



Hochschule für Technik,  
Wirtschaft und Kultur Leipzig

## Exposé für eine Masterarbeit

zur Erlangung des akademischen Grades

Master of Science

im Studiengang Informatik  
der Fakultät Informatik und Medien  
der Hochschule für Technik, Wirtschaft und Kultur Leipzig

## Einsatz von Scheduling-Tools im Data-Warehousing - eine Marktanalyse

Vorgelegt von:	Raphael Drechsler
Anschrift:	Kieler Str. 34, 04357 Leipzig
Kontaktdaten:	Tel.: +49 1525 4194262 E-Mail: raphael.drechsler@googlemail.com
Matrikelnummer:	69872
Fachsemester:	4
Erstgutachter:	Prof. Dr.-Ing. Thomas Kudrass (HTWK)
Kontaktdaten:	Tel.: +49 341 3076-6420 E-Mail: thomas.kudrass@htwk-leipzig.de
Zweitgutachter:	B.Sc. Torsten Böttcher (integration-factory GmbH & Co. KG)
Kontaktdaten:	Tel.: +49 69 25669269-0 E-Mail: boettcher@integration-factory.de
Vorauss. Abgabedatum	Dezember 2019
Datum	28. Juni 2019

# Inhaltsverzeichnis

<b>1 Problemstellung</b> . . . . .	3
<b>2 Zielsetzung</b> . . . . .	3
<b>3 Vorgehensweise</b> . . . . .	4
3.1 Vorbereitung Evaluation . . . . .	4
3.2 Kandidaten suchen . . . . .	4
3.3 Durchführung Evaluation . . . . .	5
3.4 Formulieren der Konsequenzen für Kundenszenario A und B . . . . .	5
<b>4 Voraussichtliche Gliederung</b> . . . . .	5
<b>5 Erste Literaturverweise</b> . . . . .	6
<b>6 Zeitplan</b> . . . . .	6
<b>Literaturverzeichnis</b> . . . . .	7

# 1 Problemstellung

Das Unternehmen *integration-factory* ist ein Consulting-Unternehmen und bietet für seine Kunden Lösungen in den Bereichen Business Intelligence und insbesondere Integration von Unternehmensdaten an. Gegenwärtig besteht dabei in zwei Kunden-Projekten eine Situation, in der durch *integration-factory* eine DWH-Lösung bereitgestellt wurde, welche von einer Marktanalyse von Scheduling-Tools sowie einer Betrachtung von ETL-Workflow-Optimierung profitieren könnte.

Im Folgenden sollen diese zwei Kundenszenarios beschrieben werden.

**Kundenszenario A** In Kunden-Solution A ist bereits das kommerzielle ETL-Scheduling-Tool *Control-M* von *BMC Software* im Einsatz. Mit den derzeit im Projekt umgesetzten ETL-Workflows treten Probleme mit unnötigen Wartezeiten bei Statusübergängen im Workflow auf. Es besteht der Bedarf an einer performanteren Lösung. Durch die Umsetzung dieser Arbeit sollen für diese Kundensituation entsprechende Handlungsoptionen evaluiert werden. Dies soll einerseits durch das Untersuchen der Optimierungsmöglichkeiten in *Control-M* sowie der Betrachtung von Workflow-Optimierungen und andererseits durch einen Vergleich der *Control-M*-Lösung mit weiteren Scheduler-Alternativen geschehen.

**Kundenszenario B** In Kundensolution B ist gegenwärtig kein dediziertes Scheduling-Tool im Einsatz. Eingesetzt wird das ETL-Tool *Informatica PowerCenter* von *Informatica*, welches Funktionalitäten für ein rudimentäres Scheduling von ETL-Workflows bereitstellt. Über die Integration eines Open-Source Scheduling-Tools in die Kundensolution besteht die Möglichkeit komplexere ETL-Workflows abzubilden. Durch die Umsetzung dieser Arbeit sollen für diese Kundensituation infrage kommende Open-Source-Kandidaten gefunden und verglichen werden.

