Big Data Analytics

for Smart Cities and Industries

Maribel Yasmina Santos

ALGORITMI Research Centre
Department of Information Systems
University of Minho, Portugal

Lynx lab
Discovering Invisible Knowledge

CENTROALGORITMI

Centro de Computação Gráfica

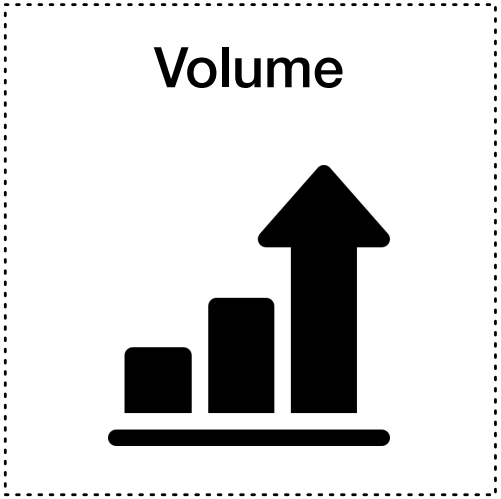
Universidade do Minho

Outline

- Characteristics and Definition
- Conceptual and Technological Developments
- Challenges for Analytics in Big Data Warehousing
- Demonstration Cases: the SusCity and iFactory projects

Big Data

- How can be characterised?



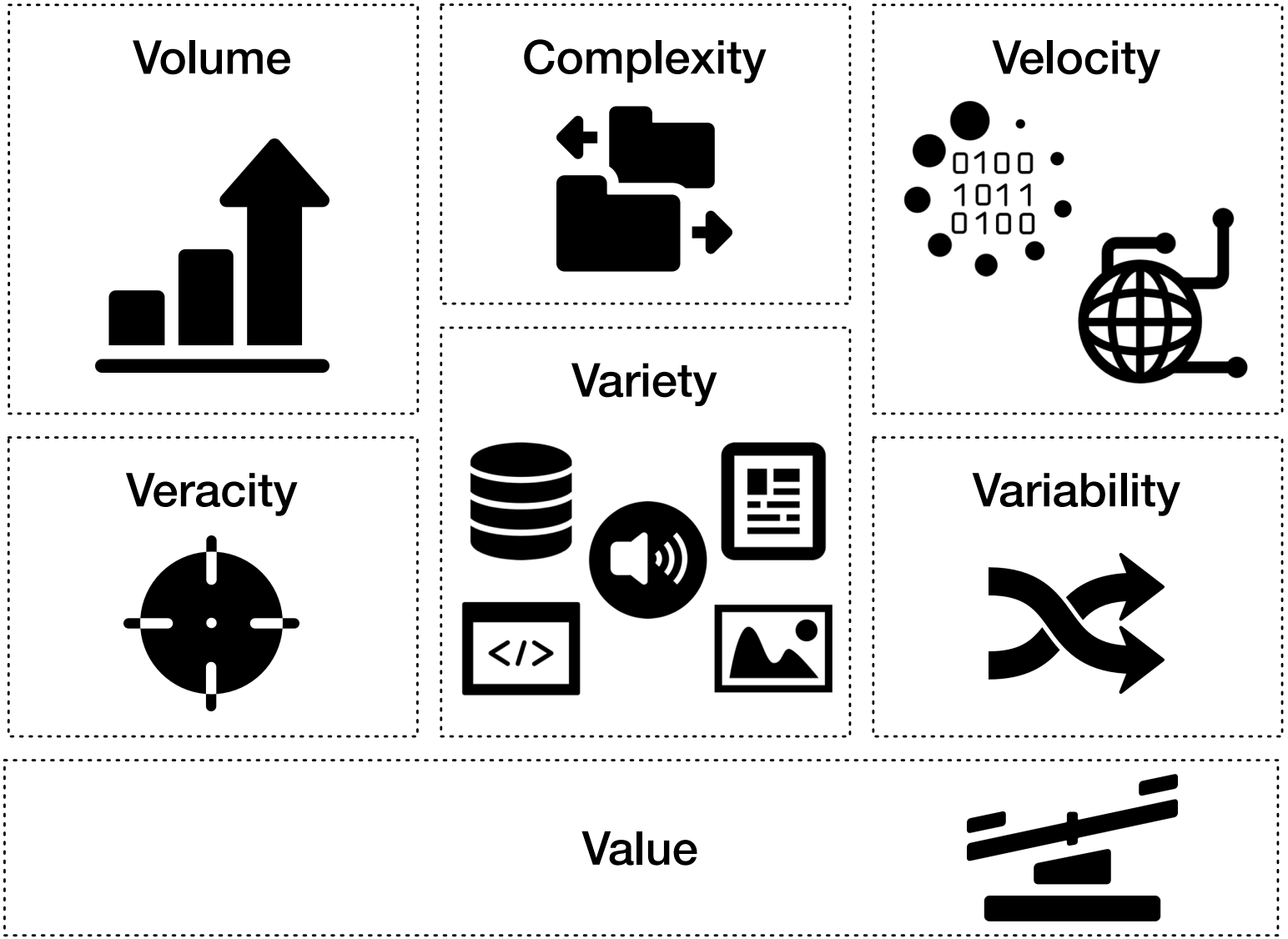
Data **available** in organisations

Data **processed** in organisations

Big Data

- Other relevant characteristics?

Big Data


Volume
Complexity
Velocity
0100
1011
0100
Variety
Veracity
Variability
Value

Big Data

- How can be defined?

From Databases to Big Data

Sam Madden • *Massachusetts Institute of Technology*

IEEE INTERNET COMPUTING

What Is Big Data?

data that's *too big*, *too fast*, or *too hard* for existing tools to process.

Big Data

- Conceptual and technological developments?
  - Conceptual

International Journal of Information Management 35 (2015) 137–144

Contents lists available at ScienceDirect

International Journal of Information Management

journal homepage: www.elsevier.com/locate/ijinfomgt

Beyond the hype: Big data concepts, methods, and analytics

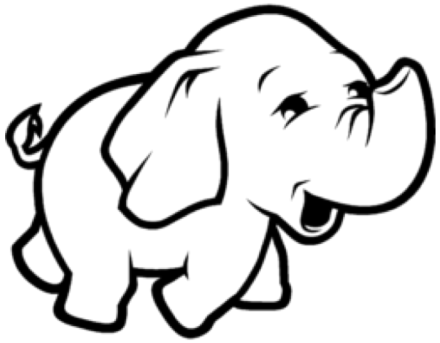
Amir Gandomi*, Murtaza Haider

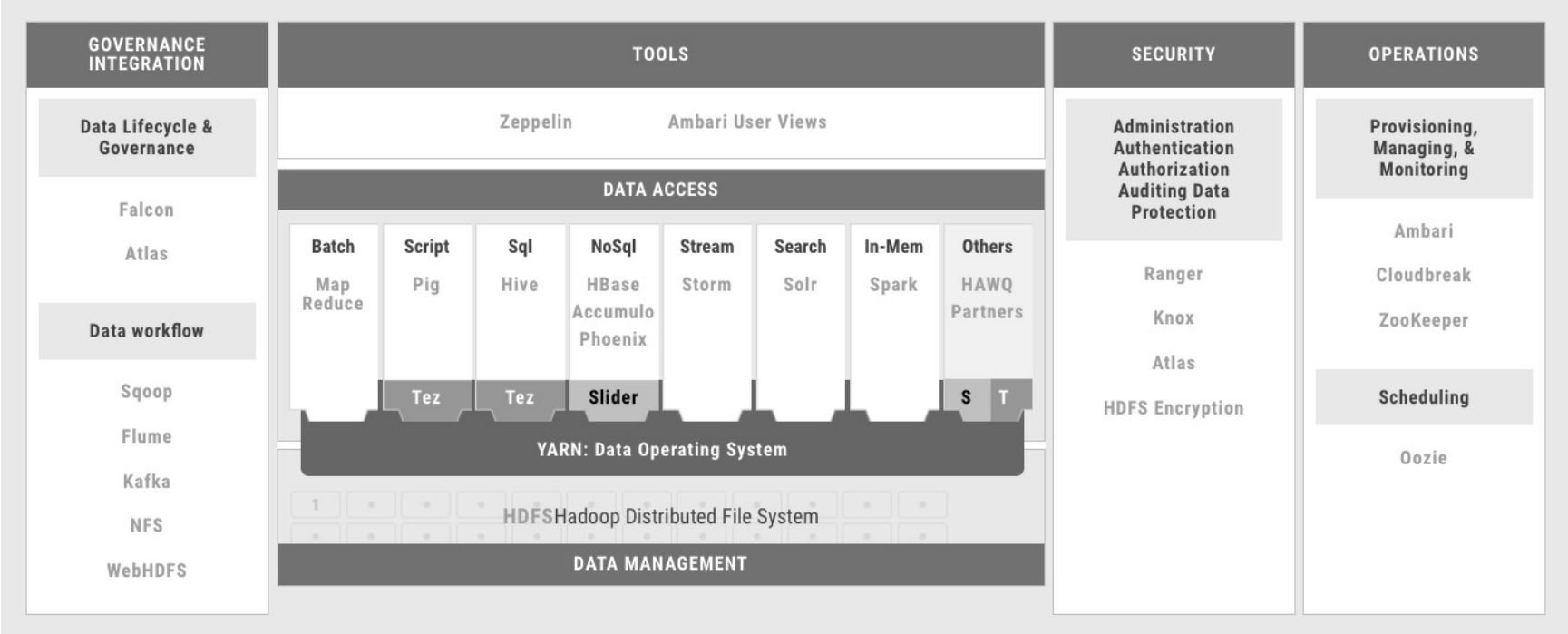
Ted Rogers School of Management, Ryerson University, Toronto, Ontario M5B 2K3, Canada

The fast evolution of big data left little time for the discourse to develop and mature in the academic domain.

