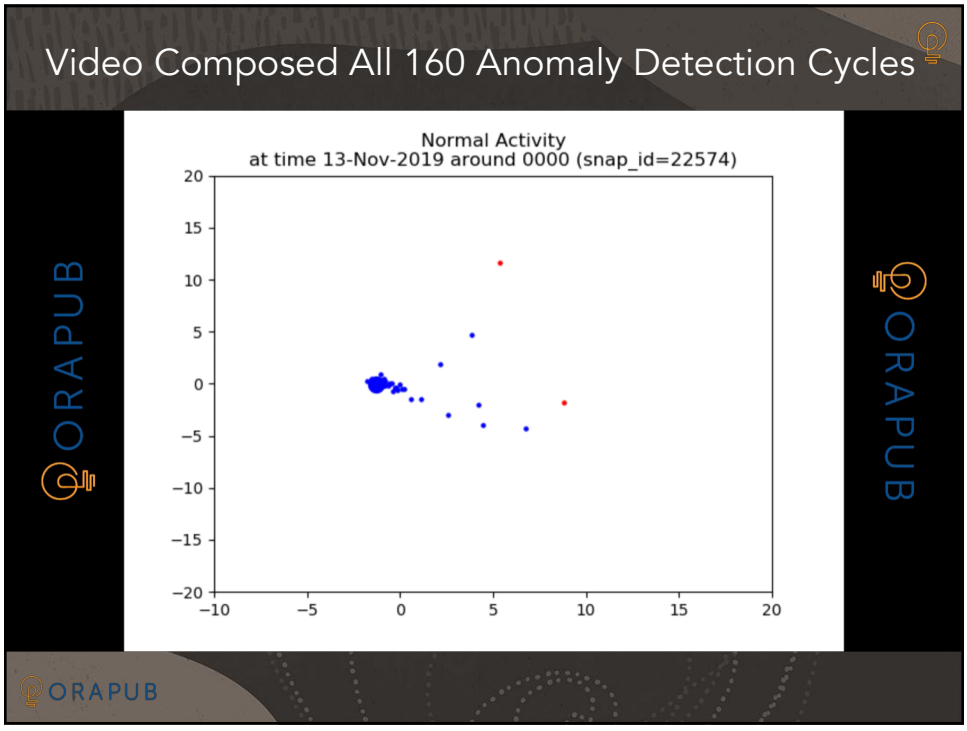
snap_id	uc_psec	aas	trx_psec
1001	2500	34.25	9.45
1002	1200	14.50	6.50

ORAPUB

This presentation was given by Craig Shallahamer (craig@orapub.com) at the December 2019 UKOUG (UK Oracle User Group). There is likely a more recent version at www.orapub.com




Demo!
ORAPUB
Demo-UML-Anom-1c.txt
# Load Python Libraries & Set Gobal Variables
import numpy as np
import pandas as pd
import matplotlib
import os
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
from sklearn.cluster import KMeans
from matplotlib import pyplot as plt
from numpy import linalg as LA
# Define objective
"""
This file contains the complete demonstration steps for how to build
and use an unsupervised single cluster machine learning model to detect
an anomalous Oracle performance situation that warrent an analyst's attention.
Data from the AWR dba_hist_sysmetric_history table is used to build the model.
"""
ORAPUB


Video Composed All 160 Anomaly Detection Cycles
Normal Activity
at time 13-Nov-2019 around 0000 (snap_id=22574)
20
15
10
5
0
−5
−10
−15
−20
−10
−5
0
5
10
15
20
ORAPUB
ORAPUB
ORAPUB

Imagine…

500+ Oracle databases gathering AWR data every 30 minutes. After each snap, your anomaly detection routine is automatically run, checking for activity that warrants a DBA to get involved.

ORAPUB

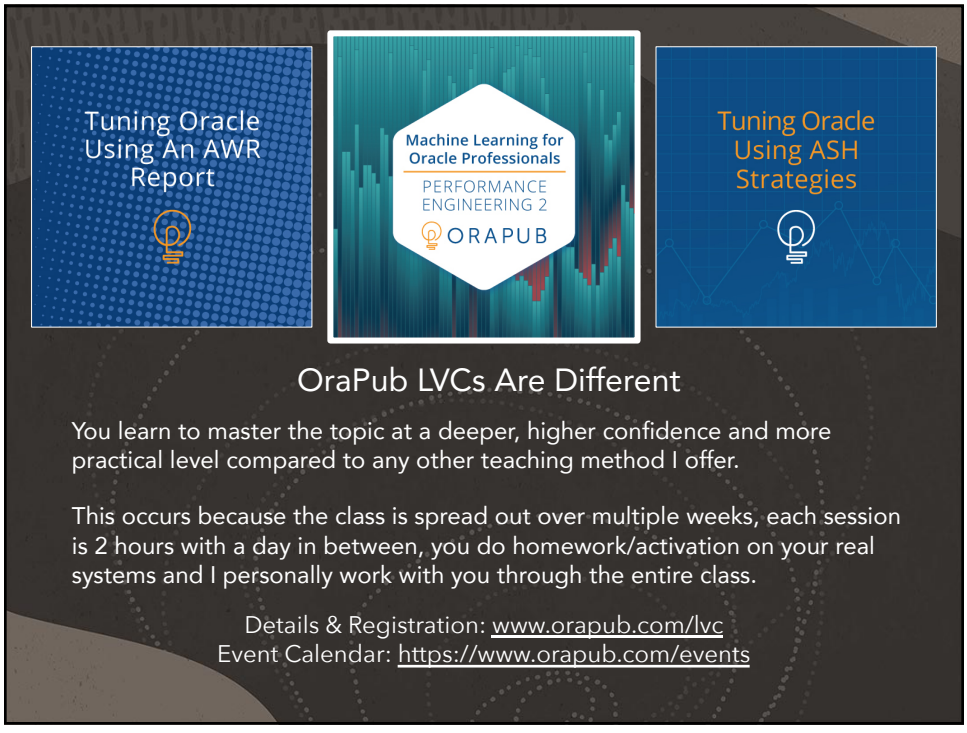
OK. I'm interested. What should I do?

ORAPUB

Want my updated slides, Python code, data set and view the demo recording?

`www.orapub.com >`
`(become a free member) >`
`Login >`
`Tools & Presentations`

ORAPUB



Thank You!

ORAPUB