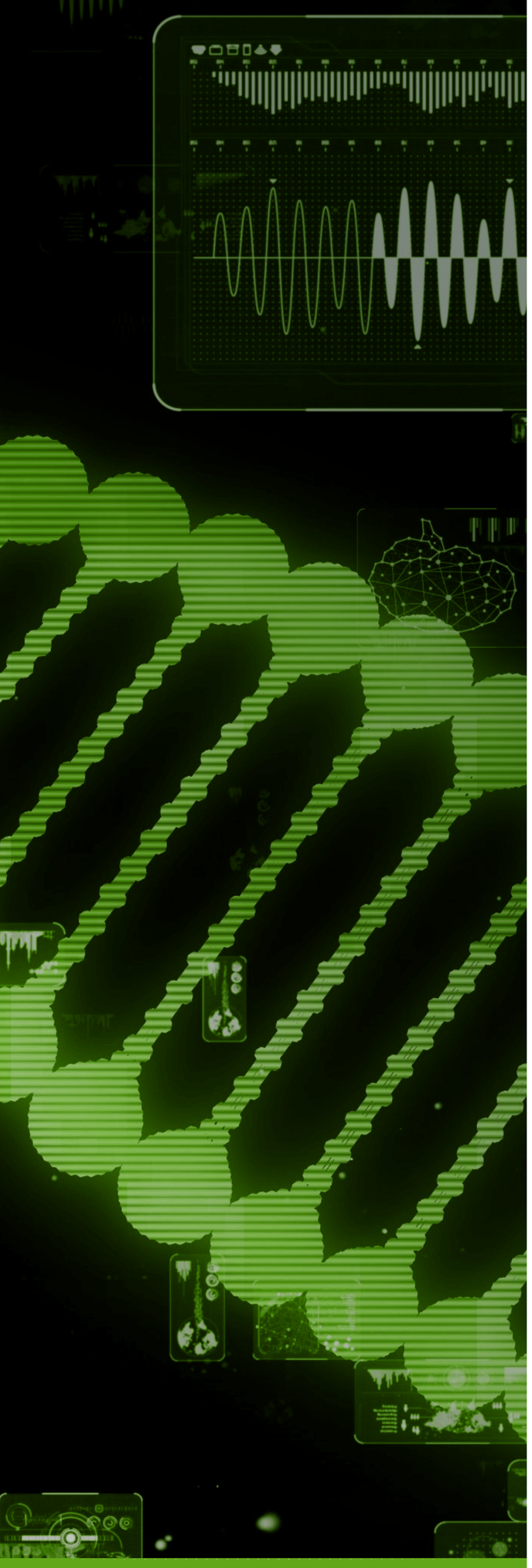
DATAXIGHT™

CLEARXIGHT

Clinical Data Transformation, Transformed.



WWW.DATAXIGHT.COM

REDUCE FRICTION, EXPAND INSIGHT

Unified Framework for Code Reusability

{ CHALLENGES }

LACK OF CODE REUSE

SQL scripts and notebooks with arbitrary Python or R code intertwine data parsing and data processing, requiring new versions for each version of a dataset.

{ SOLUTION }

REPRODUCIBLE EXECUTION

Declarative syntax and separation of the dataset specification from the processing specification allows for the same pipeline definition to be used with many datasets.

Efficient Scalability at Any Scale

{ CHALLENGES }

COST OF SCALABILITY

Code written to process files of a certain size typically has to be rewritten to process a dataset that is larger by an order of magnitude or more.

{ SOLUTION }

AUTOMATIC SCALABILITY

Declarative syntax allows decomposition of operations so that many steps can be parallelized and deployed at scale without code changes.

Reliable Testing for Better Accuracy

{ CHALLENGES }

VERIFIABILITY OF CORRECTNESS

Testing SQL queries or Python/R code for correctness is challenging, which makes it hard to ensure their correctness.