Big Data

- Conceptual and technological developments?
  - Technological





https://hortonworks.com/ecosystems/

HORTONWORKS

BIG DATA & AI LANDSCAPE 2018

FIRSTMARK
EARLY STAGE VENTURE CAPITAL

https://industrypulse.com/2018/07/01/matt-turcks-2018-big-data-ai-landscape/

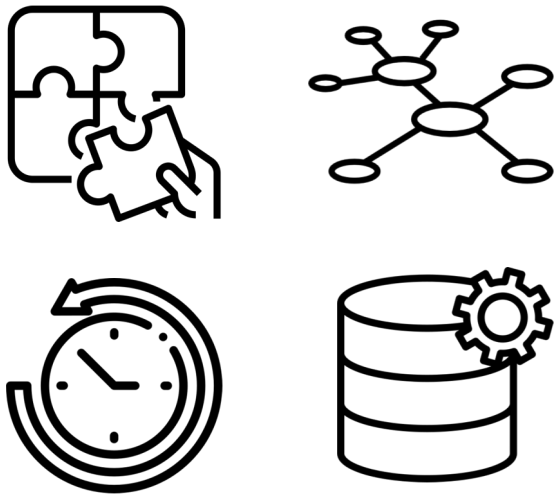
Big Data (Analytics)

- Ambiguity…
  - Lack of common technical & scientific approaches
  - Wide range of technologies
  - Lack of skills in IT teams

Big Data Analytics

Traditional Data Warehouse

Big Data Warehouse

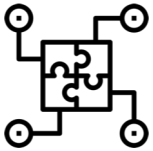


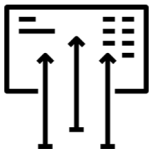
Big Data Analytics

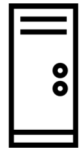
- Characteristics
  - mixed and complex analytical workloads (e.g., ad hoc querying, data mining, text mining, exploratory analysis and materialized views)
  - real-time operations (stream processing, low latency and high frequency updates)
  - scalability to accommodate growing data, users and analysis

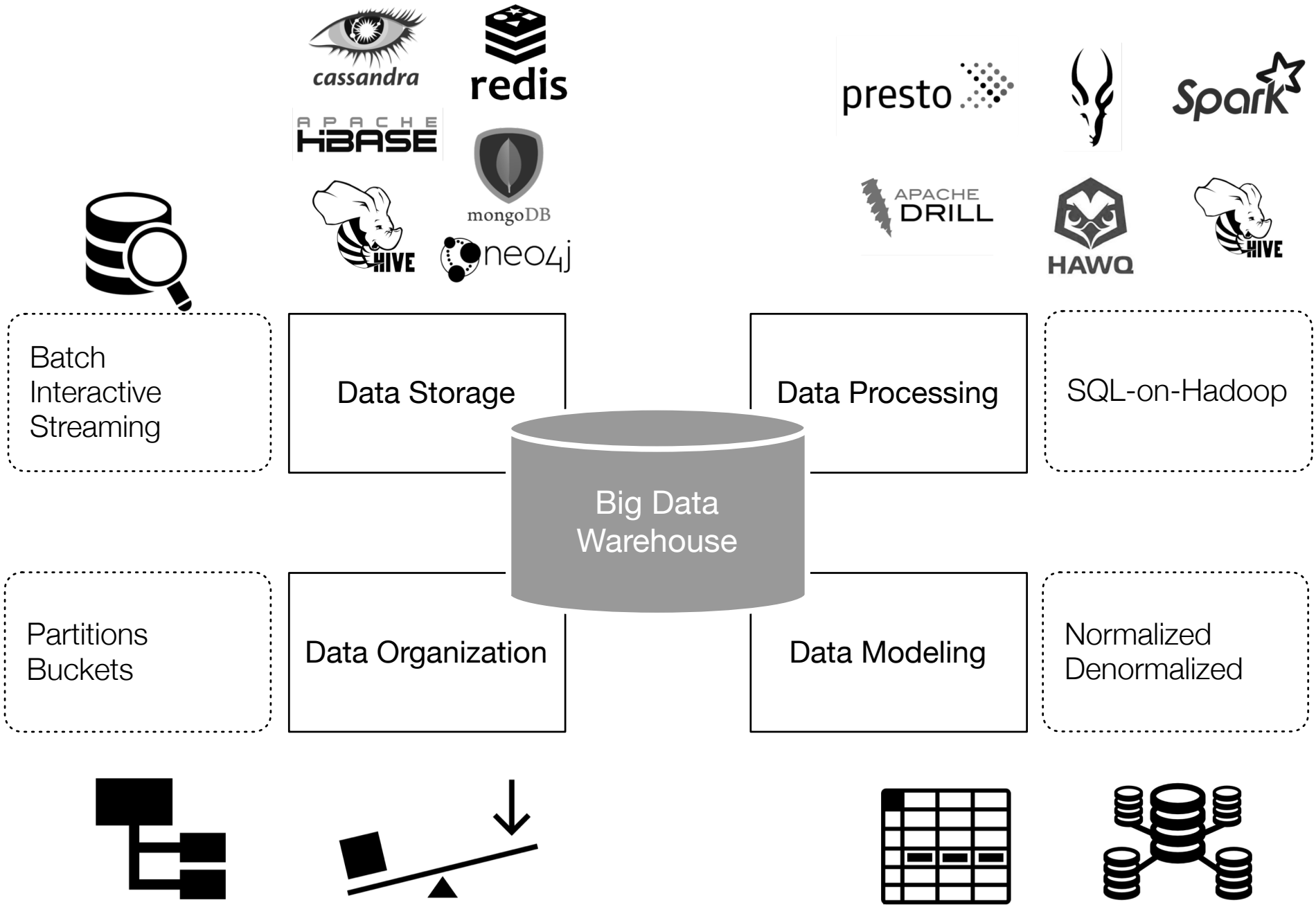
Big Data Analytics

- Characteristics
  - massively parallel processing
  - flexible storage to support data from several sources
  - use of commodity hardware to lower costs
  - interoperability in a federation of multiple technologies

