**Decision Engine Platform**

### BITS ZG628T: Dissertation

by

Pramod Kumar N

2014HT13292

# Dissertation work carried out at

## Coextrix Technologies Pvt. Ltd., Bengaluru

****

**BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE**

**PILANI (RAJASTHAN)**

November 2016

**Decision Engine Platform**

### BITS ZG628T: Dissertation

by

Pramod Kumar N

2014HT13292

# Dissertation work carried out at

## Coextrix Technologies Pvt. Ltd., Bengaluru

Submitted in partial fulfillment of M.Tech. Software Systems degree programme

Under the Supervision of

Mr.Ramesh Krishnamoorthy, CEO,

Coextrix Technologies Pvt. Ltd., Bengaluru

****

**BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE**

**PILANI (RAJASTHAN)**

November, 2016

#### 

#### CERTIFICATE

This is to certify that the Dissertation entitled Decision Engine Platform and submitted by Pramod Kumar N having ID-No. 2014HT13292 for the partial fulfillment of the requirements of M.Tech. Software Systems degree of BITS, embodies the bonafide work done by him under my supervision.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Signature of the Supervisor

Place : BANGALORE\_\_\_\_\_\_\_\_\_\_

Date : \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Ramesh Krishnamoorthy, CEO, Coextrix Technologies PVT LTD., Bangalore

(Name, Designation & Organization &Location)

iii

**Birla Institute of Technology & Science, Pilani**

**Work-Integrated Learning Programmes Division**

**First Semester 2016-2017**

**BITS ZG628T: Dissertation**

**ABSTRACT**

**BITS ID No.** **: 2014HT13292**

**NAME OF THE STUDENT** **: Pramod Kumar N**

**EMAIL ADDRESS** **: pramod974@gmail.com**

**STUDENT’S EMPLOYING** **: Coextrix Technologies, Bengaluru**

**ORGANIZATION & LOCATION**

**SUPERVISOR’S NAME** **: Ramesh Krishnamoorthy**

**SUPERVISOR’S EMPLOYING** **: Coextrix Technologies, Bengaluru**

**ORGANIZATION & LOCATION**

**SUPERVISOR’S EMAIL ADDRESS : ramesh.krishnamoorthy@coextrix.com**

**DISSERTATION TITLE** **: Decision Engine Platform**

iv

**Abstract**

In modern day applications that provide more visibility into data irrespective of domain, require collecting data from several different sources. Collected data needs to be consolidated, normalized and then presented through an application. There is a lack of correctness in the data aggregated owing to the nature of issues with the sources of data itself. Since the original data sources are doing nothing to clean up the data they provide or generally there is no standardization maintained across domains in context. Several applications are designed to clean up the data coming through various sources and allow customers to perform complex analytics on top of the data. Transforming and reconciliation process is extensively complex, but is required to create higher accuracy in the data. The transformation and reconciliation process comprises of multiple rules, interactions with multiple components in order to achieve the accuracy. In popular existing open source technologies the way the rules are written and configured is stereotype and hard to maintain owing to the quantity of rules modelled and its ability to scale in future.

There are no matured open source decision engines built using python which satisfies applications current needs, where there is an end to end feature to model and maintain rules which could interact with multiple sources and components. Accurate data has been always sought for and is the next gen thing to get custom insights irrespective of domain. The underlying need to build a configurable and maintainable platform is highly compelling, where a Decision engine facilitates connecting to various sources and components in order to write domain specific configurable rules so as to achieve valuable decision in limited turnaround time. The Goal is to build, deploy and scale the existing stereotype rules/models to a more maintainable and configurable Decision engine and also provide an easy interface and framework to model rules on the fly build, test and deploy it without any manual intervention.

Broad Academic Area of Work: **Software Architecture**

Key words: **Rules, Engine, Decision, Python, Platform, Framework, Configurable**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Signature of the Student**

**Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Date:**

**Place: BANGALORE**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Signature of the Supervisor**

Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Date:**

**Place: BANGALORE**

v

**ACKNOWLEDGEMENT**

The satisfaction that accompanies the successful completion of dissertation would be incomplete without mentioning the people who made it possible, whose constant guidance and encouragement crowns all the efforts with success.

