Real Time Prediction for an Active Risk Prevention

João Batista Lopes

April 17 2018

Know the unknown..

weDO technologies

AGENDA

1. About WeDo
2. Introduction
3. Drop Calls
4. Adaptive Thresholds
5. Solution
6. Conclusions & Future Work


WeDo

ABOUT WEDO TECHNOLOGIES

AN EUROPEAN COMPANY FOUNDED IN 2001 WITH A WORLDWIDE FOOTPRINT

10 COUNTRIES

5 CONTINENTS

600+ EMPLOYEES

COUNTRY OFFICES

WEDO **AUSTRALIA**
WEDO **BRAZIL**
WEDO **EGYPT**
WEDO **IRELAND**
WEDO **MALAYSIA**
WEDO **MEXICO**
WEDO **PORTUGAL**
WEDO **SPAIN**
WEDO **UK**
WEDO **USA**

SOFTWARE HOUSE

PORTUGAL (BRAGA)

OUR PRODUCT PORTFOLIO

COMMON ARCHITECTURE FOR ALL SOFTWARE PRODUCTS

RAID RISK MANAGEMENT

- RAID REVENUE ASSURANCE
- RAID FRAUD MANAGEMENT
- RAID BUSINESS ASSURANCE

ON PREMISES MODULES

CLOUD APPS

RAID BUSINESS MANAGEMENT

- RAID INCENTIVES
- RAID COLLECTIONS
- RAID ROAMING

ON PREMISES MODULES

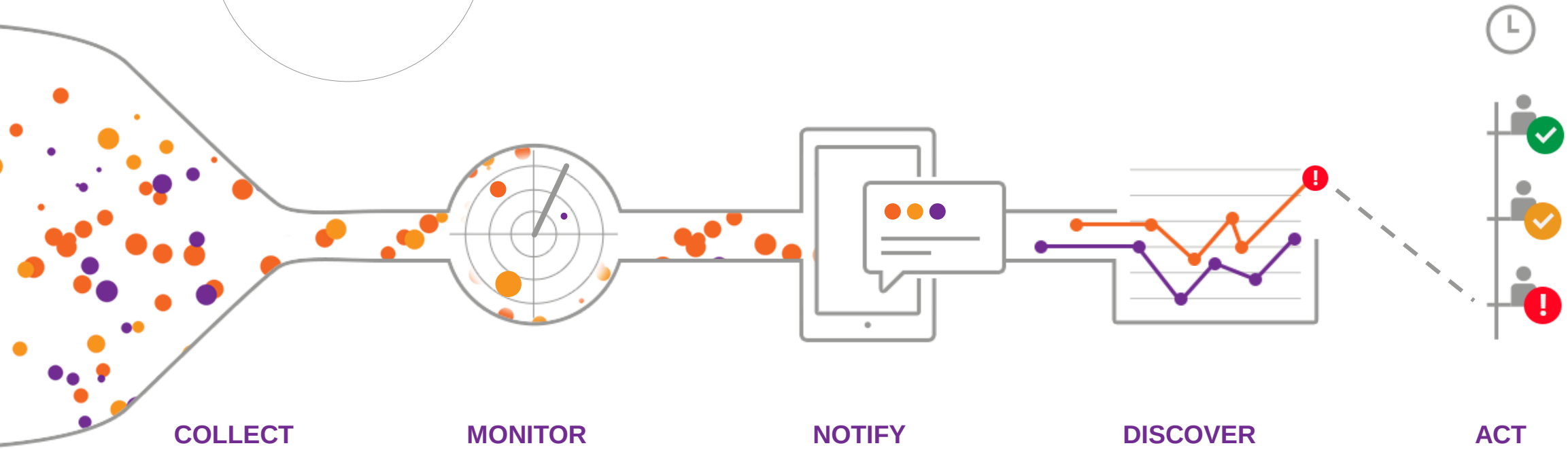
RAID BUSINESS OPTIMIZATION

- RAID OPTIMIZE

ON PREMISES MODULES

RAID END-TO-END SOFTWARE

NEXT GENERATION BUSINESS MONITORING PLATFORM



COLLECT

Smart Data Stream

(ETL & CEP)

MONITOR

Unified Validation Engine (UVE)

Fraud Management Engines

Advanced Fraud Detection (AFD)

