
Financial Fraud Detection
with Tibco® Spotfire® & Streambase®
Alycante

Business Scenario

The combination of legislation, market dynamics, and increasingly sophisticated fraud strategies is requiring institutions to be increasingly pro-active in detecting fraud quicker and more effectively.

Transaction surveillance helps to ensure orderly and trustworthy markets, where buyers and sellers participate because they feel confident in the fairness, transparency and accuracy of transactions.

Today's dynamic detection systems need to be agile, scalable and intelligent.

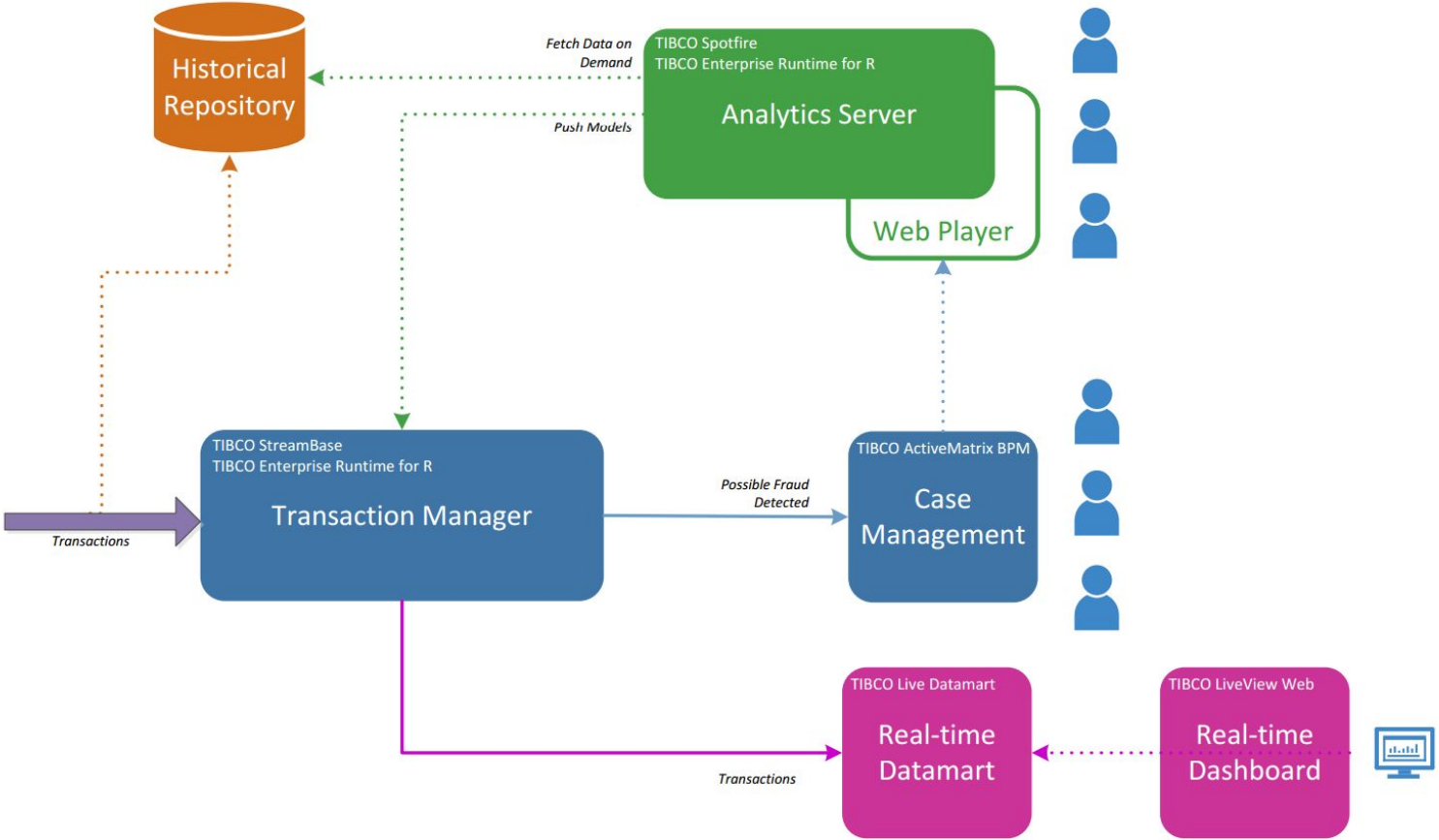
- Agile to adapt to ever evolving compliance regulation.
- Scalable to deal with ever increasing transaction volumes.
- Intelligent to detect increasingly sophisticated fraud patterns, and also to reduce false positives in the alerting stage.

