PREDIS

PREvisão DIStribuída de Consumos e Geração em "tempo real"
Load and Generation dissagregated forecast in "real time"

Lisboa, 29 de Setembro de 2017

EDP Distribuição and EDP Inovação – facts and figures

EDP Distribuição is the EDP Group's company operating in the regulated distribution and supply businesses in Portugal. EDP's distribution activity is regulated by ERSE (Entidade Reguladora dos Serviços Energéticos) which defines the tariffs, parameters and prices for electricity and other services in Portugal.

Percent of the electricity distribution network owned in mainland Portugal

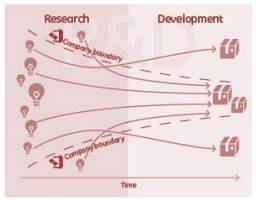
Distribution network approximate length

Number of Substations

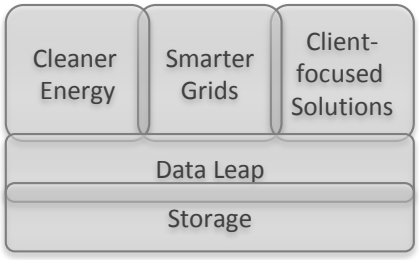
Number of Distribution transformers

Approximate number of customers served

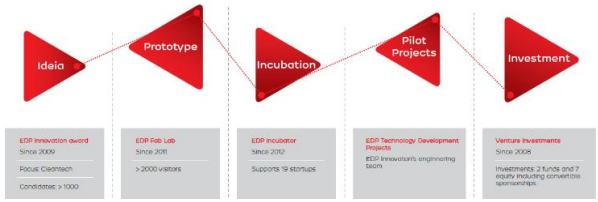
EDP Inovação is the innovation arm of EDP Group, promoting value-adding innovation within the Group by leading the adoption of new technological evolutions and practices.



Open innovation approach



5 strategic innovation areas



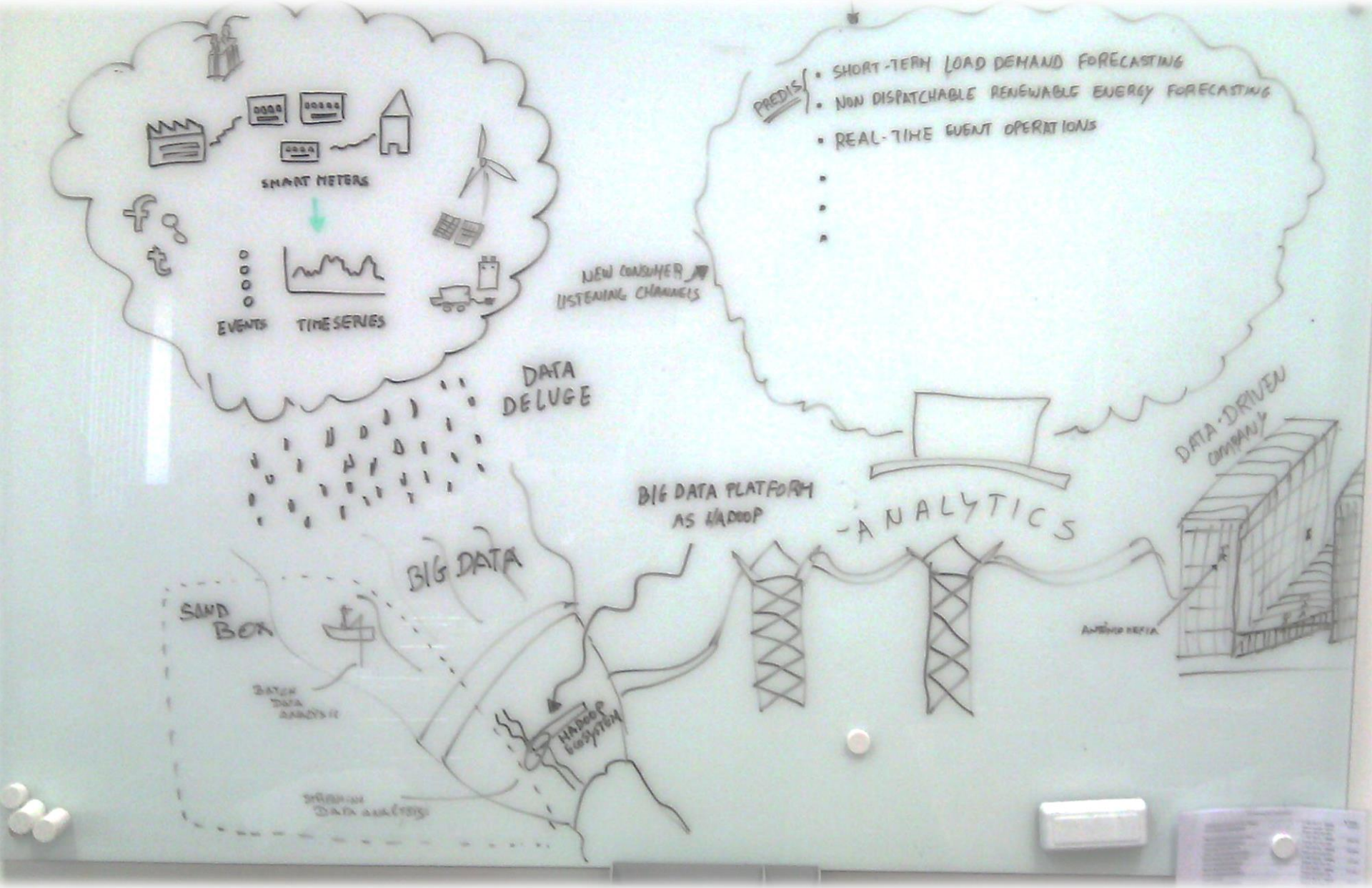
Entrepreneurship & Venture Capital ecosystem

