

Minimize data preparation bottlenecks and speed time-to-insight with Paxata on Microsoft Azure

SOLUTION BRIEF

Businesses are migrating growing volumes of data to big data clusters to support digital transformation and analytical initiatives. Data lakes are full of data collected without a clear sense of how it will be used.

Before data can be analyzed for insight, however, it must be clean, complete, and contextual. About 80 percent of analytics effort is spent bringing together data, and cleansing, shaping, and preparing data with specialized IT resources. As a result, getting data ready for analysis is tedious and expensive.

PAXATA SELF-SERVICE DATA PREPARATION SOLUTION ON MICROSOFT AZURE

With the Paxata Adaptive Information Platform running on Microsoft Azure, businesses can push control of data preparation to business users. They can work with data from any source, including unstructured data from Azure Data Lake and Azure Blob Storage, and publish it anywhere, including Microsoft Power BI, in days, versus weeks.

The Paxata platform runs natively on your Hadoop stack on Azure. Use it with HDInsight—or on Cloudera on Azure. Paxata offers powerful algorithmic intelligence, an intuitive user experience, and a smart distributed architecture that supports collaboration.



At Paxata, we turn raw data into trustworthy information at the speed of thought. Our Adaptive Information Platform enables business analysts with an enterprise-grade, self-service data preparation system to support the on-demand and ad-hoc business data needs for analytics, operations and regulatory requirements. Business analysts work within an intuitive, visual application to access, explore, shape, collaborate and publish data with clicks, not code, with complete governance and security.



Microsoft enables digital transformation for the era of an intelligent cloud and an intelligent edge. Microsoft Azure is an open, flexible, enterprise-grade cloud computing platform. Wherever your data lives, Azure services will help you unlock its potential. Services include HDInsight, a fully managed cloud Apache™ Hadoop and Apache™ Spark™ service backed by a 99.9% SLA.

FASTER INSIGHTS FOR NON-TECHNICAL USERS

Minimize analytics bottlenecks and gain insights from data faster. With the Paxata Self-Service Data Preparation application, any business user can turn raw data into trustworthy information with clicks, not code.

- Excel-like user interface requires minimal training, presenting a visual and interactive experience.
- Smart suggestions help users detect joins, overlaps, and misspellings, so you can standardize and unify data very quickly regardless of volume, variety, and source.
- Algorithmic intelligence works behind the scenes to accelerate ad hoc, operational, and predictive analytics.

MAKING SENSE OF BIG DATA

If you're running Apache Spark™ on HDInsight or Cloudera on Azure, you are already processing massive amounts of data. And while it's relatively simple to get data into a Data Lake Store or other Blob storage, you probably know that it's difficult to get it out and make sense of it—until Paxata:

- Land data without concern for schema and use Paxata to make sense of it when you are ready. As business users pull data, the intelligent ingest capability detects file type, including whether it is a semi-structured, nested file or an electronic document, and automatically flattens the file into a tabular format.
- Support massive data volumes with connectivity to Azure Blob storage and Azure Data Lake Store.

GOVERNANCE WITH SELF-SERVICE

Paxata governance controls enable you to control access to the system and deliver traceability.

- All activity is auditable and all data transformations preserve data lineage, which helps you understand the process the analyst used to make decisions about what to keep and change.
- Paxata also provides support for authentication, authorization, encryption, and usage tracking.
- Launch from a centralized and governed data catalog. Register answer sets with the data catalog.
- Re-use projects with new data sets.

“Azure customers can now leverage Paxata’s governed, self-service data preparation solution to leverage the intelligence, hybrid and enterprise-grade capabilities of Microsoft Azure – harnessing the full power of Azure’s Big Data offerings with a single click on HDInsight.”

Jason Zander,
Corporate Vice President, Microsoft Azure

ENTERPRISE GRADE

Built for scale: Powered by the Apache Spark™-based engine for in-memory high-performance, parallel, pipelined, distributed processing.

Built for the cloud: Elastic scalability to support variable workloads. Deploy from the Azure Marketplace in one click, without opening any ports, and run using native Azure services.

Secure and compliant: All requests are routed through the secure gateway on HDInsight and users also are authenticated with Azure Active Directory.

PAXATA HELPS YOU UNLEASH THE POWER OF DATA



PAXATA SUPPORTS YOUR DIGITAL TRANSFORMATION

Paxata helps you unlock the power of your data to engage customers, empower employees, optimize operations, and transform your products. As the data service layer on top of your data stores, Paxata is a critical piece in the process of modernizing your data estate.

- Support hybrid operations with ability to pull from on-premises and cloud-based data stores and take advantage of cloud-based analytics services.
- Accelerate your move to the cloud, by cleaning data as you migrate it.
- Gain flexibility of deployment. Run Paxata on your choice of Hadoop stack, either HDInsight or Cloudera on Azure.
- Write unstructured data into your data lake without concern for how it will be consumed—empower the business users who understand the data to explore it, assess data quality, and prepare it for transformative insights on their own.

