INCEDO

Hackathon 2K18 | EDH-chennai 2

DB Doc

Application Design document

Contents

[Abstract 2](#_Toc505249814)

[Solution Approach 2](#_Toc505249815)

[Design Architecture 2](#_Toc505249816)

[Project Repository: 3](#_Toc505249817)

[Connection Library: 3](#_Toc505249818)

[Metadata Library 3](#_Toc505249819)

[Data Modeller 3](#_Toc505249820)

[Tag Engine 3](#_Toc505249821)

[Document Template Library 3](#_Toc505249822)

[Publish Engine 3](#_Toc505249823)

[Business Impact 3](#_Toc505249824)

DBDoc

Flexi tool to document the Data source

# Abstract

A well understood data is the key for all quality analysis and in a real world scenario, it is often a challenge for the developer and analyst and many case for the business owners to attain this understanding, due to lack of documentation in many of the transaction data sources available today.

This utility, DB Doc, aims at documenting the various data sources, in a well flexible way so that the documents can be use by any sort of audience

# Solution Approach

To resolve the issue of documenting the DB sources and over come the restrictions of the existing tools in market, We tried to design this utility as flexible as possible so that the user have freedom to extend the usage of this tool to any number of data sources and any desired document template.

The tool make use of the Metadata available in all data sources that defines the structure of the source and also the relations between the objects.

We broke down the process of approach the data source and reading the metadata into small steps so that each stage can be designed as configurable and the user will get the flexibility to use the standards or define their own standards to define a object

# Design Architecture

Utility designed based on the common design approach followed by most of the data modular and analytics tools. It is a complete template based design with multiple components.

The utility comprises of the below components

* Project Repository
* Connection Library
* Metadata Library
* Data Modeller
* Tag Engine
* Document Template Library
* Publish Engine

## Project Repository:

It is a collection of Projects created by the user. The user can create their own project and save them as draft until they publish them. The project repository also manage the versions of projects created. The user will have the option to share the project with other users or groups for view or edit

## Connection Library:

It is a collection of all drivers and standard connection Strings (JDBC/ ODBC) templates of all common Data sources like MS SQL, MySQL, Oracle, Informix, DB2, stored as JSON

While connecting to a data source, these templates will pop in and get the variable data like DBName, UserID, Password etc to compile a complete ODBC Connection string or JDBC connection URL (whichever is preferred)

## Metadata Library

This is the collection of Standard piece of codes used to Query the metadata of a particular type of Server.

User have the flexibility to add their own codes in the library to query the legacy DBs or any new DB

These codes are connected to the type of data source and they will get invoke in the project on connecting to a particular type of data source

## Data Modeller

This component works on the collected metadata and provide the visual representation of object and object relation

## Tag Engine

Aim of this piece of utility is to map the Business terms with the commonly appear column names and create an engine that will traverse around the DB metadata and suggest the mapping business term, which further get use in the final document

## 

## Document Template Library

This library is a collection of Document templates. User can use the existing templates or create their own templates as per the need of final audience

## Publish Engine

This component will give the plugins to publish the document in various formats like doc, pdf and on various platforms like share point, embedded html codes etc

# Business Impact

The utility is mainly designed for the developers and analyst, it will reduce the tremendous effort of going through all DB objects, relations and make sense of them. It will also reduce the huge time spent on them for documenting in a traditional way and hence save lots of money and effort and also ensure improved productivity.