A concept for the visual and interactive impact analysis and simulation of data changes to enterprise metrics

Final Presentation

29.07.2014

Master Thesis Matti Maier
Index

• Motivation
• Research Questions
• Literature Review
• Market Overview
• Solution Approach
• Architecture
• Demo
• Byproducts
• Outlook
• References
Motivation

- Understanding the system
- Solving what-if questions
- Visualization of change impact
- based on KPIs and their Visualization

Industry-driven Problem
Research Questions

1) What is an appropriate concept to visualize the impact of data changes to enterprise metrics?

→ appropriate: theoretically profound, technically viable

2) How can interaction on this visualization help to improve the user’s understanding of the calculation system?
Literature Review: KPI Visualization

• Common visualization technique: **Dashboard** [1]
  – Each KPI, one graphical component
  – Supportive diagrams (charts, graphs, tables)

• Generic adequate visualization technique: **2-dimensional graph-based layouts** [2]

• Equations well represented by *node+link graphs* [3]

\[
\begin{align*}
a &= b + c \\
c &= d + e
\end{align*}
\]
Literature Review: Impact Visualization

- **Impact of influencing factors on processes** evaluated by Hao et al. [4]
  - Hyperbolic graph layout
  - Different colors and line weights show impact of each factor

- **Display delta** between an original value and a new value with size of node [3]

- Enable the user to **interact with the model** and **compare scenarios** [5]
Market Overview

**Simulation tools**
e.g. Simio, SimuLink/MathWorks, Simul8, OptQuest, OpenSim, OpenModelica

- Targeted at engineers and scientists
- Usability for user group,
  e.g. console view only, complex languages
- Not suitable for the problem

**Microsoft Excel**

- Insufficient visualization of impact
- Understanding the system difficult
  if formulas are hidden

**Powersim Studio 10**

- No comparison between original and deviated values
- No support for “Virtual Data”
Solution Approach: Basics

(1) \[ d \ast \Delta + c = s \quad \text{with} \quad \Delta = 0 \quad \text{for} \quad c \neq 0 \]

d: original value
\Delta: simulated delta
c: value replacement / constant
s: simulated value

(2) \[ D \cup V = S \]

D: original data
V: virtual data
S: simulated data
Solution Approach: Visualization

KPIs based on **Formula**
Visualization of Formula based on Spence [3]

Different **weights and colors** [4]

**Interaction** with scenario [5]
Modular Architecture
Demo

<table>
<thead>
<tr>
<th>Renter</th>
<th>Building</th>
</tr>
</thead>
<tbody>
<tr>
<td>- id: String</td>
<td>- id: String</td>
</tr>
<tr>
<td>- name: String</td>
<td>- name: String</td>
</tr>
<tr>
<td>- age: int</td>
<td>- street: String</td>
</tr>
<tr>
<td>- rent: float</td>
<td>- city: String</td>
</tr>
<tr>
<td>- utilities: float</td>
<td></td>
</tr>
<tr>
<td>- extra: float</td>
<td></td>
</tr>
</tbody>
</table>

1..* building 1
Research Questions Answered

1) What is an appropriate concept to visualize the impact of data changes to enterprise metrics?
   → Visualization of KPIs with 2-dimensional graphs
   → Visualization of data changes with size and color

2) How can interaction on this visualization help to improve the user’s understanding of the calculation system?
   → Compare scenarios
   → See where impact originates
Byproducts

**MxLDriver for Java**
CRUD operations, authentication, meta information, static/dynamic graph generation

**Documentation in Swagger [6]**
Interactive and static documentation

**Modelling Patterns**
Aggregation, Condition, Implicit Join and more
Outlook

User Validation Needed (Industry Partner)  Usability Validation Needed (User Tests)  Data Source Integration (e.g. SAP HANA)
References


A concept for the visual and interactive impact analysis and simulation of data changes to enterprise metrics

by Matti Maier
Additional Information

BACKUP
Examples within this Thesis

1. **Hotel Business**
   - On SQL Database and T39
   - Generated Data
   - Goal: Easy example for presentation

2. **EAM**
   - On T39
   - E.g. simulating the impact of additional applications in domains

Other areas of application include gross margin calculations, financial portfolio analysis, risk management evaluation and more.
Research Methodology

**Design Science** (Hevner et al. 2004)

- **Design as an Artifact:**
  - Models (e.g. architecture) and a prototype (application)
- **Problem relevance:**
  - Problem is to visualize impact of influencing factors --> solve with an application
- **Design evaluation:**
  - Descriptive Evaluation -> Scenario to demonstrate the utility
  - Testing -> Black- and Whitebox with Unit tests
  - Analytical -> fit of architecture into application landscape
    - SQL databases as common data stores, CSV for Excel as a common tool, etc.

- **Research Contributions:**
  - Visualization technique
  - Application design, e.g. data source interface

- **Research Rigor:**
  - Construction methods: patterns (e.g. composite, singleton, client/server)
  - Evaluation methods: (see above)

- **Design as a Search Process:**
  - evolution of the architecture --> refinement, extension
  - Different model storages to proof the portability and improve the interface design

- **Communication of Research:**
  - Presentations
  - Final paper
  - Swagger UI

**Systematic Literature Review** (Kitchenham/Charters 2007)

- **Review Protocol**
  - Questions
  - Resources
  - Search Terms
  - Selection Criteria
Challenges in Implementation

• **Connecting T39**
  – Result: MxL Driver
  – Specialty: Data retrieval over MxL

• **Generating Queries**
  – Different operations
  – Different data sets
  – Recursive dependencies

• **Finding the right frameworks**
  – D3.js vs. Gephi vs. GraphStream vs. ...
  – Tempo.js for JavaScript/JSON Templates
  – Validation frameworks
Performance

• **Parallelization**
  – Calculation is performed in parallel as much as possible

• **Code to Data**
  – Computation is pushed to data as far as possible

• **Connection Pooling**
  – JDBC Data Sources are connected using connection pools

• **Lightweight JSON REST Interfaces**
  – Reuse of JSON where possible
  – Thus little conversion necessary
Security

• **Input Validation**
  – from User
  – from external systems, e.g. T39
  – Counter measure against XSS

• **Session Reinitialization**
  – Prevent Session Fixation

• **Forced HTTPS Connections**
  – Forward from HTTP to HTTPS connections

• **Custom Error Pages**
  – No presence of version and server
Modelling Patterns

• Column Aggregation
e.g. sum of a column in a table

• Calculating the Average
i.e. sum / count

• Copying Node
How to copy and reuse a node?

• Condition
How to model conditions?

• Implicit Join
How to implicitly join tables?
Migration on T39

Why was this application not migrated on T39?

• Industry-driven Topic
  – Various data sources, e.g. SAP HANA

• Integration of T39 earlier this year
  – Could have lead to a delay of the thesis

• Incompatible libraries in T39
  – e.g. D3.js requires other jQuery(s)
  – When DataWidgets are completed, then integration in DW Infrastructure