

Let's Build An
Automated
Anomalous DB
Activity Detector
Using Machine
Learning

Craig Shallahamer
craig@orapub.com

ORACLE
ACE Director

ORAPUB

NYOUG

This presentation was given by Craig Shallahamer (craig@orapub.com) at the October 13, 2020 NYOUG Conference. There is likely a more recent version at www.orapub.com > Login > Presentations

Oracle performance issues typically fall into two categories. Either "I've seen this before and it's bad!" or "I've never seen this before. We better check it out!"

The good news is, a trained analyst with many years of experience can quickly do an AWR or ASH analysis.

The bad news is, **this DOES NOT SCALE!** Even an expert can't comfortably monitor hundreds or thousands of databases.

And our **RULE BASED SYSTEMS** are relatively simplistic, because they **CAN'T CAPTURE THE COMPLEXITY** and diversity of activity in a production Oracle system.

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

How to apply machine learning to quickly and automatically detect an anomalous or recognized performance situations (all before the phone rings, is what this presentation is all about.

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

Besides an introduction to machine learning in the Oracle DBA world, how to apply machine learning to quickly and automatically detect an anomalous performance situation, all before the phone rings, is what this presentation is all about.

Anomalous performance does not imply poor performance.

About Me...

- Long time Oracle DBA
- Specialize in **predictive analytics**, **machine learning** and Oracle performance tuning
- 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 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.



The banner features a teal background with a stylized sun and a person's silhouette. A white box at the top contains the URL <https://www.orapub.com/ml-ecourse>. The central text reads "Machine Learning For Oracle Professionals E-Course" above the ORAPUB logo. Below the logo, it says "FREE from OraPub.com" and "OraPub is committed to cross-training Oracle Professionals in Machine Learning".

<https://www.orapub.com/ml-ecourse>

Machine Learning
For
Oracle Professionals
E-Course


 ORAPUB

FREE from OraPub.com

OraPub is committed to cross-training
Oracle Professionals in Machine Learning

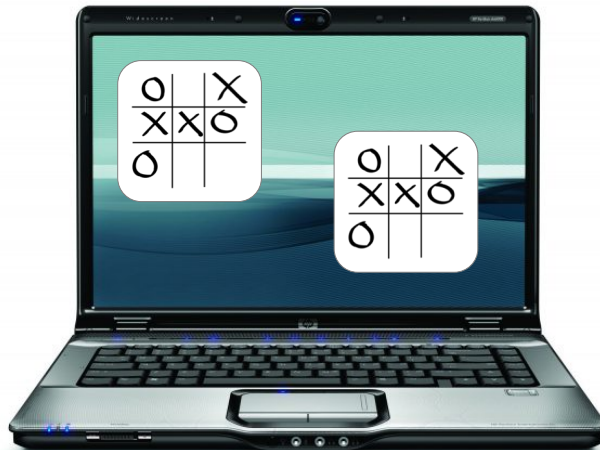
Want my
updated slides?

www.orapub.com >
Login >
Tools & Presentations



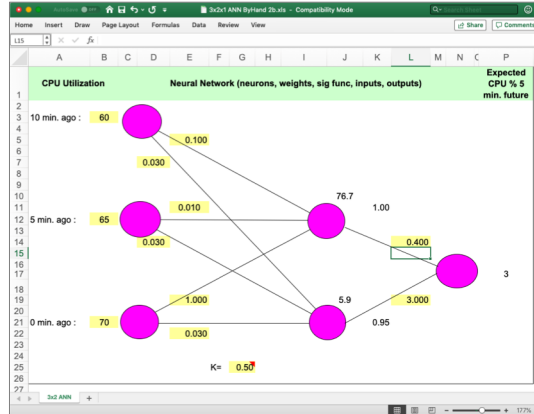


It all started in 1990



I created a program where I could play the game tic-tac-toe with a computer, but also enabled the computer to play itself... and learning through that process.

In 1997 Neural Networks...



I was obsessed is performance prediction... forecasting what was likely to happen or not happen in the future.



