J. R. Almeida, L. Coelho and J. L. Oliveira

BIcenter: A collaborative Web ETL solution

Contents

List of Figures

List of Tables

Preface

License

This document was written in RMarkdown[1] using the bookdown[2] package.

[1] https://rmarkdown.rstudio.com
[2] https://bookdown.org

1

Introduction

BIcenter is a web-based platform that allows the building and management of ETL pipelines, by non-IT users, in a multi-institution environment. Each institution manages and mantains ETL tasks and provides the resources for the execution of the associated tasks. Thus, each institution owns its private data sources, servers for ETL task execution and a task scheduler that allows periodic execution. In order to provide access and management control of ETL tasks and institutions, there are four distinct types of users:

- Administrator: Entity that moderates the platform. This actor has permissions to create and delete institutions.
- Resource Manager: Entity that manages private data sources and execution servers. This actor has permissions to create and delete private data sources and execution servers, within specific institutions.
- Task Manager: Entity that builds and executes ETL tasks. This actor can create and configure ETL tasks, within specific institutions.
- Data Analyst: This actor has permissions to inspect task execution history, namely the resulting data, execution logs and performance metrics.

1.1 Main requirements

1.1.1 Information Security

Since ETL tasks parse and handle sensitive data that belongs to a particular institution, the system must be designed and implemented taking in account these security issues, namely user authentication, access control, data protection and isolation.

1.1.2 System Reliability

Considering the periodic execution of ETL tasks, it is important to ensure that each execution is correctly initialized, started, motorized and concluded. When some fatal error occurs during an ETL task execution, the system must be able to handle the error and successfully conclude the execution.

1.1.3 Solution Scalability

Since a complete ETL tool typically encompasses a wide variety of components, it is crucial to build an agile approach to the development and integration of new ETL components.

2

To do

2.1 to do

to do