Commercial Proposal

for

**Application for identifying and outlining data quality**

**Submitted To:**

**DataLyticx**

**Submitted By:**



**Confidentiality Notice**

This document contains proprietary information of Oxyent Technologies Pvt. Ltd. No part of this document may be reproduced, stored, copied, or transmitted in any form or by any means now known or hereinafter invented, electronic, digital, mechanical, photocopying, scanning, recording or by any information storage or retrieval system, without the express consent of Oxyent Technologies Pvt. Ltd.

# 

# Company Profile

Oxyent (hereafter referred as “Company”) is an ISO 9001:2008 and DSIR (Department of Scientific and Industrial Research) recognized company based in Delhi, India. The company’s development and R&D center is located at DLF Jasola, New Delhi. The company's framework and its components incorporating industry best practices along with specialized domain knowledge helps in achieving goals of informatics system.

Oxyent’s goal and vision was, and will always be, to provide affordable, functional software and services to various domain. Our team of core members is our greatest asset. Our team consist of domain experts and a group of highly motivated young professionals who are dedicated enough to share the companies value system and to push it to new heights.

DataLyticx (hereafter referred as “Client”) is brand name of product line being offered and represented by Tech Optimization, a California Corporation (CO).

# Background and Mission Statement

Data quality-related problems cost companies millions of dollars because of lost revenue opportunities, failure to meet regulatory compliance or failure to address customer issues in a timely manner. Below are some facts stating the market opportunities in this area:

* + - * + Gartner predicts that MDM will grow at an 18 percent compound annual growth rate over five years, reaching $2.9 billion worldwide in 2014
        + The average company loses $8.2 million a year from poor data quality, according to a Gartner survey of 140 companies.
        + A recent study indicated that in the US medical business alone, companies lose over $5 Billion, annually, due to bad data

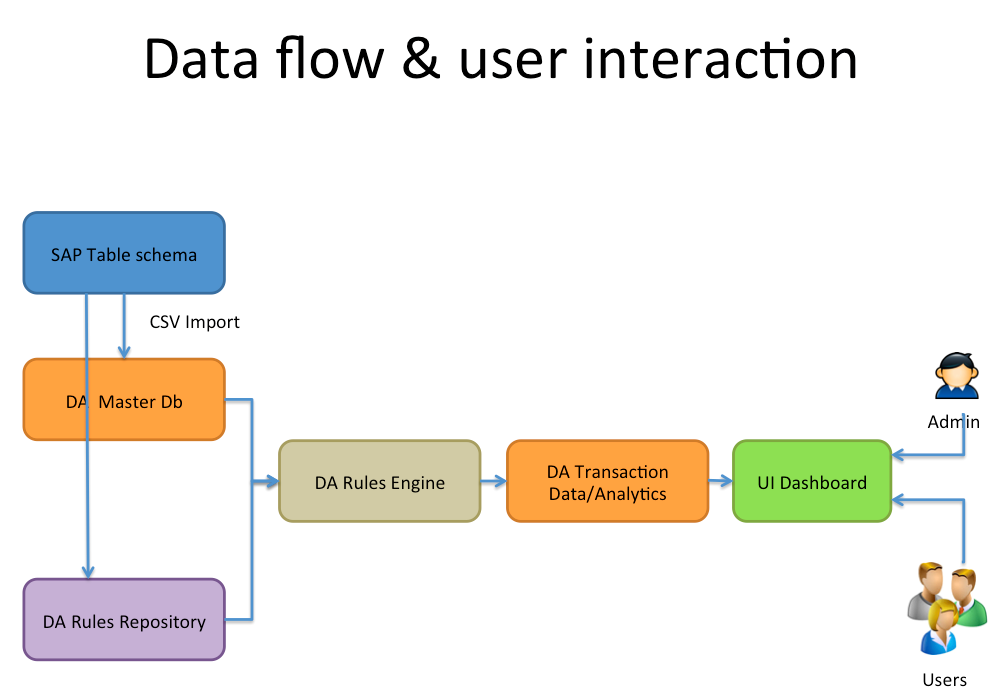
One of the key hampering reason slowing the growth of solution in this space is lack of clarity in measuring data quality and its financial impact on business.

DataLyticx team identified the opportunity in the data quality industry and decided to offer its flagship quality product defined in next section of this document. The current document intends to scope the development effort required to build first working prototype of the solution to offer its possible customer base a working concept copy. The prototype is intended to have all characterstics i.e. scalability, concurrency and multi-tenancy to ensure that path to prototype development is in line to product roadmap.

# Objectives

DataLyticx propose to improve the data quality issues present across industries by measuring data quality across pre-defined metrices. It proposes to connect data with rules defining the completeness, accuracy, integrity, relevance and validity of data. The solution should have the abilty for organizations to accept the metrics in their current format or modify/enhance existing ones and/or add new metrics to measure and improve data quality.