Coupled with the detection system, an essential component is Case Management.

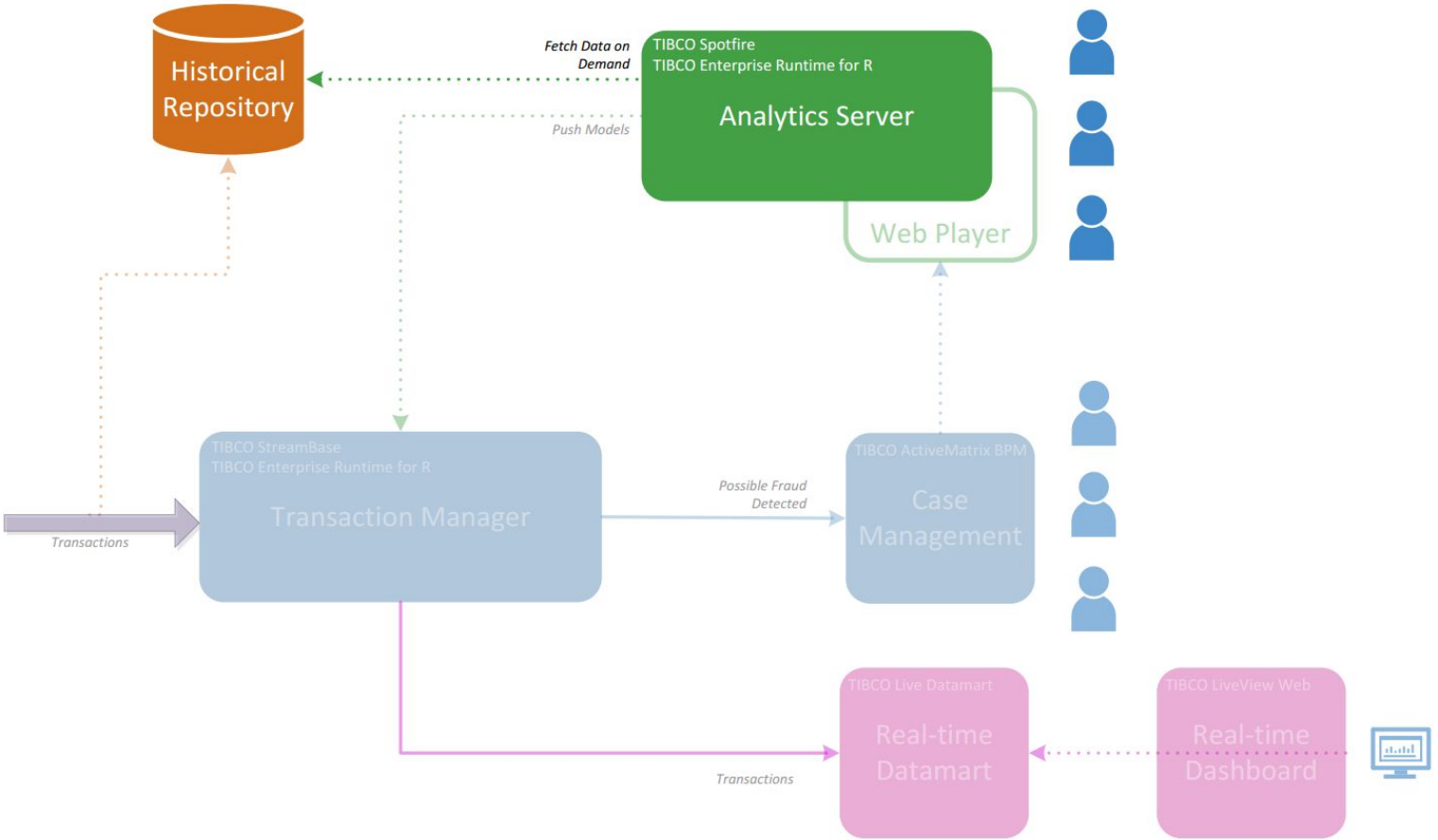
Taking an alert from when it is raised, through the investigation stage and if applicable on to the regulator requires a capable and integrated Case management tool, which can capture all the inputs from historical data, surveillance analyst notes and transaction log extracts.

