



Industry Data Management 4.0

Matthieu Dureau - David Duarte

www.linkedin.com/in/david-duarte-5b0627160

www.linkedin.com/in/dureau-matthieu-4994318





The Renault IDM 4.0 program

GROUPE RENAULT



RENAULT NISSAN MITSUBISHI

**International Company bolstered by a unique Alliance
with Nissan and Mitsubishi Motors**

KEY FIGURES 2020



2.95 M

SOLD VEHICLES



170000

EMPLOYEES AROUND THE WORLD



5

BRANDS



39

FACTORIES



3

INNOVATION LABS



6

TECHNICAL CENTERS



Renault



Dacia



LADA

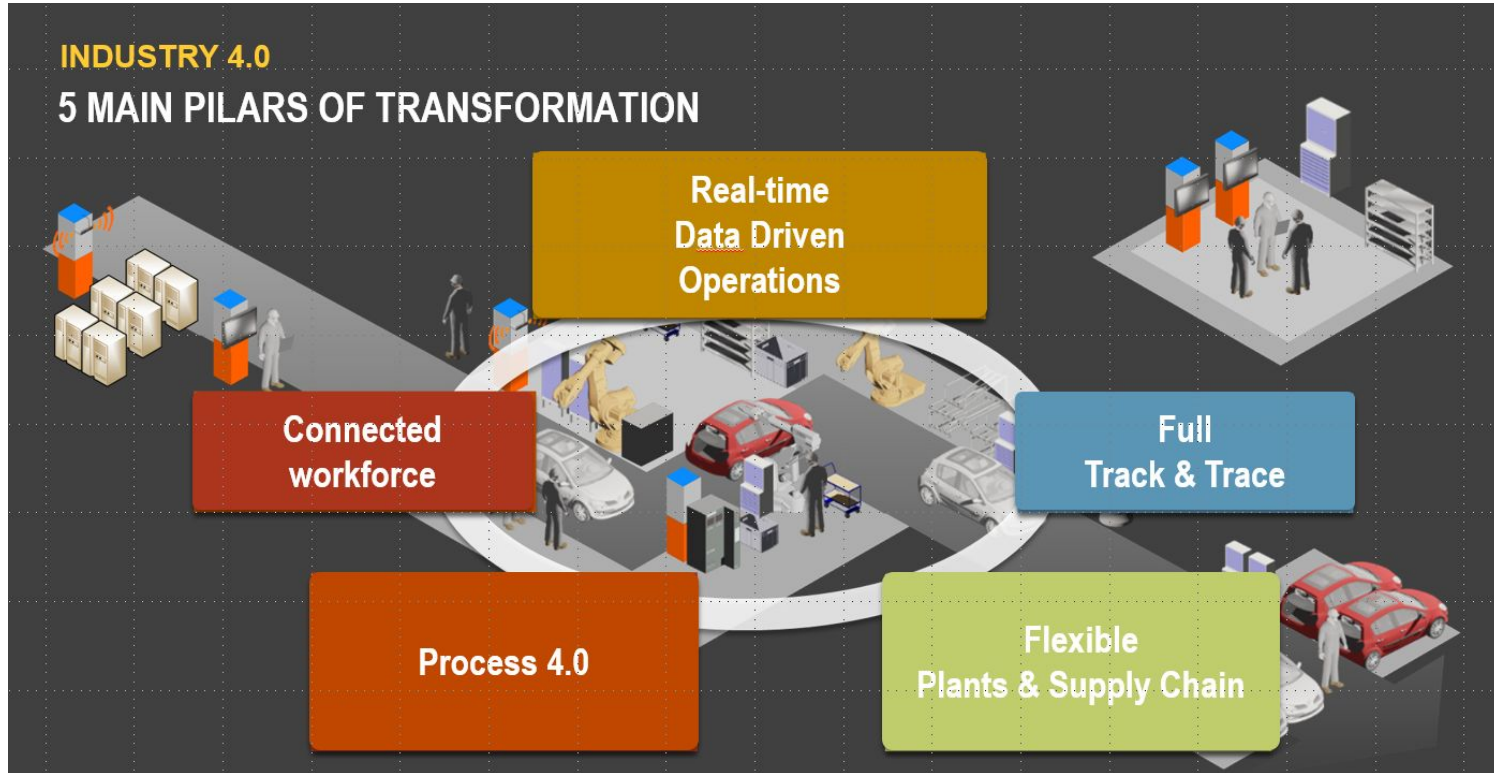


Alpine



Mobilize

The Renault IDM 4.0 program



The Renault IDM 4.0 program

INDUSTRY 4.0 DATA MANAGEMENT

The ambition of the Industry Data Platform program is to enable Manufacturing, Supply CHain and Production Engineering teams to develop quickly Analytics, AI, and Predictive applications based on a single industrial Data Capture and Data Referential architecture



Industry
Data
Exploitation

Provide products and platforms for corporate & plants users for non expert to data-scientists that allows exploration, analysis and development of predictive models, dashboards and applications on dynamic data.