NOTIFY

KPI Designer

Business Sensors

Balanced Scorecards

DISCOVER

Investigation Workbench

Smart Data Blueprints

Link Analysis

ACT

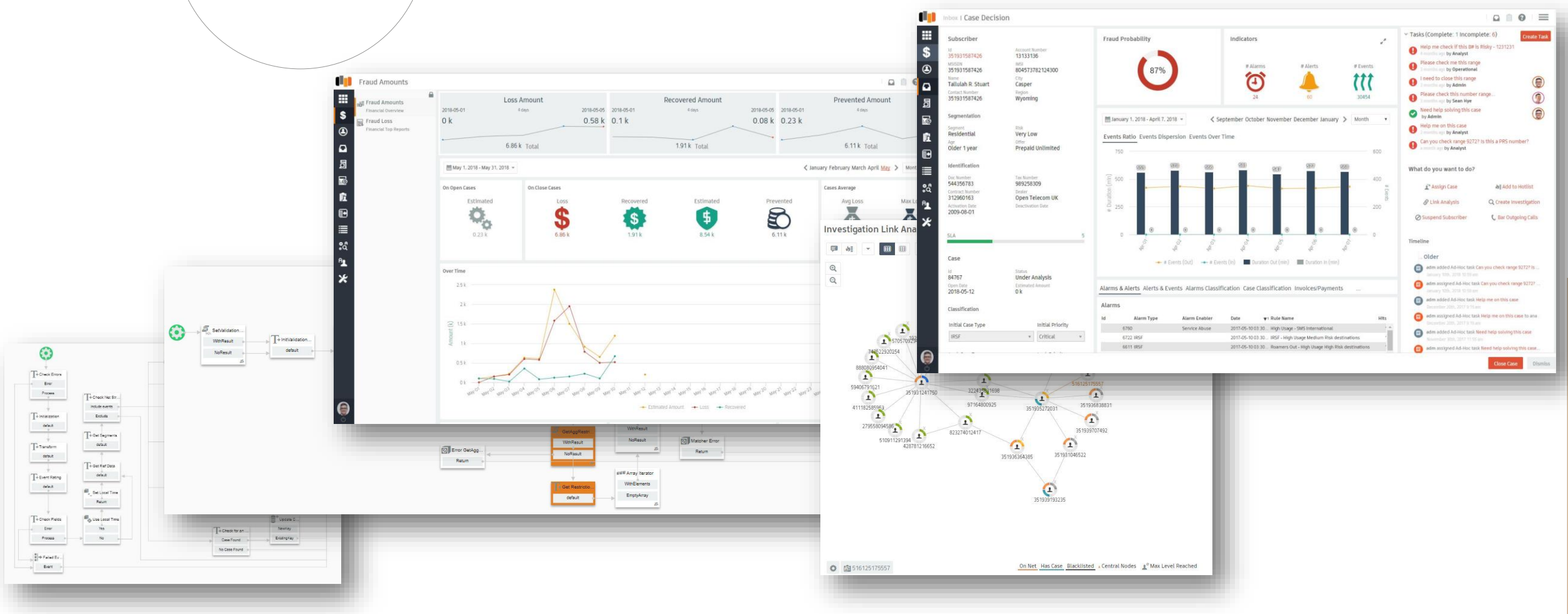
Adaptive Case Management

Web Portal

Dashboard & Reporting

RAID END-TO-END SOFTWARE

NEXT GENERATION BUSINESS MONITORING PLATFORM



COLLECT **MONITOR** **NOTIFY** **DISCOVER** **ACT**

MARKET LEADERSHIP IN TELECOM

WORLDWIDE LEADERSHIP IN REVENUE ASSURANCE AND FRAUD MANAGEMENT SOFTWARE

RAID Gartner #1

RAID analysys mason #1

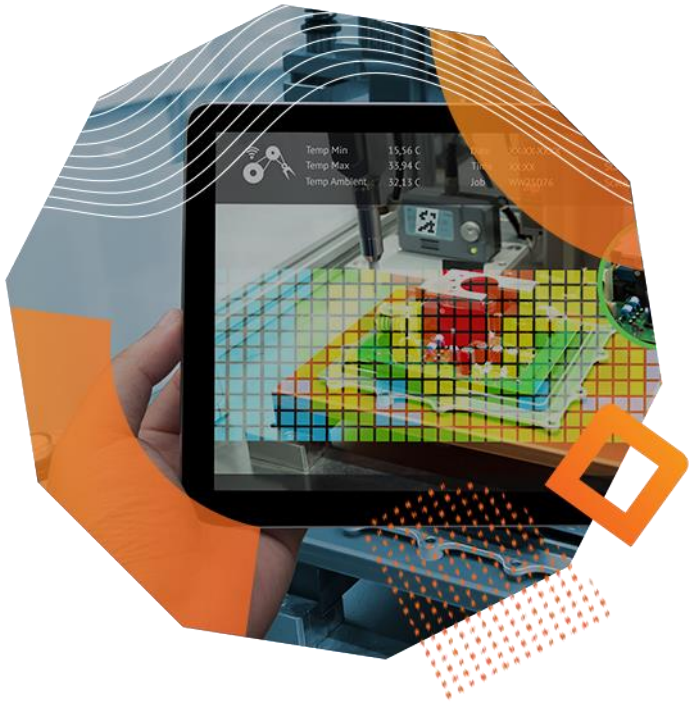
RAID Stratecast #1

Recognized market leaders by
3 independent industry reports

