PUNJAB UNIVERSITY COLLEGE OF INFORMATION TECHNOLOGY

FINAL PROJECT REGISTRATION FORM

Date: 02-11-2022 Degree Software Engineering Session 2019

No: _____

Project Title: ETL Pipeline Implementation for Demand Forecasting (Research)

Proposed Primary Advisor's Information:

Name: Dr. Shuja Ur Rehman Baig

Address: Punjab University College of Information and technology Old campus

Designation: _____ Email: shuja@pucit.edu.pk

Ph. No. Res. _____ Office _____ Mobile: 03214207445

Group Leader:

Roll No. BSEF19M039 Mobile No: 0312-4736426

Name. Muhammad Ahmad Ashfaq

	Roll Number (Group member)	Name of Group member	Email	Signature
1.	BSEF19M001	Atif Iqbal	bsef19m001@pucit.edu.pk	Atif
2.	BSEF19M010	Nimra Haq	bsef19m010@pucit.edu.pk	
3.	BSEF19M038	M. Usman Arshad	bsef19m038@pucit.edu.pk	
4.				

Supervisor Signature: _____

(For Official Use Only)

☐ **Recommended**

☐ Meeting Required: Date: _____ Time: _____ Place: _____

☐ Rejected

Remarks: _____

Project Title (Revised): _____

_____

Project Office Supervisor

ETL Pipeline Implementation for Demand Forecasting (Research)

Project Proposal Document

LEADER
Muhammad Ahmad Ashfaq
BSEF19M039

Atif Iqbal
BSEF19M001

Nimra Haq
BSEF19M010

Usman Arshad
BSEF19M039

SUPERVISOR
DR. SHUJA UR REHMAN BAIG
Department of Information Technology

Faculty of Computing & Information Technology
Wednesday, 2 November 2022

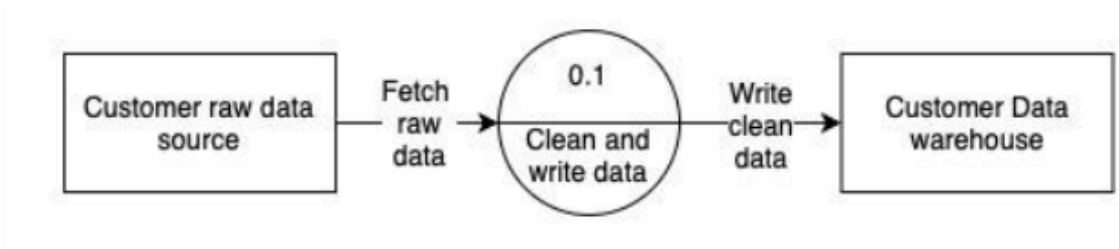
CONTENTS

PROJECT TITLE

ETL Pipeline Implementation for Demand Forecasting (Research)

INTRODUCTION AND BACKGROUND

Big data analytics in supply chain management is getting more attention nowadays. This is because it becomes easier to analyze customer behavior, trends, and demand prediction. In this research project, we will be evaluating the flow of data, as in the supply chain the data is scattered and is in different formats so we'll be looking for a method to collect that data that can be batched as well as real-time from different sources and store data at one place i.e, warehouse. We investigate the Machine Learning algorithms that can be used for data analysis and a pipeline for the flow of that data.
Introduction:
Nowadays, businesses adopt ever-increasing precision marketing efforts to remain in the competition and maintain and grow their profit margin. Forecasting models used customer data and transactions to manage product supply chains accordingly.
In ETL, extract, transform, load is a three-phase process where data is extracted, transformed and loaded into an output data container. The data can be collated from one or more sources and it can also be outputted to one or more destinations. The first step is to connect to the source of the raw data and fetch the data load using REST API services. The system processes the load, to handle data integrity issues, such as redundant data, skipped data, updated and deleted data, and change of data types for a particular field based on the schema available at the source. Final step is to load the clean data into the cloud-based warehouse.



Project Goals & Objectives

Demand forecasting helps reduce risks and make efficient financial decisions that impact profit margins, cash flow, allocation of resources, opportunities for expansion, inventory accounting, operating costs, staffing, and overall spend. All strategic and operational plans are formulated around forecasting demand.

ETL Pipeline Implementation for Demand Forecasting (Research)

Group Leader: Muhammad Ahmad Ashfaq

Project Members:

Name	Registration #	Email Address	Signature
Atif Iqbal	BSEF19M001	bsef19m001@pucit.edu.pk	
Nimra Haq	BSEF19M010	bsef19m010@pucit.edu.pk	
Usman Arshad	BSEF19M038	bsef19m038@pucit.edu.pk	
M. Ahmad Ashfaq	BSEF19M039	bsef19m039@pucit.edu.pk	

Project Goal: The aim for this research is to provide an idea about implementing the ETL Pipeline using best Environment.

