



TaranDM

Real-Time Decision Manager

September 2020

TaranDM is a new-generation decision manager which allows you to make smart real-time recommendations and decisions during customer interactions.

TaranDM is end-to-end reliable, scalable and modular solution for enterprise decision management, built with the best-performing big data technologies with no vendor lock-in. It is fully configurable and easy to deploy into any company's enterprise architecture.

Alibaba/Mail.Ru/Megafon

JV's new payment solution and eWallet service

"TaranDM solution built on open source with modern big data architecture is a great match for the systems which we develop internally."

O. Grich, CTO

Key Benefits

**Scalable**

TaranDM has been designed as highly scalable and massively parallel tool and is able to process tens of millions of clients daily.

**Big Data & Cost-Saving**

In order to provide machine learning models with the fullest information possible, TaranDM gathers data from multiple data sources. Since the external data sources are often paid, TaranDM is querying data sources in steps and the least promising cases are not queried further. Data requests are also cached so duplicate requests are not executed and paid.

**Flexible**

TaranDM is built directly on Python functionalities - any new machine learning algorithm could be easily and quickly tested, evaluated and deployed.

**Versatile Options for Deployment**

Thanks to the use of OpenAPI standard for interface specification, TaranDM is easy to deploy into client's current software architecture and connect to internal and external data sources as well as the front-end.

**Reliable**

Continuous Integration/Continuous Development software standard ensures the reliability of TaranDM 24/7. Any software maintenance is done without affecting accessibility or performance.

**Auditable**

TaranDM is fully compliant to work as a credit risk system. Every input and output is saved and every historical decision is replicable and deterministic. All changes to business rules are automatically versioned and every historical strategy is replicable.

**Reporting**

Interactive reports built on dedicated reporting data warehouse enables a user to see all relevant performance statistics and outcomes of any business strategy as well as performance of TaranDM run itself. Model development properties, such as train/test set, predictors' bins cutoffs etc. are propagated automatically into the reporting data warehouse and provide data scientists with full model-building detail.



One-Click Model Deployment

Model specification, development sample properties, predictor bins and other features related to model development are stored during modeling phase automatically and then accessible for direct deployment. This automation eliminates risk of human errors, saves data scientists' time and allows high flexibility in A/B testing.



Latest Technology Friendly

TaranDM is built as a modular tool - any module could be handled and modified on its own. Modular design also enables that any new/promising technology could be separately tested and later deployed in the respective module.



No Vendor Lock-in

All technologies used to develop, maintain, and run TaranDM have been carefully selected. Special emphasis has been placed on the latest and the best performing big data technologies and open source languages and tools. TaranDM can be integrated into enterprise's current solution with no vendor being locked in.



AntiFraud Module

Companies in financial sector face continuous thread of fraud which becomes more and more sophisticated and handling such risk is crucial especially when offering credit products online. TaranDM AntiFraud module performs set of pre-defined checks based on information such as personal data, device fingerprint, IP address, etc.



Lower IT Costs

Predictor construction and calculation, data source connection, creation of an automated decision process, deployment of a machine learning model into production - examples of tasks which have traditionally been bottlenecks in the decision management process, and often needed significant help from IT department. TaranDM was designed to reduce this dependence, and the workflow is done either automatically or by data scientists.



Data Scientist Friendly

Experienced data scientists are motivated when the software allows them to use the state-of-the-art machine learning algorithms, when they don't need to copy-paste results and other model information, and when they can easily modify strategies to maximize advantages of A/B testing. TaranDM has been designed in such way and enables data scientists to effectively build, test and compare strategies directly over the data lake. Predictors and all other model specifications are directly accessible from run-time environment.

TaranDM Workflow



There are two main ways to initialize the run of TaranDM. The first one is a **real-time trigger** (examples: applicant fills an online application for a credit product, client is at specific geolocation, client performs pre-defined behaviour in the mobile app, ...).

The second way how to run TaranDM is the batch **pre-scoring** of selected clients/leads performed either on direct request or during timeslots defined by the business user.

Each request is processed in **Strategy Selector** module, which select business strategies to be applied.

Data Source Caller then provides data from internal as well as external sources and **Model Selector** selects a machine learning model.

