Safe Harbor Statement

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described for Oracle’s products remains at the sole discretion of Oracle.
Machine Learning Turbo-Charges the Ops Portion of DevOps
CMG (Computer Measurement Group) Canada

Ben Reader
Oracle Management Cloud Specialist
April, 2017
Program Agenda

1. Defining terms
2. Why (Dev)Ops is Perfect for Machine Learning
3. Making Machine Learning Smarter
4. Q&A
Program Agenda

1. Defining terms
2. Why (Dev)Ops is perfect for machine learning
3. Making Machine Learning Smarter
4. Q&A
• **Machine Learning**
  – Machine learning is the subfield of computer science that gives computers the ability to learn without being explicitly programmed. Evolved from the study of pattern recognition and computational learning theory in artificial intelligence, machine learning explores the study and construction of algorithms that can learn from and make predictions on data.

• **DevOps**
  – DevOps (a clipped compound of "software DEVelopment" and "information technology OPerationS") is a term used to refer to a set of practices that emphasize the collaboration and communication of both software developers and information technology (IT) professionals while automating the process of software delivery and infrastructure changes.

• **Systems Management or IT Operations Management**
  – IT Operations is responsible for the smooth functioning of the infrastructure and operational environments that support application deployment to internal and external customers, including the network infrastructure; server and device management; computer operations; IT infrastructure library (ITIL) management; and help desk services for an organization.
Program Agenda

1. Defining terms
2. Why (Dev)Ops is perfect for machine learning
3. Making Machine Learning Smarter
4. Q&A
We have a problem: Dev has outpaced Ops

- Development is creating faster...
  - Low-code
  - Agile
  - Microservices
  - CI

- (Dev)Ops is promoting faster...
  - Containers
  - IaaS & PaaS
  - CD
  - Packages

- (the rest of) Ops is not moving any faster...
  - #(*^#) & ^$ (@ @ ($ $ ($ @ ) $ & $ ^ $ * & ) ! ! !
  - #(*^#) & ^$ (@ @ ($ $ ($ @ ) $ & $ ^ $ * & ) ! ! !
  - ...

Copyright © 2017, Oracle and/or its affiliates. All rights reserved.
One of Two Likely Outcomes, Both Bad

OPTION 1: Your Changes Don’t Hit Production Until Ops is Ready

Option 2: You Promote Unmanaged Code Anyway
The Reason: Ops Depends on Human Effort

Where’s the data?  What does the data mean?

It's not my code, it's your code!

It's not my machines, it's your machines!
We Can Help!  Ops Data is Perfect for Machine Learning

- Structured, Time-Series
  - User Performance Metrics
  - Server-side Performance Metrics (App & Infrastructure)
  - Configurations
  - Events/Alerts
  - Transaction Payloads
- Unstructured Text
  - Log Records

- Massive volume
- Highly patterned
- Predictable format
- Silos can be unified
- Seasonal trends
- Known sources
Algorithmic Approaches to IT Ops Data

- Structured, Time-Series
  - User Performance Metrics
  - Server-side Performance Metrics (App & Infrastructure)
  - Configurations
  - Events/Alerts
  - Transaction Payloads

- Unstructured Text
  - Log Records

ANOMALY DETECTION
CLUSTERING
CORRELATION
PREDICTION
Program Agenda

1. Defining terms
2. Why (Dev)Ops is perfect for machine learning
3. Making Machine Learning Smart for IT Ops
4. Q&A
We Know The Questions We Want To Ask of IT Ops Data

- What caused the problem?
- How is this application actually built/architected?
- How should I rebalance workloads?
- How do I prevent the problem in the future?
- Is what I’m seeing normal or abnormal?
- What do I need to pay attention to right now?
- What areas can I improve, and how?
- WHAT WILL HAPPEN TOMORROW?
Maturing Machine Learning: A Three-Step Approach

ML is not smart out of the box for every question

To make ML smarter, know the questions you want to ask, then...

1. Enhance Algorithms
2. Increase Breadth
3. Increase Depth
Working Example 1: Enhancing Anomaly Detection

- Begin with the Basics
  - Distribution Based Unseasonal Model
  - Daily + Weekly Additive Holt-Winter Modeling
  - Automatic Season Detection

- Tune Based on Validation
  - Robust to Sparse Pattern Variability
  - Robust to Small Anomalies
  - Graceful Transition from Daily-to-Weekly
  - Evaluation Model Segmentation
9x False Positive Reduction With Seasonality Enhancements

**Before**: Anomalies are out-of-band samples.

**After**: Anomalies are statistically significant out-of-band samples.

**Before**: Flagged as an anomaly due to load/measurement variability.

**After**: Computing baselines at higher scale (hourly, configurable) solves this problem.

**Before**: Weekdays and weekends are allowed to be imbalanced.

**After**: Select days to keep weekday-weekend balance.
5x Performance Boost with Data Pipeline Segmentation

Unseasonal, Daily, Weekly Models for Metric #1

Unseasonal, Daily, Weekly Models for Metric #2

Base Line Model Cache

Data Pipeline
Working Example 2: Enhancing Prediction/Early Warning

• Begin with the Basics
  – Robust Linear Regression for Unseasonal
  – Automatic Season Detection
  – Tolerance Intervals

Tune Based on Validation
  – Season Specific Trending-Uncertainty
  – Regime Change Detection
  – Seasonal Pattern Trending
  – Temporal Weighting
2x Improved Forecast Accuracy with Enhanced Trending


2. Unsupervised seasonality sometimes selects subtle highs.

3. Incorrect trend b/c small set, sparse high early in week, & most near low.

Better treatment is to trend the “pattern” not “samples”.

BEFORE

AFTER
Additional Accuracy Using Temporal Weighting

• Example: a common based approach to robust linear regression is Thiel-Sen Estimator

• Thiel-Sen can be enhanced with specific tuning that accounts for expected seasonality in the underlying (IT Ops) data set, for example:
  – Business day vs after-hours
  – Same business day, different weeks
  – Frequency/periodicity of samples
DEMO: Matured Machine Learning in Action