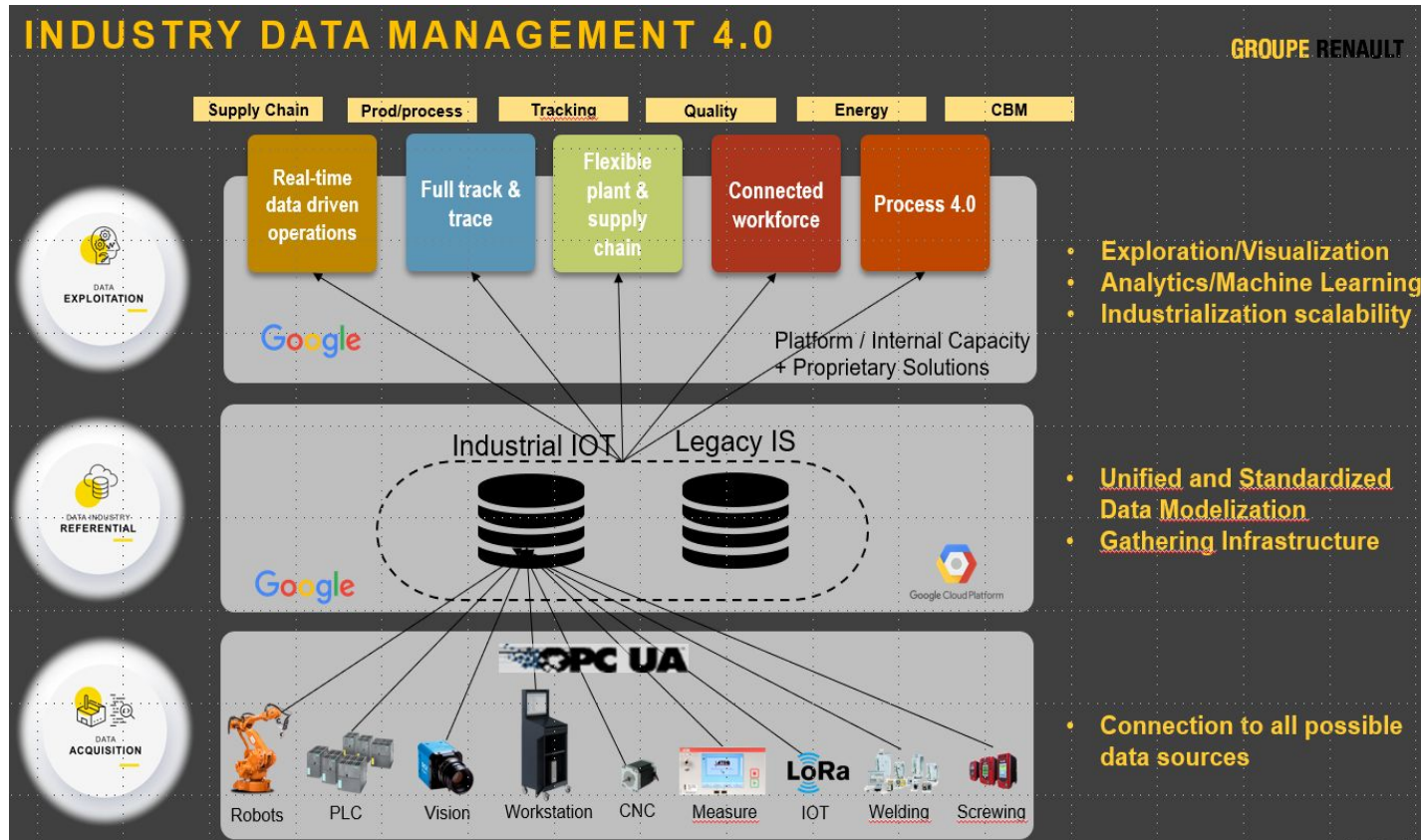
Industrial
Data
Referential

Expose contextualized and cleaned data at the right level of aggregation

Industry
Data
Capture & Publish

Collect data once, in a secured, resilient and standard way and expose it

The Renault IDM 4.0 program



TODAY

- > 45 Data Models
- > 4 Millions Avatars
- > 850 Millions Records
- 15 Plants
- >3000 Data Sources
- 1 Billion Msg/Day