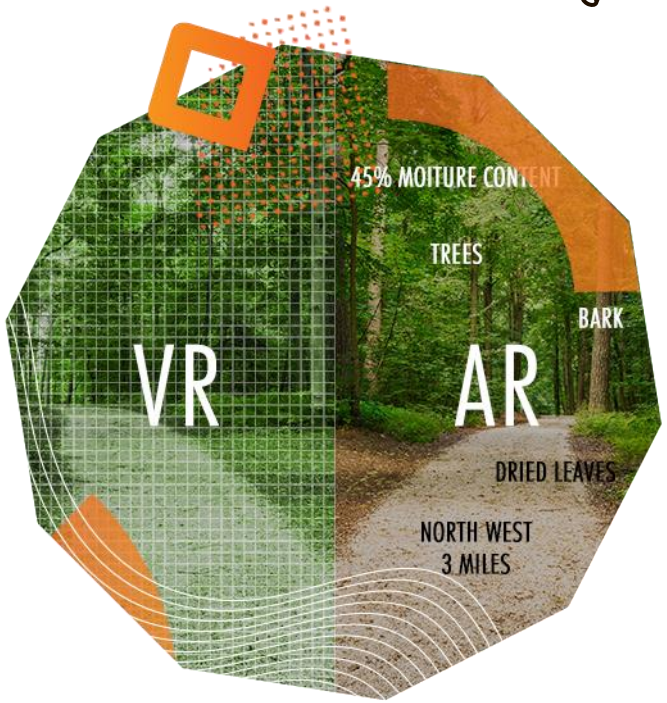
Trusted by more than 220 CUSTOMERS
in more than 100 COUNTRIES

DIGITAL TRANSFORMATION
Heading to a Digital Society

The rise of
digital twin



The blur between the
digital and physical worlds



One reality
vs.. Many realities

INTERNET OF THINGS (IoT)
WOULDN'T EXIST WITHOUT SMART SENSORS

Growing use of *smart technology* is transforming the way in which businesses **collect** & make **sense of data**

Smart Sensors allow for more **Connectivity** and **Analytics**

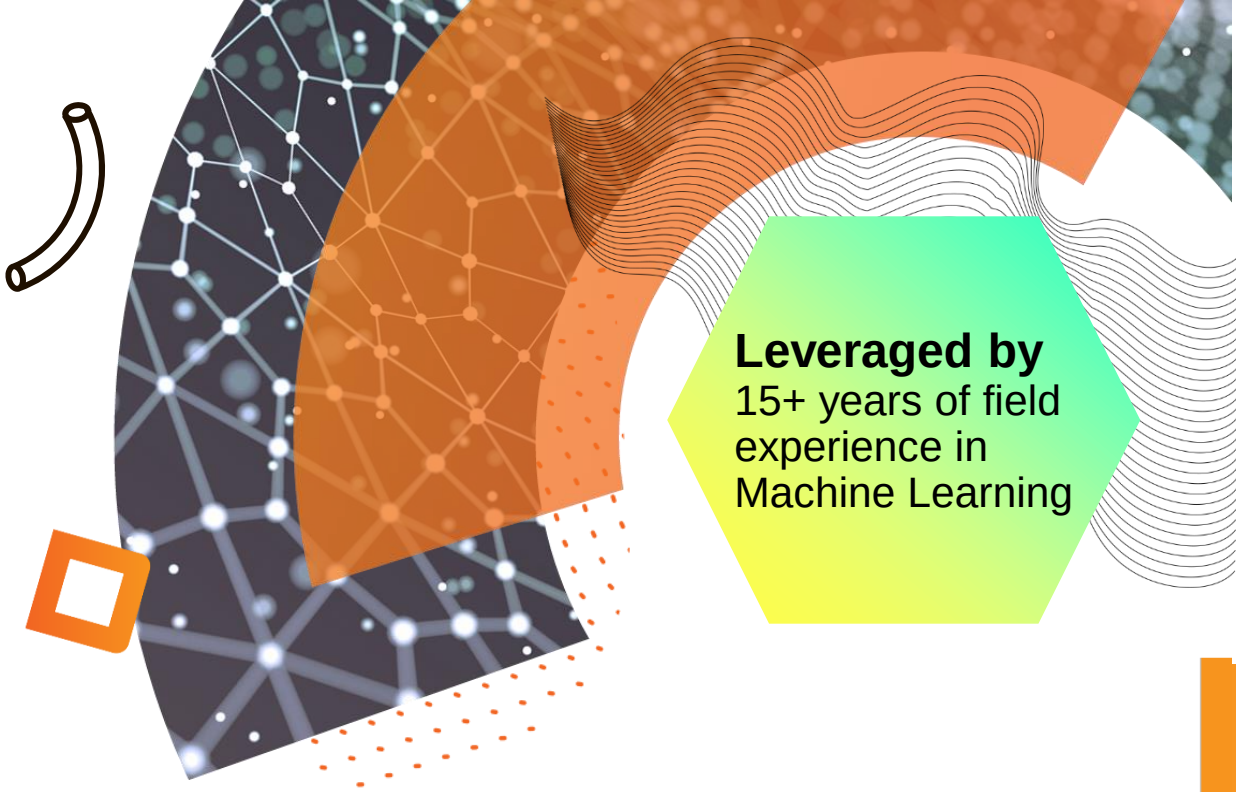
212 billion — Total # of Sensor-Enabled objects by 2020