Alycante

Financial Fraud Component Diagram



Stage 1 & 2: Adhoc Data Discovery and Model



Stage 1 & 2: Adhoc Data Discovery and Model

In Stage 1, analysts use the historical data repository to analyze previous transactions.

This need not be a single repository but could be a variety of sources that can be then synthesized and analysed within the Analytics Server.

During Stage 2 the analyst uses the data obtained to build a predictive model that can be used to classify and score transactions based on fraud. This is done using a combination of Spotfire and TERR. These could be a combination of supervised and unsupervised models with varying thresholds for escalation.

Alycante

Stage 1 & 2: Adhoc Data Discovery and Model

Instructions

These plots show additional ways of looking at how well your model is performing. All results regard the randomly chosen test set, which correspond to 20% of historical data held aside for model evaluation. This is therefore an impartial assessment of the quality of the model.

The Cumulative Gain Chart measures how well the model can identify likely Fraud cases in the transaction pool. The Y-axis shows the fraction of existing Fraud that gets detected; the horizontal axis shows the proportion of total transactions that must be investigated to detect the different levels of fraud. A curve that rises well above the diagonal line indicates an effective model. Similarly for the Precison and Recall Chart. **If you do not know much about data science, this is all you need to check**. We present more metrics for information purposes.

The ROC ("Receiver Operating Characteristic") curve measures the TruePositive and False Negative rate, which are overall measures of the model accuracy. A curve that approaches the top-left corner (close to the green lines) is considered a good model. A curve approaching the red diagonal is said to be no better than tossing a coin.

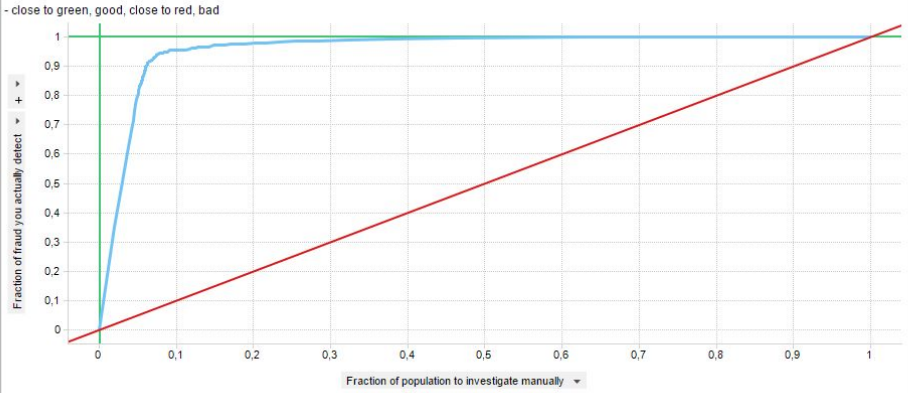
The Confusion Matrix shows how many cases in the test set had a prediction equal to Reality and how many did not. We want high values on the diagonal and low elsewhere.

The Maximum Attainable F1 table gives the value of Threshold that leads to, as the name suggests, maximum F1 results. F1 is an average between Precision and Recall and is a commonly used model quality metric. The threshold is the value of predicted probability of fraud from which we consider a line in the data to require manual investigation. **You can replace F1 with your own calculation of model quality** by right-clicking and selecting Properties, then clicking on Data and below, under Limit data using expression, replacing [F1] with your own metric.