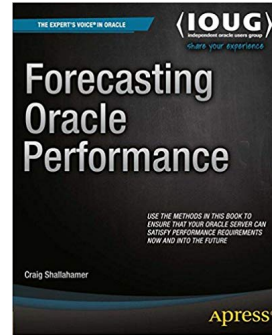
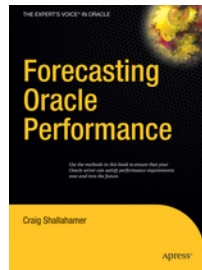
In 2001 at RMOUG

orders (#/day)	floor users (peak #)	batch reports (#/day)	number of mtrs around (#)	new prod? (yes,no)	mgt experience (poor, ok, good, great)	weather (bad, good)	perf issues? (none, some, sig, killer)	orders peak floor ok	batch reports	mgt experience	new prod? (yes, no)	weather	perf issues	input	output	line #
548	148	1500	10	no	ok	bad	none	010 010 011 0 01 0 00	0100100110110010	00	*	1	0100100110110010	00	*	1
553	254	1435	8	no	good	good	some	010 101 010 010 0 10 1 01	0101010100100101	01	*	2	0101010100100101	01	*	2
860	275	2760	12	no	good	bad	sig	011 101 101 100 0 10 0 10	01110110110000100	10	*	3	01110110110000100	10	*	3
910	150	3401	15	no	ok	bad	killer	011 011 110 101 0 01 0 11	0110111101010010	11	*	4	0110111101010010	11	*	4
250	300	743	6	no	good	good	none	001 110 001 001 0 10 1 00	0011000100101010	00	*	5	0011000100101010	00	*	5
1465	345	3290	14	no	poor	bad	some	101 110 110 101 0 00 0 01	1011101101010000	01	*	6	1011101101010000	01	*	6
140	135	734	10	yes	good	bad	none	000 010 001 011 1 10 0 00	0000100010111100	00	*	7	0000100010111100	00	*	7
1010	150	2500	10	no	ok	bad	some	100 011 101 011 0 01 0 01	1000111010110010	01	*	8	1000111010110010	01	*	8
550	234	1465	10	no	ok	bad	some	010 100 010 011 0 01 0 01	0101000100110010	01	*	9	0101000100110010	01	*	9
1243	354	1102	10	no	ok	bad	sig	100 111 010 011 0 01 0 10	1001110100110010	10	*	10	1001110100110010	10	*	10
248	73	354	5	no	good	good	none	000 001 000 001 0 10 1 00	0000100000101010	00	*	11	0000100000101010	00	*	11
576	153	1465	10	no	ok	good	none	010 011 010 011 0 01 1 00	0100110100110011	00	*	12	0100110100110011	00	*	12
865	350	1764	10	yes	poor	bad	killer	011 111 011 011 1 00 0 11	0111110110111000	11	*	13	0111110110111000	11	*	13
186	106	905	10	no	ok	bad	none	000 010 001 011 0 01 0 00	0000100010110010	00	*	14	0000100010110010	00	*	14
189	156	1100	10	yes	ok	bad	sig	000 011 010 011 1 01 0 10	0000110100111010	10	*	15	0000110100111010	10	*	15
600	1	2000	5	no	ok	good		010 000 100 001 0 01 1	0100001000010011				0100001000010011			

In 2001, after two years of rejections, I presented on ANNs at the RMOUG conference. My goal was to predict poor performance in the near future. While ANN is sound, I was never able to realize my "near future" dream. But I believe that dream could now be realized.



2007



What Is Machine Learning



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.



What's the big deal?
What has changed?



Why All The Buzz Now?