Scoring module calculates scores (or propensities) using appropriate data, machine learning model and other rules assigned to the set of strategies.

Decision module combines calculated scores with pre-defined business rules to create decisions, which are used further in the automated business flow.

TaranDM Business Usage Example

Credit Scoring

Financial institutions can significantly improve their risk management by incorporating new sources of external data into their infrastructure and extracting the value through state-of-the-art machine learning techniques. TaranDM is fully compliant to work as a credit risk system while supporting targeting clients in new ways, such as online scoring or delayed payment in online purchases.



Limit Management

Regulatory requirements, companies' long-term business goals and portfolio composition, clients' needs and their ability to repay – these all are important things to consider when managing risk limits of issued credit products. TaranDM limit management module is created specifically for this task and is automatically executed in the credit scoring service flow. Users can implement desired business logic through editable decision tables.



Credit Risk Evaluation and Approval Process

Once a loan is significantly past due, chances of repayment are limited even with highly efficient collection process and legal system. Hence, credit risk evaluation and approval process are the most important steps of credit risk management. TaranDM enables a user to execute them with emphasis on the following aspects: multiple data sources availability, access of the state-of-the-art machine learning methods, easy A/B testing of credit scoring strategy, flexible creation of predictors directly from raw data, automated scorecard deployment, explicability of used models, auditability and replicability of the whole process.



Online Scoring & Pre-Scoring

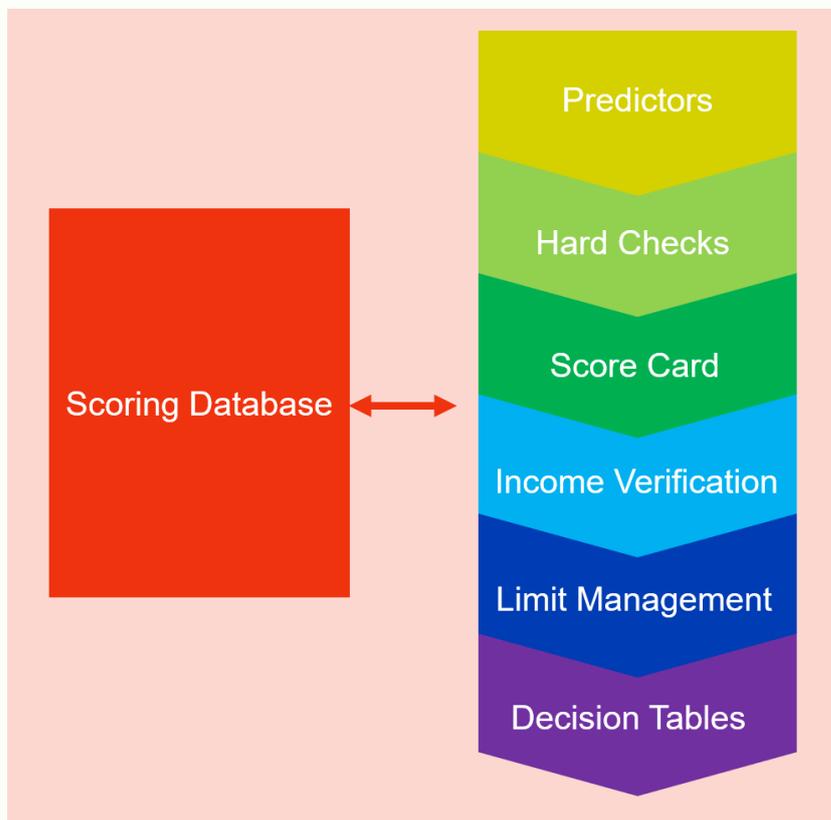
TaranDM is designed to process tens of thousands requests in the real-time on a daily basis. Scoring results are integrated in the architecture and follow-up processes (such as decision management), clients' required actions and related-data processing. Another method of usage is pre-scoring of existing set of clients (in size of tens of millions). Outcome of a batch pre-scoring is possible to be combined with propensity to buy functionality and used as an input for marketing activities (clients with the highest propensities and the best credit scores are targeted in the direct campaign).