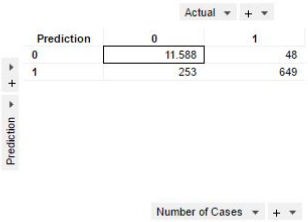
The Optimal Threshold Chart allows seeing the behaviour of an overall quality metric given a choice of threshold. Again, **you can replace F1 with your own calculation of model quality** by right-clicking on it and selecting Custom Expression. When you mark a point on this curve, you can see what fraction of the transaction pool is expected to be selected for investigation in the adjacent chart.

The bar plot on the middle right splits the transaction pool into groups based on their predicted risk, and looks at how well the actual fraud rates in each group is predicted by the model. If the bar groups are both crescent, this builds confidence in the model.

How easily can you catch a lot of fraud? (Cumulative Gain Curve)

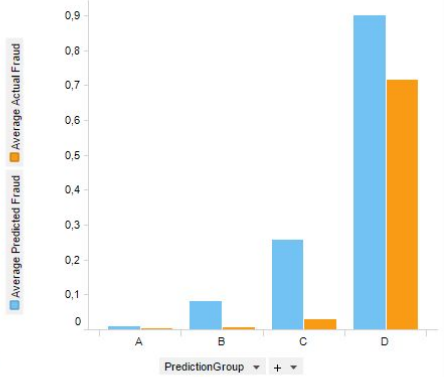


Confusion Matrix



How accurately does the model find the correct Fraud in ea..

