



From Insight to Action:
Analytics from Both Sides of the Brain

Vaz Balasingham

Director of Solutions Consulting

vbalasin@tibco.com


INTERCONNECT
EVERYTHING




AUGMENT
INTELLIGENCE

Insight to Action – from Both Sides of the Brain


Value



Grow Revenue



Reduce Risk



Increase Productivity
ROI

- Both Sides of the Brain
 - Fast & Slow
- Insight to Action
 - Visual Analytics
 - Numerical Algorithms
 - Insight Execution
- Wrap-Up / Questions
 - Come see the demos



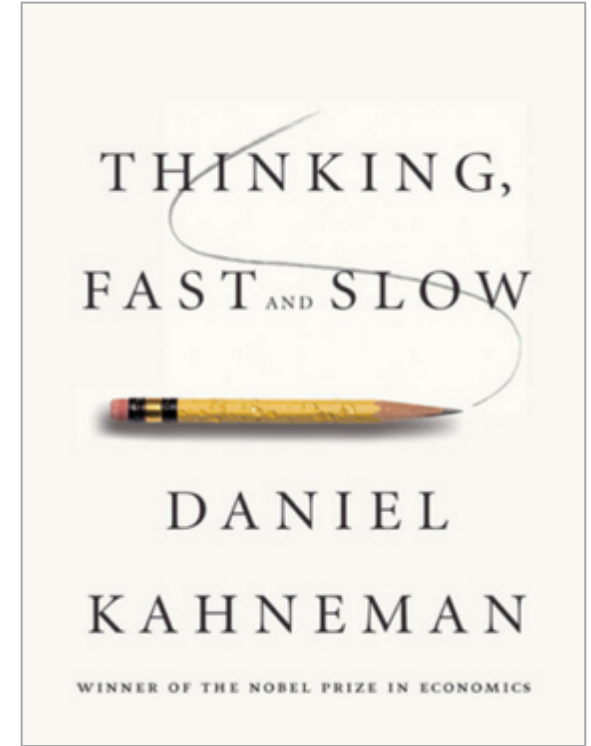
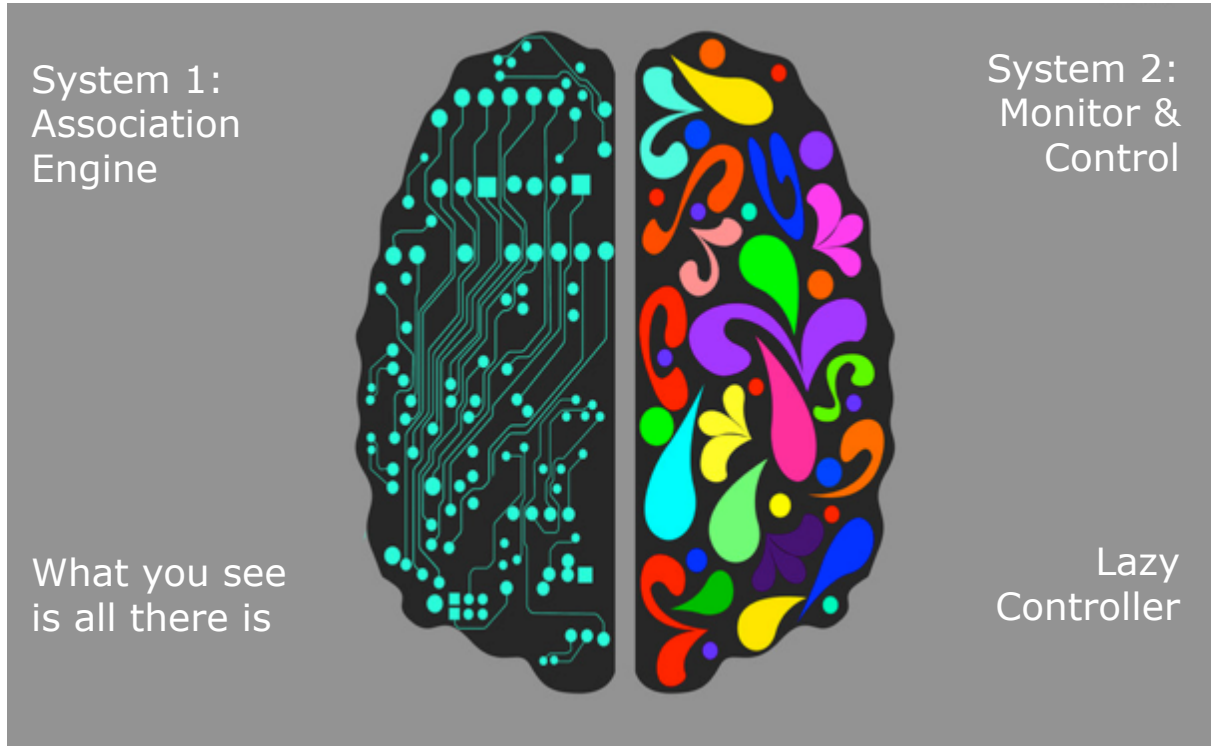
Fun Problem from Kahneman



