

AHEAD
THROUGH
KNOWLEDGE

Welcome to the World of Big Data

PROFINIT

Are you looking for ways to process ever growing volumes of data? Ways to reduce the load on data systems? Ways to analyse data flows in near real-time mode? Ways to enable complex computationally intensive analytics over millions, perhaps billions of records? Then you probably need a solution based on Big Data technologies, and we are here to help you.

WHAT IS BIG DATA?

We can interpret the term “Big Data” as referring to such kind of data that cannot be effectively handled by traditional systems such as RDBMS. Big Data must be treated with special regards to the whole data life cycle including ingestion, storage, querying, analyses, reporting, retention policies etc. It involves distributed computing capabilities and distributed storage. The nature of Big Data allows fast processing as well as easy scalability of the system.

Big Data can help you to process and analyse application logs or financial transactions, analyse records of telephone conversations or data transfers, detect fraudulent activity online, store data from millions of sensors in real time and take advantage of endless other possibilities.

The most common reason for implementing a Big Data solution is the need to add a new

system for more flexible analytics in the existing BI infrastructure. The first step usually involves building a so-called data lake that can contain practically all the enterprise’s data. The next step is to apply advanced statistics or machine learning algorithms capable of using the vast computing capacity of the Big Data cluster, thus providing an entirely new view of the data collected.

Find more information on our smart analytics experience in our [Data Science sheet](#).

Apache Hadoop is one of the most popular Big Data platforms, and it is also the option we propose in most cases. This platform supports dozens of other tools that can run “inside” it, including Apache Spark, Apache HBase and Apache Hive, to name a few. To ensure flawless cooperation, the tools are usually supplied in tested distribution packages. At Profinit, we work primarily with the two most common distributions: Cloudera® and Hortonworks®.





Profinit is based in Prague, Czech Republic. During last 20 years we have acquired top-notch knowledge in broad areas of data processing. We have participated in projects in finance, telecommunications, IoT and many others. We are willing to cooperate onsite or in nearshore mode. Let us share our experience and expert knowledge with you!

PROFINIT BIG DATA SERVICES

- › Studies and consultations
- › Testing the proposed solution on our own physical Hadoop cluster.
- › HW and SW infrastructure design according to the customer's needs
- › Evaluation of the suitability of the virtualized or physical environment.
- › Cluster installation, configuration and administration; support for operations teams.
- › Custom development using the expert knowledge of our engineers and developers.
- › Data analysis teams with latest technological and domain knowledge
- › Support with building internal Big Data competence

MODEL SITUATIONS

- › We need to process data in volumes from tens of terabytes to petabytes.
- › Our analytical tasks are exerting an excessive load on the current BI system.
- › Our current infrastructure does not have a sufficient performance to run our tasks.
- › We are looking for a data analytics system for our innovations centre / data science team, or we want to use advanced methods of machine learning.
- › We need to store and analyse vast volumes of data in real time.
- › We need to work with unstructured data such as audio or video records, texts or special formats from sensors.
- › We need to archive data or replace tape archiving.



PROFINIT EU

Tychonova 2 / 160 00 Prague / Czech Republic