Good when orange and blue bars evolve together



ROC Curve

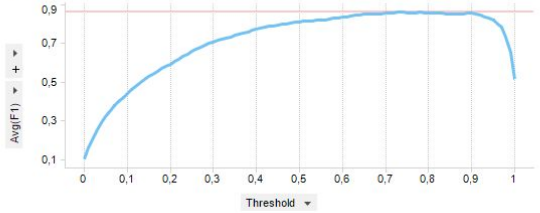
Precision / Recall Chart

Maximum attainable F1

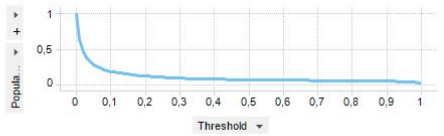
Threshold	F1
0,73	0,86

Optimal Threshold Chart

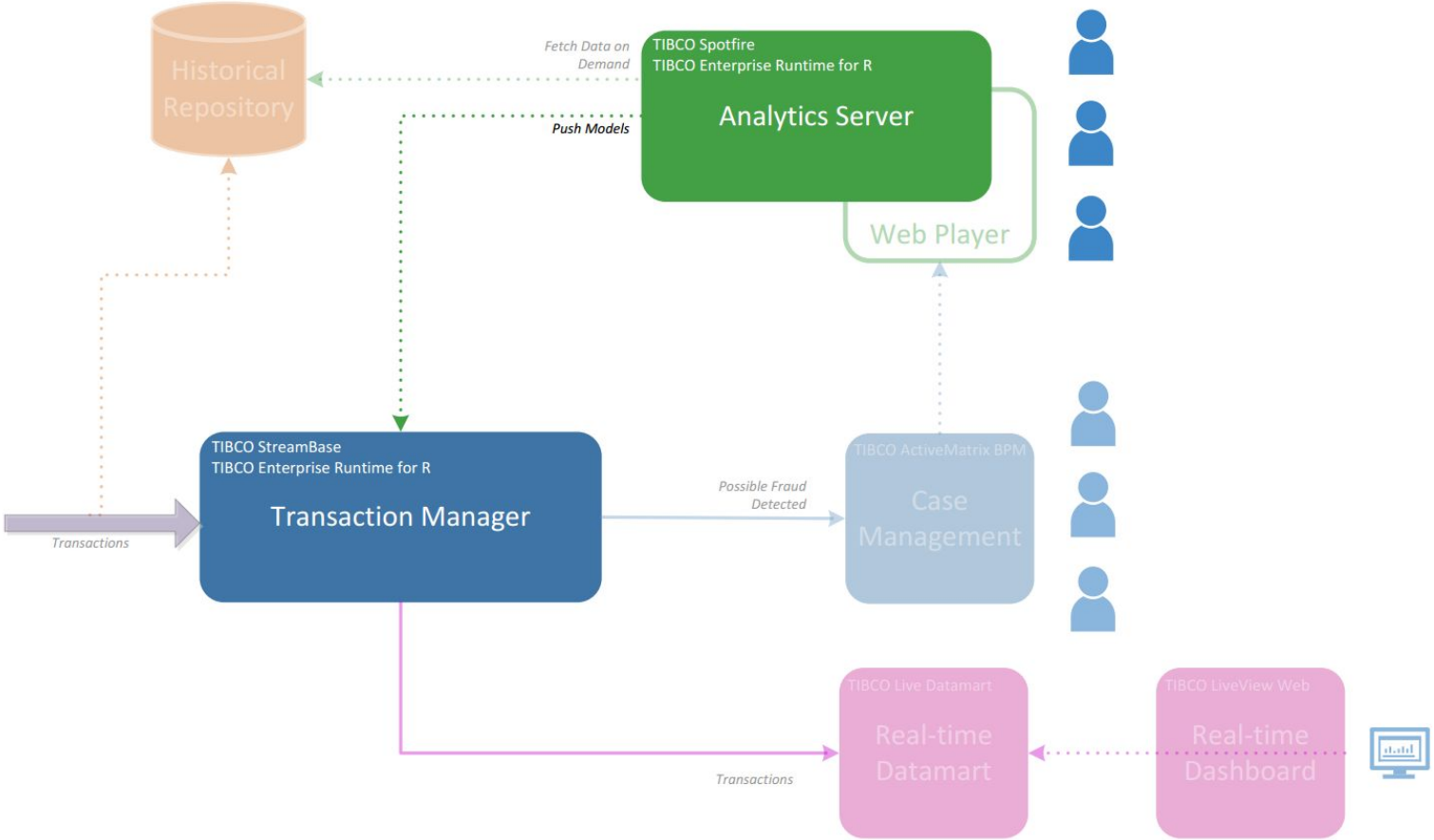
You can replace F1 with your own calculation in the Y-axis.



