

# GSAW 2018

## Machine Learning

Space Ground System Working Group

*Move the Algorithms; Not the Data!*

Dan Brennan  
Sr. Director Mission Solutions  
[daniel.p.brennan@oracle.com](mailto:daniel.p.brennan@oracle.com)  
Feb, 2018

# 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.

# 2001: A Space Odyssey

## The Dawn of Man scene

- Adoption of machine learning and “artificial intelligence” is about at this stage
- the BEGINNING!



Dawn of Man Scene in, *2001: A Space Odyssey*, produced and directed by Stanley Kubrick, 1968

# Key messages:

- Ground System ML/AI is not a unique “Space” Challenge – It’s a IT, Data Management, Analytics challenge
- Move the Algorithms – not the Data
  - Moving Data creates Platform Sprawl: Architecture Complexity, Duplicated Data, Data Latency, Data Consistency Issues, Security Exposures, and Duplicated Storage, Backup, Systems, etc/etc
- Leverage Commercial Technology – Private R&D Investment in ML is Several Orders of Magnitude more than Government investment in this field.
  - And it’s Moving Fast
- Evolve towards a combined data management + advanced analytics environment that can analyze data, perform machine learning and essentially “think”
- Don’t throw away historical Data – That’s Training Data!
- Operational ML/AI Solution must enable Timely Deployment of Analytic Models



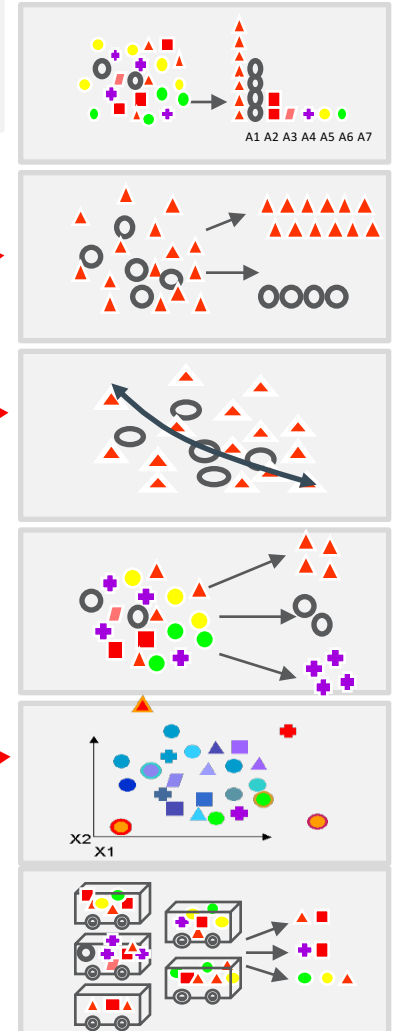
# Example of Machine Learning in Industries

- Financial
  - Enterprise Risk Management,
  - Financial Crime and Compliance
  - Credit Score/analysis
  - Customer Relationship/marketing
  - Customer Insight
- Retail B2C
  - Market Basket Analysis
  - Event Based Marketing
  - Purchased X – Recommend Y
  - Customer Segmentation
  - Customer Loyalty
  - Sales Predictions
- Industrial
  - Predictive Fault Monitoring
- Health Care
  - Illness pattern analysis
  - Patient Care & Quality Analysis
- Human Capital Management (HCM)
  - Employee turnover, performance prediction and “What if?” analysis
- Government
  - Threat Detection
  - Cyber/Trend Analysis
  - System Failure prediction
  - Computer Vision
  - Sentiment Analysis
- IT Infrastructure
  - IDAM: Real-time security and fraud analytics
  - Autonomous Database
  - Customer Support: Predictive Incident Monitoring