We are living an era where the electric sector is having the most profound changes since its creation.

The exponential growth of Distribution Generation, storage and eletrical vehicles bring increased complexity to the grid management.

The utilities will need to deal with information in a much larger scale than they were used to

There is technology that can support us to cope with this difficult context (Big Data, IOT, machine learning) but we need to leverage on it.


SMART METERS
EVENTS
TIMESERIES
NEW CONSUMER LISTENING CHANNELS
PREDIS
• SHORT-TERM LOAD DEMAND FORECASTING
• NON DISPATCHABLE RENEWABLE ENERGY FORECASTING
• REAL-TIME EVENT OPERATIONS
DATA DELUGE
BIG DATA PLATFORM AS HADOOP
ANALYTICS
DATA-DRIVEN COMPANY
BIG DATA
SAND BOX
HADOOP ECOSYSTEM

PREDIS

Inspired on the work of EDF and IBM, in 2013 EDP started to look at big data and advanced analytics, developing comparison between the performance of a conventional DataBase and Hadoop.

National Energy Consumption aggregation (with load curves) by voltage level*							
System	Nodes [#]	Cores [#]	RAM [GB]	Cluster	Readings [10^6]	Volume [MB]	Processing Time [h:min:sec]
BO (EDP)	4	96	202	Local	12 x 6	72	3:45:00
Hadoop	21	42	157	Virtual / Cloud	96 x 6	576	00:09:37

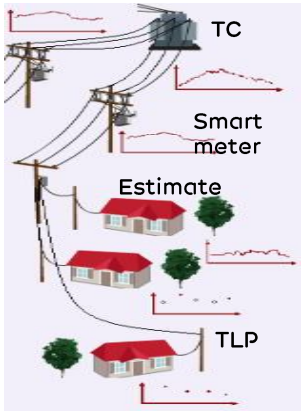
*This Proof of Concept was done in the cloud payed with a credit card and cost around $30.

Profiling + aggregation	Technology	Time	Notes
Current architecture	Oracle	Around 8h	4 Million points
SQL with Big Data	Hive, Impala	1 to 4h	Inadequate
Customized Programming without Big Data	Java	Around 5min	One machine (multi-core)
Customized Programing with Big Data	Spark	<5 min	Multi machines with Big Data Higher resilience Parallelization

Main conclusions:

- The Hadoop cluster is by nature resilient and coped with nodes failure.
- The processing times can be greatly improved over traditional arquitecture
- **There is a high need for customization**
- **The choice of the tool from the Hadoop ecosystem depends highly on the type of calculations to be made.**

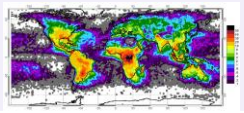
With the results obtained we set up a project called PREDIS to have load and generation forecast at an disaggregated level in real time (with 15 minutes refreshment).