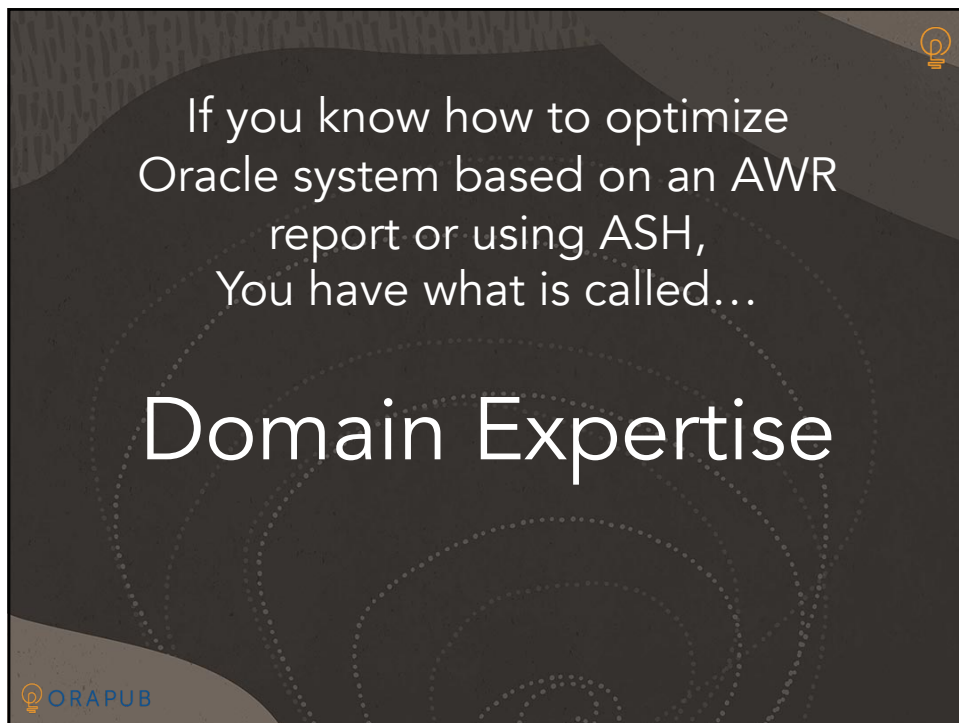
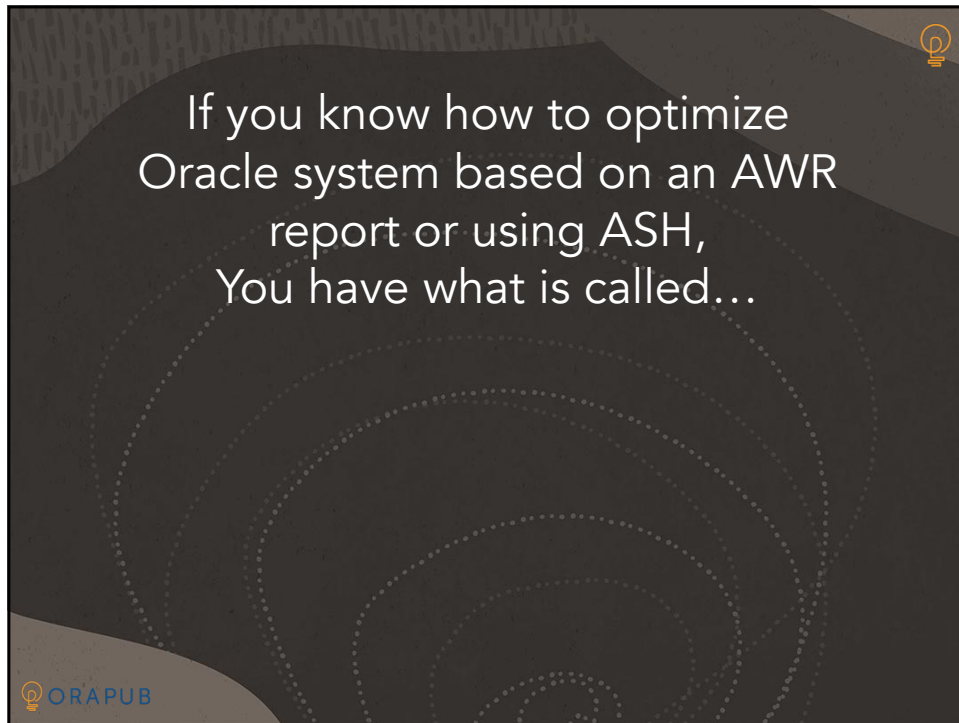
- **Management looking to implement AI** to demonstrate they are cutting edge, reduce repetitive and timing consuming tasks, and do things their competition can't.
- **Software.** Powerful, "free" and available to anyone. It's available on Windows, Mac OS X and Linux. And, the serious number crunching can be run in parallel!
- **Python** and **R** are free statistical packages. Both have powerful libraries, enabling a wide variety of machine learning tasks. Did I forget to mention they are free?
- **Oracle & ML.** Move the algorithms; Not the Data!
 - A tremendous amount of time and effort is spent simply moving data around so it can be processed.
 - All the Oracle Database data and transaction management is, of course, available.
 - Oracle is moving more and more ML processes directly into Oracle, including the Autonomous Database. These are and will be accessible through SQL commands and PL/SQL procedures.
- **Computing Power.** CPUs and GPUs (fast matrix multiplications) are crazy fast and ML software is designed to take advantage of this power. Processing can be in a cloud.
- **Lots of data** is now available. Repeat.
- **Organizational Impact.** ML is touching all business aspects in new ways... IT !
- **Domain Expertise.** **Oracle performance analysts have a unique advantage and place in this movement.**