30 billion — Sensor-Enabled Network-Connected objects by 2020

8 billion — Total World Population

PUTTING AI IN THE CENTER OF OUR STRATEGY

TO HARVEST THE VALUE OF DATA
FUELING AN **INTELLIGENT** RISK MANAGEMENT



Leveraged by
15+ years of field experience in Machine Learning

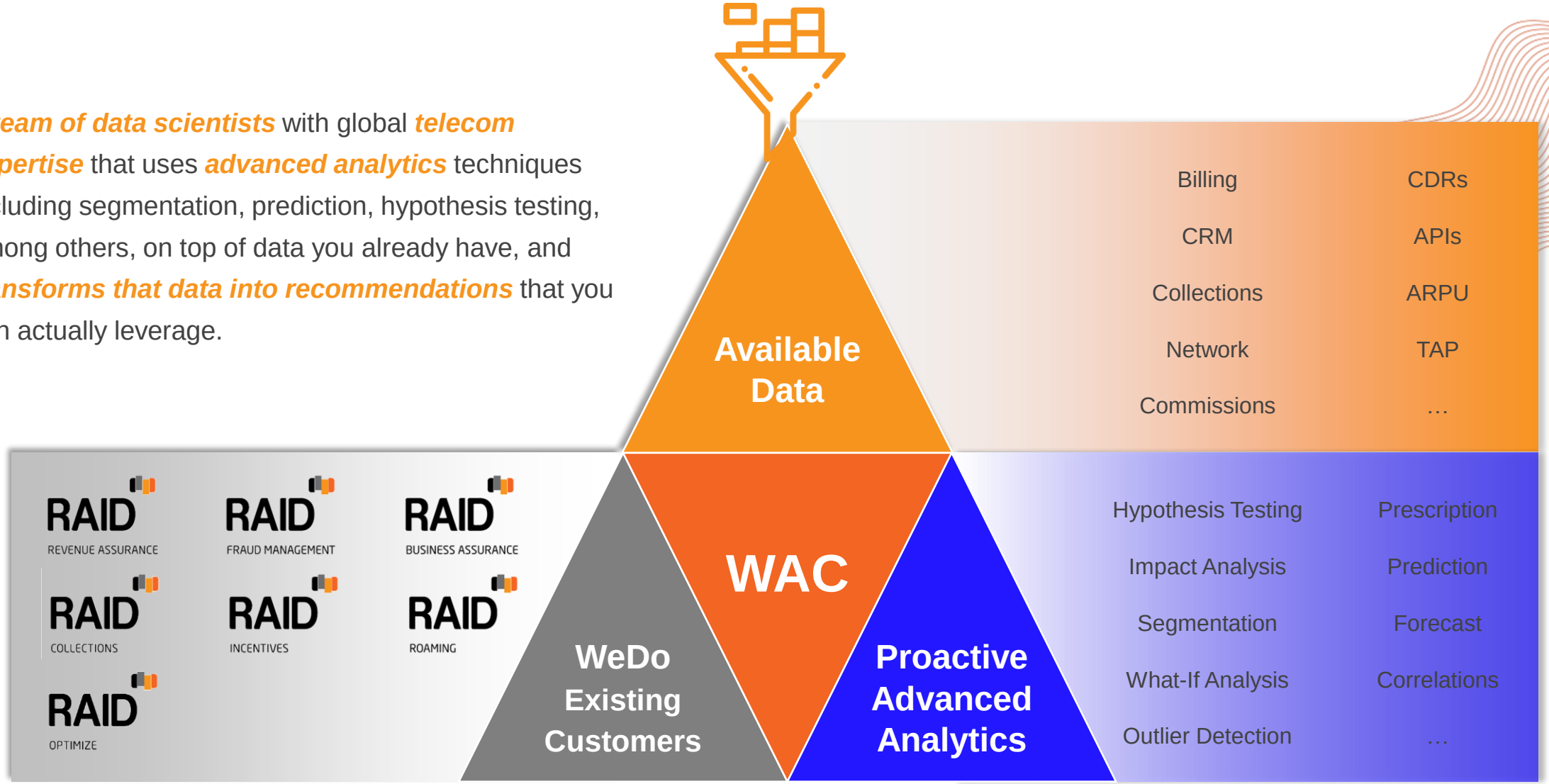
Google ASSISTANT

Hi, I'm Cortana.

IBM Watson

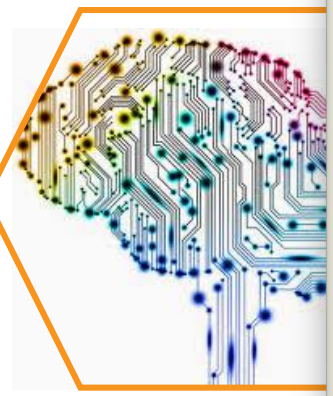
WEDO ANALYTICS CENTER (WAC)

A ***team of data scientists*** with global ***telecom expertise*** that uses ***advanced analytics*** techniques including segmentation, prediction, hypothesis testing, among others, on top of data you already have, and ***transforms that data into recommendations*** that you can actually leverage.