Grid data
SCADA
OMS
GIS

Energy data
EMS
AMI

Weather Variables



Local Variables

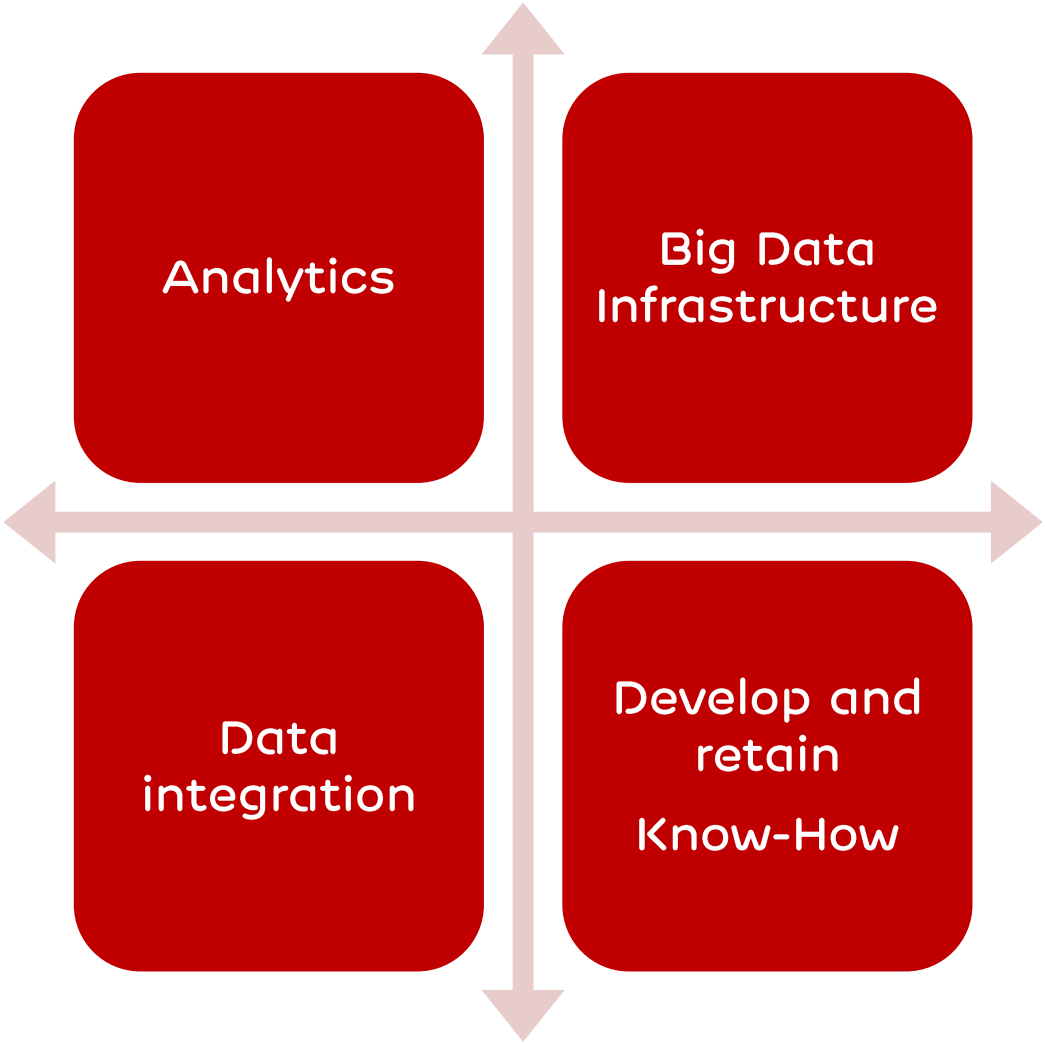
RES farms data

Disaggregated Load Forecast
Disaggregated RES Generation Forecast
PREDIS

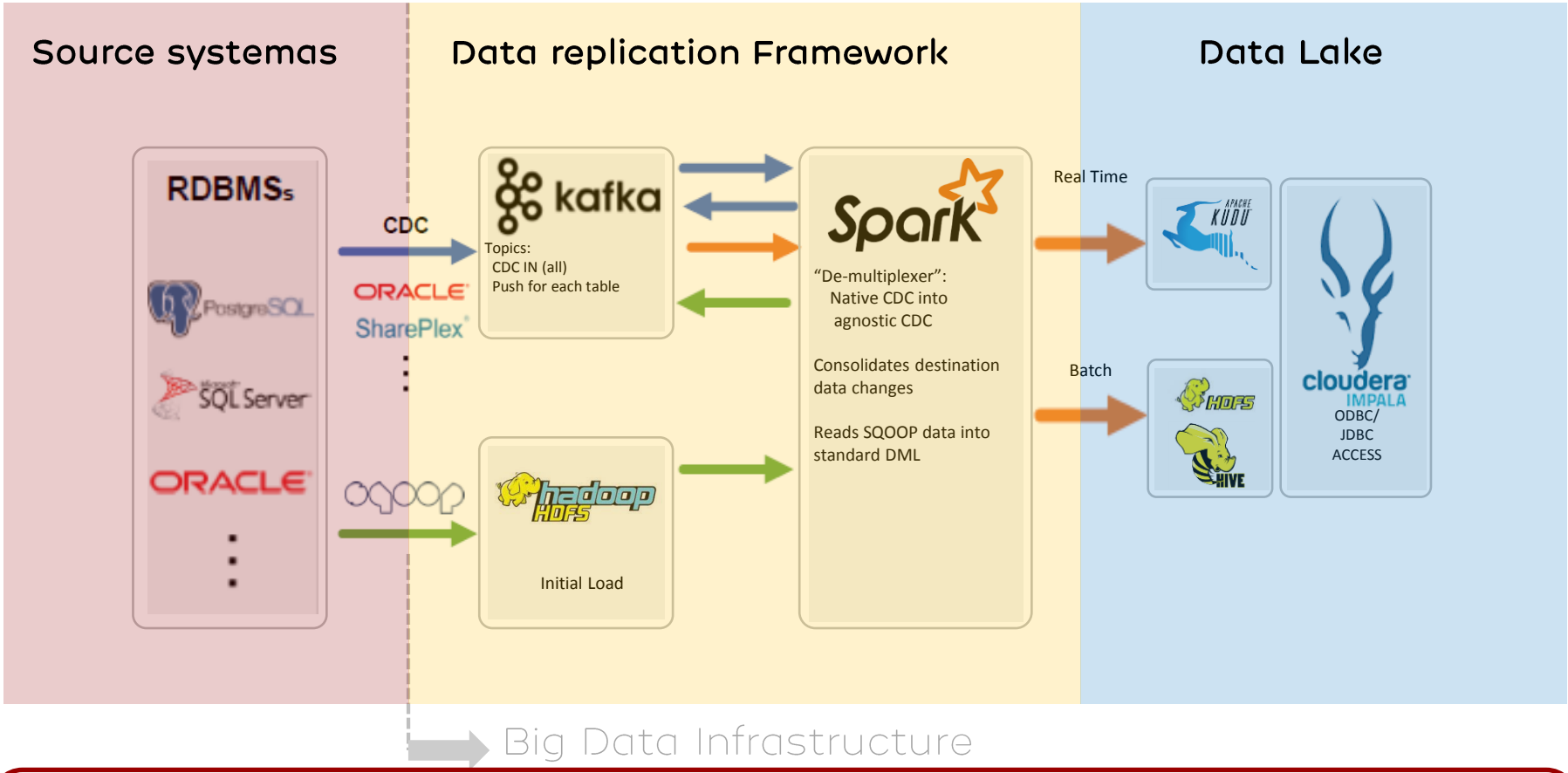
- Energy Balance
- Grid Dispatch
- Fraud
- Revenue Assurance
- Grid Planning

The project combines external and internal sources of data to forecast and to give results to support several operational processes

To reach this goals we needed to develop competences on four areas:



We had to develop an architecture to have data consolidated in Big Data Cluster in "real time"