How does this relate to
me?

I'm an Oracle
professional!





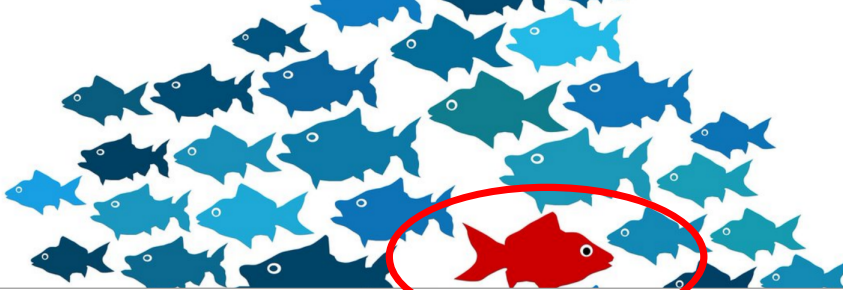


"I've seen this before!"

A group of stylized fish. Most are blue, but one is red. A red circle highlights a small group of blue fish on the left side of the group. The fish are arranged in a loose, overlapping pattern.

We call this,
"classification."

"I've never seen this before!"



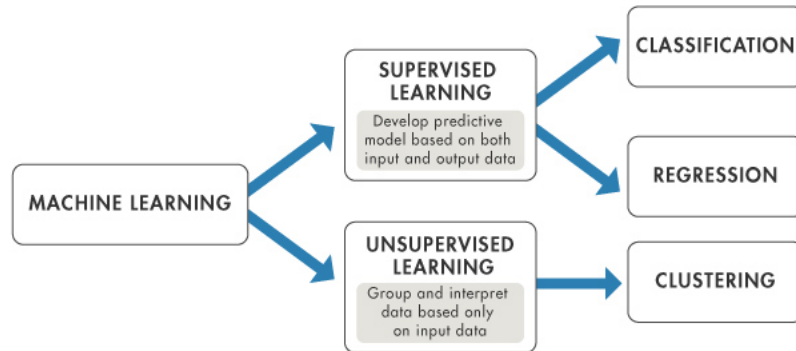
We call this,
"anomaly detection."



Supervised
Vs
Un-Supervised



Supervised and Unsupervised Learning



ORAPUB

This presentation was given by Craig Shallahamer (craig@orapub.com) at the October 13, 2020 NYOUG Conference. There is likely a more recent version at www.orapub.com > Login > Presentations

Supervised Vs Un-Supervised

snap_id	uc_psec	aas	trx_psec	perf
1001	2500	34.25	9.45	bad
1002	1200	14.50	6.50	good
1003	1150	16.50	16.50	good
1004	1250	18.50	9.50	good
1005	1300	24.50	2.50	bad

Label

Supervised

Will snap_id 1005 result in "the phone ringing?"

snap_id	uc_psec	aas	trx_psec
1001	2500	34.25	9.45
1002	1200	14.50	6.50
1003	1150	16.50	16.50
1004	1250	18.50	9.50
1005	1300	24.50	2.50

Un-Supervised

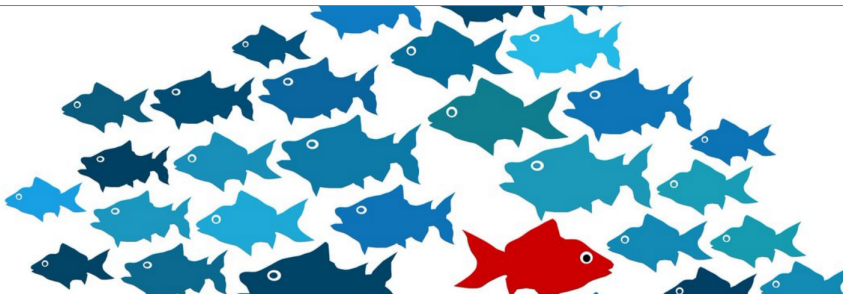
Is snap_id 1005 anomalous?

