

Data Automation Evolution

Simon Spring, Operations Director Thomas Heuer, Sales Director

August 2022

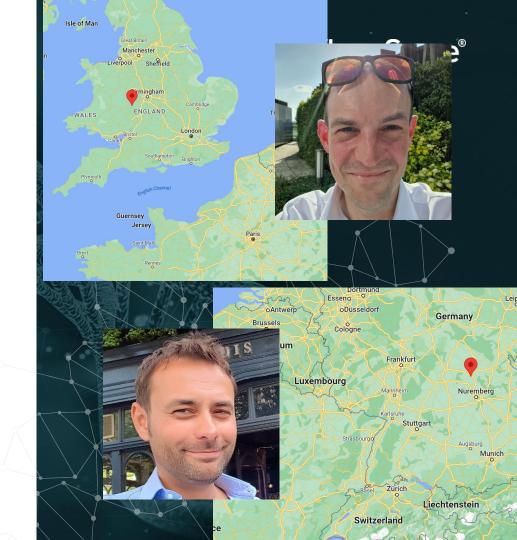
About Us

Simon Spring

- Software Development > BI > Data
 Warehousing > Data Warehouse Automation
- Based in Worcestershire, UK
 - Pronounced "Wooster sher"!

Thomas Heuer

- Rock Star > Software Sales
- Based in Bamberg, DE (1 year)



IT-Logix & WhereScape







- Introductions
- Embracing More & More Data
- Meeting the Growing Demand for
 Information
- WhereScape Roadmap
- Q&A

Automation meets the Demands



Embracing More & More Data

- Volumes, Sources & Source Formats
- Choosing the right platform & components and then evolving them to support those volumes, formats and workloads
 - File-based / DB platform / delta lake
 - ETL/ELT options
 - Scripting languages
 - Modeling-style

- Automation sits on top of your platforms and tools
- Model-driven automation means rapid onboarding of data sources into PSA/ODS, Data Lake, DV
- Switch easily & scalably:
 - Target Platforms
 - Target Modeling-style
 - Ingestion/execution tools
 - Sources (as they evolve)



Increasingly Complex Transformations

• Analysis spanning an increasing number of data sources

• Complex transformations / algorithms worked on by various teams / data scientists

- Discover sources and automatically generate target models and all ingestion/ETL code to populate them
- Extending model-level automation to other layers like information marts
- Data-driven automation means rapid prototyping great for business logic
- Embed SQL-based transformations or callable functions into data flows
- Trigger external algorithms from within DW jobs: add/update columns during staging flows
- Write your own templates to implement other languages such as PySpark, Scala functions...



More Delivery Formats and Squeezed Timescales

- Different consumers want their data delivered in multiple formats (Files, Tables, APIs) and consumers can range from people to downstream systems, AI/ML models, user apps...
- Fast and Standardised
 - Information is often needed this week not next month!

- Cost efficient
 - Affordability needs to scale

- Choose a mixture of modeling approaches (star schema/wide tables/Data Vault/normalized/nested tables)
- Hub / Spoke
 - Model automation for core layers
 - Data driven for information marts
- Iterative working
- Design once | deploy many
- One customer had 4 squads of 10 people delivering and needed to increase throughput x3 so was going to take on another 8 x 10 people!



Tested, Governed and Secured

• Tested

- Governed
 - well defined & documented

- Secured
 - Correct location / jurisdiction
 - Access controls / data masking

- Model design test automation
- Integration with test frameworks like BiGEVAL
- Jira-driven

- Data Catalog Import/Export
- Integration with Governance tools like Collibra

- Use metadata and configurable metadata tags / column-level UI fields to drive choice of
 - storage location / schema
 - masking algorithm/security context



WhereScape Roadmap





Roadmap

Themes

- Platform / Cloud
- DevOps / CI-CD
- Usability

Products

- WhereScape RED
- WhereScape 3D
- Enablement Packs



Now

Next

Later

Cloud / Platform

- REST API source
- XML/JSON/ORC/Parquet platform specific loaders
- Delta Lake / Databricks target
- Azkaban Scheduler (Linux/Windows with web UI)
- PostgreSQL metadata repo
- Object-level "Action Scripts" the next level of templating
- Snowflake Tasks & Streams
- Azure Data Factory

Usability / CICD

- RED UI Refresh: Menus
- New Installer (Enablement Pack aware)
- Advanced Authentication (oAuth)
- Data Vault sample template library
- Limits lifted (object names, DDL, transforms, jobs)

Cloud / Platform

- Streaming source connectors
- External Scheduler integration e.g. Airflow
- Azkaban containerisation
- Object-level "Action Scripts", user-extendable, enabling:
 - Git push/pull
 - Data Catalog sync
 - Procedure compilation (Snowflake, Synapse, BigQuery..)
- Qubole target

Usability / CICD

- RED UI Refresh: Workflows
- Role-based controls for features
- PostgreSQL repo migration tool
- Deployment merge options

Cloud / Platform

- External Scheduler API (for WS metadata read/write)
- Rest API access to WhereScape CLI
- Data Catalog integration with 3D
- AWS/GCP marketplace

Usability / CICD

- RED UI Refresh: Diagrams, Generated Docs
- User defined object types/subtypes
- Role-based controls for objects
- Git integration at deployment package level

Ongoing

- Enablement Pack evolution
- 3D>RED integration
- Decoupling of UI from services (for future new UI)
- Ecosystem partner integrations
- Model Conversion Rules = new levels of automation
- Performance

Follow WhereScape Automation

Sign up for release notifications

www.wherescape.com/register

Read about recent releases

www.wherescape.com/blog/

Browse our Enablement Packs

<u>https://www1.wherescape.com/support/software-downloads-documentation/enablement-packs/</u>

Check out our upcoming releases

- File Parser (XML/JSON/Parquet/Orc) Load templates for SQL Server/SQL Azure DB/Synapse September
- REST API source September
- 3D 8.6.8 October
- RED December

Join our beta programme

simon.spring@wherescape.com

