



# How to Build Your First Predictive Model in Seconds with InfluxDB and Loud ML™

AN INFLUXDATA PAPER

**Sébastien Leger**

Founder and CEO, Loud ML

April 2018

# Overview

Redmint wanted to build a highly scalable machine learning API that allows developers to build models without the algorithms' complexity and that serves as the underlying engine powering applications with predictive requirements. While demand for machine learning is growing, so are its complex and costly development requirements. To build an easy-to-use machine learning API for time series data forecasting, Redmint chose InfluxData.

Redmint's Loud ML™ machine learning API uses unsupervised learning frameworks to gain deep insights into InfluxData time series data. Loud ML allows users to select a time series from InfluxDB to train a model against the Loud ML API.



Visualizing Loud ML Predictions in Chronograf using InfluxDB as the Data Source

## About Loud ML

Redmint is a French Tech software company with expertise in data science and deep learning solutions. Its innovative Loud ML™ software framework delivers easy-to-integrate APIs for Linux, significantly reducing the time to market for smart capabilities integration into vendor applications and software stacks. The company intends to open its capital later in 2018 for Series A financing with plans to hire talent and reach new ambitious targets.

Redmint Network has established a partnership with InfluxData to develop tools by combining its own data science expertise with the Time Series Database monitoring and analytics of InfluxData.

By integrating InfluxData into its machine learning software, Redmint Network will be able to quickly consume time series data and apply machine learning, turning it into actionable data for the enterprise.

*“So the question might be: What’s a good use case for machine learning? I think, at the end of the day, it’s about predictability.”*

**Sébastien Leger, Founder and CEO**

# The Business Problem

Machine Learning is receiving widespread attention across industries and media. It has been defined as a “type of AI that learns to perform a task by analyzing patterns in data rather than by relying on rules written by people” (*Wired*) and as “crunching vast quantities of data to find patterns and make predictions without being programmed to do so” (*The Economist*). Yet the vision that inspired Redmint to create Loud ML is that “Artificial Intelligence is the New Electricity” (as data scientist Andrew Ng has written)—a new power source that everyone will eventually use. The main benefit of AI, which is built on top of Big Data, is to understand data.

Machine learning adoption is set to grow because it can solve difficult problems that computer science was unable to solve in the last decade. This capability represents a new era in data science and technology, creates new opportunities for all industries, and is disrupting all sectors which invest heavily in R&D. Recognizing that potential, Redmint wanted to open deep learning to a large community without requiring developers to master the required data science and math skills. Deep learning is part of a broader family of machine learning methods based on learning data representations, as opposed to task-specific algorithms.

Redmint set out to create a machine learning API that would enable the creation of a predictive model in seconds and that is designed to work with InfluxDB Time Series Database. Through its API, Redmint wanted to enable developers to rapidly integrate machine learning capabilities for use cases where pattern-based predictability is a goal, across industries such as fraud detection (cybersecurity), predictive maintenance (IoT), customer churn prediction (ecommerce), and user-centric behaviour analysis (retail).

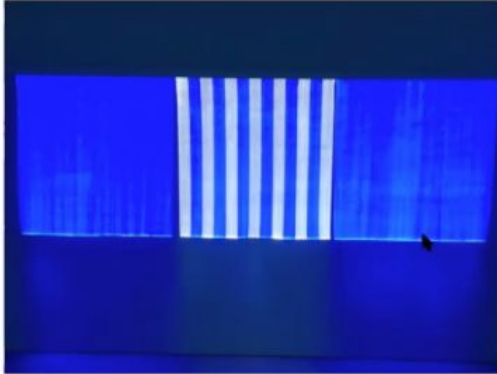
# The Technical Problem

Redmint’s API had to answer two questions arising from time series data and applicable to many use cases:

- The forecasting question: Given a single series and a history of that data, can you use that history to forecast a trend? Trends could include health checks to minimize incident response times and forecasting to handle capacity planning.
- The comparison question: Given a large number of series, how do you create clusters (data point groupings) to compare patterns, manage data easily, and identify cluster trends over time? Such predictions would enable drilldowns into advanced use cases, such as predictive maintenance; fraud and anomaly detection; and recommendation engine enhancement.

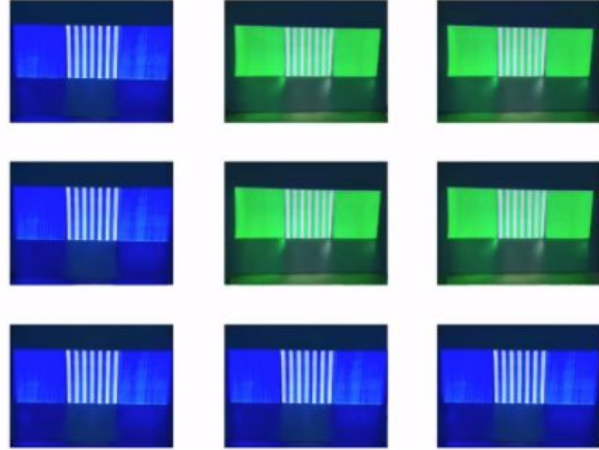
## Two Time Series Questions to Be Solved Using ML

**1. If given a single series, how do you forecast its values?**



(Daniel Buren, 2012 – 2014  
Triptique électrique)

**2. If given N series, how do you compare them?**



Machine learning is the art of teaching an algorithm to perform a highly specific operation. Neural networks and deep learning require lots of calculation, processing power, and data, as well as clever management of that data. Without enough data, the ML model will not be able to converge and therefore will not be able to find an optimal solution. With sufficient data, however, unsupervised learning with neural nets can reduce 60,000 series, for example, to 64 classes.