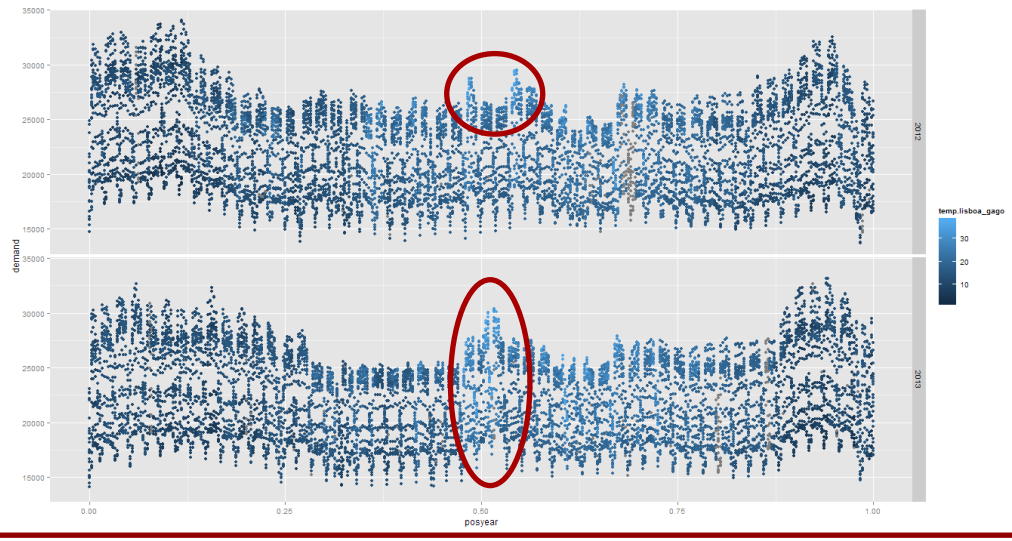


This architecture replicates the operational systems in the Big Data infrastructure, processes and consolidates the information in real time without impact to the source systems

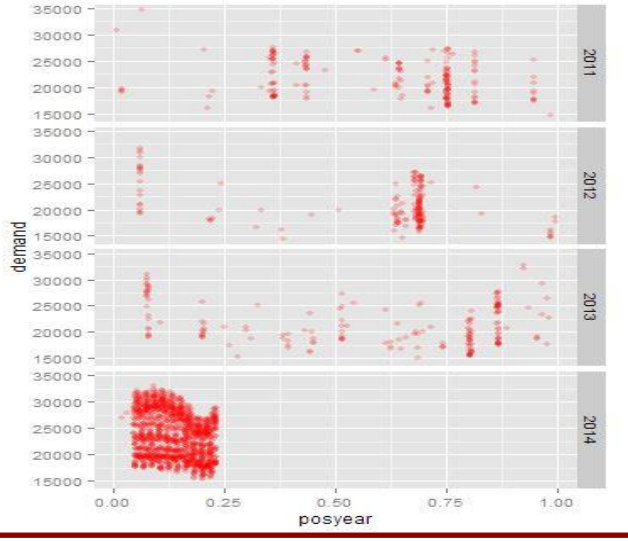
For the analytics phase we have a 3 step process for the development of the forecast algorithms.

DATA DISCOVERY → MODELATION → OPERACIONALIZATION

Temperature effect in national load for two years



Missing temperature data for the last 4 years

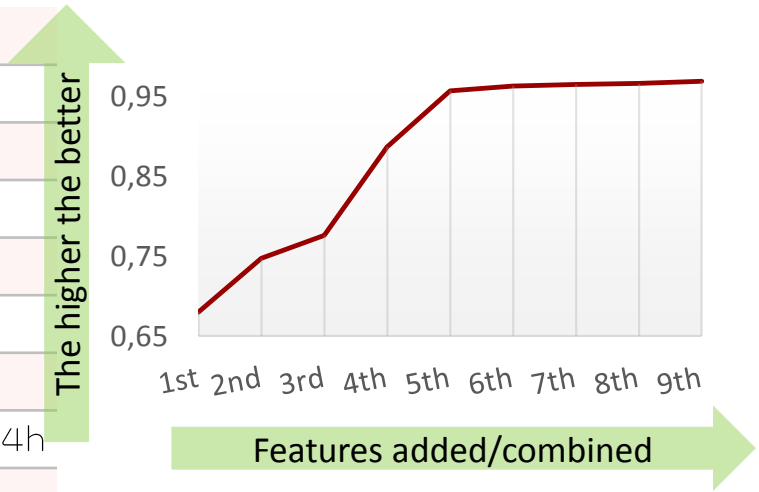


We used R (open source) that allow us to easily analyse data in several dimensions, check for data quality and find correlation between data.

