Lumada Analytics: Adding Portability and Scale to Custom Workflows
First, a Question.. What kind of ‘Custom Workflows’ are we talking about?
Answer: Workflows involving machine learning

Machine Learning has gained tremendous velocity over the past decade
What are the drivers fueling ML adoption?
Predictive Analytic techniques widely used today
ML techniques that have been popular over the past decade
Deep Learning Neural Networks

Framing the next decade of ML and Industrial IoT

Achieving Portability and Scale with ML
What are some of the core challenges?
• Precision, reliability, stability
• Support for heterogeneous techniques, languages, tools
• Reduce human involvement at scale

Lumada Analytics and Pentaho
A perfect match for going both deep and wide
What are the Drivers Fueling ML Adoption Today?

- **Cloud Computing**
  - Ubiquitous access to computing power
  - Elasticity on-demand
  - Commodity pricing
- **Exponential increase in data generated**
  - Need strategy to deal with data flooding
  - Search queries, social media, consumer purchases, financial transactions, consumer IoT, industrial IoT, etc.
- **Accumulation of massive datasets**
  - Search queries, social media, consumer purchases, financial transactions, credit history, consumer IoT, industrial IoT, etc.
- **Track record of successful ML implementations in industry**
  - Fraud detection, data security, financial trading, personalization, recommendation, digital assistants, etc.

Rhetorical Question: How much of the data generated is retained? If so, for how long?
1. Simple Methods: Basic predictors, Linear estimation, Historical Patterns, Heuristics, etc.

2. Regression Analysis
   • Linear regression
   • Logistic regression
   • Poisson regression
   • Least Square Regression

3. Machine Learning- Many different techniques to choose from… let’s double click on this…
ML Techniques Popular over the Past Decade

‘Big Ten’ - ML Algorithms Widely Used Today

1. Decision Trees
2. Naïve Bayes Classification
3. Least Square Regression
4. Logistic Regression
5. Support Vector Machines
6. Ensemble Methods
7. Unsupervised Learning
8. Principal Component Analysis
9. Singular Value Decomposition
10. Independent Component Analysis

Source: http://www.kdnuggets.com/2016/08/10-algorithms-machine-learning-engineers.html
Will Deep Learning Dominate the Next Decade?

- Deep Learning is a term most often associated with Artificial Neural Networks (ANN)
- Multi-layer Neural Networks show tremendous promise for discovering and learning patterns in large datasets
- Has gathered significant industry mindshare - but not expected to be dominant in Industrial IoT
- However, perhaps ANNs are less than ideal for many industrial challenges requiring ML
  - Challenges include: Explaining causality, Imposing real-world limitations, catastrophic forgetting, stability of output over time, etc.
What Does Old-guard Academia Think…

...about the potential of Deep Neural Networks?

Tom Mitchell  
CMU

Ray Mooney  
U of TX

Jude Shavlik  
U of WI

Track how technology is transforming work

What machine learning teaches us about the brain

Natural-Language Video Description with Deep Recurrent Neural Networks

CS 638 - Building Deep Neural Networks (Spring 2017)
Machine Learning by Tom Mitchell (1997)

**Tom Mitchell**

**CMU**

<table>
<thead>
<tr>
<th>Format</th>
<th>Price</th>
<th>Prime Price</th>
<th>Used Price</th>
<th>Rent Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hardcover</td>
<td>$89.83 - $226.48</td>
<td>$89.83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paperback</td>
<td>$32.01</td>
<td>$32.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other Sellers</td>
<td>from $85.00</td>
<td>from $85.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**ISBN-13:** 978-0070428072  
**ISBN-10:** 0070428077

*In Stock.* Ships from and sold by Amazon.com. Gift-wrap available. Prime | FREE One-Day

Get it tomorrow, Sept. 20. Order within 14 hrs 43 mins and choose One-Day Shipping at checkout. Details

**More Buying Choices**

- **16 New** from $199.71
- **38 Used** from $85.00
- **2 Rentals** from $89.83

56 offers from $85.00

[Add to Cart](#)
1. A decade dominated by the ubiquitous application of ML techniques that are already proven to work
2. A ‘one size fits all’ or unified approach to ML for Industrial IoT proves illusive
3. A diverse set of predictive analytic techniques and platforms will be relevant in the decade ahead
4. Standardization will be less important than finding something that works (..many IIoT outcomes are easily quantified and will be)
5. Deep Learning variants are on the frontier of research
What are some of the core challenges?