Training is data and computation intensive: it requires specialized and dedicated hardware. Finding the best model to fit a given data set is therefore time-consuming. Without the right hardware or remote compute power, the most difficult problems require months of computation and the results may not meet expectations.

## The Solution

*“Loud ML really works out of the box with InfluxDB.”*

### Why InfluxDB?

Redmint chose to integrate InfluxDB into Loud ML because predictions can occur over time (time is often a key dimension in prediction), and InfluxDB is a purpose-built open-source Time Series Database. InfluxDB had the properties—high throughput ingest, compression, real-time querying—to enable Redmint to quickly consume time series data and apply machine learning. InfluxDB also had built-in downsampling and data retention policies.

Loud ML makes it easier for developers to prepare, train, and deploy machine learning models for predictive analytics with InfluxDB. It enables developers to get smart and automated insights from time series data:

- Developer-friendly: Loud ML has used the latest technologies to remove the complexity of machine learning with TensorFlow.
- Multiple use-cases: Developers get the most out of their existing time series data, such as log files, application and performance metrics, network flows, and financial or transactional data.
- Infinite data volume: The licensing model puts no restriction on the volume of data that can be ingested since machine learning is a data-hungry application.

## The Features of Loud ML

Redmint built Loud ML to be a versatile API that empowers enterprises to integrate actionable data science into their applications with minimum development time. With its intelligent prediction and embedding tools, Loud ML negates the need to wait for system usage observations before taking smart auto-scaling decisions.

The Loud ML framework will allow organizations to:

- Dynamically scale and drive smart load-balancing decisions for VMWARE and/or KVM virtual resources according to *predicted* load, enabling cloud and hosting providers to reduce operating costs while delivering high-quality services and response times to their users
- Embed smart intent-based decisions in network services and NGN equipment when network traffic patterns indicate that streaming congestion will impact customer experience, permitting service providers to scale and reduce operating costs
- Spot anomalies in e-commerce purchasing patterns, and automatically send smart alerts which filter out noise, ensuring e-commerce companies receive the most relevant alerts when things go wrong in the customer's journey
- Predict changes in customer purchasing history, enabling retail companies to forecast demand with optimal accuracy

The Loud ML APIs are easy to integrate and available to a large community, first being offered in the InfluxData solution. They will accelerate the introduction of autonomous deep learning into a wide range of applications spanning multiple industries. Launched in partnership with InfluxData, Loud ML seamlessly integrates with InfluxDB:

- Loud ML supports any function that InfluxDB does (such as mean, max, average, count, style, navigation, spread). You can use what you've already created in InfluxDB, then apply trainings onto your various time series and visualize the results in Chronograf.

## Visualizing the Results in Chronograf



- By training time series, it will learn how the data evolves over time.
- Training will output the model accuracy level: a percentage. The higher the accuracy, the better in order to use the model in production.
- After training, you can perform inference. This means your model can repeat the operations that it knows (or have been discovered through training) using brand new data. Unsupervised learning involves inferring a model to describe hidden structure from unlabeled data.
- Loud ML's unique feature description language (DSL) can be used to drive personalization, recommendations, user behavior, fault and incident detection.

To increase training accuracy in Loud ML, Redmint recommends:

- Ensuring you have sufficient data since low accuracy could simply mean insufficient data
- Setting your `max_evals` (max evaluations) to your preferred level depending on the time you can wait to improve accuracy
- Experimenting with Loud ML to find the right values that fit your data (the right `bucket_interval`, `span`, `interval`, etc.)

## Trial and Error to Identify The Right Values

```
File Edit View Search Terminal Help
{
  "type": "timeseries",
  "name": "page-hits-model",
  "default_datasource": "http_datasource",
  "bucket_interval": "30m",
  "span": 6,
  "use_daytime": true,
  "offset": "30s",
  "max_evals": 100,
  "interval": "1m",
  "default": 0,
  "threshold": 50,
  "features": [
    {
      "name": "count_page_hits",
      "measurement": "hit",
      "metric": "count",
      "field": "page_url"
    },
    {
      "name": "avg_page_duration",
      "measurement": "hit",
      "metric": "avg",
      "field": "duration"
    }
  ]
}
```

## Technical Architecture

*“The metric can be any metric that InfluxDB already supports. It’s an aggregation function, and we’re really mapping the values to the capabilities that are supported by the database.”*