I express sincere gratitude to Mr.Ramesh Krishnamoorthy, CEO, Coextrix for being kind enough to provide me an opportunity to work on a project in this organization and also being my mentor.

I am thankful to Mr. Ganesh Guruswamy, CTO, Coextrix for providing valuable ideas, suggestions, facilities, co-operation and his encouragement at all moments of my approach towards tackling the problem.

Last but not the least, I would like to express my love and gratitude to my beloved family and friends, for their understanding & motivation, through the duration of this project.

vi

**Table of Contents**

Certificate………………………………………………………………………………………………………………………………iii

Abstract…………………………………………………………………………………………………………………………………iv

Acknowledgment………………………………………………………………………………………………………………….vi

Table of Figures ix

Chapter 1: Introduction 1

1.1 Data Source 2

1.2 Problems and Challenges with Data 2

1.3 Data Transformation and reconciliation…………………………………………………………………………………….….3

1.4 Software Architectures and Design Patterns 4

Chapter 2: Project Overview 5

2.1 Existing System 5

2.2 Problems with Existing System 5

2.3 Available Software Solutions and their issues…………………………………………………………………………….5

2.4 Proposed System…………………………………………………………………………………………………………………………………..5

2.4.1 Rules……………………………………………………………………………………………………………………………………………………..5

2.4.2 Rule Orchestration.……………………………………………………………………………………………………………………………..5

2.4.3 Decision Engine …5

**Table of Figures**

Figure 1: Knotted rules in Existing System 4

Figure 2: Proposed System Decision Engine 6

**GLOSSARY**

REST – Representational State Transfer

API – Application Program Interface

CSV – Comma Separated File

MSEXCEL – Microsoft Excel

**Chapter 1: Introduction**

**1.1 Data Sources**

Data has been most sought for in the modern era to gain deeper and meaningful insights into ones business. The insight from data is required irrespective of which domain one is working on, in order to gain such insight data has to be curated from multiple sources. The data source [1] is can be typically a connection set up to a carious computer databases running as a server or it could be as simple as a file (CSV, text, MSEXCEL, etc.) or it could just be a stream of data coming in via API or live feed.

**1.2 Problems and Challenges with Data**

The volume, correctness and consistency pose a major challenge with the data collected from various sources has to be processed before they can be analysed. Some businesses require domain specific incorporations as well to meet the data quality issues. Some of the common data related problems and challenges are as follows.

1. Poor data quality such as noisy data, dirty data, missing values, inexact values, inadequate data size and poor representation.

2. Integrating conflicting or redundant data from different sources and other forms like audio, video and images, geo data, text, social, numeric, etc.

3. Proliferation of security and privacy concerns by individuals, organisations and governments.

4. Unavailability of data or difficult access to data.

5. Efficiency and scalability issues to effectively extract the information from huge amount of data in databases, dealing with huge datasets that require distributed approaches, dealing with non-static, unbalanced and cost-sensitive data.

6. Constantly updating model to handle new incoming data, processing large, complex and unstructured data into a structured format.

**1.3 Data Transformation and Reconciliation**

Owing to multi natured problems with collected data, it has to be transformed and reconciled. Typically along with all standards process available for transformation and reconciliation, business use domain specific rules and knowledge to transform and reconcile the data so they are analysis ready.

Data transformation is the way of converting data from one format to another; generally this involves conversion of format of a source system into the required format of a new resultant system. The typical process involves converting documents, but data

**Page 1 of 6**

Conversions at times involve the conversion of a program from one computer language to another in order for the program to run on a different platform. The usual reason for this data migration is the adoption of a new system that is completely different from the previous one. [4]

In real world, data transformation involves the use of a special program that can read the data’s original base language, determine which language the data that must be translated for it to be usable by the new program, and then move ahead to transform that data.