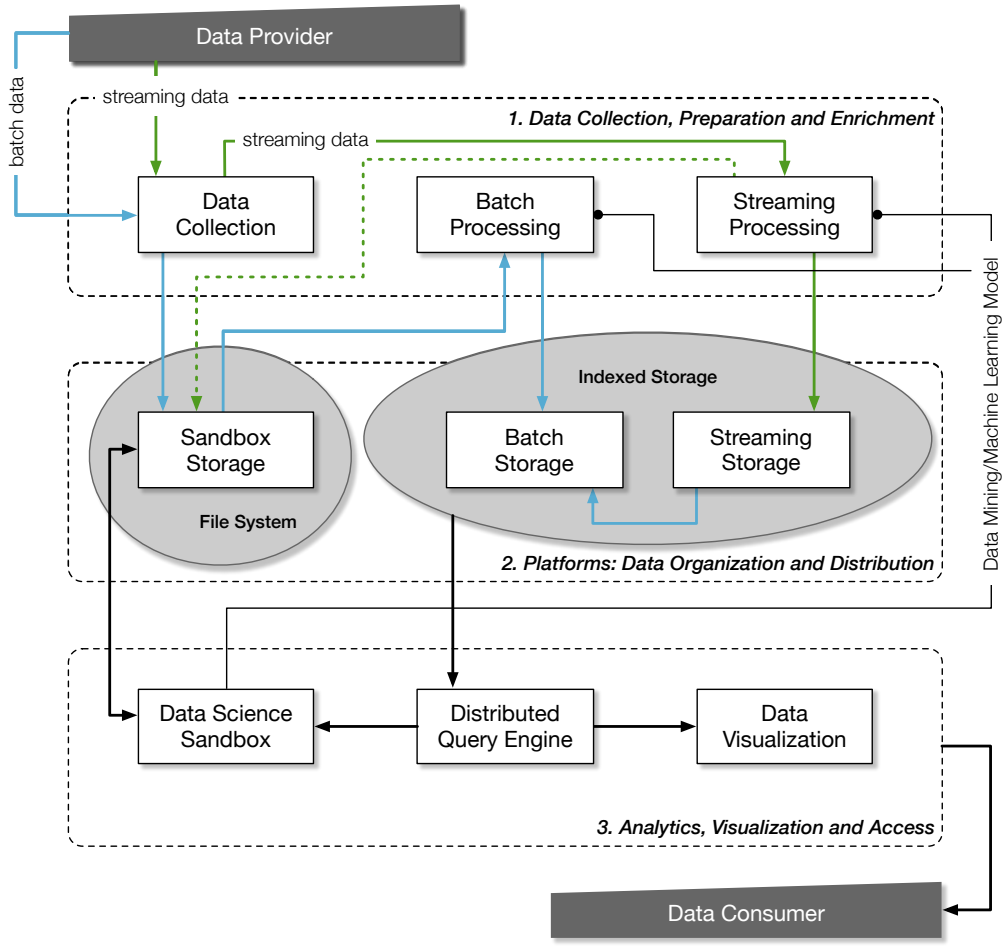






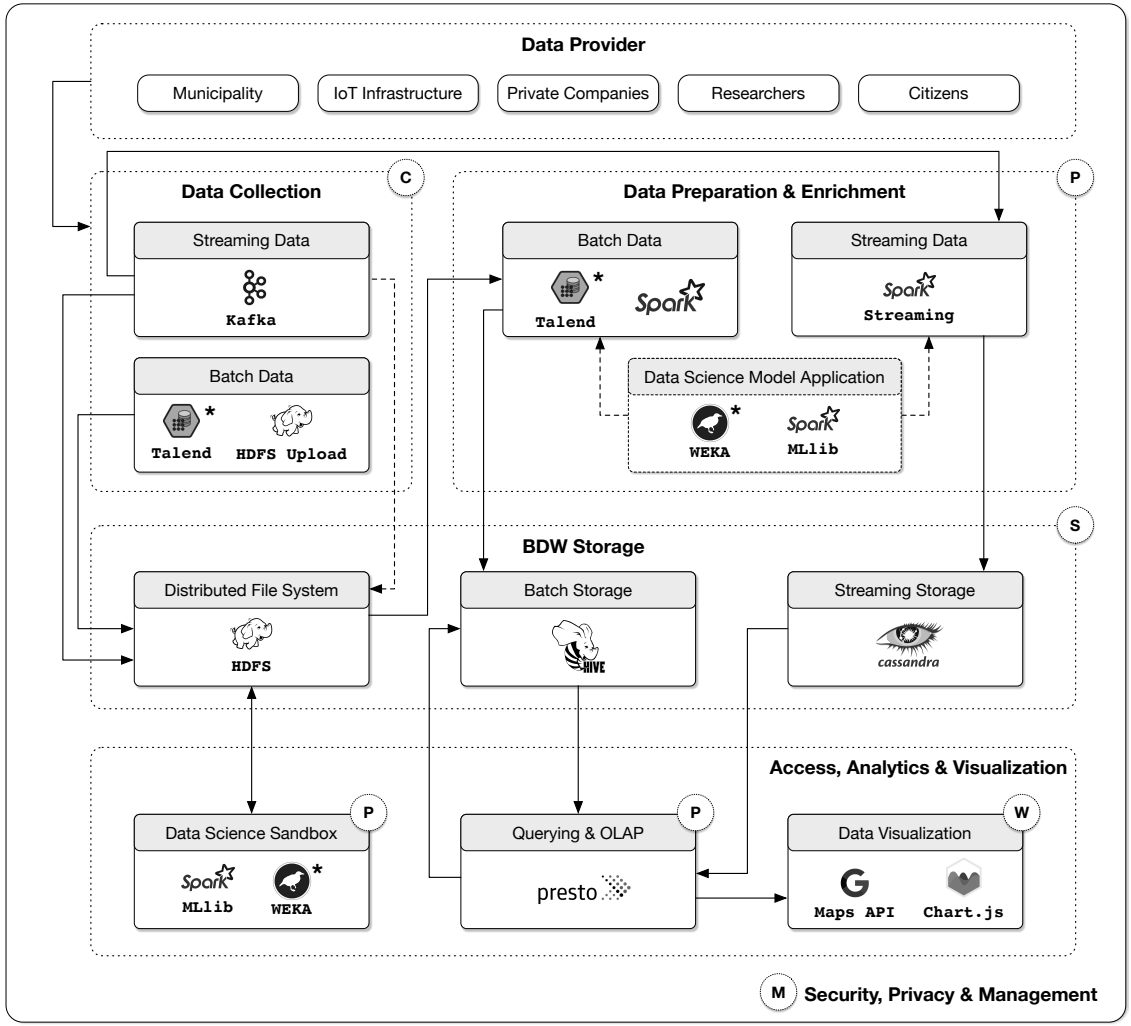

cassandra
redis
APACHE HBASE
mongoDB
HIVE
neo4j
presto
Spark
APACHE DRILL
HAWQ
HIVE
Batch
Interactive
Streaming
Data Storage
Data Processing
SQL-on-Hadoop
Big Data Warehouse
Partitions
Buckets
Data Organization
Data Modeling
Normalized
Denormalized

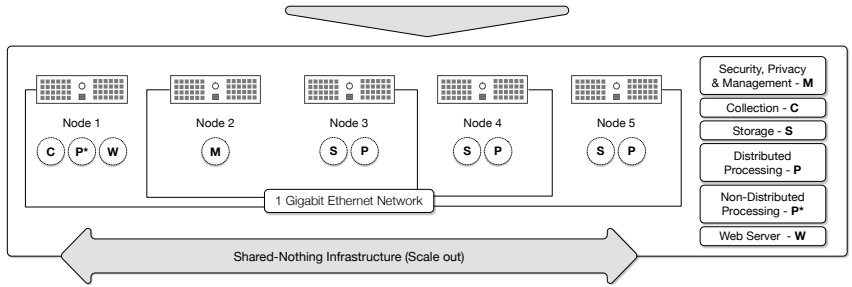
BDW Architecture



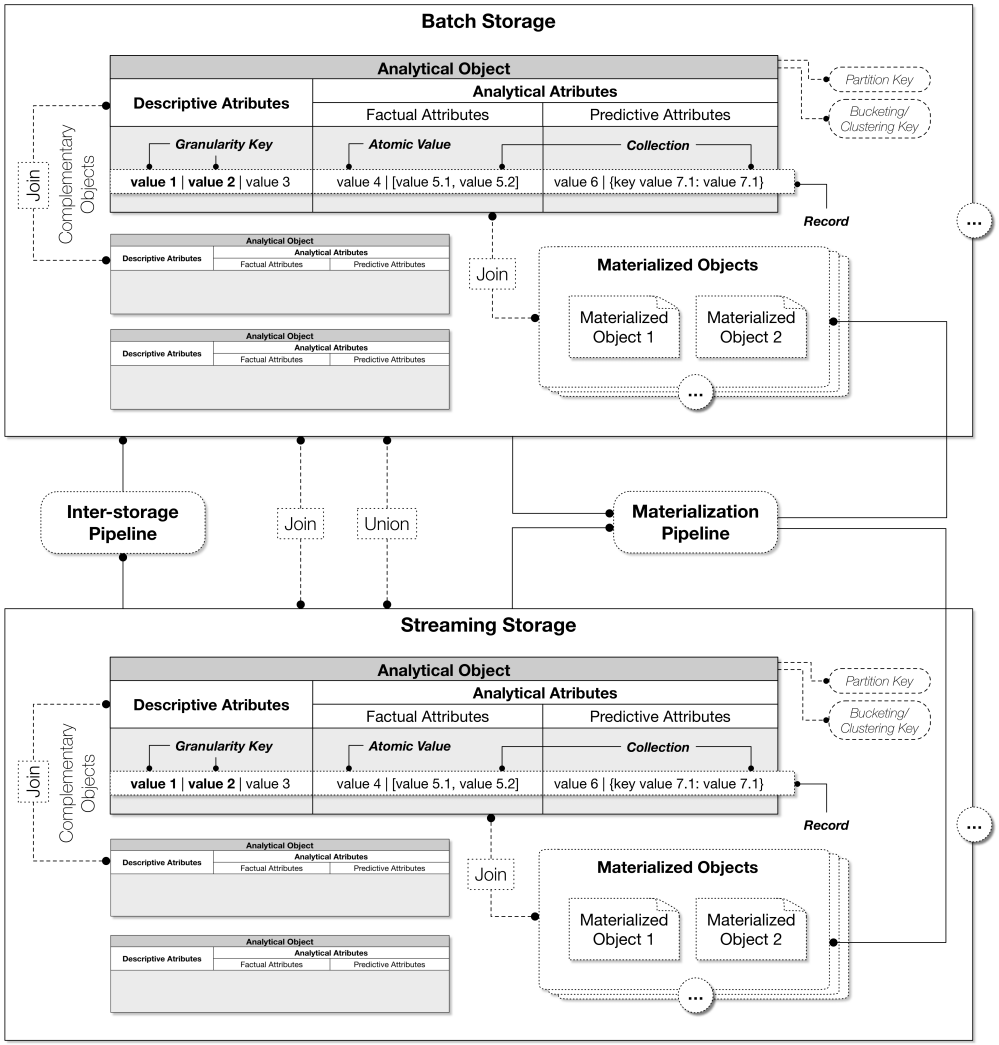
Costa, C., C. and M. Y. Santos, "Big Data Warehouses for Smart Industries", in Sherif Sakr and Albert Y. Zomaya (Eds.), Encyclopedia of Big Data Technologies, Springer, 2018.

