



# Big Data Analytics



## What is it about?

The **Big Data Analytics Starter Kit** provides a broad overview on how to process and extract valuable information from big data in an automatic, semi-supervised manner. As more and more data become available, it seems increasingly necessary to find methods that can efficiently process, analyze and display the information they contain. Experience and reliable guidance are required to navigate through the panorama of techniques and methods available in data analysis. We offer a starter kit to address both theoretical and practical aspects of extracting highly valuable information from big data, so that they can easily be used in downstream analyses.

## Why do i need this?

- I have processes which collect **large amounts of data** (either as batch or streams).
- My **data sources** are **heterogeneous**, and they do not follow a fixed structure.
- I spend far **too much time** in **data tagging** and **preparation**.
- I need to **extract valuable information** that needs to be passed on to downstream production steps.
- I do not want to **estimate the data life** cycle, nor I want to manually adjust my storage requirements.

- I neither have **time nor money** to get the minimum amount of information regarding analytical methods.
- I need to **identify potential issues** before failures occur in my processes/services.
- I prefer to **dynamically visualize** and interpret the results without interrupting the data flow.
- I want to **take automatic decisions** based on the processed data.

## What do i get?

- **Increased value** of the information extracted from your data by using the most performant semi-supervised methods for data feature extraction using services like Amazon Athena, Amazon Kinesis, or Amazon Redshift.
- **Guidelines** (decision trees) on how to proficiently choose the right analytic method for your data from a vast collection of ready-to-use techniques.
- **Knowledge and practical examples** on how to generate automatic data-driven decisions to early detect underlying issues in upstream processes/services and integrate them in solid pipeline using AWS Data Pipeline.
- **Templates to automatically set and manage** data life span based on storage policies.
- **Consulting and hands-on experience** on a pilot project using your own data.

## Solution Design and Overview:

The number of methods available for big data analysis is continually increasing as is their complexity. Here we offer a valuable starting point for deepening your knowledge on the topic, efficiently applying modern analytic techniques in real case scenarios, and integrating them in your everyday workflow.

### We offer the following solution:

- **An introductory course** on big data analytics with **step-to-step guides** on how to choose the best performant analytic strategy (2 days)
- **Technical presentation** of real user case applications of big data analytics using AWS, including limit cases and points of failure (1 day)
- **Development of a made-to-measure pilot project** based on a small snapshot of your data **and integration of the analytic pipeline** in your process/service workflow (to be defined with the customer)

## At what price and what is the value?

This package is structured such as you can benefit from it twice. First, we will guide you through the theoretical aspects related to the abundance of methods currently available, and second, we will show you how to address the needs of deepening the understanding of your data with our made-to-measure pilot solution.

### Pricing:

→ On Request

