ubiquis
BUSINESS INTELLIGENCE

Tales from the road

Modular ETL and testing framework

Nelson Sousa / Luis Silva
nsousa@ubiquis.co.uk / lsilva@ubiquis.co.uk

ubiquis
BUSINESS INTELLIGENCE

The Project

- Losch is a network of car dealerships based in Luxembourg.
- They have a SQL Server application to handle inventory, sales, orders, catalog, etc.
- They need analytics

ubiquis
BUSINESS INTELLIGENCE

The Problem

- ~30 identical or near identical DB
- No primary keys
- No FK/PK relationships
- No data quality enforcement
- No update timestamps or other CDC stuff

ubiquis
BUSINESS INTELLIGENCE

The Solution

- ❑ Kettle, of course.
- ❑ Modular framework, deal with 1 table at a time
- ❑ Stage data with some intelligence
- ❑ Set up for multi-environment execution
- ❑ Track it all with Git

Testing

- ❏ Reusable
- ❏ Extensible
- ❏ Usable for test and CI
- ❏ Able to test orchestration
- ❏ Able to run with different datasets on different environments.

ubiquis
BUSINESS INTELLIGENCE

Testing

- ❑ Git tracks code
- ❑ Feature branches for development
- ❑ Jenkins polls all branches
- ❑ Tests committed with code
- ❑ Bash scripts run tests, merge to master and deploy

ubiquis
BUSINESS INTELLIGENCE

Current scenario

- ❑ 50+ atomic tasks to E, T, and L data
- ❑ 4 fact tables, 20+ dimensions
- ❑ 3 remotes (dev, test, prod)
- ❑ 200+ tests (at least 2 per task)
- ❑ Code tested and deployed daily

ubiquis
BUSINESS INTELLIGENCE

To do

- ❑ Finish building fact tables (2 or 3 to go)
- ❑ Finish building dimensions (10-20 to go)
- ❑ Bug fixing: newly discovered data quality issues
- ❑ Convert tests to PDI jobs
- ❑ Handover to support team

ubiquis
BUSINESS INTELLIGENCE

Questions?

ubiquis
BUSINESS INTELLIGENCE

Thank you!