

Data Toolkit

BigDataStack Software Component developed by UbiTech

The Data Toolkit is the component which takes care to design an end-to-end Big Data application graph and create a common serialization format in order that it is feasible to execute valid analytics pipelines. It supports Spark and Spark Machine Learning (ML) library, but it may also be extended to support other machine learning frameworks such as indicatively Tensorflow and Scikit-learn .

Input

High level graph from Process Modelling component.

Initial TRL

TRL 3

Output

Fine grained, enriched and grounded execution graph.

Final TRL

TRL 5

Data Toolkit

End Users

- Data scientists,
- Data analysts,
- Data engineers

Key Features and Benefits

- Automation,
- Online graph composition constraint enforcement

Standards involved in the development of the component

None

Implementation in BigDataStack Use Cases

All three BigDataStack Use Cases, however, it is not limited to specific UCs as it adopts a generalized approach which can be applicable to other cases too.

Differentiators from competitors in the market

Data Toolkit can be deployed over any type of cloud computing operating system, including OpenStack, Microsoft Azure, etc.