



The Power of the Clinical Data Lakehouse

Author: Moulik Shah, Founder and CEO, MaxisIT

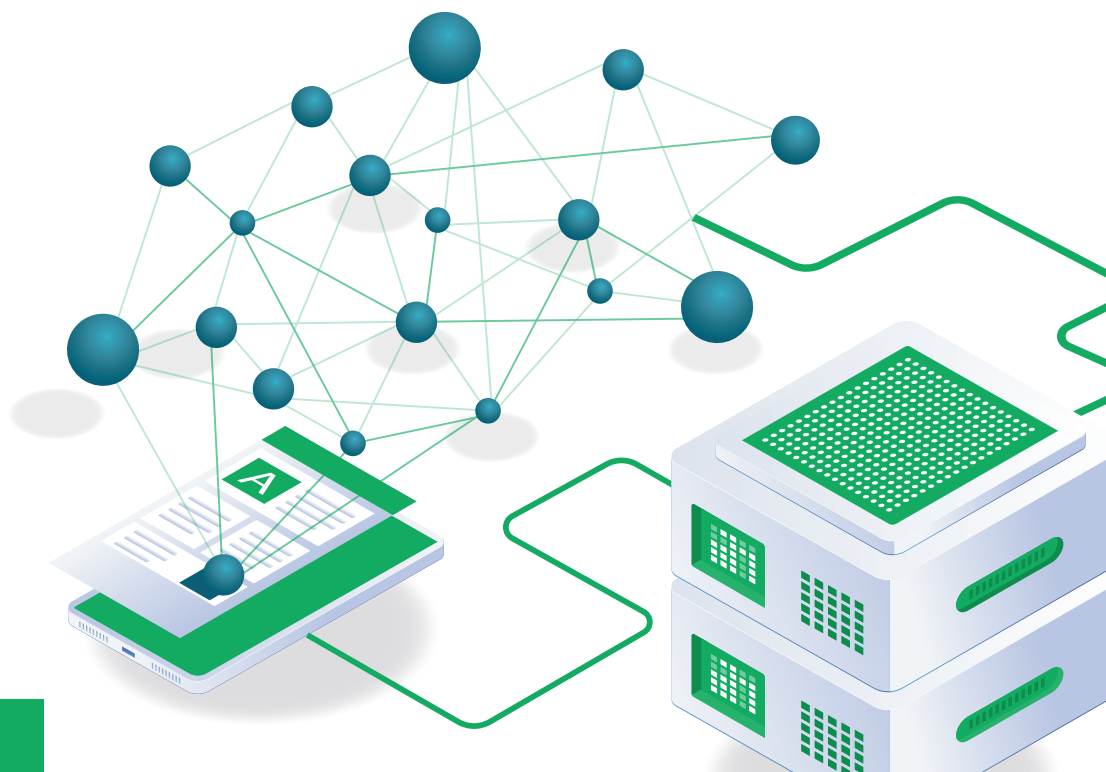
How can a data lakehouse optimize clinical data management?

The Power of the **Data Lakehouse** for **Clinical Trial Data Management**

From traditional clinical data to electronically captured patient-reported outcomes and wearables data, today's clinical trials capture more data from more sources than ever before. This variety gives clinical trial sponsors the tools to conduct more innovative trial designs, but it also requires more data management resources.

While data managers and other individuals tackle more challenging workloads, traditional clinical data management approaches and technology have failed to keep up with the increasing volume, variety, and velocity of clinical trial data. Systems capable of storing, recording, and managing higher quantities of data are required. The systems must also have the capability to analyze all that data because currently, data management teams can only use a fraction of the data that's stored.

A data lakehouse approach enables data management teams to not only store vast amounts of data but more easily manage and analyze that data. By realizing the full power of the data, clinical trial sponsors can gain comprehensive and interwoven insights that lead to more efficient clinical trials, as well as more effective programs, trial designs, and products.



Missed Data Opportunities



The average 12-month clinical trial with 2,000 data points generates up to **three million data** points. Given the number of times data are handled, that means many more opportunities for error. Reducing the opportunity for error via improved data management results in higher-quality data.¹

What's more, only a fraction of these data points are analyzed. In a 2022 webinar, Wendy Morahan, IQVIA Senior Director, Clinical Data Analytics, said that about 20% of clinical trial data are analyzed. The rest sit in a data warehouse, unused.² The reason so much data goes unused stems from inefficient clinical data management system as well as slow adoption of data analytics tools.