WHY PAXATA

Faster time to value

Paxata helped a bank speed reporting by 90% and improve data quality by 52%. The 150-year old, multi-national bank, operating over 70 different markets, was able to accelerate reporting on financial crimes and compliance and reporting.

Paxata also helped a sovereign wealth fund detect transaction irregularities faster by reducing processing time for AML Surveillance from 21 days to 1.

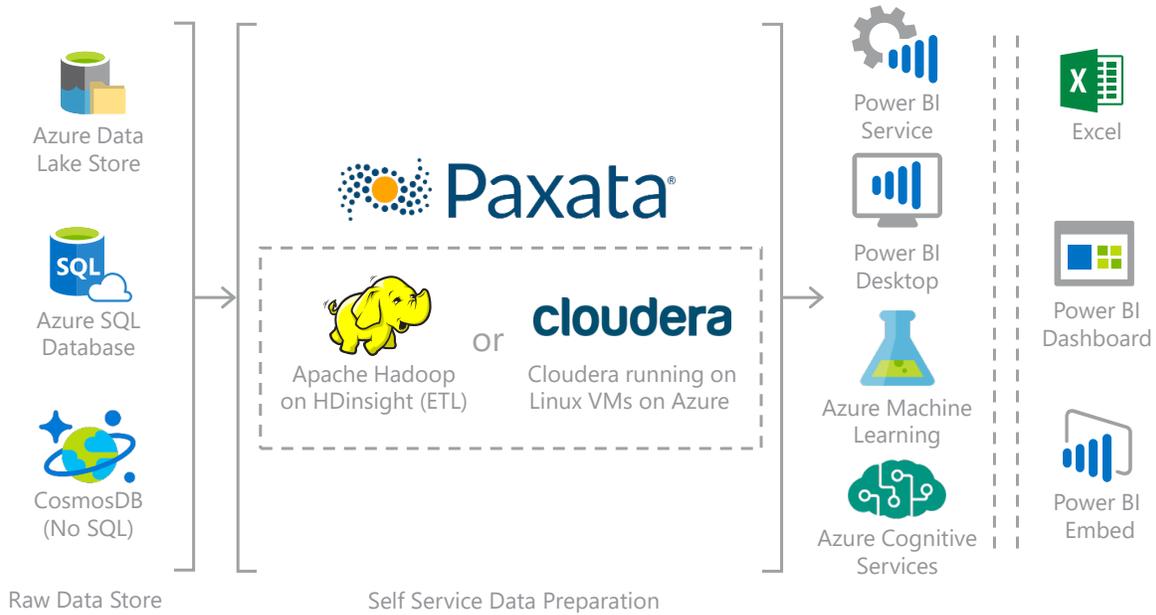
Cost efficiencies in data cleanup

A large telecom company migrated its data in 3 months, saving approximate \$13 million in staffing and consulting costs. Paxata's automatic validating and one-click profiling processes enabled the company to clean the data as they migrated it.

More value from data

A healthcare company used Paxata to reduce the process to uncover insurance fraud from 300 days to 90, with the help of some of Paxata's interactive features, including Explore, Clean & Shape, and Intelligent Ingest.

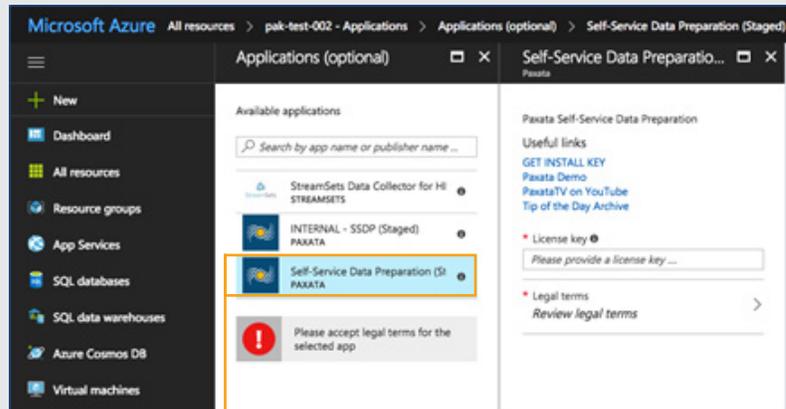
PAXATA IS CENTRAL TO YOUR MODERN DATA ESTATE



Paxata as part of an Azure solution architecture. Business users can pull data from any source in any format, and prepare data for analysis and consumption using the Hadoop stack of their choice.

GET STARTED TODAY

With Paxata, you're just a quick install away from accelerating your time to insight in a secure, scalable, data preparation environment running on Microsoft Azure.



Deploy with one click from Azure Marketplace

TRY PAXATA FREE FOR 30 DAYS ON AZURE

FIND IT ON THE AZURE MARKETPLACE OR VISIT [PAXATA.COM/ALLIANCES/MICROSOFT](https://paxata.com/alliances/microsoft)