Technological Infrastructure





Costa, C., and M. Y. Santos, "The SusCity Big Data Warehousing Approach for Smart Cities". In Proceedings of International Database Engineering & Applications Symposium (IDEAS'17), Bristol, United Kingdom, 12-14 July, 2017.



Analytical Objects

- Fully denormalized structures
- Nested structures (e.g., arrays, maps)
- Descriptive attributes = dimensions
- Analytical Attributes = facts and predictions

Materialized Objects

Unified **Batch and Streaming**

- Same modelling approach
- Integrate both in the same query
- Possible due to SQL-on-Hadoop

Costa, C., and M. Y. Santos, "The SusCity Big Data Warehousing Approach for Smart Cities". In Proceedings of International Database Engineering & Applications Symposium (IDEAS'17), Bristol, United Kingdom, 12-14 July, 2017.

Costa, C., C. and M. Y. Santos, "Big Data Warehouses for Smart Industries", in Sherif Sakr and Albert Y. Zomaya (Eds.), Encyclopedia of Big Data Technologies, Springer, 2018.

…advance the science of urban systems modeling and data representation supported by urban "big data" collection and processing, with the double objective of enabling and demonstrating new services that explore economic opportunities associated with the transition to sustainable urban systems.

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

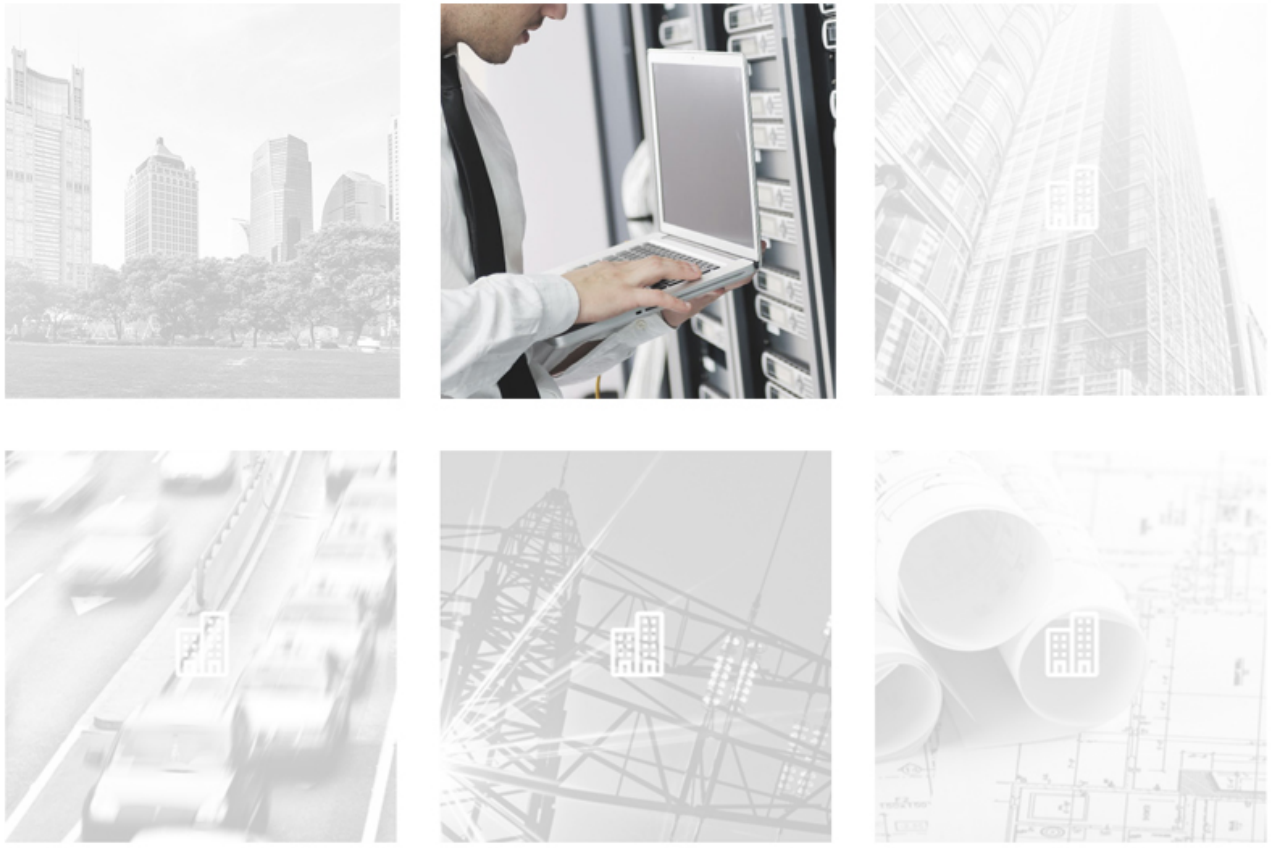
INESCPORTO LABORATÓRIO ASSOCIADO · IST-ID Associação do Instituto Superior Técnico para a Investigação e Desenvolvimento · Massachusetts Institute of Technology · Universidade de Coimbra · Universidade do Minho · Ciências ULisboa · idmec · R&D NESTER · iTds · LNEG · ADENE Agência para a Energia · edp distribuição · NOVABASE · IBM · Câmara Municipal Lisboa