“A bat and a ball cost \$1.10 in total. The bat costs \$1.00 more than the ball. How much does the ball cost?”

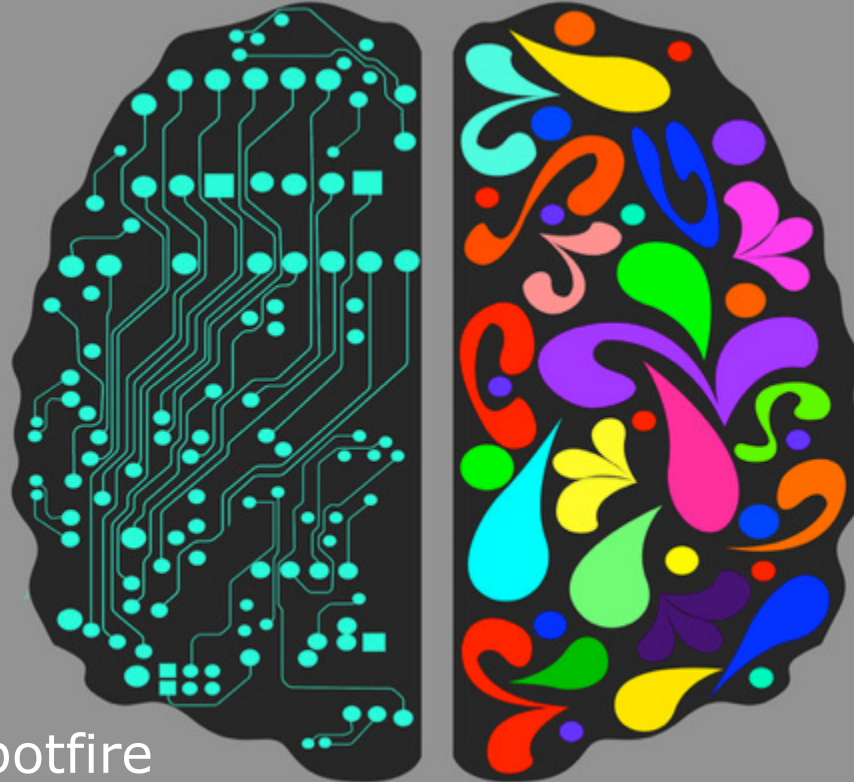


Themes: Thinking Fast and Slow... from Both Sides of the Brain