### InfluxDB Deployment in Loud ML

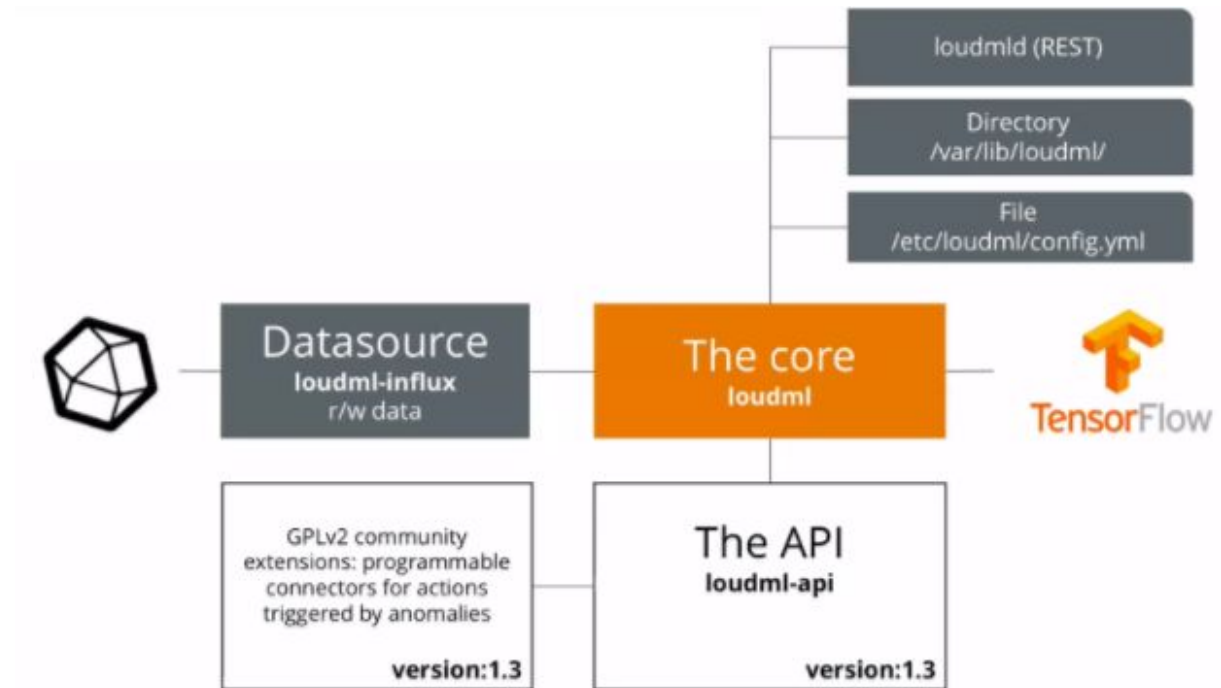
#### 4. Write points in a new measurement



The dataflow is as follows:

- Loud ML reads and extract data from your databases to train your machine learning models and then output predictions.
- When Loud ML outputs predictions, it writes the points in new measurements into the same database from which the data is being pulled.
- If you want to run anomaly detection, you will find and get notifications in real time.

### Inside Look at Loud ML Architecture



- Loud ML is based on TensorFlow, which Redmint decided to use because they found it to be a future-proof framework.
- The data source is a backend that allows reading and writing data from the database (InfluxDB). This object is expressed in YAML format and defined in the configuration.
- At the top of the flowchart are configuration files that Redmint uses.
- At the bottom of the flowchart is the API. Since there are many ways to send anomaly notifications when you want to take action (email, SMS, Rest API or SNMP traps) to control an equipment, Redmint provides users the flexibility and freedom to choose the preferred notification method by opening up this API, which will be JPL V2.

What's Next for Loud ML?



Redmint plans to incorporate the feedback it has received from the InfluxData community into Loud ML future releases. It will be adding, to its roadmap, the requests that it's receiving because it believes this is the best way to drive its software forward.

## Results

*“There’s already a community of InfluxDB users who are getting their hands on Loud ML. There’s already good feedback from that community.”*

Loud ML facilitates deep learning usage as it is data source agnostic (it will connect to all major NoSQL databases). The APIs are well-documented with a CLI and REST endpoints so it can be controlled remotely. With Loud ML, Redmint has enabled developers to address the top 4 machine learning challenges:

1. **Availability:** Loud ML democratizes access to machine learning. The Loud ML Developer Edition is free (available for Loud ML for Debian 9 and Red Hat 7, and soon as a docker container). Loud ML is also available for purchase under a subscription license to providers who wish to integrate predictive and unsupervised machine learning capabilities with their own products or services.
2. **Affordability:** Enterprises who have attempted a DIY machine learning solution know it is expensive, can take 6-18 months, and involves extensive trial and error. Loud ML saves them the time and cost investment and frees them to focus on their core business.
3. **Reliability:** Achieving reliability for software in production is hard. Loud ML applies all best practices, implements code review, continuous integration and delivery, and automated testing.
4. **Trust in the algorithms:** The issue of trust in the algorithms is an ongoing debate.

Loud ML is helping developers and organizations convert high-velocity information and massive big data into knowledge. Through Loud ML, AI is now accessible to everyone and the community is growing fast.

Built to mine intelligence from InfluxDB time series data, Loud ML is fulfilling its mission of being the “disruptive machine learning API”.