SusCity Project

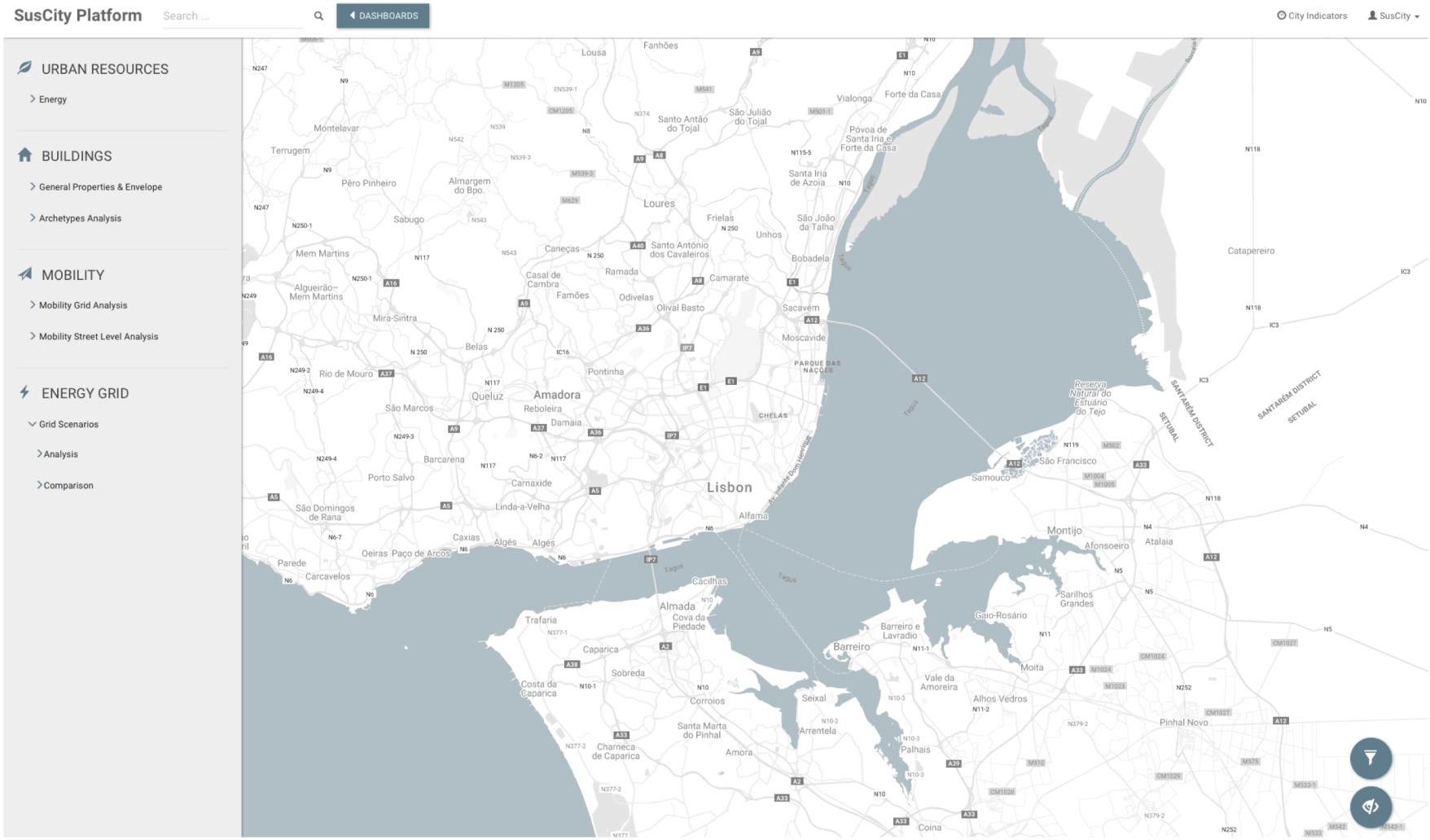




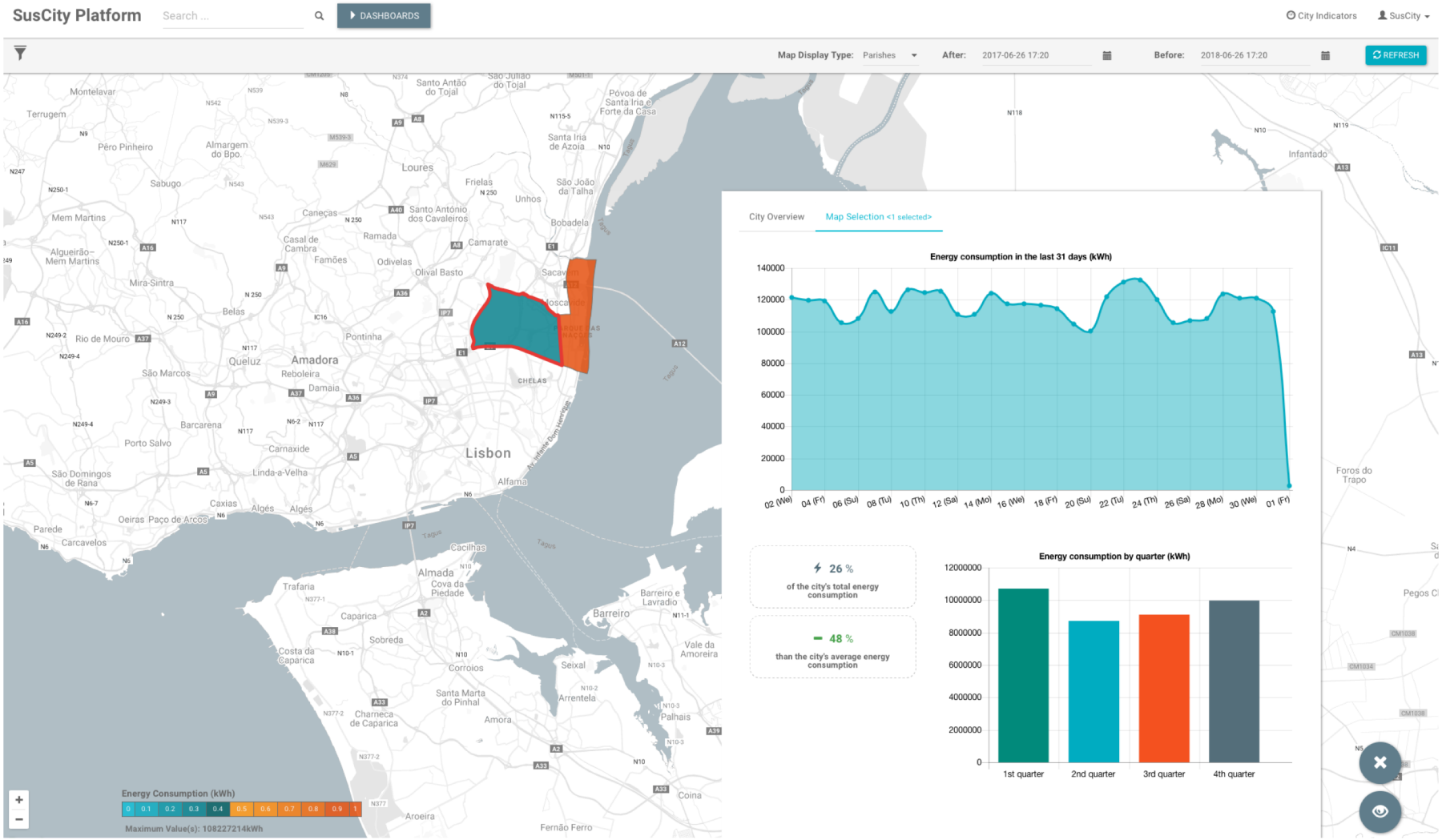

Information Services and Data Processing Platform



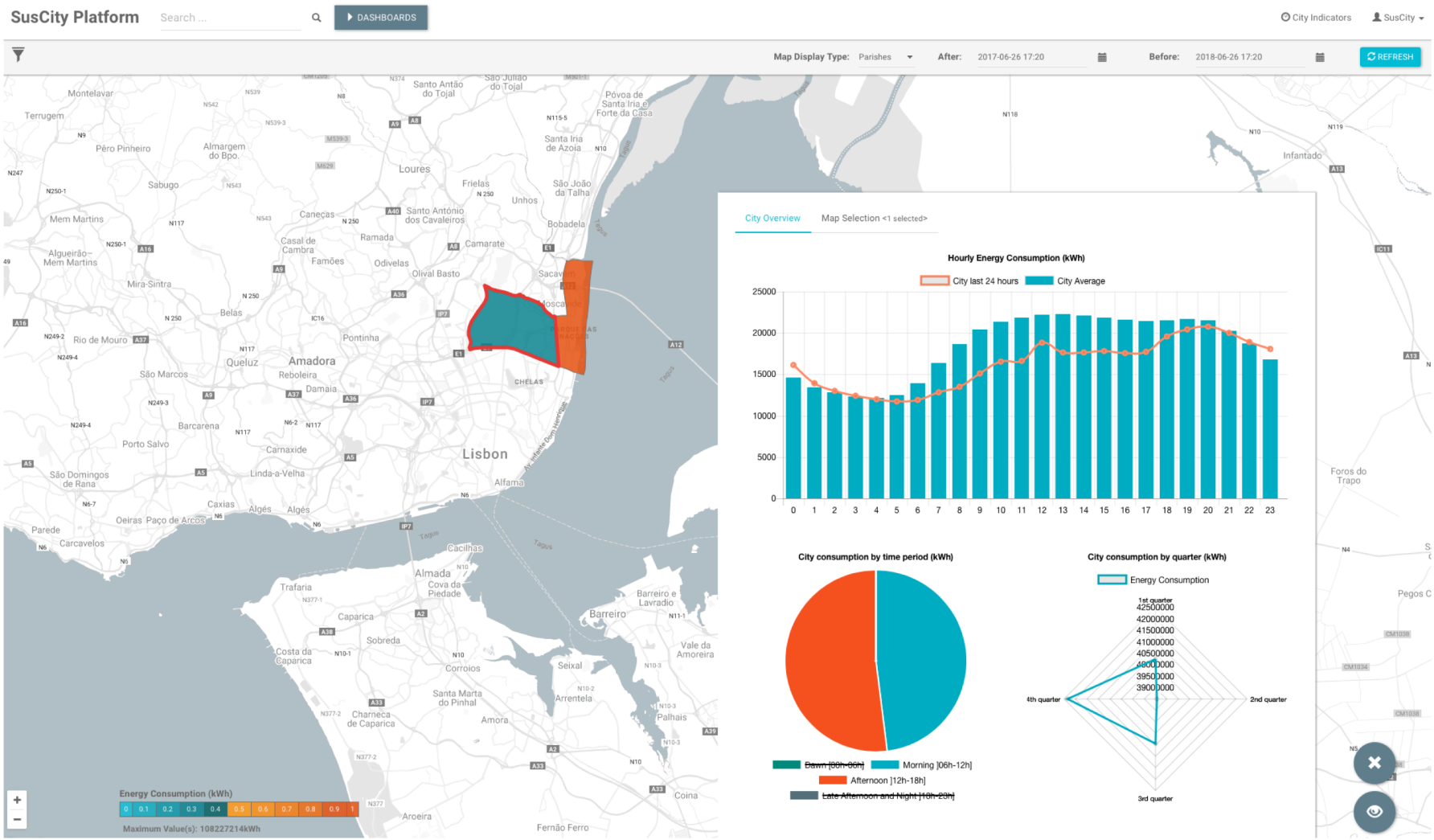
Analytical Dashboards



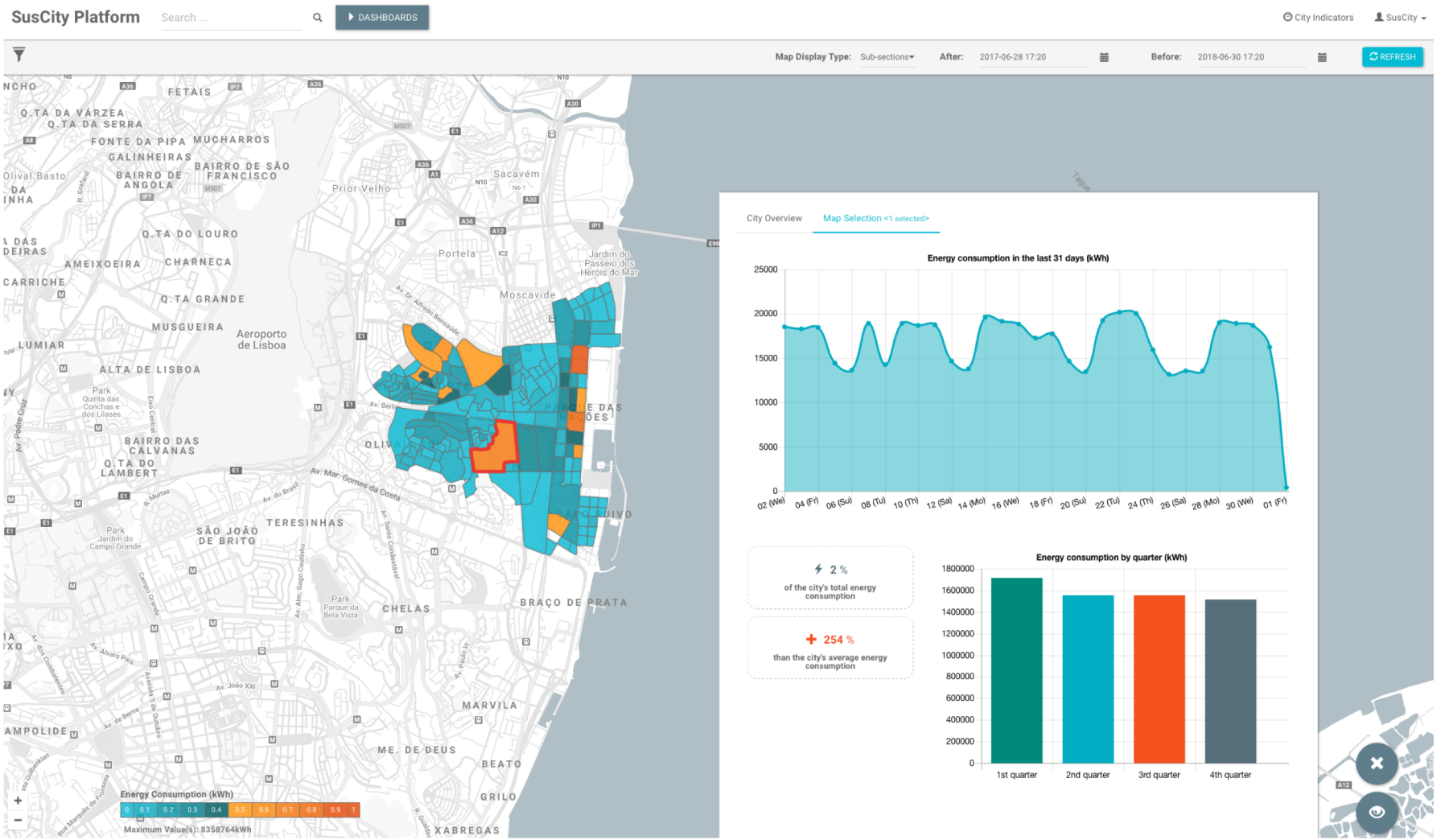
Analytical Dashboards



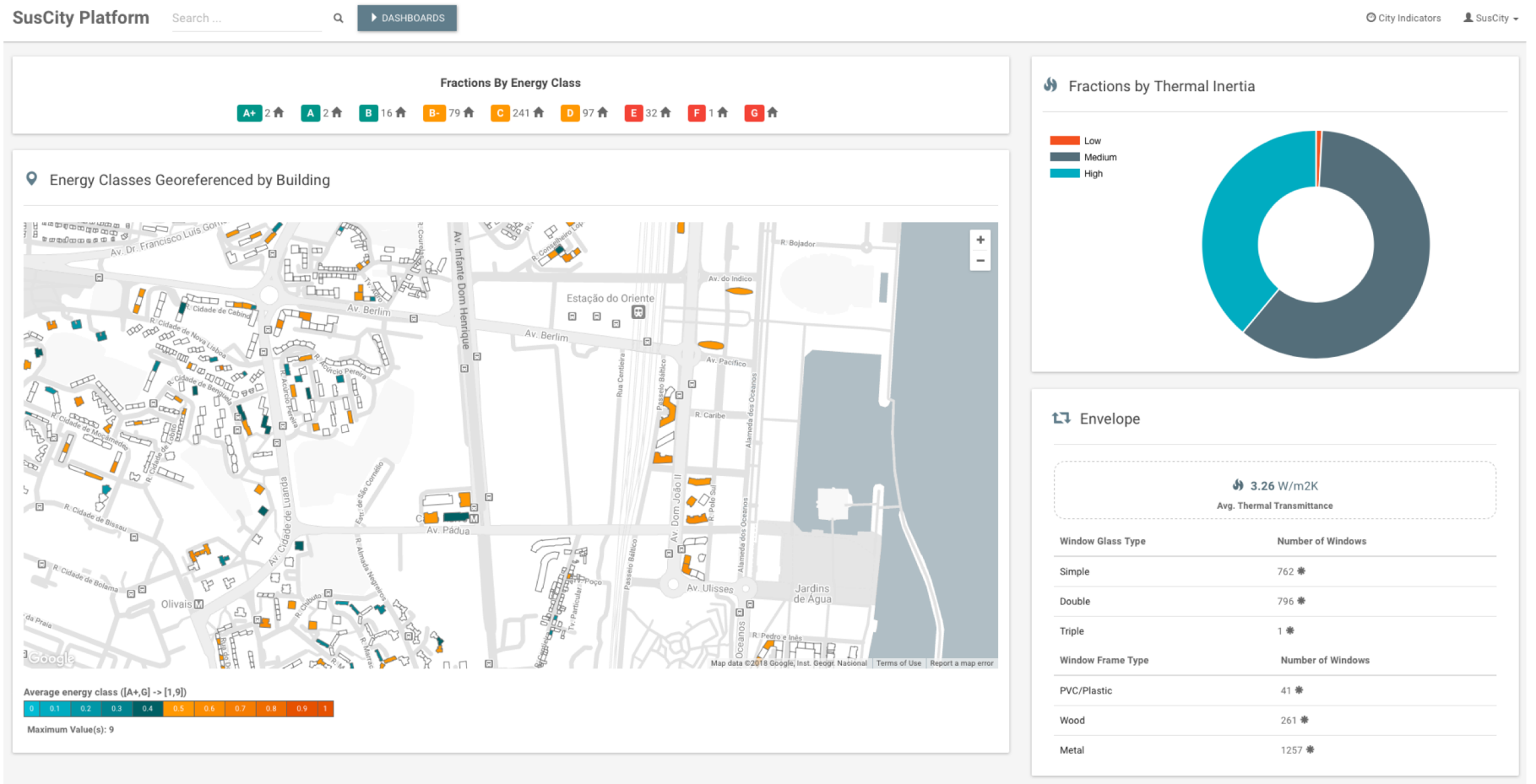