Themes: Thinking Fast and Slow... from Both Sides of the Brain

System 1:
Association
Engine



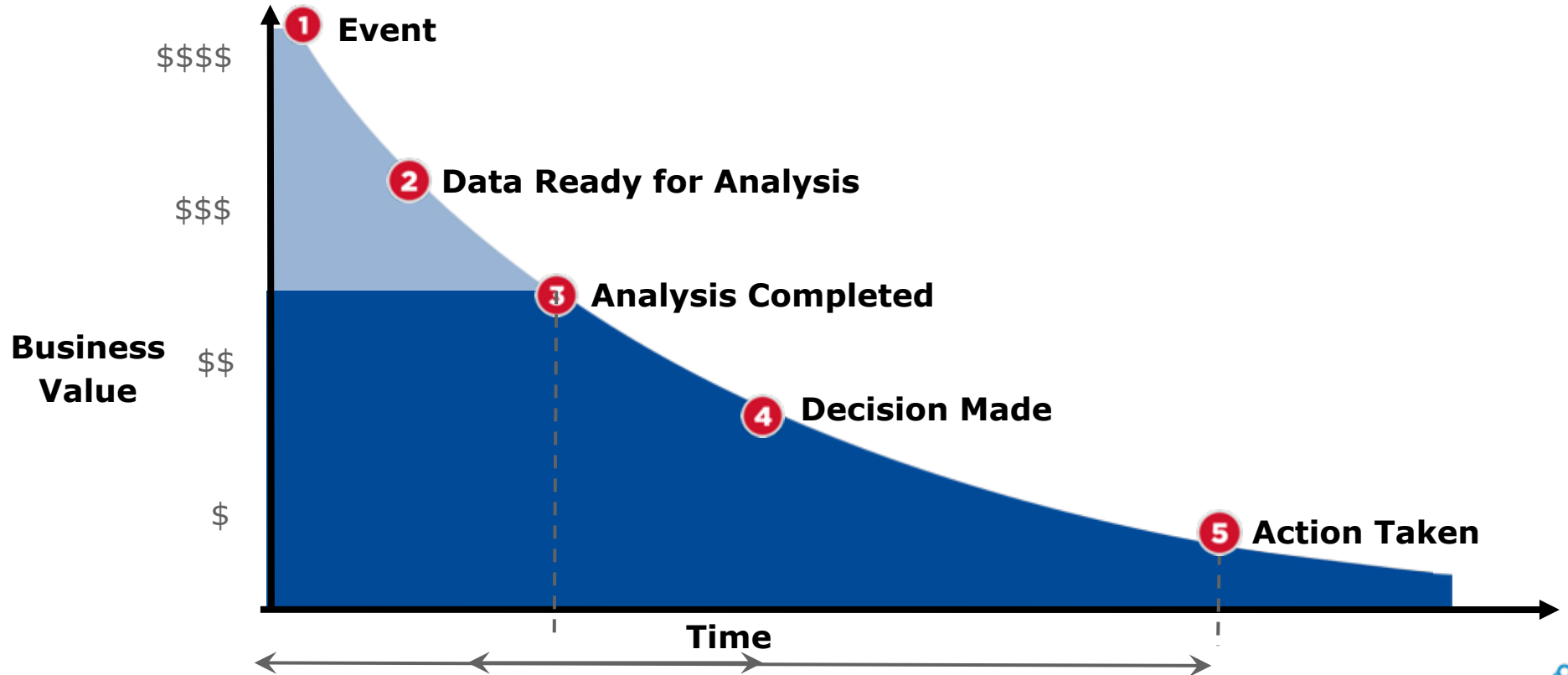
System 2:
Monitor &
Control



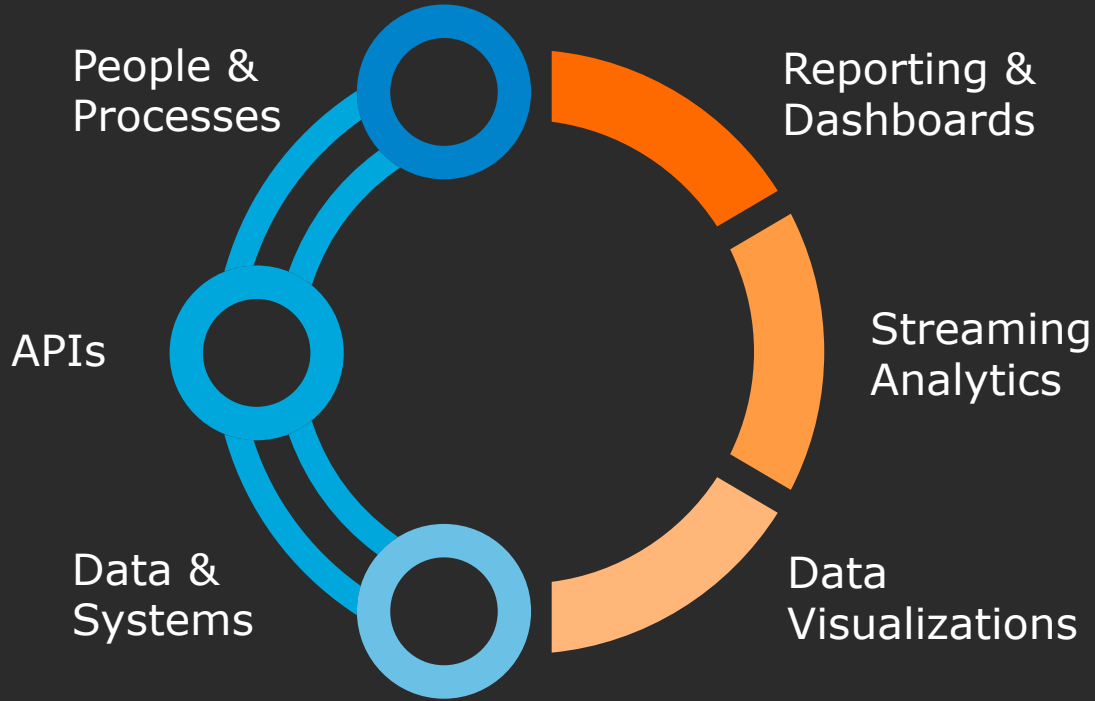
Spotfire