Data Transformation consist of two phases:

Data Mapping: This involves assignment of elements from the source system toward the destination to capture all transformations that commonly occur. This is not trivial when there are complex transformations like many to one or one to many rules for transformation that need to occur.

Code Generation: The creation of the actual transformation program. The resulting data map specification is consumed to create an executable to be executed on computer systems. [5]

The Reconciliation process consists of validating, computing, generating data mathematically so the integrity of the data in picture is validated to most correct extent from available assumptions and resource.

**1.4 Software Architecture and Design Patterns**

Software architecture refers to the fundamental structure of structures, the discipline of creating such structures, and the documentation of these structures. These structures help in describing about the software system. Each structure consists of software elements, relations between them, and also properties of both put together as well, along with rationale for the introduction and configuration of each element.

The architecture of a software system is similar to the architecture of a building. The software architectures comprises of following activities.

Architecture supporting activities

Software architecture supporting activities are done during core software architecture activities. These activities help a software architect to do analysis, synthesis, evaluation and evolution. For example, architect curates knowledge, make decisions and document during the analysis phase.

Knowledge Management and Communication is the activity of exploring and managing knowledge that is important for designing software architecture. A software architect does not

**Page 2 of 6**

Work alone. They get details, functional and non-functional requirements and design contexts, from various intended stakeholders; and provide outputs to these stakeholders. Software architecture knowledge is often tacit and is held in the minds of stakeholders.

Design Reasoning and Decision Making is the activity which evaluates design decisions. This activity is important to all three of the core software architecture activities. It consists of gathering and associating decision contexts, formulating design decision problems, finding solution options and evaluating positives and negatives before making decisions. This process occurs at various levels of decision, while evaluating significant architectural requirements and software architecture decisions, and software architecture analysis, synthesis, and evaluation.

Documentation is key activity of making note of all the design generated during the software architecture process. A system design is described using several views that frequently include a static view depicting the code structure of the system, a dynamic view depicting the actions of the system during execution, and a deployment view depicting how a system is placed on hardware for execution.

**Page 3 of 6**

**Chapter 2: Motivation**

* 1. **Introduction**

Every organization will have business to do a lot of different things, each of these business span across different domains and need rules to maintain their business generally the rules are very enormous in nature and are if there are no business rules it is very chaotic and the intended output is very poor in nature and results in more work.

Business organizations in order to tackle this rules use the help of skilled domain experts or go in leveraging specific automated software to suit their needs, sometimes there is both combination of software and human effort.

Few organizations use software to meet their needs as there are few tools available to help them automate their requirements. The major issues are when there is change in the business rules, making these changes come in to action is a very cumbersome task.

* 1. **What are Business Rules?**

A business rule is a rule that defines or constrains some aspect of [business](https://en.wikipedia.org/wiki/Business) and always results in the final answer being either true or false. [[6]](https://en.wikipedia.org/wiki/Business_rule#cite_note-BRG1-1) Business rules goal is to assert business structure or either to control or influence the behaviour of the business. For instance, a business rule might state that if the price and volume petrol is right to be lifted from a terminal, another one would be to select a most good pick of supplier from list of preferred suppliers and supply schedules.

Business rule are usually informal or even unwritten, documenting the rules clearly and making sure that they don't conflict is an indispensable activity. When carefully steered, rules can be used to aid the organization to better achieve goals, remove obstacles to market growth, reduce expensive mistakes, foster communication, comply with legal requirements, and increase drastically customer loyalty.

All the business rules fall into any of the following four categories:

* Definitions:

These define the actual domain specific definitions how data needs to act. These are generally the high level concepts which are fundamentals.

* Facts:

These are atomic level calculations which define how the data has to be or how it has to be calculated or formulated

* Constraints:

The constraints define how far the data can be or how they are quantitatively controlled.

* Derivations:

These are derived from facts and generally the most common rule which interleave into the facts to form more complex formulations.