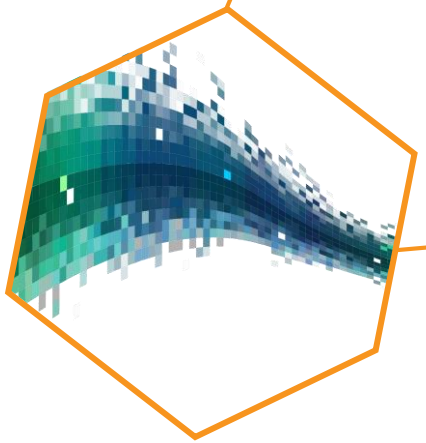


ADAPTIVE THRESHOLDS
FOR MAXIMUM E

Powered by
Machine Learning in Hadoop



Predicted Behavior & Threshold



Real Time Behavior

Alarms on exceptions

RT behavior monitoring

Adaptive threshold setting / suggestion

Predicted RT Behavior

Smart alarms and case generation

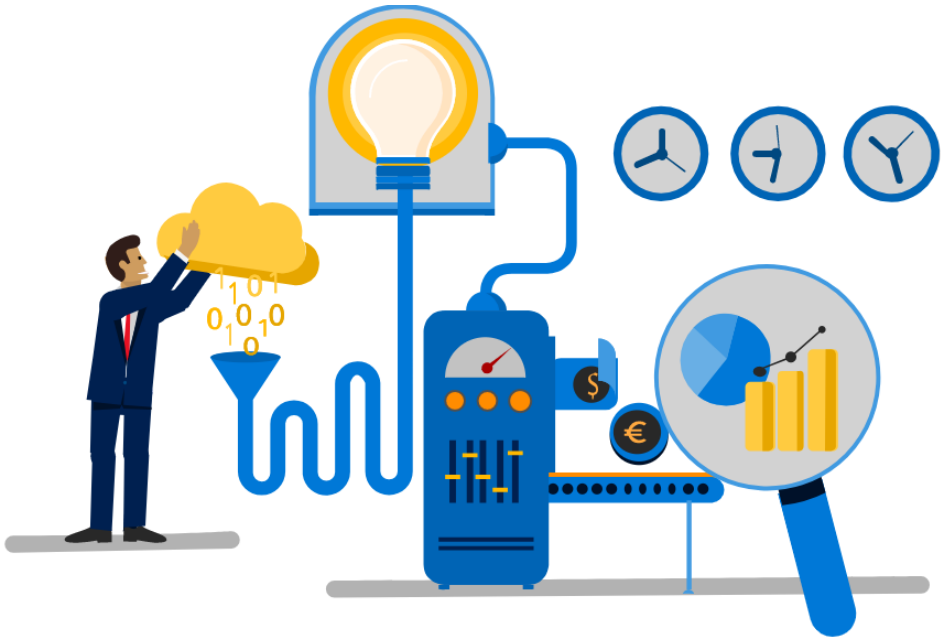
1

Introduction

Introduction

Motivation

- **Artificial Intelligence (AI)** solutions are being deployed everywhere
- Advances in **data availability and processing** changed what is possible to accomplish in real-time
- Making a predictive analytics **application real-time** is a today question of infrastructure
- Big data and AI solutions are today the key basis of competition, underpinning new waves of productivity growth and innovation



Introduction

Goals

- How can WeDo help risk management teams, on defining the best revenue assurance and fraud alerts thresholds, for each business reality?
- Can these thresholds rules be adjusted dynamically, without human intervention, and deliver the right outcomes?
- Could these predictive models be updated/fit automatically using an on-going process with high volumes of data?

2

Drop Calls

Drop Calls

Industry

- **Cell phones** are used more than ever and have increasing impacts in our social and business lives
- New-generation smartphones and data applications support simultaneous voice and data services