Themes: Insight without Action has Little Value



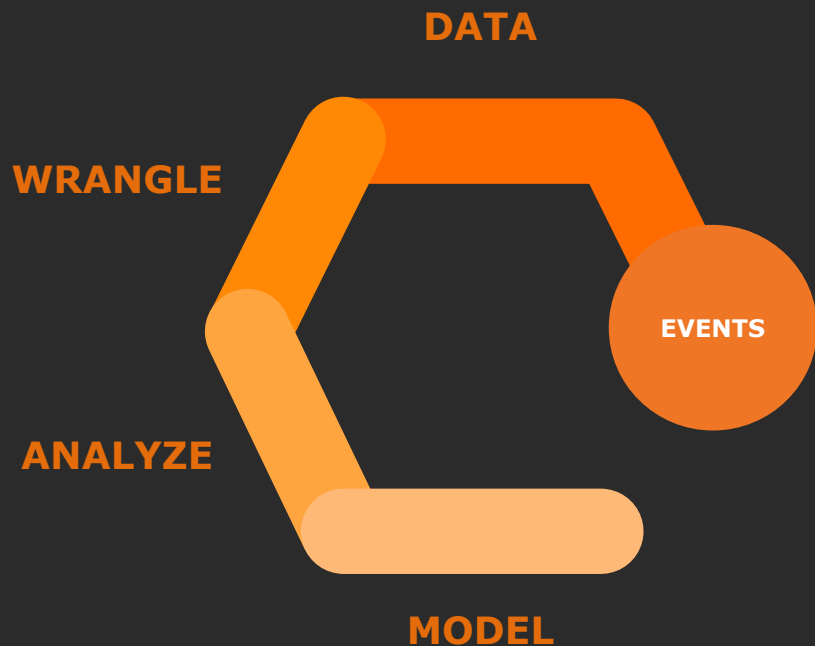
INTERCONNECT EVERYTHING



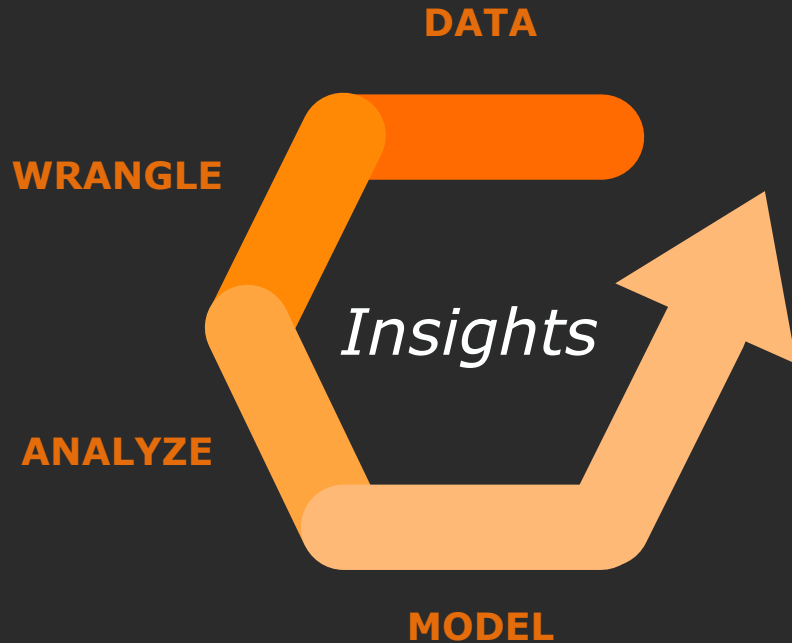
The TIBCO Insight Platform



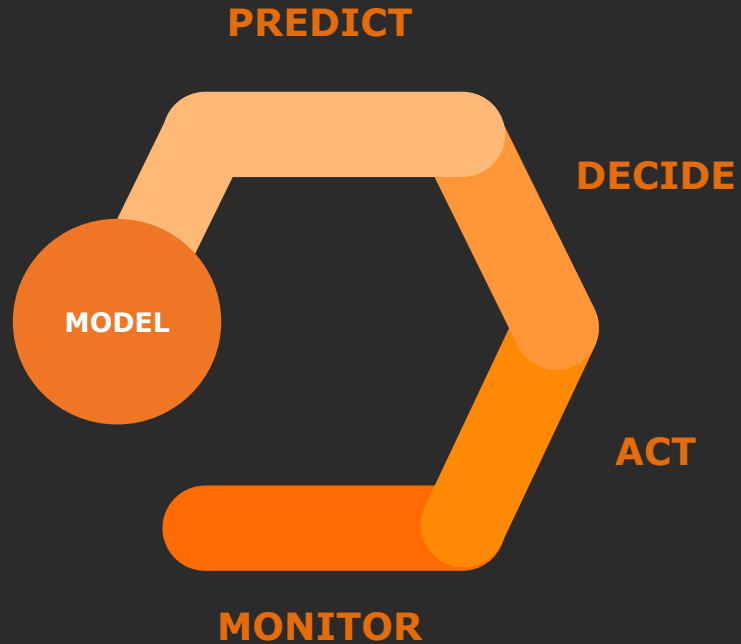
The TIBCO Insight Platform



