Setup and Implementation of an Automated Testing Pipeline for a DataOps Use Case

Bachelor's Thesis (T3201)

presented to the
Department of Computer Science

at the
Baden-Wuerttemberg
Cooperative State University
Stuttgart

by
Oliver Rudzinski

submitted on
September 7th, 2020

Project Period (CW)	25/2020 – 36/2020
Matriculation Number, Course	`5481330`, TINF17A
Training Company	Hewlett Packard Enterprise
Internship Company	DXC Technology
Project Supervisor	Dipl.-Ing. Bernd Gloss
University Supervisor	Jamshid Shokrollahi, Ph.D.

Erklärung

Declaration of Authorship

Ich versichere hiermit, dass ich meine Bachelorarbeit mit dem Thema:

I hereby declare that I am the sole author of this bachelor's thesis on the topic:

Setup and Implementation of an Automated Testing Pipeline for a DataOps Use Case

selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

and that I have not used any sources other than those listed in the bibliography and identified as references.

Ich versichere zudem, dass die eingereichte elektronische Fassung mit der gedruckten Fassung übereinstimmt.

I further declare that the electronically submitted version of this thesis is identical to the printed version.

_____, _____ _____

Ort Place | Datum Date | Unterschrift Signature

Abstract

Contents

List of Acronyms

BI	Business Intelligence
CI/CD	Continuous Integration & Deployment
DWH	Data Warehouse
ELT	Extract-Load-Transform
ETL	Extract-Transform-Load
DAMA	Data Management Association
MBA	Market Basket Analysis
ML	Machine Learning
SPC	Statistical Process Control
VCS	Version Control System

List of Figures

List of Tables

List of Source Code Excerpts

1 Introduction

1.1 Relation to Project Environment

1.2 Project Scope

1.3 Task Definition

1.4 Chapter Overview

2 Theoretical Backgrounds

3 Testing Framework Design Process

4 Analytics Pipeline DataOps Enablement

4.1 Actual State Analysis: MBA Data Analytics Pipeline

4.2 DataOps Enablement Requirements

4.3 Architecture Design

4.4 Modus Operandi

4.5 Implementation

5 DataOps Testing

5.1 Testing Strategy

5.2 Testing Architecture Design

5.3 Implementation

6 Solution Evaluation

7 Conclusion