

Lightweight DWH Data Analysis for SMEs

Lukas Dötlinger, Manuel Penz, Markus Reiter & Stephanie Widauer

University of Innsbruck, Innsbruck, Austria

Outline

- Motivation
- Methodology
- DWH Needs for SMEs
- Review of DWH Services
- DWH Services
 - Segment
 - Panoply
 - Tableau
 - Snowflake

Motivation

- Increasing digitisation results in collection of more data
- SMEs are more commonly using digital systems
- They therefore need a proper way to organise their data

- Determine the specific needs for a DWH
- Summarise key factors and challenges

- Research for different DWH solutions targeted at SMEs
- Comparison and classification of those systems
- Check of suitability for SMEs

- **Heavy reliance on spreadsheets**
 - growing number of spreadsheets across multiple departments is unmanageable combining them all together
- **Overwhelming spreadsheets**
 - spreadsheets can reach the point where no data can be added anymore
- **Takes too much time to wait**
 - one spreadsheet per department means other employees have to wait for the newest version

DWH Needs for SMEs (contd.)

- **Discrepancies on data and reports**
 - different third party information in every department's spreadsheet could lead to data discrepancies after combining all together
- **Too much time invested in creating reports**
 - auto update the DWH System after data in sources changes – departments can rely on always correct up to date Data

- **Software as a Service (SaaS) / Data Warehouse as a Service (DWaaS)**
 - cost-effective pay-as-you-go model
 - straightforward setup
 - no hardware necessary, can be deployed instantly
- **Review & Comparison of DWaaS Products**
 - What do they have in common?
 - How do they differ?
 - Are they suitable for small businesses?

Services we have compared:

- Segment
- Panoply
- Tableau
- Snowflake

Segment

- customer data platform
- plans
 - **Free:** 2 data sources, 1,000 API calls, 300+ integrations
 - **Team:** unlimited data sources, 10,000-100,000 API calls (\$120-\$1,125/month)
 - **Business:** 100,000+ API calls, custom pricing
- in total 98 data sources and 650 data destinations
- programming languages or ReST APIs as data sources allow integration of virtually any system
- no analytics capability, meant to simplify data collection and distribution by managing all data flow in one place

- Data visualisation & data reporting with help of machine learning
- Tableau Software Inc. offers a tool for every step in the analytics/business intelligence (BI) cycle:
 - **Transactions Stored** - no actual database - data is stored in *.tde* or *.hyper* files
 - **Data Analysed**
 - Data Preparation - *Tableau Prep Builder*
 - Analysis & Explore - *Tableau Desktop & Tableau Public*
 - **Insights Shared**
 - **Decisions Made**
 - **Outcomes Monitored**
 - Safety & Security - *Tableau Server & Tableau Online & Tableau Mobile*
 - Collaboration – *Tableau Server & Tableau Online & Tableau Mobile*
 - Stability – *Tableau Server & Tableau Online & Tableau Mobile*
 - Data centre or Cloud – *Tableau Server & Tableau Online & Tableau Mobile*