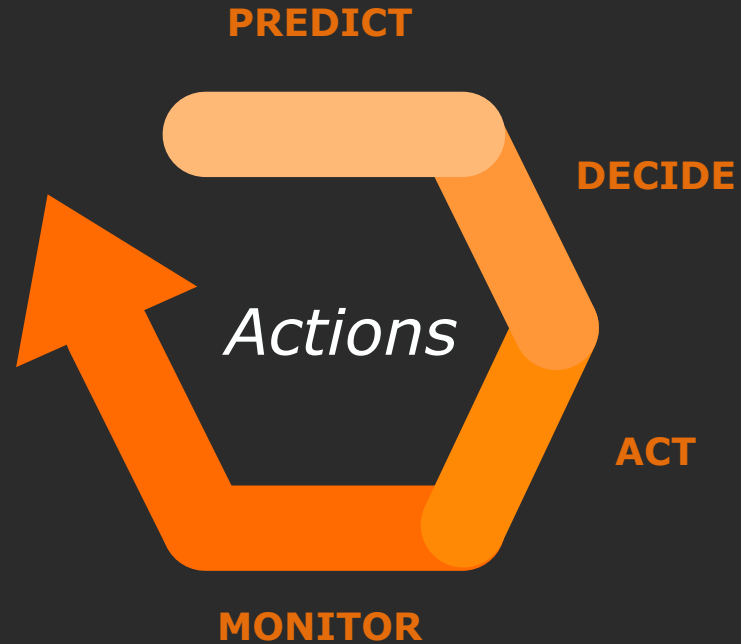
The TIBCO Insight Platform



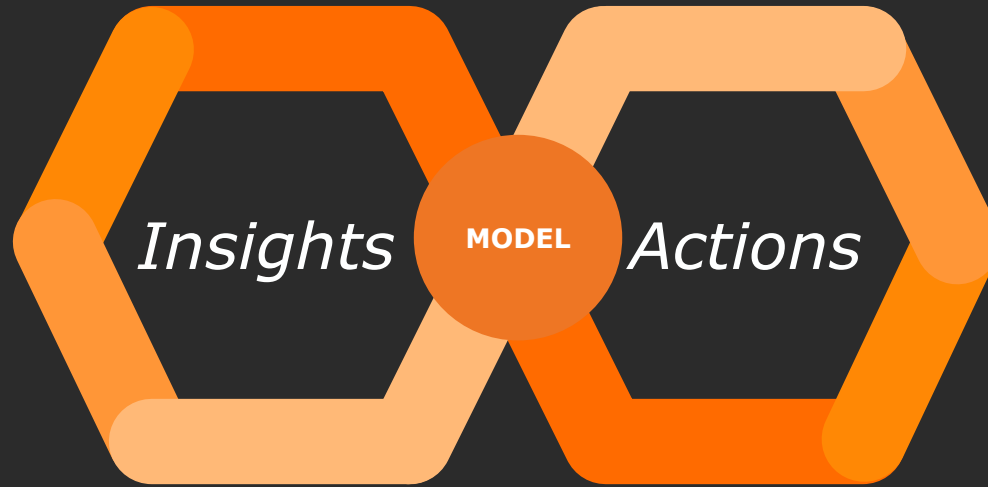
The TIBCO Insight Platform



The TIBCO Insight Platform



The TIBCO Insight Platform

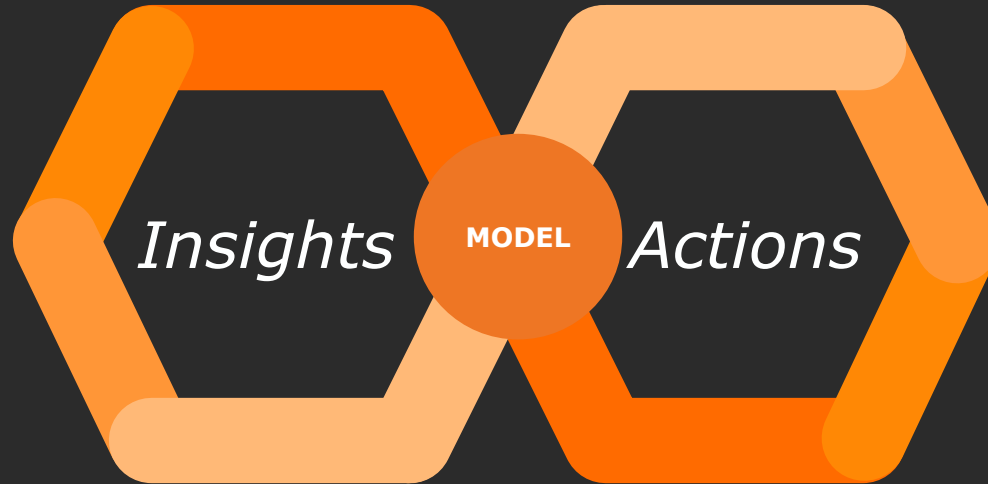