The load forecast model was developed internally using R on the Big Data cluster

DATA DISCOVERY → MODELATION → OPERATIONALIZATION

Iteraction	Variable
1	24h lagged load
2	temp. combined w. time of day
3	48h lagged load
4	day of week
5	public holidays
6	intra-day effect dependent on the day type
7	day of the year
8	24h lagged temp. + min and max temp. of last 24h
9	days offs before Christmas and Carnival



The national load model was the starting point for load disaggregated forecast in Substations and Distribution transformers

With the results obtained we are preparing to deliver the results to some systems/projects that need this information

DATA DISCOVERY → MODELATION → **OPERATIONALIZATION**

Load forecast:

- For the next 3 days for each 15 min
- Updated daily
- Based on Load diagrams and weather forecast

Processing power *

Step	Added time	Paralel time
Calibration	30s x 39828 = 332h	**~4h30m**
Forecast	20s x 39828 = 221h	**~3h**

*Using 46% of the cluster processins power since it is Shared with other processes needed for EDP DIST

Forecast results

Type of asset	# Grid Elements	MAPE (mediana)
Dtransf	39.075	**12.9%**
Substation	753	**9.8%**

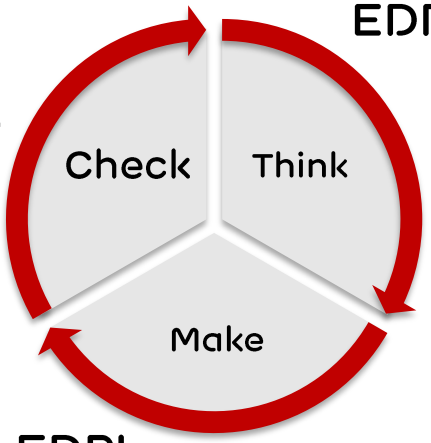
We have also developed wind and photovoltaic forecast algorithms that are being operationalized

All of this has been made with a very close cooperation between EDP Distribuição, IT and EDP Inovação teams in an iterative process

Multidisciplinar team looks at PREDIS forecasts (EDPI, DGE, DDC, DAT, DPL, DOD):

EDPD

- Evaluate the forecast quality Vs business needs



EDPD+EDPI

- Identify improvement areas
- Incorporate business know-how:
  - Operational Grid configuration
  - Determine existing systems information
  - Etc.

EDPI

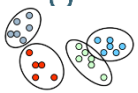
- Incorporate new data
- Develop new models

Knowledge sharing has been essential for the development of the forecast models

We have also been working with Universities to help us to explore and improve the models

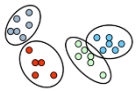
IST

Short term graph modeling



Master thesis for load profiles clustering

Nova School of Business and Economics



Short term project for load profiles clustering in Substations and Distribution transformers

FCUL

Master thesis in Centralised PV forecast

Master thesis in decentralized PV forecast

Master thesis in Wind forecast

With these works we are looking for new explanatory variables, improve models, include grid dynamics, automate data exploration and data modeling

To expose PREDIS results to business users we have develop several interfaces so that they can use this data and help us improve the results