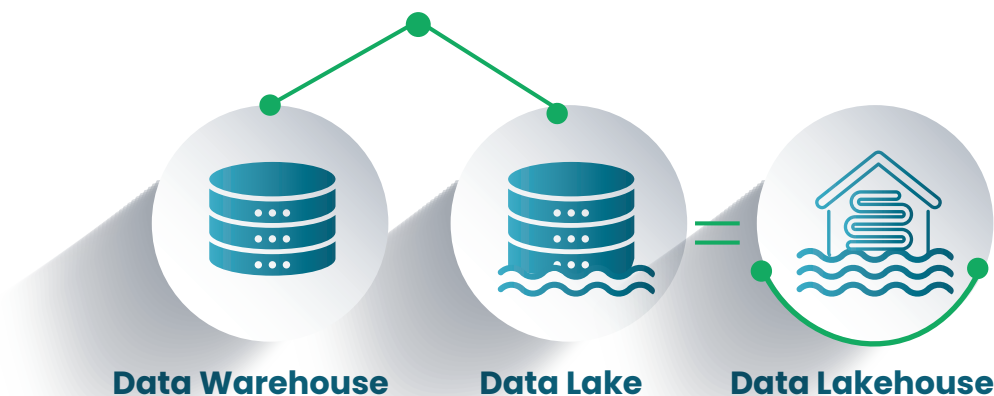
The data lakehouse approach makes it easier to integrate data sources. It also enables more seamless implementation of artificial intelligence-based and machine learning-based tools. With easier integration and availability of advanced technology, data managers and scientists can make better use of the millions of data points at their disposal.

1. <https://www.ncbi.nlm.nih.gov/books/NBK224576/>

2. https://globalmeet.webcasts.com/starthere.jsp?ei=1558373&tp_key=87268835lc

What is a **Data Lakehouse?**

IBM describes a data lakehouse as a data platform that combines the best of a data warehouse and a data lake into one data management solution.³



A data warehouse is a relational database designed for analytics. It's good for crunching numbers and analyzing trends, but it has limits. A data warehouse is also more expensive than a data lake.

A data lake stores all types of data, but those data are not formatted for analytical use. That's an extra step that takes a lot of time from experienced data scientists.

A data lakehouse combines elements of both the warehouse and the lake. It leverages the cloud to store structured, unstructured, and semi-structured data. Data are stored in a data lake in their raw form until it's time for analysis. And unlike data housed in a warehouse, it doesn't require formatting—a big time-saver.

As a cloud-based architecture, a data lakehouse enables seamless data integration, storage, transformation, and analysis. It also scales with the needs of the company for cost and capacity savings.

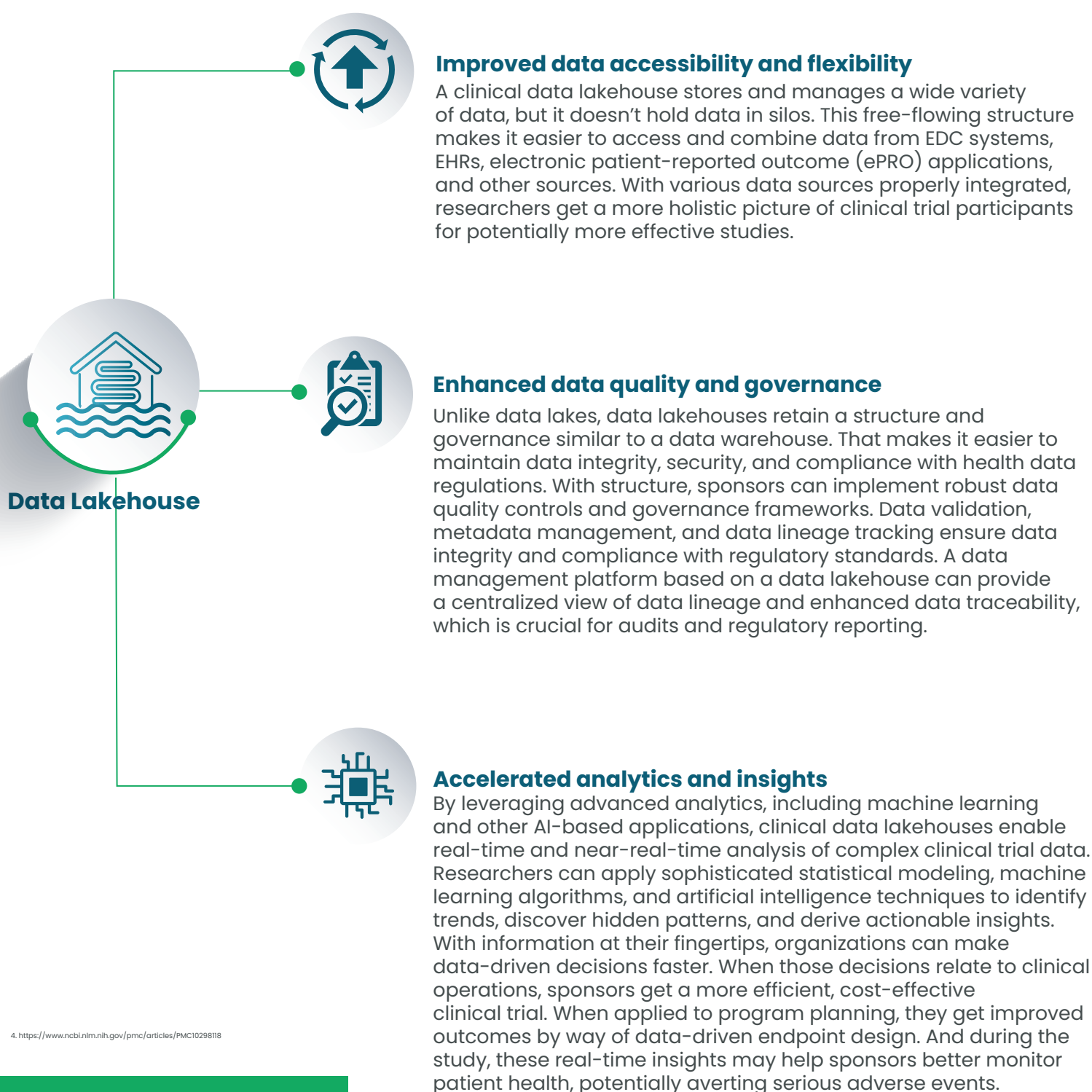
3. <https://www.ibm.com/topics/data-lakehouse>