- Service providers are **migrating current networks to next generation networks** (Ex., LTE) in order to support more demanding services such as VOIP and high resolution video
- These factors together contribute to a **higher risk of delivering a quality service**.

Drop Calls

Opportunity

- **Service quality** has direct influence on customer experience and satisfaction.
- **Drop Calls** are today the major reason of complaints by phone users.
- It is of great interest to **quickly identify** and understand the causes for dropped calls.
- While certain number of dropped calls are expected, there are others that are unexpected, and may be caused by system malfunctions.

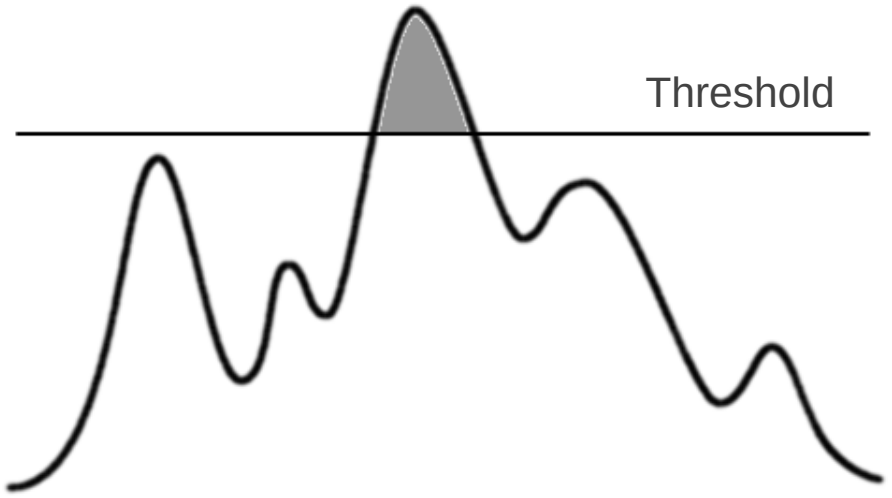
- In Europe, it's frequent a drop call rate below 1% of total voice calls. But in other countries, like India for instance, the average drop call rate can reach 4.73%!

3

Adaptive Thresholds

Adaptive Thresholds

- **Thresholds are basically borders**, which separates the normal or expected behavior from the abnormal or problematic behavior.
- **Most of the solutions** in Revenue and Business Assurance do not automate their threshold settings.
- **Important business seasonality patterns** that need to be different throughout the day, week and month are ignored.



- **Automatic derivation, configuration and adaptation** of statistically meaningful thresholds on the parameters values is essential for self-managed systems.