DataLyticx application will allow user to upload table schema in CSV format. In the first version the SAP based table schema will be used. The system will allow at run time to build entities in MySQL database based on which rules will be defined. System will allow transaction data to be loaded to help user relate data quality impact on business.

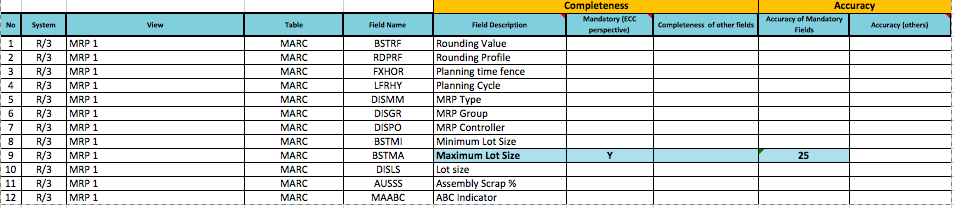


**Figure 1: Data flow in the proposed system**

Different user groups (i.e. Administrator or users) will interact with system at different touch points to either define rule metrics to either define or evaluate data quality.

Following section outlines major tasks that will be supported by proposed system:

1. Entities Definition
   1. Entities:
      1. Various entity types
         1. Plant
         2. Material Group
         3. Material
         4. Plant Type
      2. Different areas of business
         1. Logistics/MRP
         2. Sales
         3. Purchasing
         4. Storage
         5. Scheduling
         6. Manufacturing
   2. Data Quality definition for fields and entities used in rules (per customer)
      1. Mandatory fields
         1. Defining mandatory fields
      2. Completeness
         1. All fields are filled out
         2. In Manufacturing, within Material Group level, 30 fields out of 100 fields needs to have a value for data to be complete.
      3. Accuracy
         1. Check if plant is type=fab
         2. Check if time zone=PST
         3. Check if zip code=non zero
         4. Out of 30 fields which are complete, check for accuracy. For example Plat type is one of the field. It can be fab, assembly or test. These are the only three allowed values.
         5. Field level completeness should be calculated.

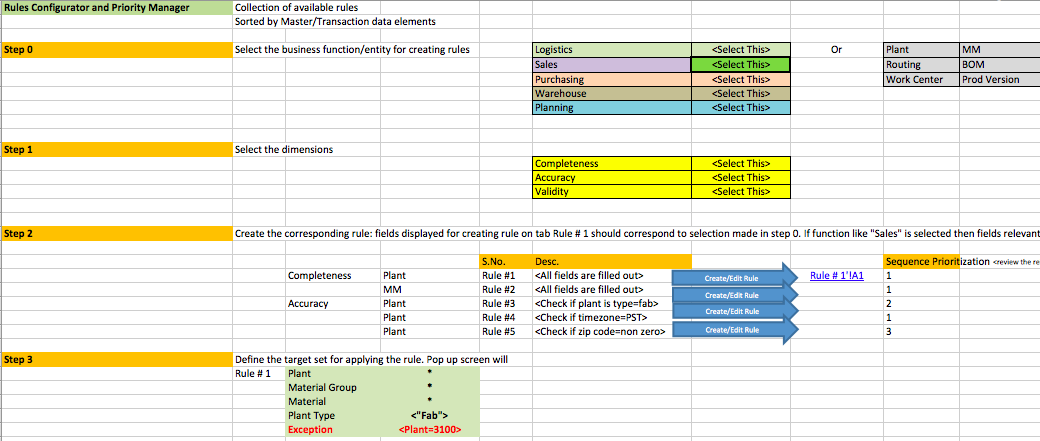


**Figure 2: Example of field mapping with SAP ECC**

* 1. Transaction data

1. Rules Definition
   1. Business rules for entities
      1. Select area
         1. Logistics/MRP
         2. Sales
         3. Purchasing
         4. Storage
         5. Scheduling
      2. Within each area there is heirarchy of rules – for example:
         1. Sales Org
         2. Sales Area
         3. Disti Channel
         4. Plant
         5. Exception
      3. Section of data quality for which this rule applies would be selected
         1. Completeness
         2. Accuracy
         3. Integrity
         4. Relevance
         5. Validity
      4. Rule definition
         1. Fields
         2. Type:
            1. All fields are filled out.
            2. Check if field type is equal to certain value.
            3. Rule Salience/prioritization
      5. Select the area or hierarchy on which this rule will apply
         1. Plant
         2. Material Group
         3. Material
         4. Plant Type