Benefits of a Clinical Data Lakehouse

Data lakehouses are good options for clinical data management teams because of their flexibility, scalability, and relatively low cost. But unlike a vast data lake, which works best as a repository for new data, data lakehouses have structure and governance suitable for clinical research.

A study published in MDPI Healthcare discussed a data lakehouse-based framework to manage health-related data. The framework streamed data from various sources, fused the data semantically, applied algorithms to resolve duplicates, and handled the history of events. The framework sped up the clinical trial analysis process, produced more robust trial analysis, and provided a more holistic view of the question being studied.⁴

In addition to potentially richer, faster analysis, key lakehouse benefits for data managers include:



4. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10298118>

● Scalability and cost-efficiency

Unlike data warehouses, which become less efficient as they grow, clinical data lakehouses can handle vast amounts of data with optimal efficiency. Because compute and storage are not coupled together, data teams can access data storage while using different computing nodes for different applications. A cloud-based infrastructure allows lakehouses to expand and contract based on demand, which helps keep costs down. Further, sponsors can leverage a pay-as-you-go model. That helps reduce upfront infrastructure costs because you only pay for the resources you consume.



Stored data can get misconfigured, raising the risk of a security breach. Look for a clinical data lakehouse that incorporates robust security measures to protect sensitive patient information. Security and privacy features to look for include role-based access controls, data encryption, and data anonymization techniques for GDPR and HIPAA compliance. Sponsors may also want to implement technologies such as differential privacy, which protects individual privacy while enabling meaningful analysis.

Data privacy and security ●

How to Keep Your Data Lakehouse **from Falling into the Swamp**

Much like any large home, without proper maintenance, a data lakehouse can fall apart. Without organization, proper governance, and maintenance of data quality, your clinical data lakehouse can deteriorate into a clinical “data swamp.”