{ SOLUTION }

RIGOROUS TESTING

Declarative processing specification enables testing of individual processing steps separately, with separate unit tests for each operation.

Knowledge Retention Across Teams

{ CHALLENGES }

KNOWLEDGE LOSS

Changes in team membership can lead to the loss of tribal knowledge about the subtle semantics of SQL scripts or Python/R code.

{ SOLUTION }

KNOWLEDGE SHARING, ENHANCED COLLABORATION

ClearXight's decomposition of pipeline steps into operations ensures consistent documentation of functionality and potential for reuse across teams.

Platform Independence and Compatibility

{ CHALLENGES }

PLATFORM DEPENDENCIES

SQL, Python, and R code is often tied to specific storage and compute platforms. Porting the same transformations to another system requires significant code changes.

{ SOLUTION }

PLATFORM-AGNOSTIC FRAMEWORK

ClearXight abstracts the user from execution environment and the storage system. This makes data pipelines developed with ClearXight more portable.

Simplified Maintenance and Documentation

{ CHALLENGES }

COMPLEX INTERPRETATION, DOCUMENTATION AND MAINTENANCE

Arbitrary code tends to grow long, eventually making it hard to understand, document, and maintain.

{ SOLUTION }

EASY INTERPRETATION AND MAINTENANCE

ClearXight's specifications are intuitive and written in simple and clear syntax, easy for interpretation, documentation, and maintenance.

COMPARISON OF DEIDENTIFICATION WORKFLOW

An example of how ClearXight simplifies data engineering.

{ TRADITIONAL DEIDENTIFICATION WORKFLOW }

- Retrieve data from the data storage platform. { STEP 1 }
- Load the data. { STEP 2 }
- Iterate over each table in the dataset. { STEP 3 }
- Iterate over each PHI column in the data table. { STEP 4 }
- Write code to deidentify each column. { STEP 5 }
- Check the consistency of deidentified information across tables. { STEP 6 }
- Save the deidentified data to the workstation. { STEP 7 }
- Upload the deidentified data to the data storage platform. { STEP 8 }

STREAMLINED WORKFLOW

FROM **8 STEPS** TO **3 STEPS**

✓ FAST, AUTOMATED, COMPLIANT

{ CLEARXIGHT NO-CODE DEIDENTIFICATION WORKFLOW }

- { STEP 1 } Specify the dataset location and PHI columns in the dataset in a data graph.
- { STEP 2 } Run the data graph on the ClearXight HIPAA deidentification module.
- { STEP 3 } Share the deidentified dataset with your collaborator the same day.

REQUEST A TRIAL

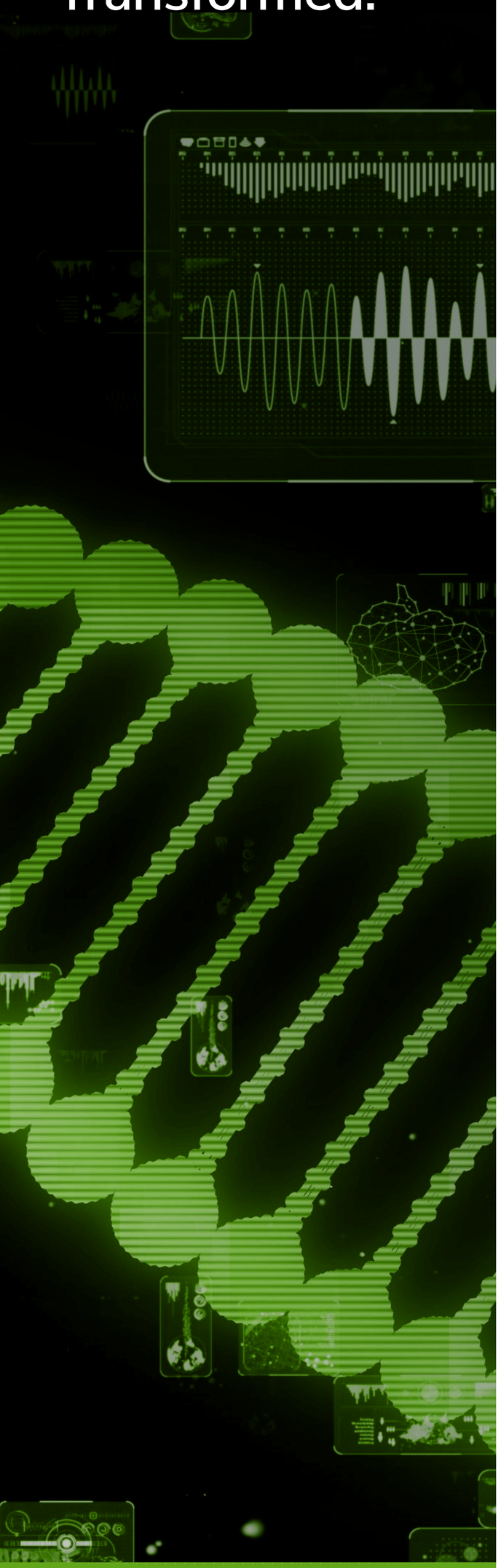
Contact us today to discuss your project at solutions@dataxight.com.