ORAPUB

Un-Supervised Learning With One Cluster



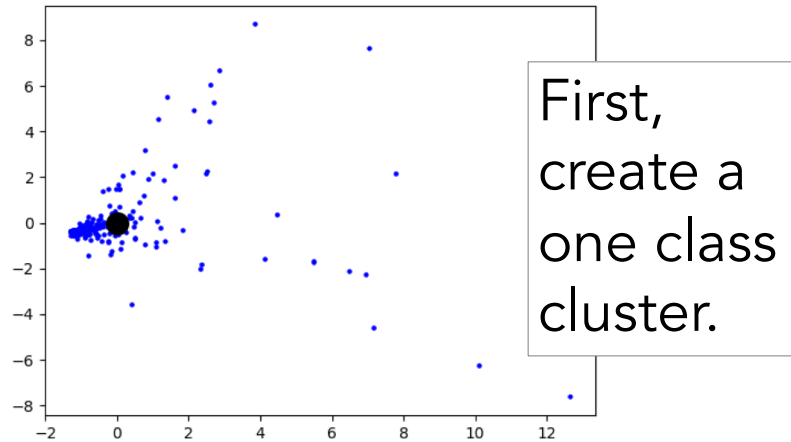
"I've *never* seen this before!"



We call this,
"anomaly detection."

Anomaly Detection

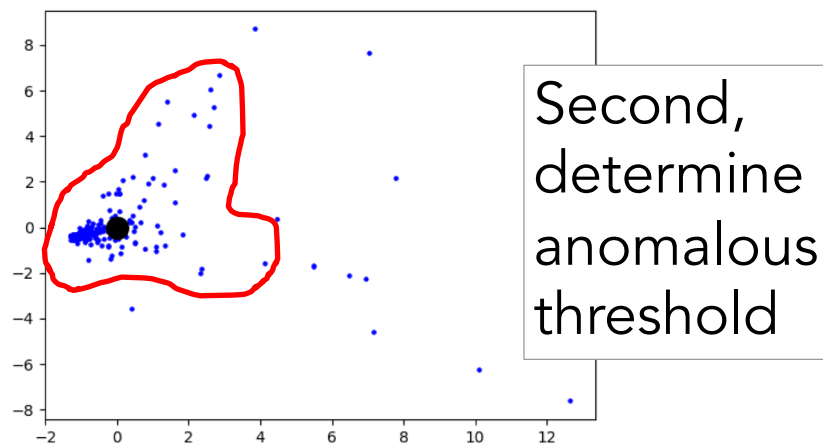
(I've never seen this before!)