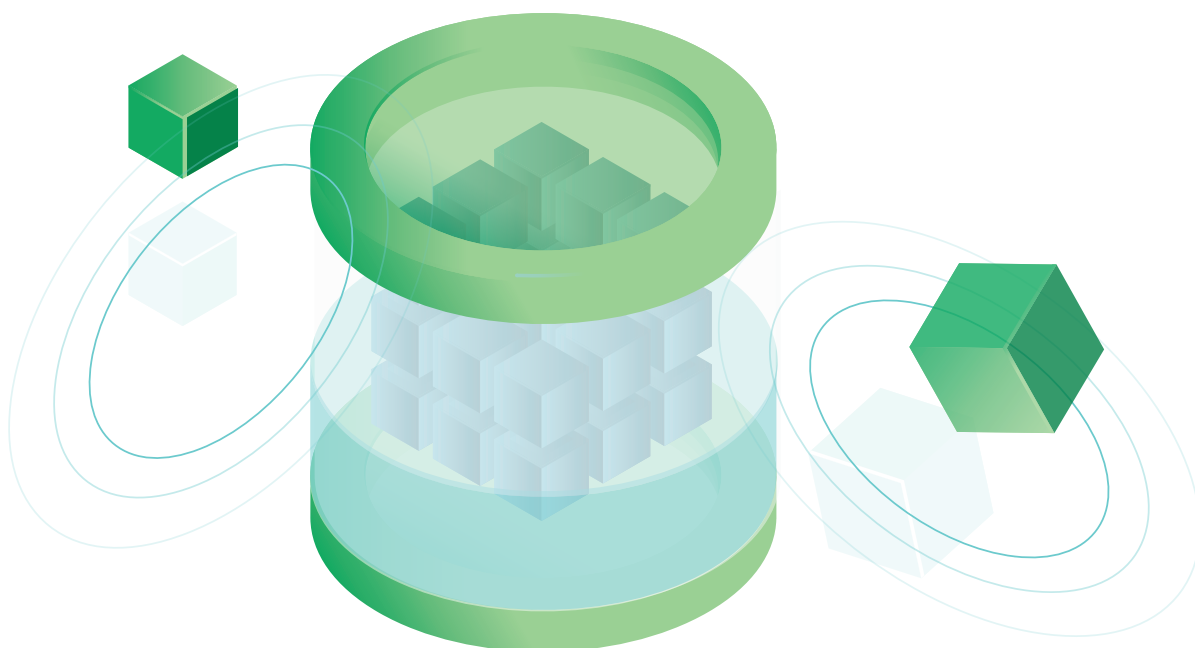
In a swamp, data become difficult to use. To keep your data lakehouse intact, establish and follow data governance policies and procedures. A few starting points include the following:

- **Metadata management**

Systematically capture metadata to provide context and make data understandable and easier to use.

- **Data cataloging**

Use a data catalog to keep track of what’s inside the data lakehouse and make it searchable for users.



- **Data quality checks**

Regularly check and validate data to ensure accuracy and consistency.

- **Data access management**

Control and monitor who has access to the data.

Not only does this help keep data managed, but it also keeps it secure and compliant.

- **Data lifecycle management**

Regularly review and purge outdated or unnecessary data in accordance with company policy.

Conclusion

A clinical data lakehouse presents a tremendous opportunity to streamline clinical data management. Its scalability, flexibility, and relative affordability can help data management teams overcome data integration challenges while also allowing them to apply advanced AI-based techniques to these data.