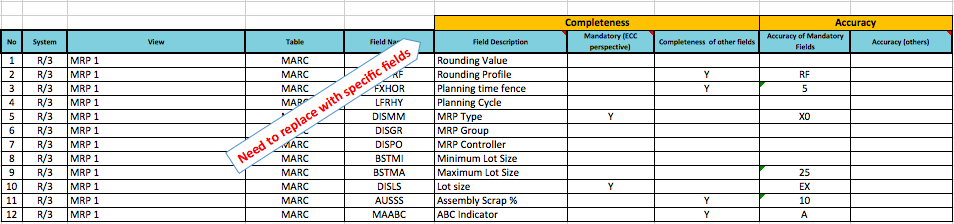
Below snapshot explain the above process via diagram:



**Figure 3: Rule definition steps**

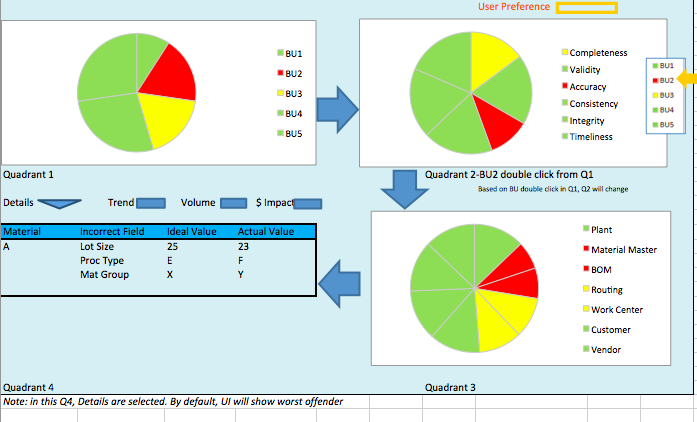
* 1. Defining rules of hierarchies
     1. Plant
     2. Material Group
     3. Material
     4. Plant Type

1. The solution should have the ability to define/change the defined metrics and also add new metrics which then would be applied by rules.
2. The rules should be classified (grouped) based on the type of analytics.
3. User should be able to use 1+ metrics to manage the data quality and don’t need to use all 5.
4. User Interface for data quality
   1. Reporting (charts)



**Figure 4: Quality calculation indexes**

* 1. Dashboard
     1. Aggregation/disaggregation based on a hierarchy
        1. Threshold definitions: User should be able to define these threshold values.
           1. Red – below 50%
           2. Orange – 51 - 70
           3. Yellow – 71 – 90
           4. Green – 90+
        2. Impact of Data Quality on business
           1. X % of Sales data has data accuracy of 50%
           2. Alerts
           3. Trends



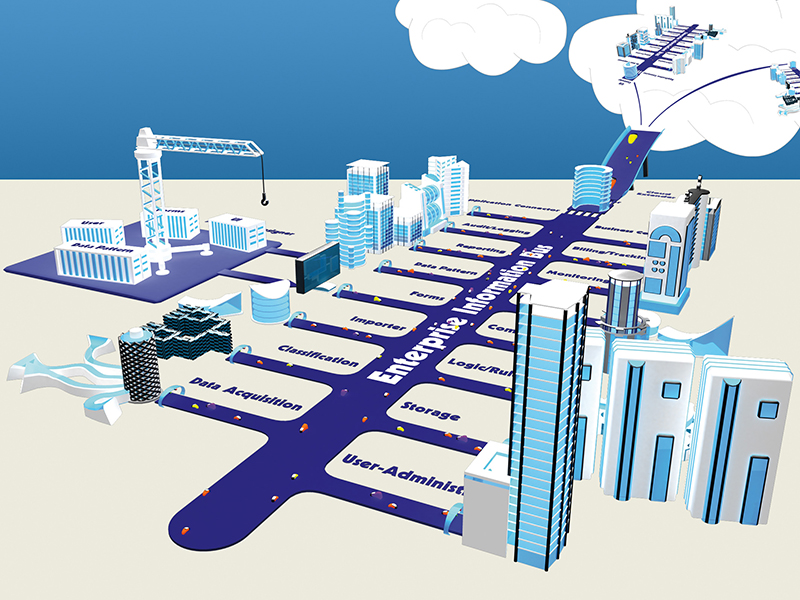
**Figure 5: Quality representation dashboard**

* 1. Master import

1. Quality Representation Dashboard- the dashboard needs to be designed in way that user is able to double click on any part of the pie chart to navigate to the next graph e.g. BU2 is Red in Quadrant 1. Double click on red pie should lead to Quadrant 2 Pie chart. Subsequent clicks should lead to other quadrants.
2. Master Data Import
   1. CSV data load
   2. Adherence to entity definition in Step 1
   3. Possibility to build entities on load of CSV file.
3. Briefing books will be provisioned for users
4. Trends over period of time will be generated in dashboard.
5. Volume – transaction data – look at the scenario 3
6. MySQL DB Schema
7. AWS instance setup
8. Transaction data integration to show impact on ROI
9. Clickstream integration to capture client information
10. Following definition of

# Architecture

Oxyent enterprise service bus platform will be used as middleware to allow scalability in future. Below diagram shows the bus component architecture highlighting its service oriented orchestration and separation of concerns via well-defined components. Open source technologies are used to handle specific application needs via components.