Objectives:

Sr.#	
1	Data Acquisition
2	Analyzation of Data
3	Discussion and Findings
4	Propose solution

Project Success criteria: Investigated the best environment for ETL implementation and Demand forecasting.

Organization Address: Punjab University College of Information Technology, University of The Punjab, Allama Iqbal Campus, Shahrah-e-Quaid-e-Azam (The Mall), 54000 Lahore, Pakistan

Type of project: Research

Suggested Project Supervisor: Dr. Shuja Ur Rehman Baig

Approved By:

Date:

LITERATURE REVIEW

The concept of data pipelines is fairly recent and the advancements in this particular domain have been enhanced with recent developments in cloud infrastructure and cloud storage. These are the few innovations in the corresponding field of data pipelining. In 2009, a research was conducted on ETL Technology and it was based on the following principle. The initial software programs that encourage the first stacking and the occasional refreshment of the warehouse are usually known as Extraction-Transformation-Loading (ETL) forms. There were sure restrictions to this, the extraction of information despite everything stays a difficult issue, for the most part because of the shut nature of the sources, streamlining and resumption issues and nonattendance of a benchmark is preventing future research.

Demand Forecasting of supply chains methodologies are widely used in many industries and can be broadly classified into two main categories:

Traditional Statistical Models:

Traditional models basically use the historical sales data provided by the organization to forecast future demand of the product. These models use time series data for forecasting. Further in this paper three most commonly used traditional forecasting methods are discussed.

Machine Learning Models:

Machine learning is based on algorithms that can learn from data over the time in autonomous fashion. In demand forecasting this technique allows us to predict the amount of product to be purchased in a defined future period. Compared to traditional statistical models ML provides more accurate forecasting, identifies hidden patterns in data, increases adaptability to changes and creates a robust system.

RESEARCH DESIGN

The primary research method for this study is literature review and conceptual modeling. We'll investigate the different platforms that can be used for implementing the ETL pipeline i.e, Apache provides different environments for implementing the Extraction and Transformation of the data.First we will analyze which is the best method for extraction of dataset.After selecting one,we will explain the methodologies of the best practice.Then we will find which is the best method for transformation of Data.After selecting this platform,we will research on the best non-structured data warehouse.Then we will find machine learning algorithm that can best fit on the project for analyzing the data for Demand Forecasting.After all these platform chosen,we will make conceptual modeling of how these platform will work together to achieve the final product.

PROJECT MANAGEMENT

PROJECT NAME	PROJECT LEAD	PROJECT START DATE	PROJECT END DATE	TODAY'S DATE
ETL Pipeline Implementation for Demand Forecasting (Research)	Mohammad Ahmad Ashfaq	18-06-2022	01-06-2023	03-11-2022

Individual columns represent weeks.

				MONTH	OCTOBER					NOVEMBER					DECEMBER					JANUARY - FEBURARY				
				WEEK START DATE	18	20	21	23	30	3	13	20	27	-	4	11	18	25	-	1	8	15	22	29
ACTIVITY	% DONE	START DATE	END DATE		1	2	3	4	5	6	7	8	9	-	10	11	12	13	-	14	15	16	17	18
Initial phase																								
Ideation	100%	18-10-2022	20-10-2022																					
Specifying use cases	100%	21-10-2022	23-10-2022																					
Research on topic	100%	24-10-2022	29-10-2022																					
Selecting and finalizing the technology	100%	30-10-2022	02-11-2022																					
Project proposal	100%	02-11-2022	03-11-2022																					
Research phase-1																								
Extracting data	**0%**	07-11-2022	07-12-2022																					
Analyzing data	**0%**	08-12-2022	08-01-2023																					
Documentation	**0%**	09-01-2023	09-02-2023																					
Findings and reviews	**0%**	10-02-2023	10-03-2023																					

PROJECT NAME	PROJECT LEAD	PROJECT START DATE	PROJECT END DATE	TODAY'S DATE
ETL Pipeline Implementation for Demand Forecasting (Research)	Muhammad Ahmad Ashfaq	18-06-2022	01-06-2023	03-11-2022

Individual columns represent weeks.

				MARCH					APRIL					MAY					JUNE				
			WEEK START DATE	2	9	16	23	30	6	13	20	27	-	4	11	18	25	-	1	8	15	22	29
ACTIVITY	% DONE	START DATE	END DATE	1	2	3	4	5	6	7	8	9	-	10	11	12	13	-	14	15	16	17	18
Research phase-2	0%																						
Extracting data	0%	11-03-2023	11-04-2023																				
Analyzing data	0%	12-04-2023	30-04-2023																				
Documentation	0%	01-05-2023	13-05-2023																				
Findings and reviews	0%	14-05-2023	03-05-2023																				
Finalization																							
Final review	0%	01-06-2023	20-06-2023																				
Final submission	0%	21-06-2023	29-06-2023																				