• Precision, reliability, stability
  • Are you OK with a system that is right 99% of the time?
  • How much do you value day-to-day stability?
  • Are you willing to trade a chance for greater precision for some lack of stability?

• Support for heterogeneous techniques, languages, tools
  • Can you afford to live in a ‘walled garden’ or do you need great flexibility?

• Find ways to reduce human involvement at scale
  • What level of automation do you need to be successful at scale?
  • Can you afford to have a human in the loop?
**Target a platform that supports:**

- Embrace and extend existing predictive analytic methods in use today
  - *Support a range of established ML techniques*
- Integrate readily with established & emerging Open Source Software projects
  - Avoid living in a ‘walled garden’
- Support integration of data sources and analytic outcomes with other Enterprise applications
  - *Data Blending from different sources is a core competency*
Lumada: a Composable, Modular platform

- Lumada Analytics runs on top of Lumada Foundry
- Lumada Studio runs on top of Analytics (and other things)
- Custom Solutions are built on top of everything else
Lumada Industrial IoT

Platform

Custom Solutions

Solution Cores

Machines

Humans

Business Systems

Edge

Core

Analytics

Foundry

Catalog

RT
1. **Heterogeneous Engines**: Analytics Platform should support diverse analytics requirements

2. **Scalable**: Analytics Platform has to be (“just add water”) scalable in cloud, on-prem and hybrid environments

3. **Machine Learning**: Analytics platform should power machine learning pipelines (run, train, build model, retrain, test, deploy/publish)

4. **Extensible**: Extending the Analytics Platform should be easy, painless and friction-free
Lumada Analytics: Portability and Scale

Lumada Analytics is a multi-engine framework that enables the use of a common workflow specification to support data processing pipelines that are portable & reusable. Solution Templates (JSON file format) are used to capture data processing workflow specifications.

The Lumada Analytics demo will demonstrate how to Train, Deploy and Run ML-based analytic models in 60 seconds.

Train → Deploy
Run

We’ll also look out how complex Analytic Workflows can be captured in a portable, reusable specification that can be deployed to dozens or even hundreds of hosts.

P(i) = f(x) + f(y) + f(z)

Analytics Workflows
Input (a) --> P(i) --> Output (b)

Analytics Pipelines

Analytics Functions

Analytics Engines

Python

R

Pentaho

HSDP

PySpark

Java

Node Red

Others
Pentaho Analytics: Integrate and Analyze

Lumada Analytics
Analytics Workflow
Input (a) --> $P(i)$ --> Output (b)

Analytics Pipelines
$P(i) = f(x) + f(y) + f(z)$

Analytics Functions
$f(x)$

Analytics Engines
Python | R | Pentaho | HSDP
PySpark | Java | NodeRed | Others

Data Integration
• Visual workflow designer
• Orchestrate ETL at scale
• Comprehensive data source library

Data Preparation
• Visual transformation designer
• Adaptive execution with Spark
• Integrated scripting for ML use cases

Visual Analysis
• Integrated data exploration
• OLAP analysis
• Production reporting and dashboards

“NewCo” Foundry
Pentaho Data Integration and Analytics

- Visual Data Transformation and Exploration
  - Access and Inspect
  - Cleanse and Transform
  - Blend and Persist
Wireless Data Analysis Example

- Analyze Communication Link Health
  - Cellular communication infrastructure
Lumada Analytics: Management Console

- Managing Workflows at Scale