Credit Risk Reporting

TaranDM reporting component gathers all the important data, stores them in dedicated reporting database and displays all aggregations and metrics in set of credit risk reports using a business intelligence platform (e.g. Tableau). Such reports allow their users to assess portfolio (or its subsets) across different risk measurements, usefulness of different data sources, approval rates, credit repayments in different scenarios and overall results of long-term business strategies.

TaranDM Credit Scoring - Core Modules



Predictors calculation

Modern machine learning algorithms are designed to take advantage of hundreds of predictors. Calculation, grouping and automated maintenance of all predictors is crucial for further smooth analysis and meaningful decision outcomes. These steps happen in the Predictors module. Only the predictors, which are necessary and available for each specific request, are calculated and as a result, the computation power needed is efficiently reduced. Moreover, the predictors' code is shared between analytics and production layers.

Hard Checks

Hard Checks module includes specific rules which are applied to data first, before any modeling or data requests take place (e.g. requirement for a minimum age when applying for a credit product). This module saves computational power and reduces financial costs as there are less requests for external data.

Score Card

Scorecards can be developed in any programming language (e.g. Python) and encapsulated as an object that contains all needed information like predictors' binning and transformation, descriptive data about development sample, data about model performance etc. The deployment itself is one click process.

Income Verification

Income verification block is responsible for calculating verified income of the applicant. This information can be collected either from credit bureau or based on other sources (e.g. internal bank data, pension funds). Multiple verified incomes are aggregated into one final value which is then compared with the income provided by the applicant.

Limit Management

Limit Management module calculates maximum limit available for an applicant with respect to multiple factors such as product type, risk grade etc. Maximum limits for separate scenarios are defined in decision tables.

Decision Tables

Implementing key decisions through the use of decision tables increases replicability, transparency and availability of the business rules. These tables are used to control which action should be taken based on given conditions. TaranDM decision tables are editable files which are evaluated dynamically and which are able to handle multiple dynamic conditions. Output of a decision table is processed automatically as it is integrated within the business flow. Moreover, each decision is available as a predictor for any subsequent scoring run or decision module.