ORAPUB

Anomaly Detection

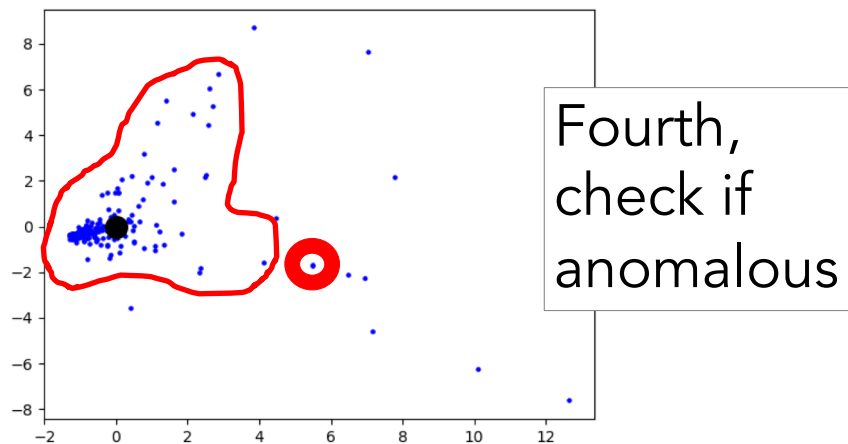
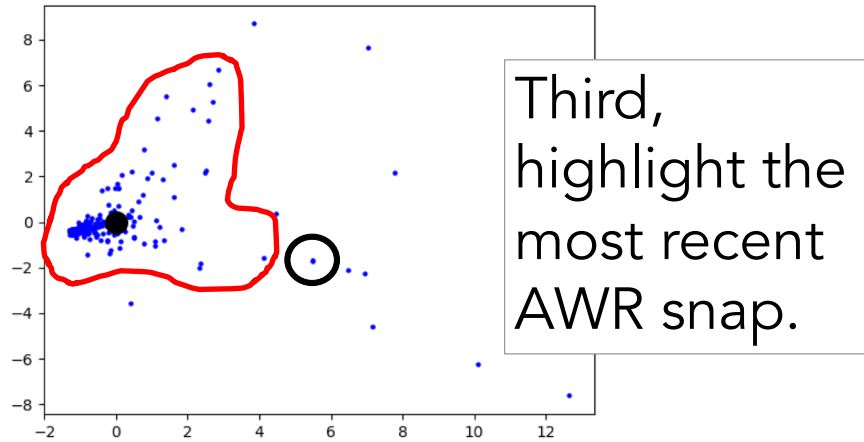
(I've never seen this before!)



ORAPUB

Anomaly Detection

(I've never seen this before!)





Let The Typing Begin

To Set Up Your Machine Learning Environment:
[blog.orapub.com/20200107/
how-to-setup-your-machine-learning-sandbox-environment.html](https://blog.orapub.com/20200107/how-to-setup-your-machine-learning-sandbox-environment.html)

You can watch the video on the presentation page:
www.orapub.com > Login > Tools & Presentations



G o i n g D e e p e r

Resource listing

- **OraPub Membership for premium & exclusive content**
 - "How To" Webinars - two each month. Over 110 recorded!!
 - Video Seminars - any device, any time, high quality
 - Learning paths, mentoring, assessments and certificates, priority response
 - 20% LVC discounts
 - You're at the "top of my inbox"
 - SLACK forum exclusively for paid members
- **Live Virtual Classroom (LVC) Training**
 - **Machine Learning For Oracle Professionals**
 - Tuning Oracle Using An AWR Report
 - Tuning Oracle Using Active Session History (ASH) Strategies
 - Core Truths For Oracle Professionals
- **Toolkits** - Many tools available at orapub.com
- **Craig's Blog & Website**
- **Presentations** - www.orapub.com
- **Books:** Oracle Performance Firefighting. Forecasting Oracle Performance.



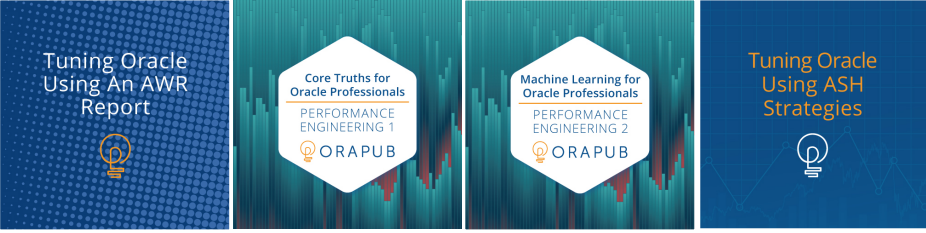
<https://www.orapub.com/ml-ecourse>

Machine Learning
For
Oracle Professionals
E-Course

 ORAPUB

FREE from OraPub.com

OraPub is committed to cross-training
Oracle Professionals in Machine Learning




OraPub LVCs Are Different

You learn to master the topic at a deeper, higher confidence and more practical level compared to any other teaching method I offer.


This occurs because the class is spread out over multiple weeks, each session is 2 hours with a day in between, you do homework/activation on your real systems and I personally work with you through the entire class.

Details & Registration: www.orapub.com/lvc
Event Calendar: <https://www.orapub.com/events>




Machine Learning for Oracle Professionals

PERFORMANCE ENGINEERING 2



and more
pr.
aks, each session
n on your real
ass.
nts



Let's Build An
Automated
Anomalous DB
Activity Detector
Using Machine
Learning

Craig Shallahamer
craig@orapub.com