Proportion of Cases Flagged per choice of Threshold



Stage 3: Deploy Models

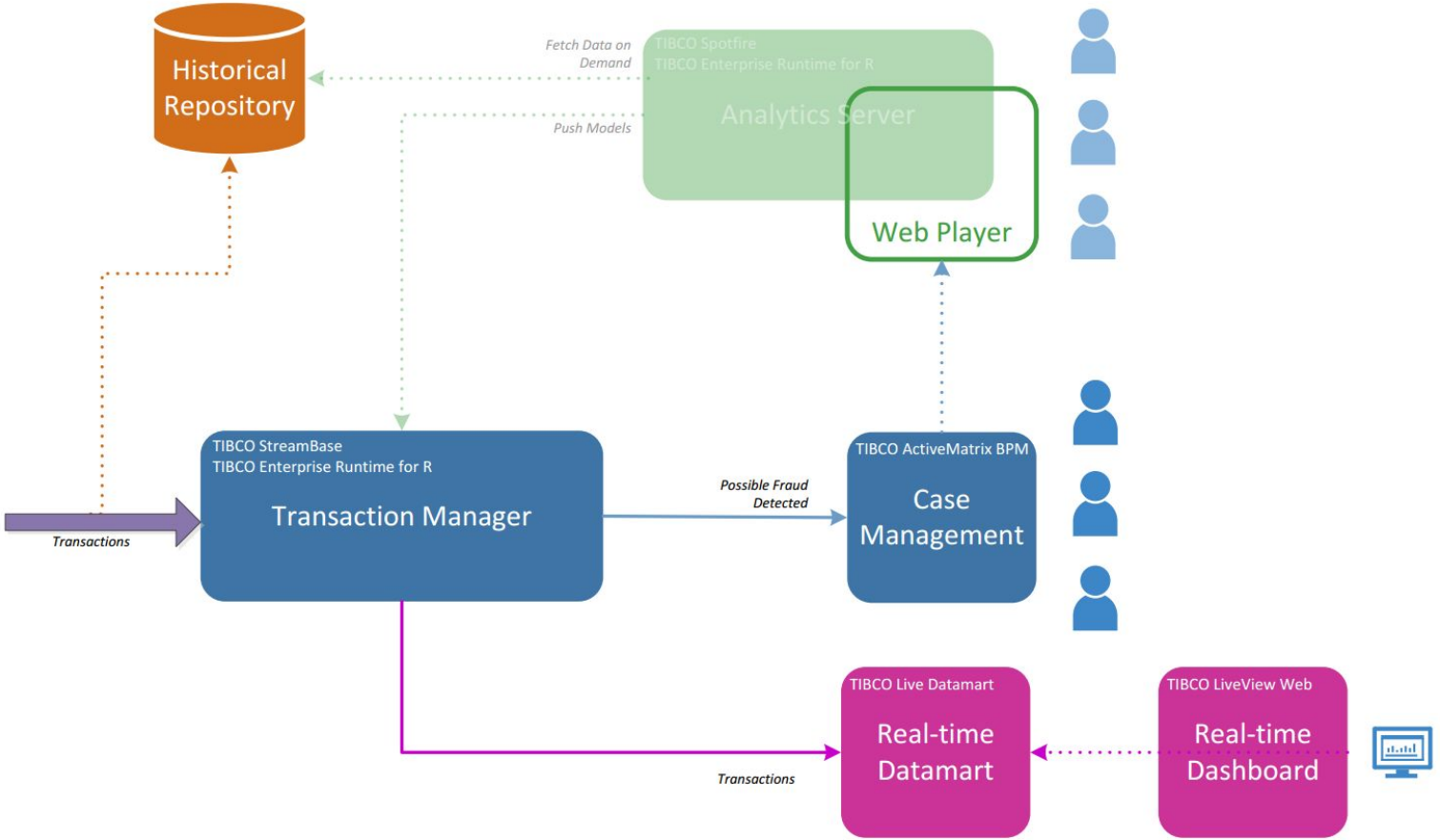


Stage 3: Deploy Models

In Stage 3 the models are hot deployed into the Transaction Manager by an analyst using the Analytics Server.

Once deployed the models are then available immediately for new transaction classification.