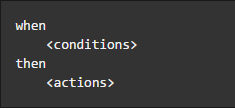
* 1. **Problems associated with Business rules**

The rules are generally chaotic though they leverage people and automation software there are a lot of problems associated with the Business rules. The most important problem is adding and changing business rules. There is always a major technical surgery required. There is a need for an application which provides consistent enforcement and constant even efficiency gain. The major technical surgery required for each change causes slow rate of change resulting in high costs.

Generally the rules are in the hands of the developer. Whatever the creation, modification, update, deployment and testing is generally done as a whole process. So the bottom line of the problem is to find a solution which is more flexible and moves the change and modifiability overhead to a non-developer or a business consultant.

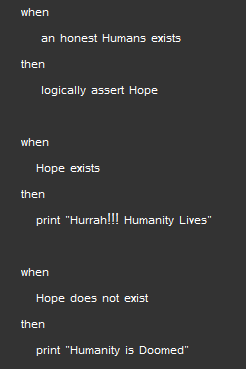
* 1. **Rule Engine**

The concept of Rule Engine is very ambiguous in nature. It could be any system that makes use of rules, in any form, in order to leverage upon it so that it can be applied to data in order to produce meaningful outcomes. The common instance of rule engine includes simple systems like that of a form validation and dynamic expression engines which work on a stream of data.

****

**Figure : Rules Skeleton**

The rule engine typically has a lot for rules similar to the skeleton shown in figure shown above, that are executed based upon a hierarchical structure defined by another entity / user. The end goal of the rule engine is to help produce meaningful outcomes. The main end goal of the rule engine is to provide a scalable environment to build rules and make use of them. The figure shows a typical inference with the help of a rule engine. The figure explains an outcome of either how one can take a decision of living of a humanity or if humanity being doomed by taking a call, based on if honest human exists. The series of when and then conditions helps the rule engine to decide the outcome. The initial rule is to assert when honest humans exist. If the assertion is true then the rule engine will print a message “Hurrah!!! Humanity lives”, if the assertion is false then the rule infers doom of the humanity by printing a message “Humanity is Doomed”.



* 1. **Advantages of Rule Engine over if… then structure**

**2.5.1 Declarative Programming**

Rule engines emphasis is on "What one has to do" not "How once can do it".

The major advantage of this by leveraging on rules makes it easy to give solutions to difficult problems and consequently have these solutions verified since rules are much simpler to read than code.

**2.5.2 Logic and Data Separation**

Your data is contained in domain objects, the logic is contained in the rules. This is fundamentally removing the OO coupling of data and logic together, which can be boon or bane depending on once point of view. The advantage is that the logic can be much easier to maintain as and when there are changes in the future, as the logic is emulated as rules. This especially a true advantage if the logic is cross-domain or multi-domain interdependent.

**2.5.3 Speed and Scalability**

The separation of data and logic to one’s domain object data help in speed, scalability and also especially efficient when one has datasets that do not change entirely.

**2.5.4 Centralization of Knowledge**

Making use of rules, one can create a repository of knowledge in the form of rules which is executable. This makes it a single point of truth, for instance business policy ideally rules are so readable that they can also double up as documentation.

**2.5.5 Tool Integration**

Tools such as PyCharm [features] and in future maybe extensions to Web based User Interfaces provide ways to edit and manage rules and also get instant feedback, validation and content aid. Auditing and debugging tools are also incorporated as part of this. The rule when implemented on a platform can also connect to multiple sources thus making it a one stop solution for all the integration and consumption.

**2.5.6 Explanation Facility**

Platform effectively provide an explanation facility by providing log of the decisions made by the rule engine along with the reasons describing why the decisions were made, thus helping the end user formulate for rules and take control of the randomness.

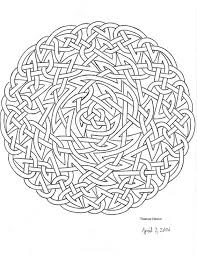
**2.5.7 Understandable Rules**

The object models and Domain Specific Languages that model ones problem domain, can set oneself up to write rules that are very similar to natural language. They provide themselves to logic that is more understandable to non-software-developers, business analysts, and Domain experts as they are laid down in their language.