Analytical Dashboards



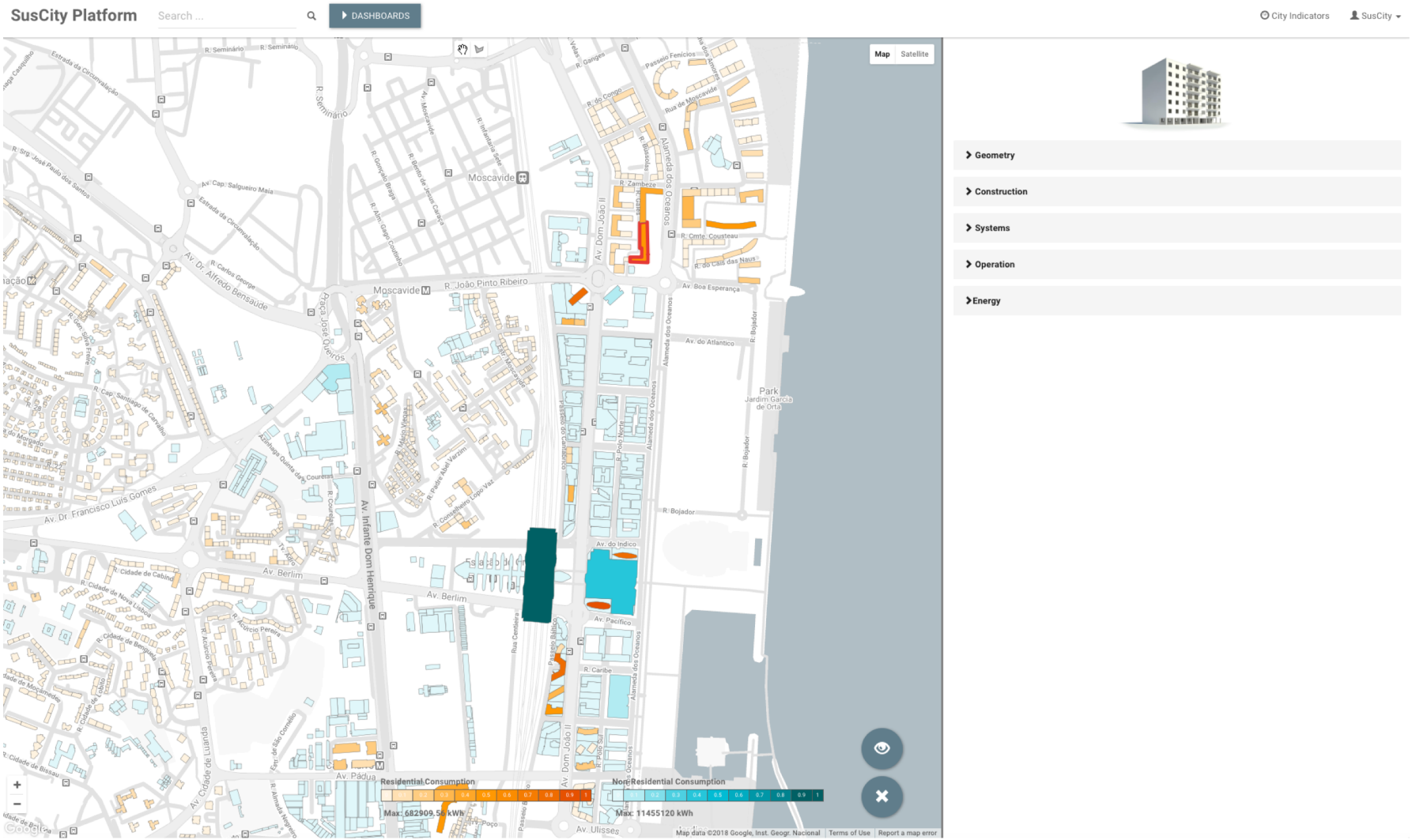
Analytical Dashboards



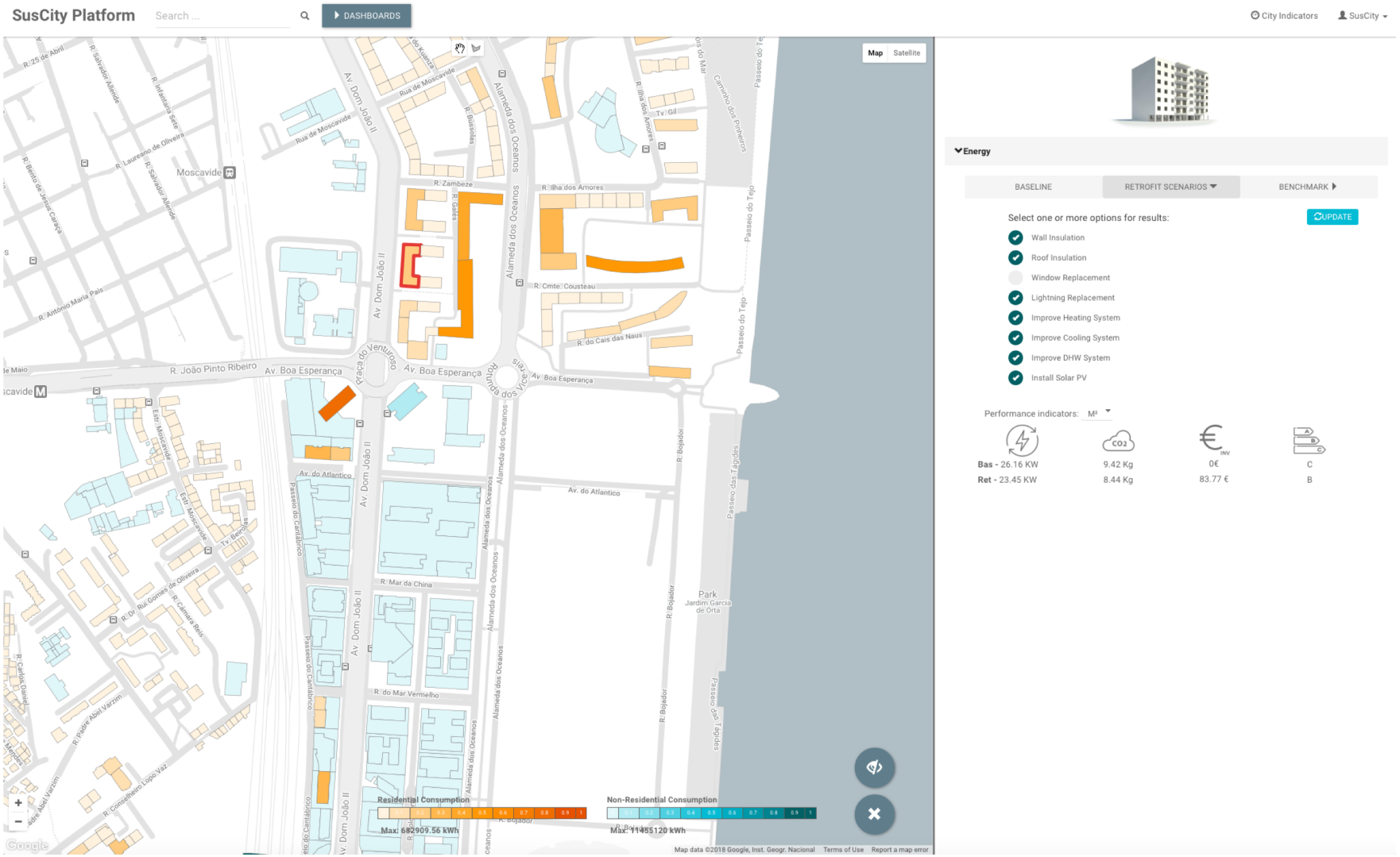
Analytical Dashboards



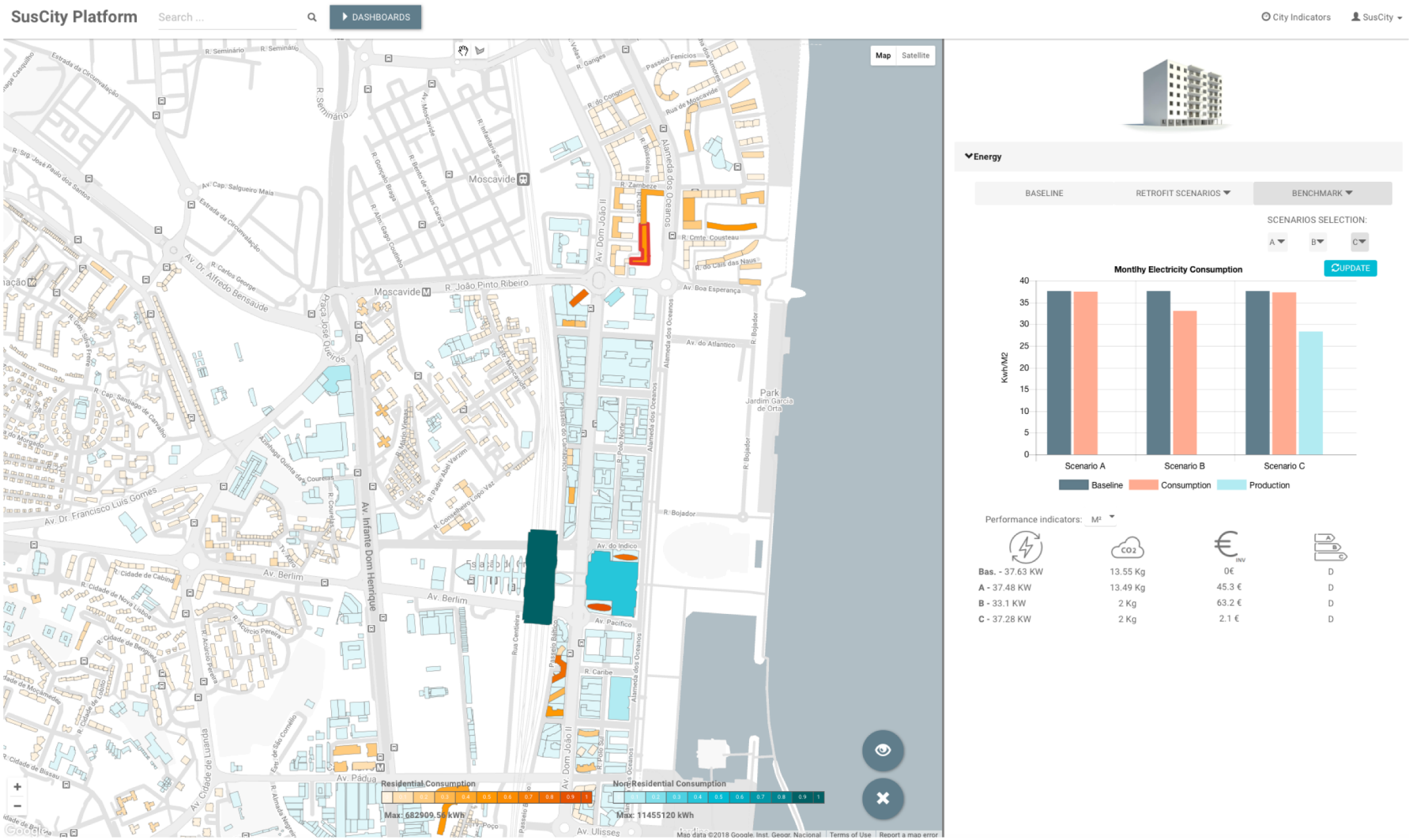
Analytical Dashboards



Analytical Dashboards

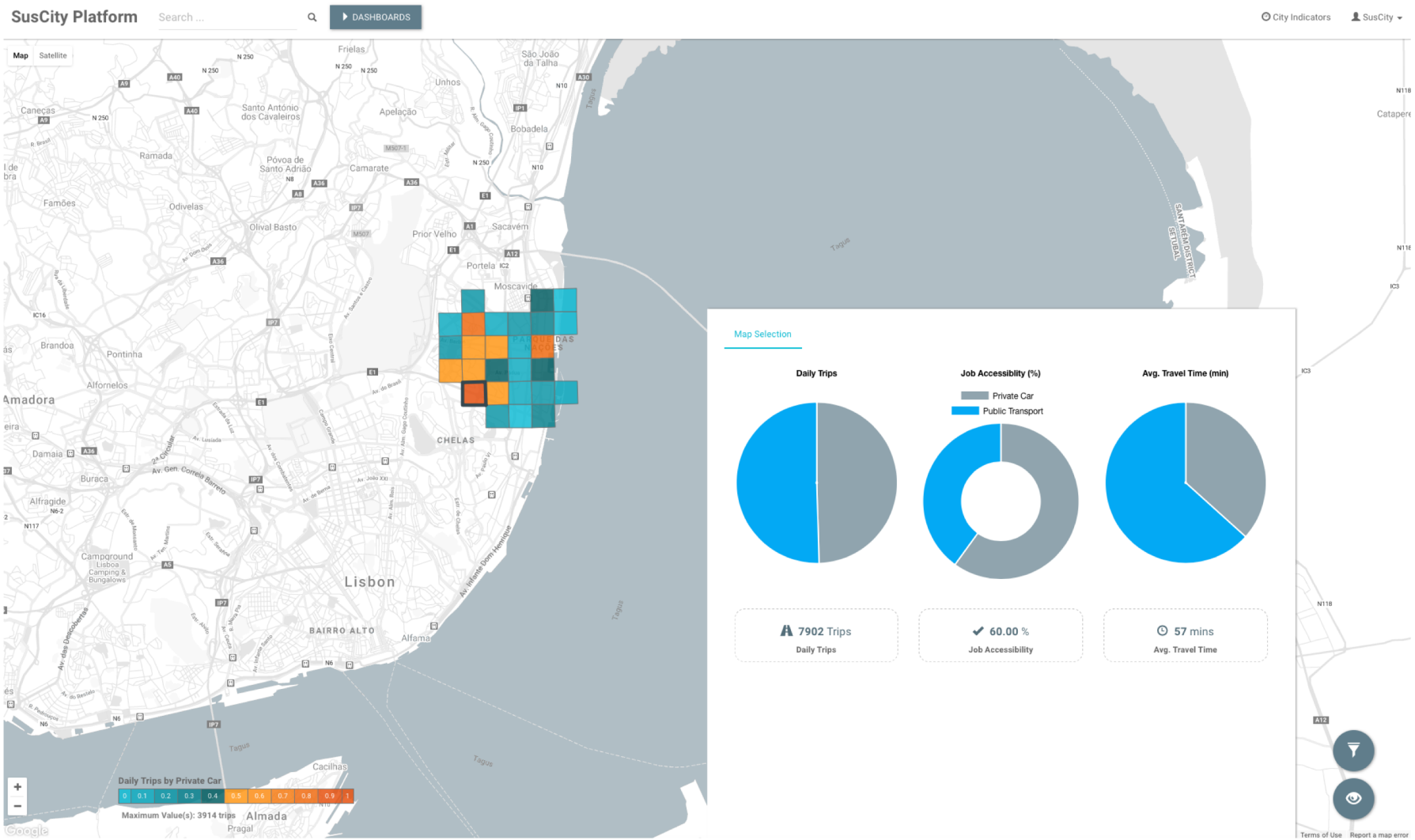


Analytical Dashboards



Analytical Dashboards

Analytical Dashboards