Stage 4: Classify and Investigate



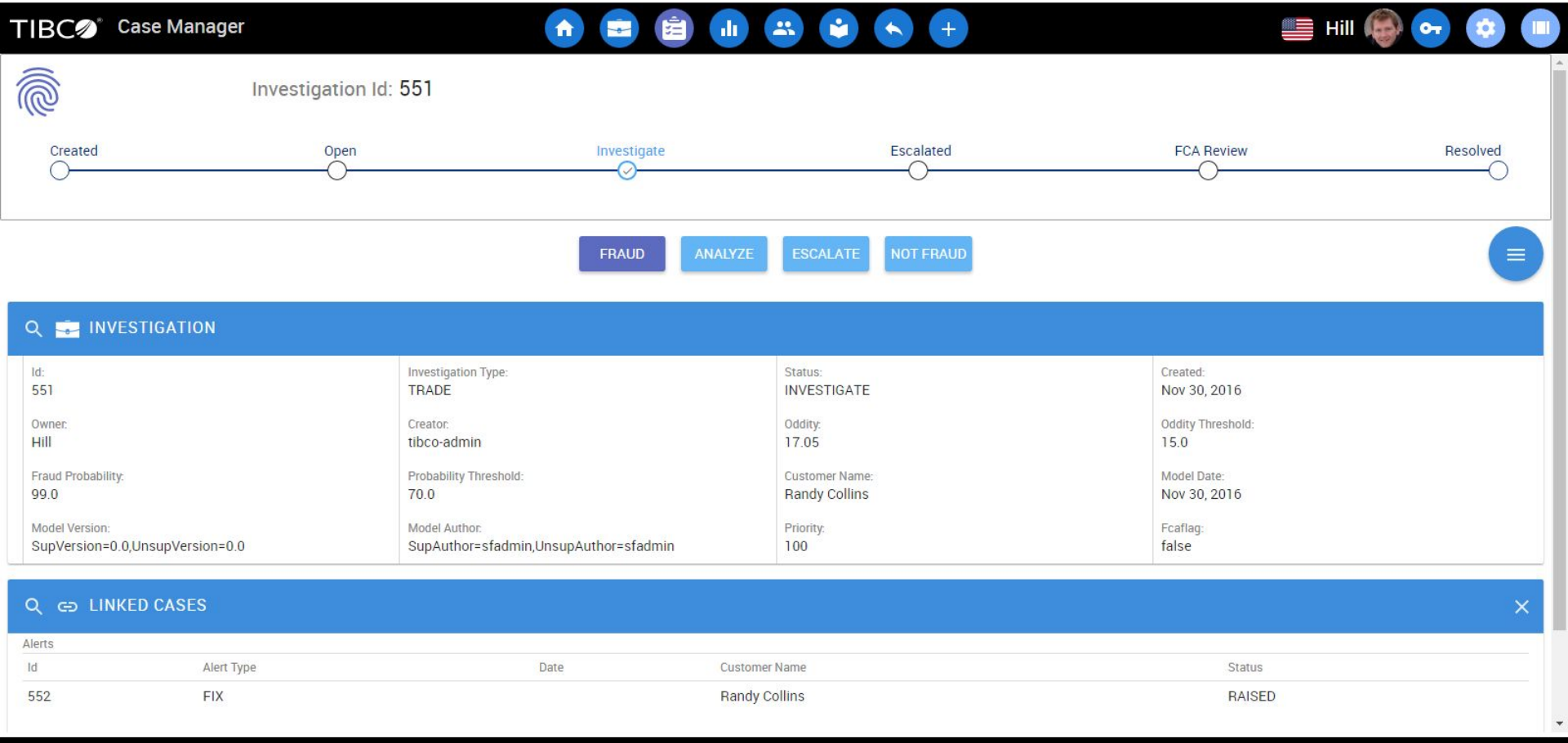
Stage 4: Classify and Investigate

In Stage 4 transactions are flowing through the system in real-time.
They are being stored in the Historical Repository and also being scored by the models within the Transaction Manager. The results of this scoring is flowed through to the Real-time Datamart for display on the Real-time Dashboard.