# 2 Zielsetzung

Aus der Problemstellung ergeben sich im Wesentlichen die folgenden Punkte für die Zielsetzung.

**Hauptziel: Marktanalyse von ETL-Schedulern** Es soll eine Marktanalyse für ETL-Scheduler durchgeführt werden, um einen Überblick über die zur Verfügung stehenden Optionen zu gewinnen. Da die Suche nach einem adäquaten Scheduling-Tool primär durch Szenario B getrieben ist, liegt der Fokus dabei vorrangig auf Open-Source-Tools. Um einordnen zu können, was ggf. ein Austausch des Schedulers in Szenario A bedeuten würde, sollen jedoch auch *Control-M* und ggf. weitere kommerzielle Scheduler-Tools in die Analyse mit einbezogen werden. Ebenso sind die rudimentäre Scheduling-Funktionalitäten von *Informatica* in den Vergleich mit einzubeziehen, um den Gewinn an Features durch eine Integration eines dedizierten Scheduling-Tools in

die Kunden-Solution B bemessen zu können.

**Integration des Gewinner-Kandidaten in die Kunden-Solution** Je nach Ergebnis der Analyse wird ggf. eine Anpassung der entsprechende(n) Kunden-Solution ein weiteres Ziel dieser Arbeit.

**Untersuchen von Optimierungsmöglichkeiten in den einzelnen Tools** Im Rahmen der Evaluation sollen Optimierungsmöglichkeiten bezüglich der Performance innerhalb der einzelnen Tools untersucht werden. Das schließt die Untersuchung mit ein, ob sich das in Szenario A bestehende Wartezeit-Problem in *Control-M* reduzieren bzw. beheben lässt.

**Untersuchen von Optimierungsmöglichkeiten für aktuell umgesetzte Workflows** Ggf. können bestehende, komplexere Workflows auf Optimierungsmöglichkeiten bezüglich Parallelisierung und Sequentialisierung untersucht werden. Komplexere Workflows sind dabei in Szenario A zu finden.

## 3 Vorgehensweise

### 3.1 Vorbereitung Evaluation

Zunächst sollen die theoretischen Grundlagen besprochen werden. Dabei soll auf die Themen ETL-Workflows, deren Einordnung in den DWH-Prozess eingegangen und die Anforderungen an einen ETL-Scheduler zusammengetragen werden. Anschließend sollen die speziellen Kundensituationen betrachtet und sich daraus ergebende, zusätzliche Anforderungen aufgenommen werden.

Anhand der gesammelten Anforderungen erfolgt das Aufstellen von Bewertungskriterien, Mindestanforderungen und Bewertungsmaßstäben, sowie die Gewichtung der Kriterien nach dem Scoring-Modell.

Für die durchzuführenden Performance-Messungen sollen im Rahmen der Vorbereitung die in den Kundenszenarios implementierten Workflows betrachtet und repräsentative Referenz-Workflows abgeleitet werden. Ggf. werden bei der genauen Betrachtung der bestehenden Workflows Optimierungsmöglichkeiten bezüglich Parallelisierung und Sequentialisierung sichtbar.

### 3.2 Kandidaten suchen

Im Anschluss an die Vorbereitung erfolgt die Suche von kommerziellen und Open-Source Scheduler-Tool-Kandidaten für die Evaluation. Diese sollen in der schriftlichen Ausarbeitung vorgestellt und bereits an dieser Stelle gegen die zuvor definierten Mindestanforderungen geprüft werden.

### 3.3 Durchführung Evaluation

In der Evaluation werden für alle Kandidaten die vorbereiteten Bewertungskriterien anhand der definierten Bewertungsmaßstäbe bepunktet. Die Teil-Ergebnisse werden in einer Bewertungsmatrix festgehalten.