The TIBCO Insight Platform

TIBCO Spotfire

Matlab

TIBCO Streambase



TIBCO Live Datamart

Smart Visual Analytics

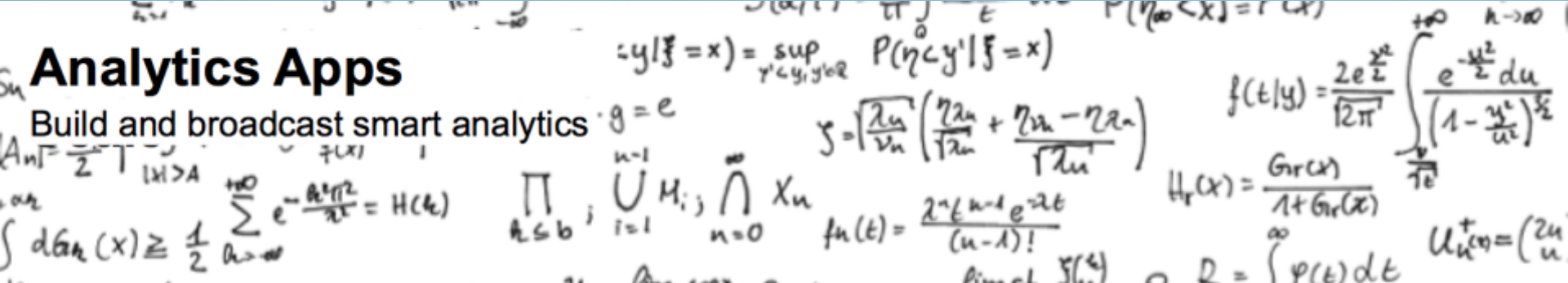
Be first to insight, first to action