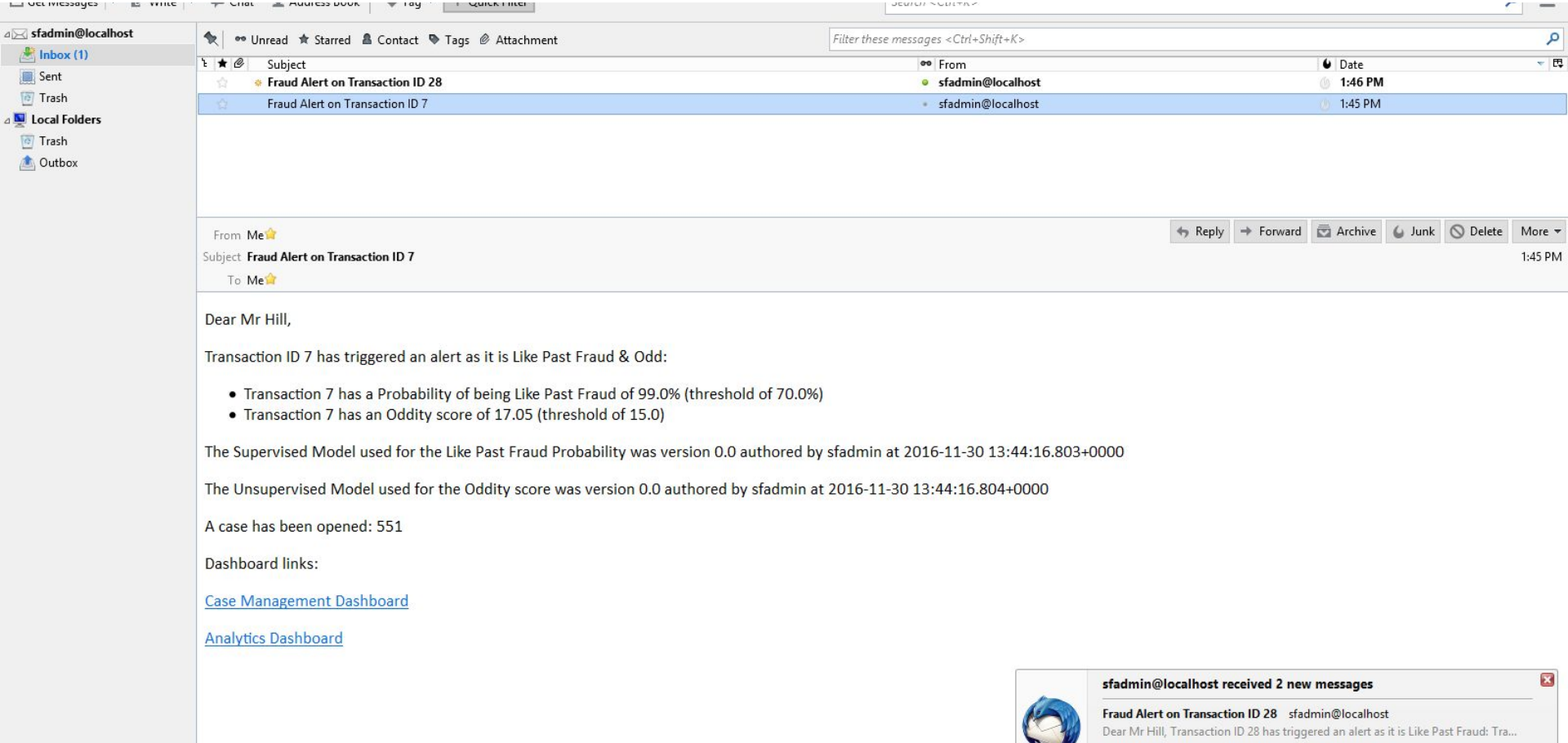
If the models indicate that a transaction is possibly fraudulent, then the Transaction Manager will route the transaction to Case Management for further investigation. The investigators can make use of the data available in the Analytics Server through the Web Player component. The results of the analysis will be recorded for future use.

Alycante

Case Manager: Investigation



Case Manager: mail alert



LiveView: real time monitoring

Benefit and Business Value

Faster time to resolution

- Data discovery, statistical model creation and integration within a single tool
- Dashboards that integrate data across databases (historical and reference data) alongside information from the real-time processing, users can filter, slice and dice, zoom in and out of data in order to determine if the alert needs further investigation or can be discarded as a false positive

Accelerate the adoption and rolling out of surveillance

- Surveillance projects can span multiple assets, data and scenarios that can go from monitoring abuse in cash equities to complex schemes involving foreign exchange (FX) and derivatives
- With connectivity to over 150 options, including Bloomberg, BM&FBovespa, Currenex, EBS, FIX, FXall, Hotspot, Interactive Data, and Thomson Reuters, all your data feeds can be included
- A graphical flow based development reduces complexity and increases collaboration

Flexible Surveillance Scenarios

- The predefined abuse scenarios adhere to the specifications published by regulators; however it is common that each institution needs to tailor not only the thresholds that trigger the alerts but also exclude certain events from the analysis altogether
- Perform correlation to detect complex potential abuses on one asset type attempting to influence a related instrument of another type.

Alycante

Financial Fraud Detection, a reference architecture that shows how Predictive Analytics, Streaming Analytics and Business Process Management can collaborate to support the full livecycle of end-to-end fraud detection in financial organizations

TIBCO™

Alycante