ConceptMix:
Self-Service Analytical Data Integration Based on the Concept-Oriented Model

Alexandr Savinov
Database Technology Group
Technische Universität Dresden, Germany
Data Commander - http://conceptoriented.com
PROBLEM

- **Variety** of data sources: one aspect of the big data problem
  - **Integrate**: data sources have to be mashed up to produce the desired result

- **Data wrangling** (curation, munging, scraping) – the most tedious part of the overall analysis process
  - **Transform**: refactor the structure of data (schema)

- **Original data** does not have data the user needs
  - **Analyze**: new attributes have to be computed

**Challenge**: How to simplify operations with data so that the tool can be used by non-IT users?
PRODUCT VISION

Data sources

ConceptMix: self-service data integration, transformation and analysis tool
ConceptMix is column-oriented rather than cell-oriented
Data is defined by column formulas (4) rather than cell-formulas
Drag-n-drop a source column (1-3) with automatic recommendations

Formula bar

= COUNT( this <- (Orders) -> (Customers) )

<table>
<thead>
<tr>
<th>Category</th>
<th>Total Amount</th>
<th>Customers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drinks</td>
<td>50.000</td>
<td>876</td>
</tr>
<tr>
<td>Electronics</td>
<td>10.543</td>
<td>356</td>
</tr>
<tr>
<td>Garden</td>
<td>3.826</td>
<td>84</td>
</tr>
<tr>
<td>Toys</td>
<td>23.82</td>
<td>1.539</td>
</tr>
</tbody>
</table>

ConceptMix: Self-Service Analytical Data Integration Based on the Concept-Oriented Model
Alexandr Savinov, DATA 2014, 31.08.2014
Key enabler: concept-orientation:

- Concept-oriented model of data (COM)
  - Unified model: simple and natural representation
  - Partially ordered set
  - Functional approach
- Concept-oriented expression language (COEL)
  - No joins, no group-bys, no formal logic
  - Simple and expressive analytical operations
  - Algebra of functions
- Column-based data processing model
  - Fast analytical operations with data (analytical database)
  - Column is a function

More info: [http://conceptoriented.org](http://conceptoriented.org)
Goal: define a new set in terms of existing sets and functions

Two operations

- Product: all combinations of greater sets
- Project: all outputs of some function

\[ \text{StatCat} = \text{PRODUCT}\left(\text{Status stat, Categories cat}\right) \]

\[ \text{Categories} = \text{Products} \rightarrow \text{cat} \]
Goal: link to sets using existing functions

```
Double Products = TUPLE (
    String kind = this.type,
    Integer num = this.no,
)
```
AGGREGATION

- Parameters:
  - Fact set
  - Grouping function
  - Measure function
  - Aggregation function

ConceptMix: Self-Service Analytical Data Integration Based on the Concept-Oriented Model

Alexandr Savinov, DATA 2014, 31.08.2014
CONCLUSION

● Novelties:
  ◆ Unified data model and expression language are used
  ◆ Column formulas as opposed to cell formulas for derived data

● Advantages of ConceptMix (Data Commander):
  ◆ Ease of use: radically simplifies analytical data integration; kills complexities when manipulating data
  ◆ Fast time-to-value: from months to minutes
  ◆ Lower IT costs: move the burden of authoring BI contents to the end users
  ◆ Increase motivation; more convenient consumption of data

● Future work:
  ◆ Assistance engine: recommending mappings, relationships, sources
  ◆ Selection propagation and inference for interactive analysis

● More info: http://conceptoriented.org