Regulatory and Compliance Data Solution for Banking and E-Commerce

Virendra Deshmukh

Data Engineer

Abstract: *A Regulatory focused scalable data Platform to build data model for any financial Institution where rapidly growing business generates huge transaction or profile data per day and company has to oblige any kind of Audit conduct by internal/external policy team or government body. This would be comprehensive and strategically built analytical platform to serve both purpose analytics and Audits.*

1. Compliance: A mandatory org of Enterprise

"Compliance is mandatory activity for Bank or financial institution and early strategic implementation would be beneficial step."

Regulatory implementation involves metrics of customers life cycle attributes which represent sensitive personal information which is required by government and its authorized agencies.

Considering global bank can have diverse customers in different countries but everywhere bank has to comply and implement legalization like KYC, FATCA etc.

Hence a generic platform required to support global business.
Early implementation also leads to company loyalty and save from huge fine.

Compliance is no more SILO processand risk of loss in current business era has increased. Adequate information about customers is primary responsibility of enterprise in competing world

2. Scalable Architecture
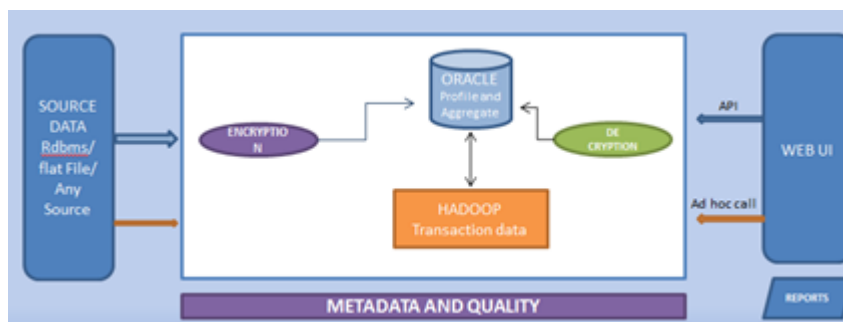
An Advance Scalable Architecture Can Serve Billions Of Records Processing."

The Proposed architecture is proven solution for huge data processing and scalable to rapidly growing business. RDBMS Oracle system has capability to perform web based analytics based on aggregated data points stored. Hadoop (HIVE) capable of providing Transactional data in batch mode. Encryption/Decryption can be handle on the fly with API based call.

3. A Centralize metadata based approach to maintain data quality and reconciliation.

Re-Conciliation
Collect all source, stage and target data points. Generate rules to compare source, stage and target data in form of script, etl, or program based and create daily report.

Quality Aspects
Generic quality implementation I.e. format, mandatory fields, not null. Drill down specific cases and implement as rules which should execute daily run and check the data, result would be alert, or hard stop process or re-initiation of process after remediation.
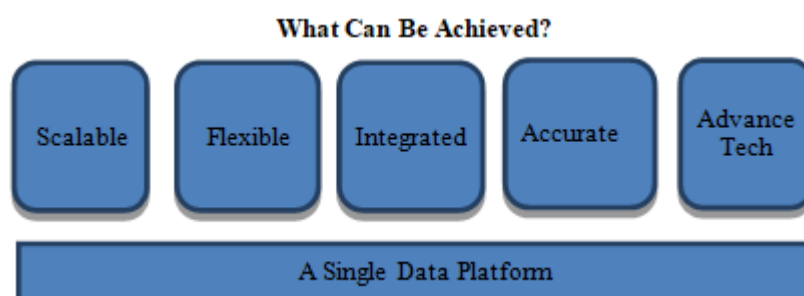
A Flexible Analytical Solution
"The Technology stacks which has Hadoop and Oracle, provide the capability of Huge Data Processing in distributed manner, huge data space and RDBMS capability of Oracle maintain the integrity and consistency in data element It serves the analytical purpose and Hadoop serves for transactional data or history of customers behaviors."

Flexible Mode of Delivery	1. Reports	2. Query based data availability	3.Web UI for end customer / Analytical tool	4.Web based search tool
	1) Daily/monthly scheduled reports	1. Aggregated data points can be queried by Internal analyst through direct database view layer.	1) A web based application which can be capable of showing metrics at aggregate level	1. A wild search to identify a given customer. It can be enable with given appropriate criteria and filter i.e. any dimension associated with a customer or any PII information Or particular document no etc…
	2) Ad-hoc reports		2) A range of criteria to list out customers	
	3) Repository of all reports in one place		3) Transactional data available in form of row level in batch mode	

4. Challenges and How Do We Solve?

- Diverse geographical location and country specific policy brings complexity to provide solution for regulatory dataset.
- A banking product workflow implementation and process work differently into various countries to comply with its policy and governance.
- Sensitive/PII/PCI data element which comes under very secure policy before storing.
- Regulatory solution involves third party data integration or government agencies.
- Policy applies on low level of granular records of customers. Which leads huge data storage?
- Government needs all segments of customers, it brings huge data sets into the system and only scalable and robust system can be used to achieve this.

- Segment based approach to land all source data to stage and then process and merge the data in to target tables.
- Table row level security to expose the data element to specific location based.
- Encryption and decryption of PII/PCI data on API call.
- A robust and scalable infrastructure platform capable to process huge dataset with data consistency and integrated datasets.
- Discreet storage for customer's Profile and Transaction attribute to build high performance promising System.
- Pluggable solution/design to scale any new countries incoming data to pipe in to system.
- Web UI based platform for non-technical or policy users to perform audit or any discovery on data in easy manner.

What Can Be Achieved?

Scalable | Flexible | Integrated | Accurate | Advance Tech

A Single Data Platform

Author Profile

Virendra Deshmukh is a DATA Enthusiastic and analyst. 6+ years of experience in finance and banking domain. Implemented Data solution for various Regulatory and RISK management project.