**Chapter 2: Project Overview**

* 1. **Existing System**

The existing system consists of multiple nests of conditional statements and rules in order to reconcile the data to take decisions. The domain specific rules written are stereotype and interlock themselves and difficult to scale them by adding new rules. The natural tendency of any software is to undergo changes in the due course of time. The changes if any are very difficult to incorporate and integrating and creating the new rules and propagating through the system are very complex in nature. Due to infeasibility the original system design cannot be incorporated due to compliance issues. The goal of the existing system is to take data as input from various sources such as CSV, text files, Database systems and data feeds etc. apply certain domain specific rules to the data and give the output of the file and also state the different anomalies so that the business analysts can take decisions.



**Figure : Knotted rules in Existing System**

* 1. **Problems with Existing System**

There are lot of complex and nested conditional statements which are very difficult to maintain and scale it, any new addition should undergo rigorous integration testing and there is lot of resources, time and cost involved. There are many loose ends that needs to be joined together so various components have more visibility into each other.

* 1. **Available Software Solutions and their Issues**

Modern technologies have spanned across various business aiding them at different levels. Business Rule engines form the foundation in modelling multiple business rules. There are lot of matured technologies available as proprietary software. Few of the available popular

**Page 4 of 6**

softwares are Drools, InRule etc. which are expensive and learning curve is huge. Since our existing systems components consist of only open-soures technologies, integrating them with available proprietary software requires lot of effort and huge licencing cost.

**2.4 Proposed System**

The rules grow exponentially and the business analysts want to explore innovative way to model rules on the fly and also add them without any manual intervention with little scripting background they want to leverage open source options where they can have a lot of flexibility and no licensing cost. The decision engine platform provides end to end features so as to satisfy all the current application needs to be more configurable , scalable and automatically deployable to get insights. The whole systems leverages the rules design pattern and completely built using open source technologies.

The proposed system will have the following components

* Rule repository
* Rule Orchestration
* Decision Engine

**2.4.1 Rule Repository**

The Rule Repository consists of multiple modules which can be modified, deleted and incorporated on the fly. The rules in this repository is easily configurable and for the foundation for the Rule orchestration.

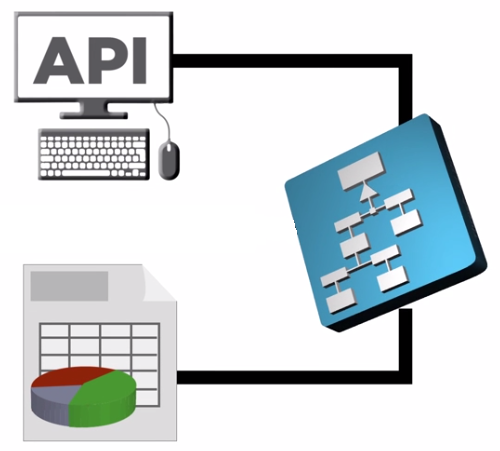
**2.4.2 Rule Orchestration**

The Rule Orchestration consists of meta class which whose modules and attributes get generated on the fly depending on which got modified, deleted and incorporated on the fly in the rule repository. The rules in repository is made ready for the Decision engine to consume.

**2.4.2 Decision Engine**

The decision engine forms the crux of this whole platform. It uses the rule orchestration class to execute the required domain specific rules on the data got either from the API or a CSV. Figure 2 shows the high level flow of how multiple rules can be applied on streams from an API or CSV.

**Page 5 of 6**



**Figure 2: Proposed System Decision Engine**

**Page 6 of 6**

**References**

1. Data Source Wiki, https://en.wikipedia.org/wiki/Data\_source
2. Computer File Wiki, https://en.wikipedia.org/wiki/Computer\_file
3. Data Stream Wiki, https://en.wikipedia.org/wiki/Data\_stream
4. Definition, https://www.techopedia.com
5. Data Reconciliation-https://en.wikipedia.org/wiki/Data\_validation\_and\_reconciliation
6. <https://en.wikipedia.org/wiki/Business_rule>
7. https://www.jetbrains.com/pycharm/features/