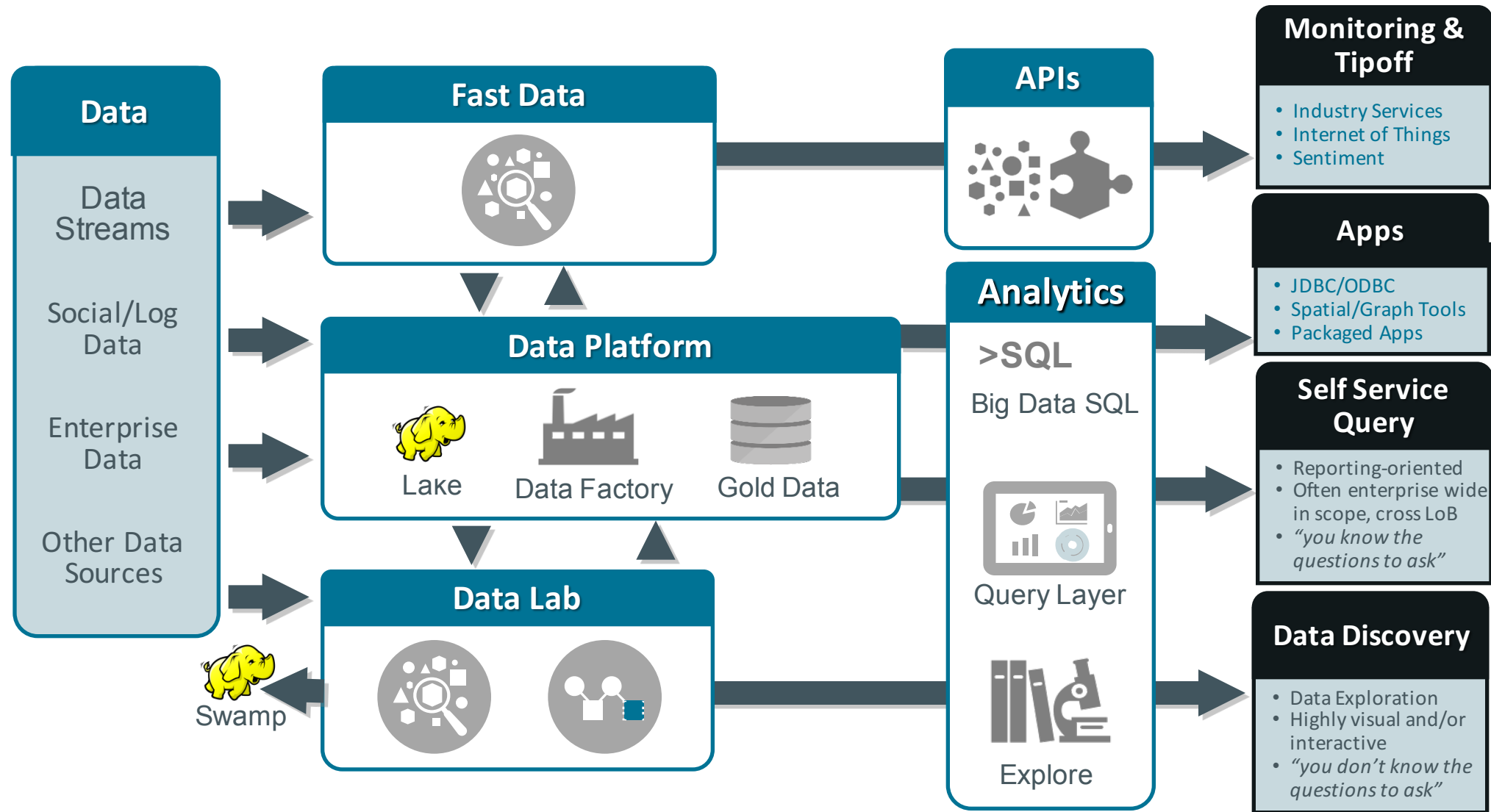
# What is Machine Learning, Data Mining, Predictive Analytics?

