Can we make complex easy?
A sneak preview at the next generation of Data Transformation tools

Thorsten Reitz, wetransform GmbH
Agenda

Cloud Transformation

Collaborative Development

New paradigms for Transformation Definition
The Basics: Declarative Mapping
The Basics: Declarative Mapping

- Easy to author & re-use
- Fast & Scalable Execution
- Any data structure & format
Current state

hale studio 3.3.1

hale connect 1.3.0
Collaborative Development

„Github for Transformation Projects“
Alignment + Cell Discussions

- **Goal:** Establish clear context for discussions (questions, suggestions)
- Can be at any level of the project (Alignment, Retype, property)
- Notifications, ...
- Available (hale connect 1.2.0+)
Versioning

- **Goal:** Track changes and variants (branches), be able to revert to earlier versions
- Understand changes to each alignment cell
- Uses a Git Backend that can also be accessed via Git or GitHub
- Available (hale connect 1.2.0+)
Task + Solution Workflows

- **Goal:** Provide traceability of tasks and assignments
- Multiple solutions can be proposed
- Solutions can be reviewed and accepted (merged)
- **Upcoming** (hale connect 1.4.0)
Cloud Transformation

„Give me all the data“
Why Cloud Transformation?

- **Goals:**
  - Transform large data sets fast
  - Transform multiple data sets in parallel
  - Perform Transformation online for network datasets

- **Available in hale connect 1.3.0**
  - Public Cloud deployment

- **Interesting also as a separate Product?**
Transformation with Amazon EC2 Container Service (ECS)

- ECS: Manage tasks and services based on Docker containers on a cluster of Amazon EC2 instances

- Task definition: Docker image, environment variables, required resources, port mappings, volumes, ...

- Transformation cluster: (For now) up to 4 c4.large instances (2 vCPUs, 8 ECUs, 3,75GB RAM)
  - Optimized for data processing (fixed ECUs)
  - Max. 2 transformation tasks per instance
Cluster scaling with Amazon CloudWatch and AWS Lambda

- **CloudWatch**: Monitoring for AWS cloud resources w/ support for alarms, events and logs

- **Lambda**: Execute functions based on events, no management of cloud resources necessary

Transformation cluster:

- CloudWatch alarms: ECS cluster memory reservation above 75% or below 25%

- Lambda functions that add EC2 instance to cluster or remove one from it and get triggered by the previously defined alarms
Cluster scaling with Amazon CloudWatch and AWS Lambda

- **Protect instances** with running transformations from scale-in

- **On-Demand Mode:**
  - Daytime: One instance always ready for transformations
  - Night time/weekend: Cluster scaled down to zero instances
  - New EC2 instance started on transformation request
  
  ➔ Longer wait time for transformations (~2min)
  
  ➔ Savings
Deploy cloud resources with AWS CloudFormation

- Define cloud resources and permissions in **YAML deployment descriptor**
  - Parameters
  - YAML extensions e.g. variable replacement, if/else

- External resources e.g. Lambda functions are uploaded to S3 buckets

- Determines **change sets** on updates and only redeployed updated resources

- Executed from jenkins on repository updates
The result
User Interface

„One baby step at a time“
Goal: Reduce repetitive cells (in non-object oriented models)

- Don’t work on individual elements of a schema
- Work on all elements that match criteria
  - (Qualified) Name
  - Type/Binding
  - Value

- Coming in hale studio 3.4.0
Research: Structured Model Mapping

- **Goal:** Focus on one alignment cell and its immediate context at a time
- Task-driven process
- Touch first (hale on a phone 😊)

- **Coming 2018**
Research: Linked Maps

- **Goal**: Provide an user experience that matches linked spatial data structures and provides added value to end users
- **Coming in hale connect 2.0.0**
Research: Direct Instance Manipulation Mapping

- **Goal:** Define the transformation using objects, not abstractions
- Schemas uses „behind the scenes“
- One main view – the map
- Task-driven process
- Touch first

- **Coming ...**
Questions? Feedback?
+49 6151 155 408

info@wetransform.to
www.wetransform.to

www.linkedin.com/company/wetransform-gmbh
https://twitter.com/tr_xsdi