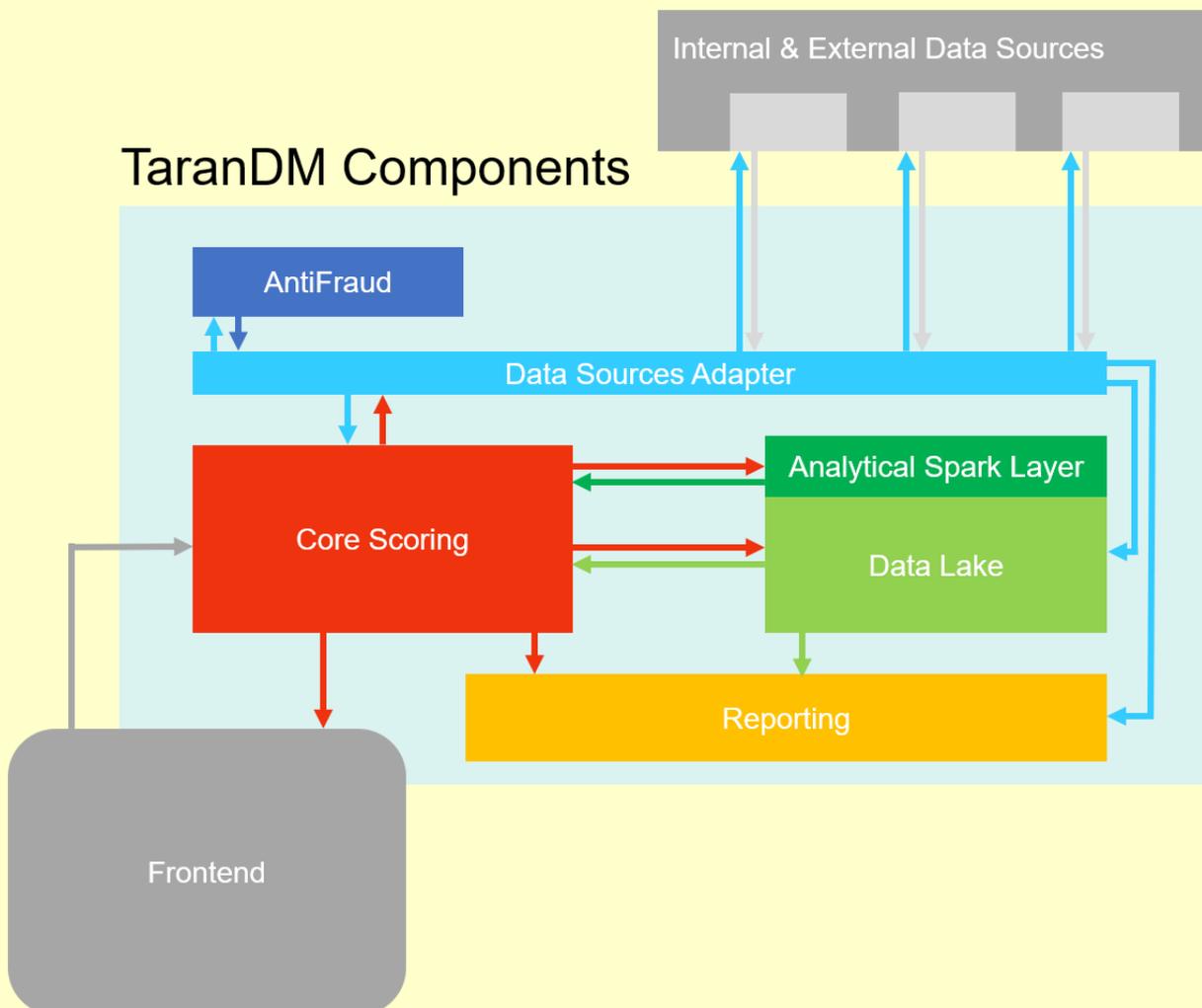
Scoring Database

Scoring process is fully auditable in its inputs, outputs and all intermediate data. Every historical decision is replicable and deterministic. State of each application is continuously stored and workflow can be temporarily stopped (e.g. if user selfie is needed) and later resumed from that particular stage without the need to reconstruct the input data.



TaranDM Technical Overview & Key Components

Programming languages, tools, concepts, standards, and principles which have been used to build TaranDM had been selected with strong emphasis on efficient and high-speed decision processing, scalability, modularity, and flexibility. TaranDM consists of five components, where each of them could be used separately and each of them could be independently built-in and connected with enterprise's current IT solution.



Core Scoring

Strategies are easy to configure and saved in human readable json files. Default strategy templates for various business problems are pre-configured and re-usable. Their adjustment and customization by the business owners is user friendly and maintenance is effortless.

Data Sources Adapter

Data Source Adapter provides platform for two-way communication between the Core Scoring component and any external/internal data source. It uses OpenAPI standard interface specification and thus can handle all types of structured and unstructured data. Data Source Adapter is equipped with cache functionality which enhances speed and reduces costs (e.g. no duplicate requests to a credit bureau).

AntiFraud

AntiFraud is a separate module which has been created to mitigate the fraud risk in real-time online processes. Credit applicant's personal data, device information and fingerprint and IP address data are collected and multiple concentration and cross checks are run in order to prevent fraudulent applications, which often come in a form of a series of requests from a single IP address or the same device. Since AntiFraud is a standalone module, concentration and cross checks can be built on a technology stack suitable for solving a demanding task of real time aggregations.

Data Lake with Analytical Layer

All inputs, outputs and other unstructured and structured data generated during TaranDM run are stored in a highly scalable Hadoop Data Lake. Analytical Spark layer, which is a part of this component, enables data scientists to work in the familiar environment of Jupyter notebooks and develop new models using raw data directly from the Data Lake (as well as other predictors and components of TaranDM). Such models could be easily propagated to the Core Scoring component, including their predictors, and used in production.

Reporting

TaranDM Reporting component uses Kafka to gather structured and unstructured data from different data sources such as Data Lake, scoring run, decisioning, external data sources etc. The data are then processed and stored in structured form in a dedicated data warehouse. Another part of the Reporting component are predefined templates for a BI tool (e.g. Tableau) which are used to visualize stored data. Reporting enables a user to access information about decision process outcome, KPIs, strategies comparison, stability and performance of models, external data sources benefits etc.