Data processing requirements



DATA

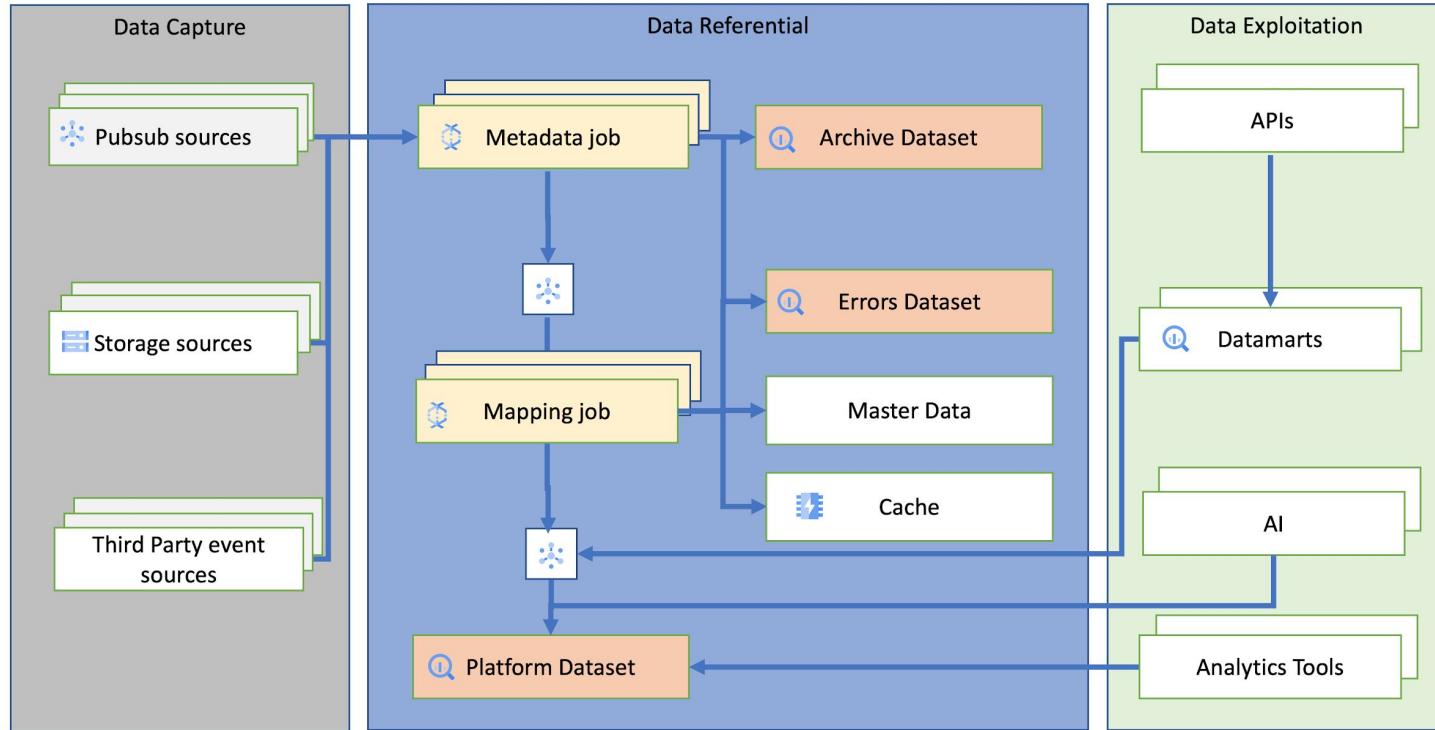
- Several data sources to ingest
- Several formats to handle
- High volumes of data to process
- Depending on source, ingestion with batch or in streaming



OPERATIONS

- Ingestion job monitoring
- Alerting on errors
- Data replay mechanisms on errors

Architecture





Dataproc to Dataflow : Why ?

1

Dataproc Best Use Case

Lift and Shift

2

Anticipated streaming (unified model Beam for batch & streaming)

Write Once, Run It for batch
and stream

3

Total Cost of Ownership

- Reduce cost of dedicated underused Dataproc cluster for streaming 24/24 - 7/7
- Better cost identification with easier job labelling



Dataproc to Dataflow : Why ?

4

**Availability of Java
Developers**

5

**Easiness of
development/test**

Use dedicated unit test/end
to end pipelines

6

Connectivity

Availability of IO connectors

Dataproc to Dataflow : Why ?

7

Easiness of Beam updates

Change the dependencies,
no need to recreate an
updated cluster as with
Dataproc

8

Easiness of operation/scaling

- Easy to monitor with dedicated alerting
- Efficient autoscaling

9

Error management

Multiple sinks : one for processed
data, one for errors



Lessons learned

BEST PRACTICES

**Make sure you
integrate end to end
data consistency
during ingestion
process**

**Don't forget to
implement
streaming failures
detection**

**To restart after
connection failures
and fix potential
corruptions, you can
replay data**

Next steps



Fine-tune for
our batch
& streaming
needs



Backup Dataflow
error handling
with long-term
archiving of our
datasets

Our advice for big data processing: go with Dataflow!

Results & Impacts

1

Reached
OPEX higher
stability

2

Improved
End to end
testing

3

Fine-tuning
thanks to
flexibility to use
batch and
streaming

4

Able to handle
a wide range
of industrial
events types
(OPC-UA, ...)

5

Scalable
platform for
future growth