DATAXIGHT™

WWW.DATAXIGHT.COM

CLEARXIGHT

Clinical Data Transformation, Transformed.



ClearXight is a modern data engineering framework tailored for transforming real-world data, enabling data scientists and engineers to simplify, standardize, and automate their data transformation workflows while ensuring compliance with industry regulations. By significantly reducing the time spent on data engineering, ClearXight empowers faster initiation of data analysis, streamlines ML model development, and accelerates the delivery of actionable insights.

FOCUS ON CLINICAL DATA

ClearXight efficiently processes tabular data, enabling users to specify table properties and relationships between columns across various datasets. Real life use cases include:

{ HIPAA DEIDENTIFICATION }
Ensure compliance by removing sensitive patient data using hashing and consistent date shifting, as required by the HIPAA Privacy Rule.

{ SYNTHETIC DATA GENERATION }
Generate synthetic datasets that preserve the statistical distribution of clinical elements.

{ COHORT SELECTION }
Define complex study cohorts from clinical datasets, including those requiring longitudinal analyses.

{ POWERFUL MAGIC }
Use your own specialized data transformation function as a module in the ClearXight framework. So easy it's magical!

HOW CLEARXIGHT WORKS

ClearXight allows users to define data transformation steps via execution graphs and dataset relationships via data graphs. ClearXight applies the execution graph to the data graph in a scalable and repeatable manner, agnostic of cloud platform.

{ EXECUTION GRAPH }
Operations for reading, processing, and outputting data.

{ COMPUTING ENGINE }
Support for tools like Ray & Daft.

{ DATA GRAPH }
Properties of tables and relationships between columns.

CLEARXIGHT DATA ENGINEERING FRAMEWORK

DATA MANAGEMENT AND CENTRALIZATION

DATA ANALYSIS AND VISUALIZATION

AI/ML MODELING

- ✓ ACCELERATE INSIGHTS
- ✓ IMPROVE DATA QUALITY
- ✓ ENSURE COMPLIANCE
- ✓ ENABLE RAPID PROTOTYPING
- ✓ ACCELERATE INSIGHTS
- ✓ ENHANCE COLLABORATION
- ✓ REDUCE COSTS

OUTPUT

{ DATAXIGHT'S CUSTOMER } → { CLEARXIGHT } → { DEIDENTIFIED AND HARMONIZED DATA } → { BENEFIT }

ClearXight's declarative syntax for describing data and transformations helps reduce data engineering time and enables self-service data engineering for non-technical scientists to subject matter experts without software engineering expertise.

REQUEST A TRIAL

Contact us today to discuss your project at solutions@dataxight.com.