**Fig 6: Enterprise service bus and components**

Below is list of components needed for the solution and associated open source solutions:

* 1. Rules Component – Drools
  2. Database – Hibernate
  3. Rendering/Charts – ZK
  4. Data Analytics - Hadoop/Mahout
  5. Import/Export – Apache POI

Oxyent has rete based rule engine leveraging both production and working memory to identify matching rules and handling their interdependencies. Below diagram shows high level view of Rete based rule algorithm



**Fig 7: Rule component managing facts and rules in real time mode**

# Timelines

|  |  |
| --- | --- |
| **Tasks** | **Effort in weeks** |
| 1. Entities Definition (SAP based)    1. Entities for plants, location etc.    2. Data Quality definition for entities (per customer)    3. Defining hierarchies    4. Transaction data | 1. 2 Weeks |
| 1. Rules Definition    1. Business rules for entities    2. Rules for data quality       1. Completeness       2. Accuracy       3. Integrity       4. Relevance       5. Validity | 1. 2-3 weeks |
| 1. User Inerface for data quality    1. Reporting (charts)    2. Master import | 1. 2 Weeks |
| 1. Master Data Import 2. CSV data load    1. Adherence to entity definition in Step 1    2. Possibility to build entities on load of CSV file. | 1. 1.5 weeks |
| 1. MySQL DB Schema | 1. 1 week |
| 1. AWS instance setup | 1. 1 week |
| 1. Transaction data integration to show impact on ROI | 1. 2 week |
| 1. Clickstream integration to capture client information | 1. 1 week |
| 1. The solution should have the ability to define/change the defined metrics and also add new metrics which then would be applied by rules. | 1. 1 week |
| 1. The rules should be classified (grouped) based on the type of analytics. | 1. 0.5 week |
| 1. User should be able to use 1+ metrics to manage the data quality and done need to use all 5. | 1. 0.5 week |

Total Effort (in weeks) = 15.5

Proposed Team (3.5)

1. One Tester
2. One Java Senior Engineer
3. One UI expert (ZK)
4. One part time architect/project manager

# Scope &Outcome

The primary outcome of this project is to build a working solution that customers can use. This will be offered both on cloud (multi-tenant on AWS) or on premises. Customer will be offered to use this service for 2-3 months for free or charge. Support and upgrades post release of prototype are not part of this scope.

The development path is broken into two paths:

Phase 1: To be completed in ~ 10 weeks and offered to customers for feedback.

Phase 2: 4-6 weeks and will include refinements based on customer feedback.

* Minimum effort estimated = 15.5 week
* Maximum effort estimated = 16 week

# Costing & Timelines

The total cost is detailed below:

Phase 1: 8 Lakhs

Phase 2: 4 Lakhs

* Project cost to not exceed 12 lacs (or 20K USD).

# Intellectual property

The term "Intellectual Property" as used herein is broadly defined to include the developed solution, inventions, copyrightable works, and tangible research property. All Intellectual Property will belong to the DataLyticx.

# References

1. Demo \_Flow Working Sessions\_Back Up 07092013.xlsx.
2. Business Plan - Datalyticx Ver 3 – O.pptx
3. Discussion diagram between Jessie and Harpreet.

# Payment Terms:

* Client shall pay Oxyent based on the following
  1. 20 % of phase 1 cost in advance.
  2. 60 % of phase 1 payment at the end of phase 1
  3. 20% of Phase 2 at the commencement of Phase 2
  4. Balance of phase 1 cost, upon completion of Phase 2
  5. Balance of Phase 2 cost, 4 weeks after completion of phase 2.
* Client will ensure that Oxyent will receive payment within 7 days of getting the acceptance certificate.

# Term and termination:

This agreement is effective from 16th day of June 2014 and valid till termination of it. Either party may terminate the same by serving two month written notice to the other party.

# General:

1. In case there is any difference of opinion arising due to any interpretation or execution of any of the Terms and Conditions, the same shall be subject to the jurisdiction of the appropriate court at New Delhi, India and California, USA.
2. These terms constitute the entire agreement between the parties relating to subject matter hereof and supersede any previous agreement, communication, representation and understanding, verbal or written, between the parties with respect to the subject matter hereof.