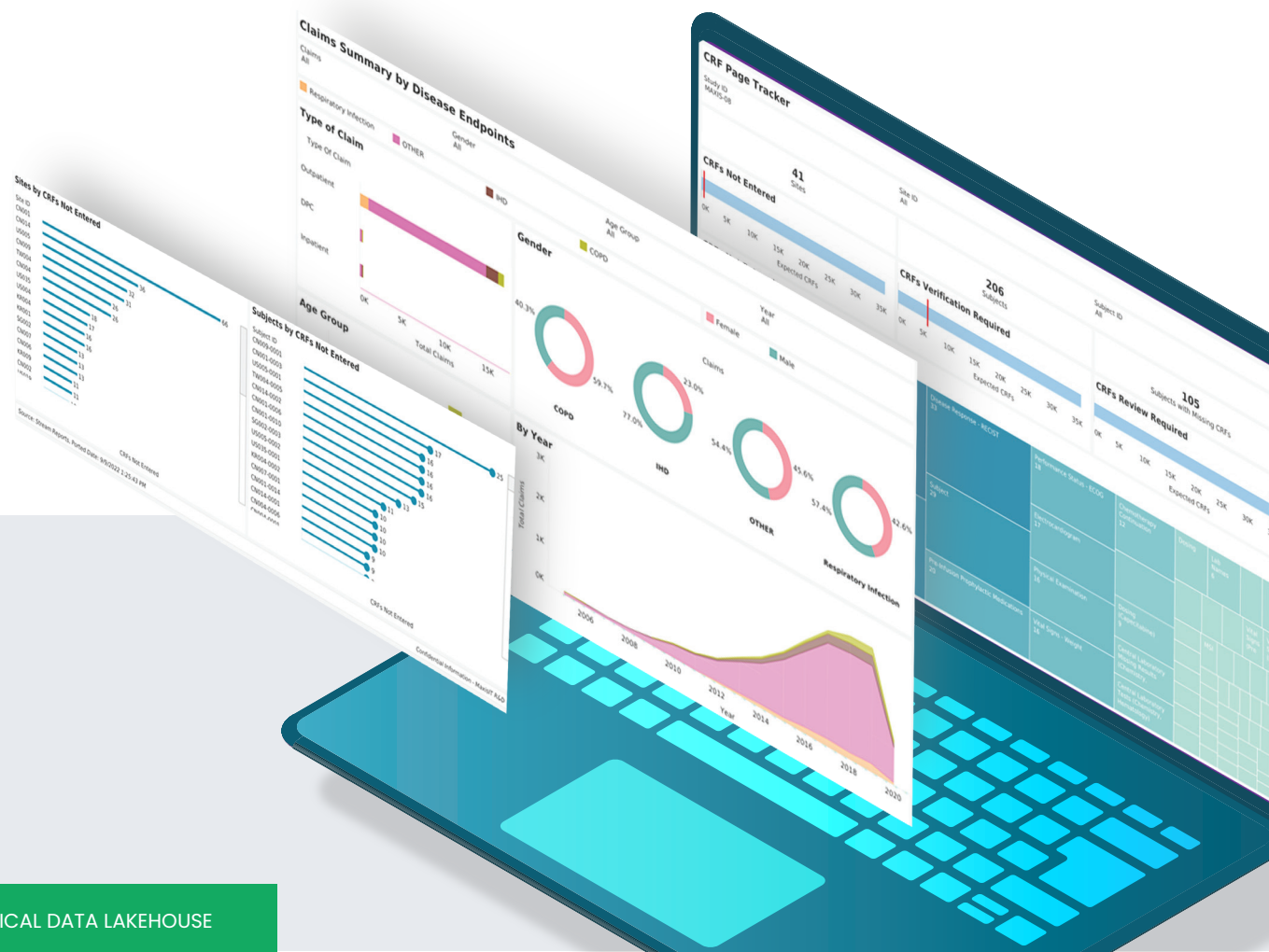
For sponsors and CROs that want to make better use of the billions of data points collected in their clinical trials, the data lakehouse allows them to do more with more.

MaxisIT's platform modernizes clinical data strategy by employing clinical data lakehouse architecture that holds immense potential for transforming clinical data management in the pharmaceutical and life sciences industry.

By leveraging this innovative approach, MaxisIT catalyzes the digital transformation of clinical trials and empowers clinical trial sponsors and CROs to overcome the challenges associated with managing and analyzing vast amounts of clinical trial data.

By leveraging **MaxisIT's Clinical Data Analytics Platform**, which integrates data lakes and data warehouses within one unified architecture, clinical trial teams get improved data accessibility, enhanced data quality and governance, accelerated analytics, scalability, cost-efficiency, and robust data privacy and security.

By embracing, MaxisIT's Clinical Data Analytics Platform powered by clinical data lakehouse architecture, sponsors and CRO organizations are now able to unlock the full value of their clinical trial data, thus driving innovation and ultimately advancing the discovery and development of life-saving treatments and therapies with improved outcomes.



About MaxisIT

MaxisIT's purpose-fit and intelligent clinical data analytics platform helps improve clinical trial performance, mitigate risk, and optimize clinical outcomes. We provide a centralized and reliable source of truth for diverse data types from various sources, giving life sciences companies real-time insight to shorten cycle time and increase return on investment.

Incorporating an end-to-end clinical data pipeline from intake to visualization, MaxisIT's solutions are powered by AI/ML and metadata-centric approaches. Our impressive portfolio of over 3,300 clinical trials and an unparalleled 100% customer retention rate affirm the quality and reliability of our services.



Moulik Shah
Founder & CEO, MaxisIT

Moulik Shah is a passionate healthcare technology entrepreneur and the visionary CEO of MaxisIT, where he has been at the forefront of leveraging technology to transform pharmaceutical and life sciences clinical trials.

His dedication to improving patient outcomes and his thought leadership in patient-centricity, patient diversity, interoperability, and real-world-data-led collaboration have been at the core of his vision of an integrated healthcare ecosystem based on effective use of data and analytics platforms.


He has been instrumental in driving innovation and progress in the industry. Under Moulik's leadership, MaxisIT has become a leading provider of clinical data and analytics which is driving real-world impact in the pharmaceutical and life sciences clinical trials.




www.maxisit.com

connect@maxisit.com



 510 Thornall Street, Suite 180,
Edison, NJ 08837

 +1 732-494-2005, ext. 135