4

Solution

Solution

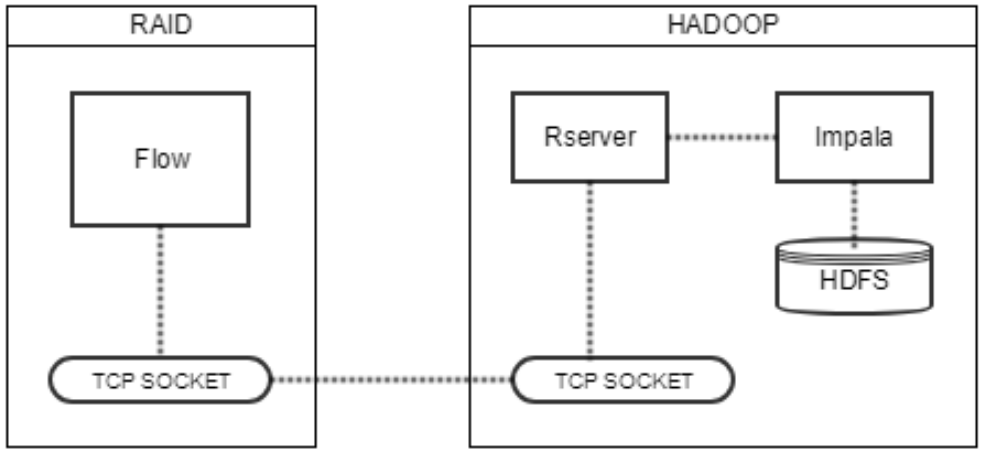
Context

- Real data from operator containing **30 days of history** (360 Million of desegregated events) with ~5300 distinct cell id's
- Time Series modeling in R
- Hadoop cluster for storage
- RAID as central coordinator for configurations and UI

❖ Based on historical total calls and drop calls indicators, predict future calls (total and drop) at a cell id level

Solution

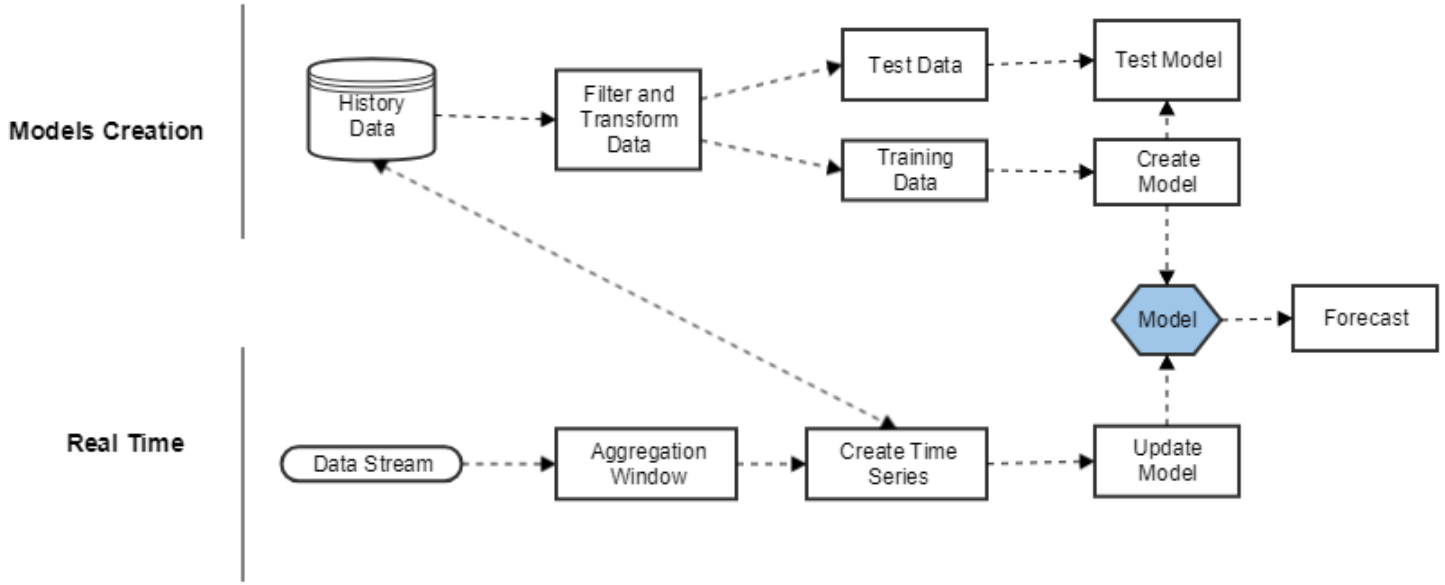
High Level Architecture



- Data processing and event storage in Hadoop
- Code with the modeling instructions is generated on the RAID side. It is sent and executed in RServer through a TCP Sockets connection
- The instance RServer uses the parallel query engine Impala to quickly fetch the data and execute the modeling tasks

Solution

Process Flow



- **The model creation event** is executed at the beginning before the system is in production or when the model update task returns a model with low accuracy. (~16 Seconds)
- **The real time event** is an on-going event where the data stream received is aggregated by a defined time window and sent to the time series creation task. (~ 2 Seconds)

Solution

Model Details

Multiple seasonality, hour-of-day and day-of-week, was modelled with:

- TBATS function with native ETS model (Error, Trend, Seasonal) and parallel configurations
- ARIMA with external regressors in the form of fourier terms to account seasonal behavior and parallel configurations

Model **Fitting strategy/algorithm** was tested with:

- Conditional Sum of Squares (CSS)
- Maximum Likelihood (ML)
- Conditional Sum of Squares, followed by Maximum Likelihood (CSSML)

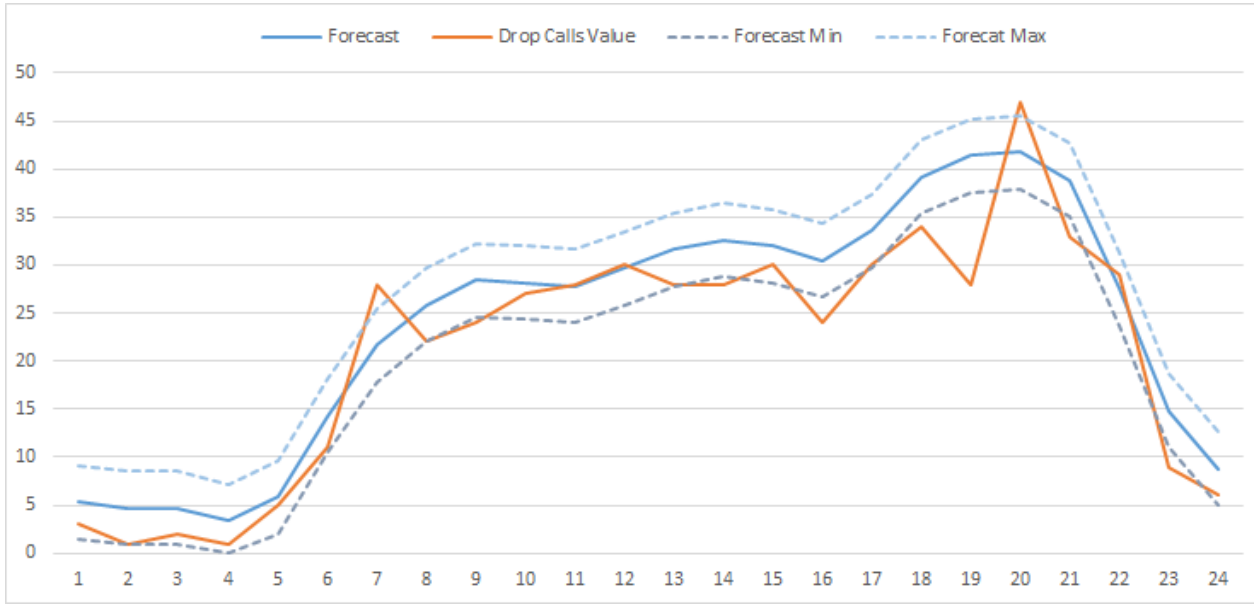
Solution

Model Details

We created an algorithm to test different models parameters and find/automate the best model according to the following accuracy measures:

- **Akaike Information Criteria (AIC)** was used to find the best fit model
- **Root-Mean-Square Error (RMSE)** was used to measure the differences between values predicted by the model and the values observed

This chart illustrates the resume of drop calls values for one cell station at Monday. Upper and Lower confidence intervals of 95% are also visible.



Solution

Results

1. CSS fitting strategy is the best for a real-time solution. Model results are very similar and the other strategies, ML and CSSML, have much higher processing cost (minimum 3 times more).
2. Although ARIMA with external regressors has better performance when forecasting up to 12 hours, it requires 5 times more CPU
3. TBATS has better accuracy with higher forecasting horizons and has best processing performance

- Each configuration can be useful depending in the business needs. For our use case, TBATS function with CSS fitting strategy was the best for a real-time solution.
- For purpose of this work, Upper and Lower confidence intervals of 95% were used for Adaptive Thresholds rules

4

Future Work

Conclusions and Future Work

1. We achieved great performance values when using a distributed Hadoop cluster. (data manipulation and model creation/update)
2. We were able to create an automatic process capable of updating a model with on-going data, maximizing the concept of real time.
3. We learned how Adaptive Thresholds solutions can be used as a suggestion or hybrid system.

- WeDo is currently working on the integration of Adaptive Thresholds solutions in its products portfolio, both on-premises and as Cloud Services (https://raid.cloud/)
- Real-time technologies are being evaluated and integrated in WeDo products (Ex., Apache Kafka messaging)

THANK YOU

joao.blopes@wedotechnologies.com

Know the unknown...