Die Evaluation wird primär auf die Performance der Tools ausgerichtet sein. Die zuvor definierten Referenzworkflows werden dabei mit den jeweiligen Tools umgesetzt und anschließend Laufzeitmessungen vorgenommen. Bei der Umsetzung der Workflows in den jeweiligen Tools können dabei voraussichtlich weitere Kriterien, wie z.B. intuitive Benutzerführung und Bedienbarkeit, bewertet werden. Weiterhin sollen die Tools auf Möglichkeiten untersucht werden, die Laufzeiten der umgesetzten Workflows zu optimieren. Sind alle Kriterien für alle Kandidaten bewertet, erfolgt die Auswertung der Bewertungsmatrix und somit das Feststellen des Gewinner-Tools.

### 3.4 Formulieren der Konsequenzen für Kundenszenario A und B

Aus dem Ergebnis der Evaluation sowie den gewonnenen Erkenntnissen sollen die Konsequenzen für die Kunden-Szenarios A und B abgeleitet werden. Ggf. erfolgt eine Umsetzung der Beschriebenen Anpassung/Änderungen in der Kunden-Solution.

## 4 Voraussichtliche Gliederung

Aus der geschilderten Vorgehensweise lässt sich die folgende voraussichtliche Gliederung ableiten.

1. Einleitung
  - 1.1. Ausgangssituation, Problemstellung und Ziel
  - 1.2. Vorgehensweise
2. Vorbereiten Evaluation
  - 2.1. Einordnung ETL-Workflows in DWH-Prozess
  - 2.2. Anforderungen an einen ETL-Scheduler
  - 2.3. Anforderungen aus kundenspezifischen Begebenheiten
  - 2.4. Bewertungskriterien und Bewertungsmaßstäbe
  - 2.5. Referenz-Workflows
3. Scheudling-Tool-Kandidaten
4. Durchführung Evaluation
  - 4.1. Kandidat 1
  - 4.2. Kandidat 2

- 4.3. ...
- 4.4. Control-M
- 4.5. ETL-Tool: Informatica PowerCenter
- 5. Auswertung
  - 5.1. Auswertung der Evaluation
  - 5.2. Auswirkungen auf Kundenszenario A
  - 5.3. Auswirkungen auf Kundenszenario B
- 6. Fazit

## 5 Erste Literaturverweise

Generelle Literatur zu DWH: z.B. *Schnider, Dani, Jordan, Claus, Welker, Peter, and Wehner, Joachim. Data Warehouse Blueprints : Business Intelligence in Der Praxis. 2016. Web. [1]*

Literatur zur Optimierung von ETL-Workflows (Parallelisierung und Sequentialisierung): *Karagiannis, Vassiliadis, and Simitsis. SScheduling Strategies for Efficient ETL Execution. Information Systems 38.6 (2012): 927-945. Web. [2]*

Vergleich von Scheduling-Tool-Features über Informationsmaterial der Anbieter: z.B. *Cisco Tidal Enterprise Scheduler Data Sheet[3]*

Nutzwertanalyse nach Scoring-Modell: z.B. *Nöllke, Matthias. Entscheidungen Treffen : Schnell, Sicher, Richtig. 5. Aktualisierte Auflage. ed. München: Haufe-Lexware GmbH & KG, 2011. Web.[4]*

## 6 Zeitplan

gegenwärtig offenes To-Do.

## Literatur

- [1] D. Schnider, *Data Warehouse Blueprints : Business Intelligence in der Praxis*. 2016.
- [2] A. Karagiannis, P. Vassiliadis, and A. Simitsis, “Scheduling strategies for efficient etl execution,” *Information Systems*, vol. 38, no. 6, 2012.
- [3] I. Cisco Systems, “Cisco tidal enterprise scheduler data sheet.” [https://www.cisco.com/c/en/us/products/collateral/cloud-systems-management/tidal-enterprise-scheduler/c78-636900-00\\_cte\\_scheduler.html](https://www.cisco.com/c/en/us/products/collateral/cloud-systems-management/tidal-enterprise-scheduler/c78-636900-00_cte_scheduler.html), 2019. [Online; Stand 26. Juni 2019].
- [4] M. Nöllke, *Entscheidungen treffen : Schnell, sicher, richtig*. München: Haufe-Lexware GmbH & Co. KG, 5. aktualisierte auflage. ed., 2011.