*Automatically* sift through **large amounts** of data to find **hidden patterns**, **discover new insights** and **make predictions**

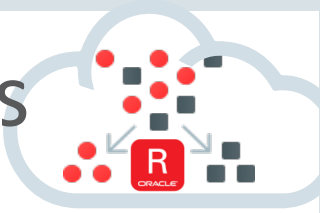
- Identify most important factor (*Attribute Importance*)
- Predict some customer behavior (*Classification*)
- Predict or estimate a value (*Regression*)
- Find profiles of targeted people or items (*Decision Trees*)
- Segment a population (*Clustering*)
- Find fraudulent or “rare events” (*Anomaly Detection*)
- Determine co-occurring items in a “baskets” (*Associations*)



# Oracle Conceptual Data Analytics Platform



# Oracle's Machine Learning & Adv. Analytics Algorithms



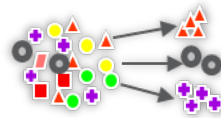
## CLASSIFICATION

- Naïve Bayes
- Logistic Regression (GLM)
- Decision Tree
- Random Forest
- Neural Network
- Support Vector Machine
- Explicit Semantic Analysis



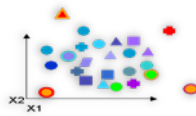
## CLUSTERING

- Hierarchical K-Means
- Hierarchical O-Cluster
- Expectation Maximization (EM)



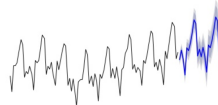
## ANOMALY DETECTION

- One-Class SVM



## TIME SERIES

- Holt-Winters, Regular & Irregular, with and w/o trends & seasonal
- Single, Double Exp Smoothing



## REGRESSION

- Linear Model
- Generalized Linear Model
- Support Vector Machine (SVM)
- Stepwise Linear regression
- Neural Network
- LASSO



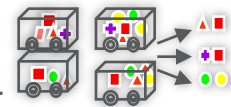
## ATTRIBUTE IMPORTANCE

- Minimum Description Length
- Principal Comp Analysis (PCA)
- Unsupervised Pair-wise KL Div
- CUR decomposition for row & AI



## ASSOCIATION RULES

- A priori/ market basket



## PREDICTIVE QUERIES

- Predict, cluster, detect, features

## SQL ANALYTICS

- SQL Windows, SQL Patterns, SQL Aggregates



## FEATURE EXTRACTION

- Principal Comp Analysis (PCA)
- Non-negative Matrix Factorization
- Singular Value Decomposition (SVD)
- Explicit Semantic Analysis (ESA)

## TEXT MINING SUPPORT

- Algorithms support text type
- Tokenization and theme extraction
- Explicit Semantic Analysis (ESA) for document similarity



## STATISTICAL FUNCTIONS

- Basic statistics: min, max, median, stdev, t-test, F-test, Pearson's, Chi-Sq, ANOVA, etc.



## R PACKAGES

- CRAN R Algorithm Packages through Embedded R Execution
- Spark MLlib algorithm integration



## EXPORTABLE ML MODELS

- C and Java code for deployment



# Potential ML/AI Ground System Resiliency Use Cases

- Premise of the Working Group
  - Platform Telemetry Analysis
  - Anomaly Detection/Prediction
- Global Ground System
  - Optimized Worldwide Comm Planning/Scheduling
  - Constellation Orbital Management
  - Anomaly Analysis/Prediction
    - MOC, Backup MOC, Comm Relay & Tracking Sites
    - Uplink/Downlink RF System Fault
    - Pedestal System
    - IT Fault Analysis/Prediction
  - WX degradation/re-plan Prediction
- Ground Facility
  - Anomaly Detection
    - Power Plant, Cooling, etc
- Product Processing
  - Automated Exploitation
  - Anomaly Detection
- Human Element
  - Employee turnover, performance prediction and “What if?” analysis

# Machine Learning, Analytics and Clouds—Oh My!



## Summary

- Machine learning, predictive analytics & “AI” have become *must-have* capabilities
- Separate islands for data management and for data science don’t work
- *Move the Algorithms, Not the Data!*
- Need to evolve towards a combined data management + advanced analytics environment that can analyze data, perform machine learning and essentially “think”
- *Leverage Extensive Commercial R&D and Investment*
  - Avoid Opportunity Costs of duplicating COTS capabilities



# ANALYTICS AND DATA SUMMIT 2018

All Analytics. All Data. No Nonsense.

March 20-22, 2018

[www.analyticsanddatasummit.org](http://www.analyticsanddatasummit.org)



We've changed our name! Formerly called the BIWA Summit with the Spatial and Graph Summit.

Same great technical content – great new name!