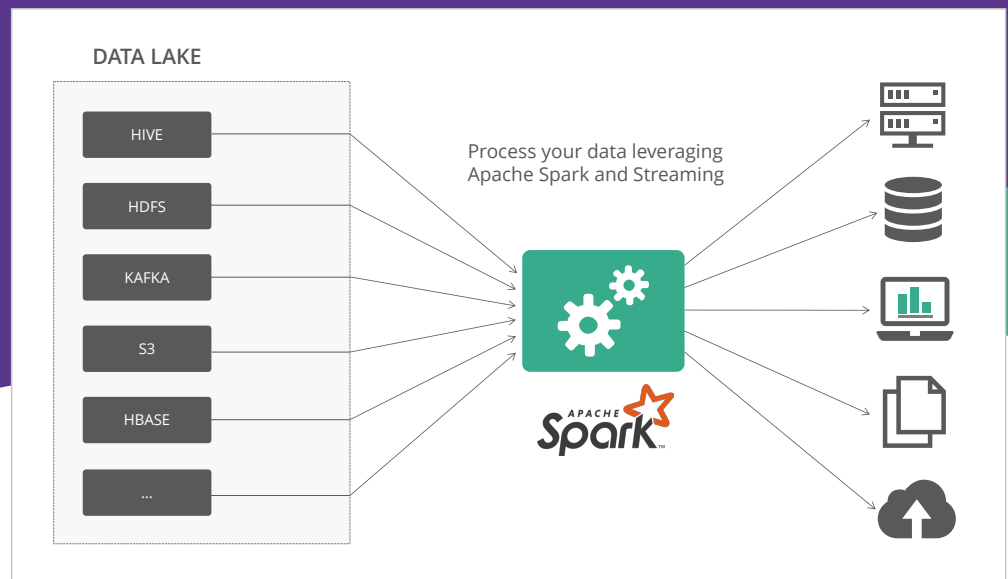


Big Data Processing & Data Integration

Rich data integration and transformation capabilities for data engineers, data scientists, DevOps, and business users.



Ataccama ONE Big Data Processing & Integration is architected for high performance and scalability, offering rich, built-in features that respond to a range of data transformation needs. Use Ataccama ONE to quickly understand and explore data in your **data lake**, rapidly process enormous volumes of data, and conduct a detailed data quality analysis before executing necessary transformations.

Ataccama ONE Big Data Processing covers the entire **data integration, ingestion, transformation, preparation, and management process**, including data extraction, import to a data lake and Hadoop, cleansing, and general processing. This ensures data is ready for further analytics at the right place and time, and in the right format.

All configuration happens through an **intuitive interface**, allowing business users to process and shape data without the need for Hadoop-specific knowledge such as MapReduce or Spark.



Gartner Peer Insights

"Our complex big data implementation found that the tool had the configurations to scale up."

– Data Quality Manager in the communications industry

Read more client reviews

gartner.com/reviews

Use Cases

- General data processing & integration
- IoT & streaming data integration
- Data Catalog & Business Glossary for data lakes
- Transformations, aggregations, data enrichment, and more
- Matching on Hadoop
- Quality control in transactional and analytical applications through a business-friendly interface
- Cleansing & unification in system migrations
- Quality assurance in software integration projects
- Data quality improvement in address and contact information
- Data cleansing & unification for client identification purposes
- Profiling, validation & correction of incomplete records as part of data integration project analysis
- Detecting inconsistencies & irregular patterns for fraud prevention and more
- Data preparation for further analytical use

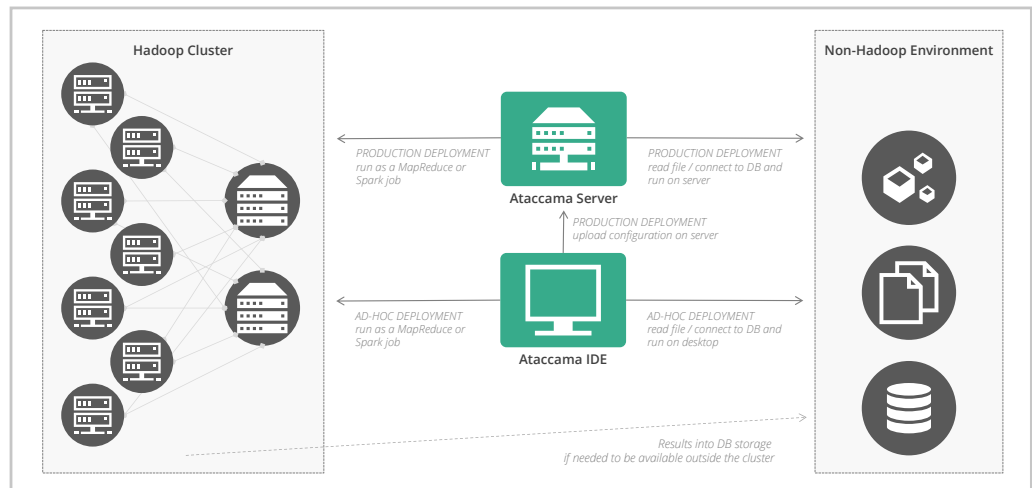
Features

- **Seamless migration** between local and big data environments, as existing configurations can be run on any environment without any changes or need for recompilation.

- **Support for elastic computing/big data**

processing on demand: Profile, process, and cleanse your data on automatically provisioned clusters with support for Azure HDInsight, Amazon EMR, Google Dataproc, Databricks, Cloudera, Hortonworks, and MapR clusters. MapReduce, Spark, and Spark 2 engines are utilized.

- **Data lake ready:** Integrate, transform, and enrich your data with external sources, with support for HDFS, Azure Data Lake Storage, Amazon S3, and other S3 compatible object storages. AWS Glue Data Catalog, Hive, HBase, Kafka, Avro, Parquet, ORC, TXT, CSV, and Excel are also supported.
- **Support for IoT and Spark Streaming,** including streaming integration with Apache Kafka, Apache NiFi, and Amazon Kinesis.
- **Unique profiling:** Enjoy rapid data analysis and advanced semantic profiling functionality.
- **Advanced core functionality:** A set of algorithms (capable of hierarchical unification by identifier keys, irrespective of internal data structures) can perform approximate matching during record unification.
- **Rich data integration and data preparation capabilities** for data engineers and data scientists. Profile, assess, transform, and join your datasets in a data lake and in the cloud.



- **Hadoop MapReduce and Apache Spark native support:** All calculations and processing are executed **directly on a cluster**, with no need to remove any data from Hadoop. Based on cluster characteristics, the solution is automatically translated into a series of MapReduce jobs, or directly utilizes Spark. Ataccama ONE supports all major Hadoop distributions.

Get started

Ask for a demo

See what we can do for your big data project. Get in touch to see a full demo of Ataccama ONE.

info@ataccama.com

Try one-click Data Profiling

Free account. Instant insights.

one.ataccama.com

Look for us in the latest Magic Quadrants

- Gartner Magic Quadrant for Data Quality Tools
- Gartner Magic Quadrant for MDM Solutions

Ataccama ONE: AI-Powered Data Curation for Digital Transformation



Data Discovery
& Profiling



Data Catalog
& Business Glossary



Data Quality
Management



Master & Reference Data
Management



Big Data Processing
& Data Integration

Discover more platform modules at ataccama.com