PREDIS results sharing

To integrate in other systems

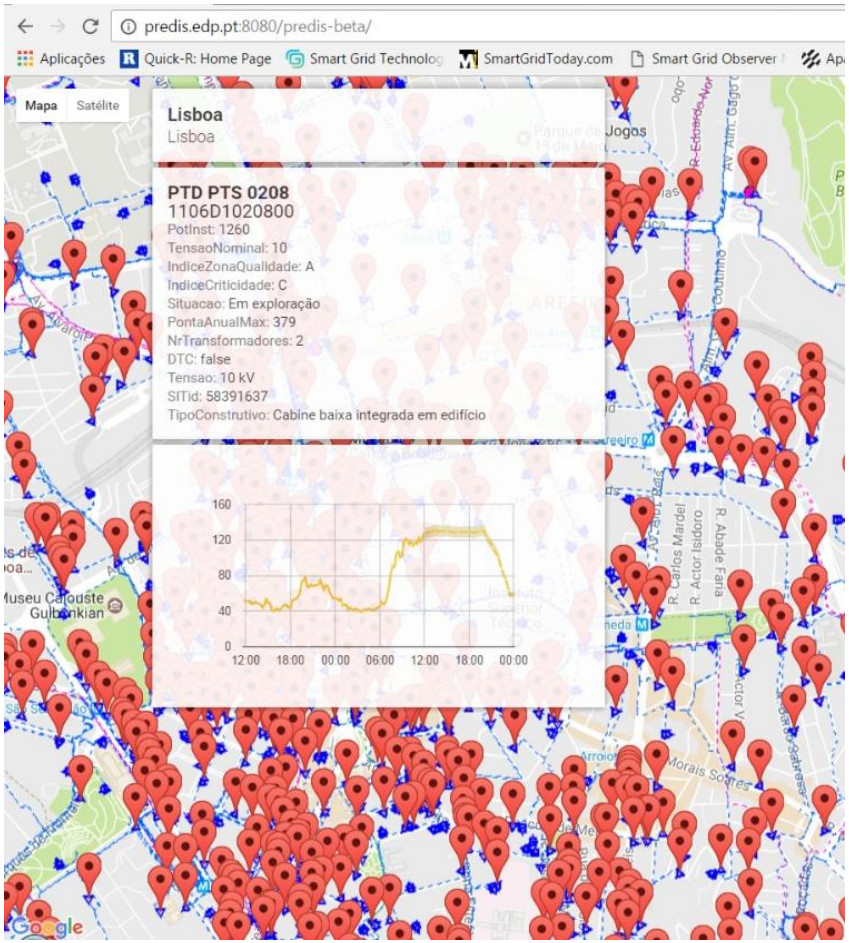
- API (application programming interface)

For "data scientists"

- PREDIS stats e PREDIS sessions
- PREDIS R pakage

For end users

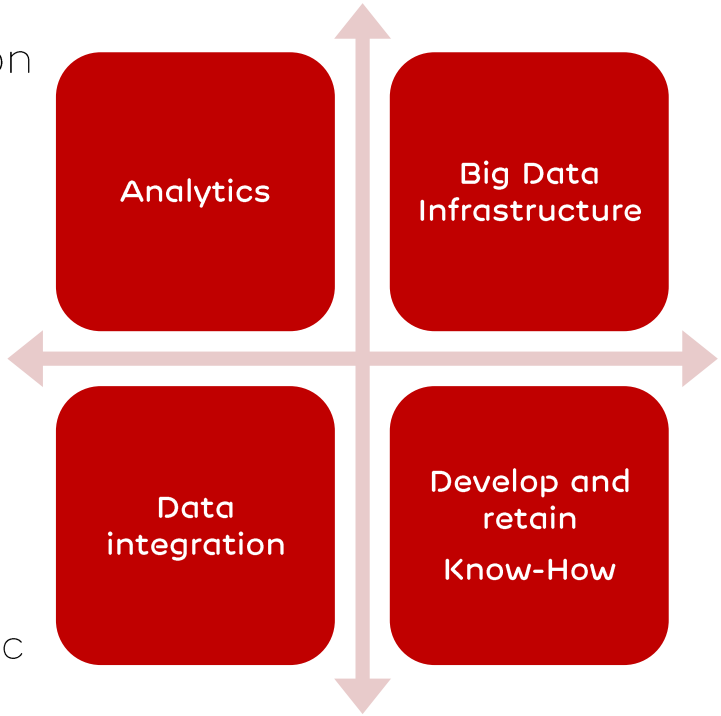
- PREDIS on Google Maps



PREDIS project has been helping EDP to develop knowledge on the four áreas that previously mentioned

Main results:

- Daily Load forecast for 40k grid elements with a granularity of 15mins
- A functional Big Data Infrastructure supported on open source tools
- Mostly internal development, allowing to retain knowledge and promote a data driven culture.



Next Steps:

- Increase the number of forecasted points. Reach all HV/MV grid elements (~80k)
- Support the dynamic tariffs pilot
- Develop na Energy balance considering grid dynamic and use this information to improve forecast
- Update the forecasts every 15min.
- Extend the forecast for the ~6M grid elements

Obrigado!