COMPETE 2020
PROGRAMA OPERACIONAL COMPETITIVIDADE E INTERNACIONALIZAÇÃO

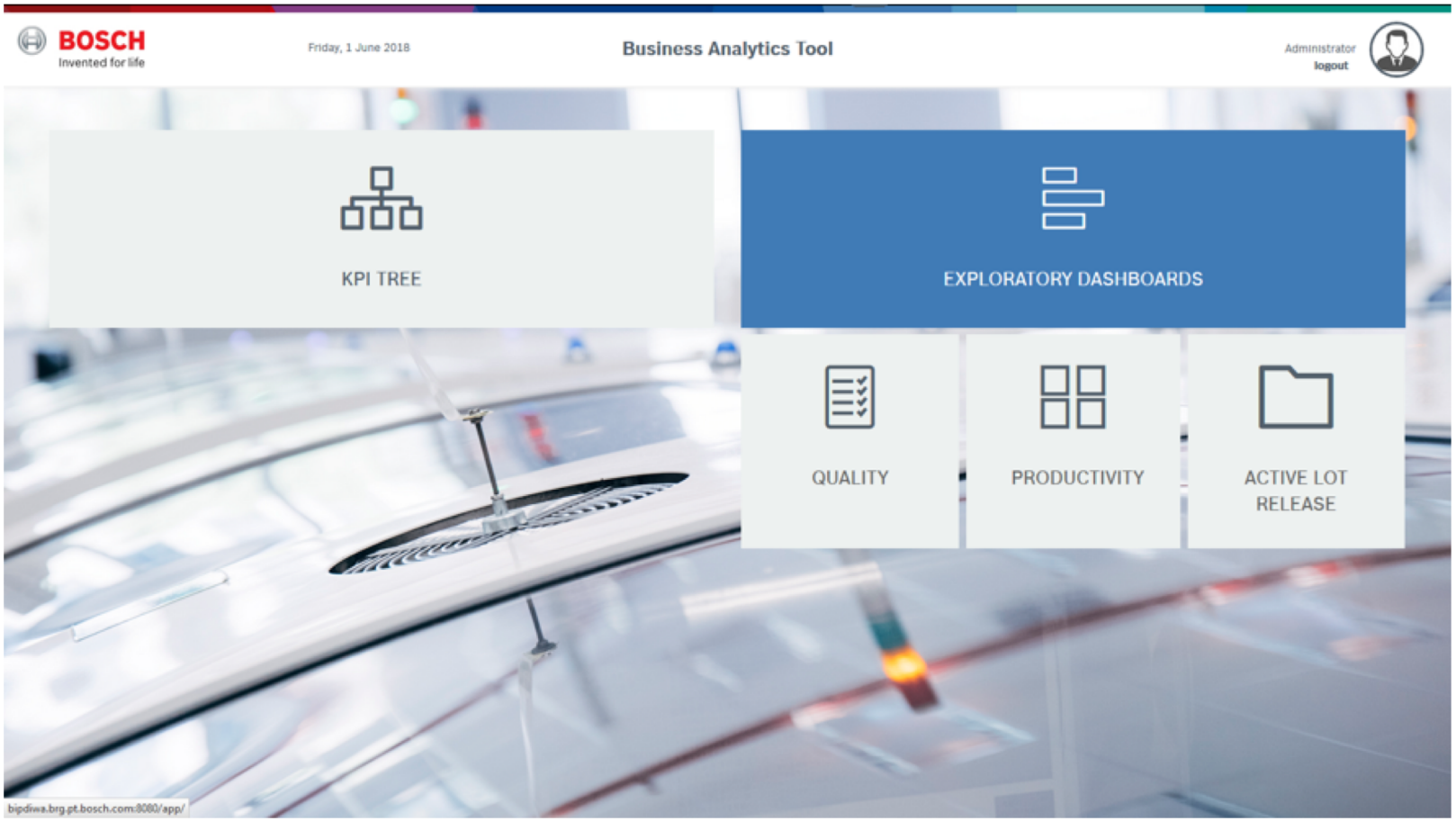
PORTUGAL 2020

UNIÃO EUROPEIA
Fundo Europeu
de Desenvolvimento Regional

Universidade do Minho

BOSCH
Invented for life

Analytical Dashboards



Acknowledgements

- COMPETE: POCI-01-0145- FEDER-007043 and FCT – *Fundação para a Ciência e Tecnologia* within the Project Scope: UID/CEC/00319/2013 (ALGORITMI)
- SusCity project, MITP-TB/CS/0026/2013
- iFactory project, POCI-01-0247-FEDER-002814

- http://www.flaticon.com
- http://www.onlinewebfonts.com

Lynx lab
Discovering Invisible Knowledge

CENTROALGORITMI

Centro de Computação Gráfica

Universidade do Minho