Visual analytics is like a bicycle for your business mind.

Analytics Apps

Build and broadcast smart analytics



Streaming Analytics

Continuous algorithmic awareness and automation



#1. Smart Visual Analytics

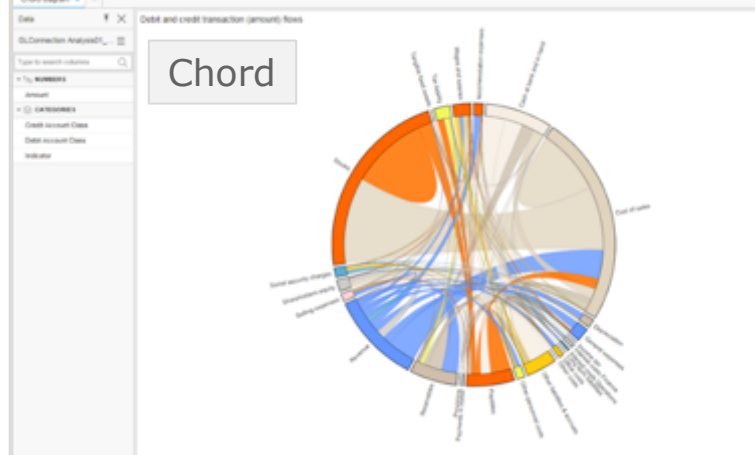
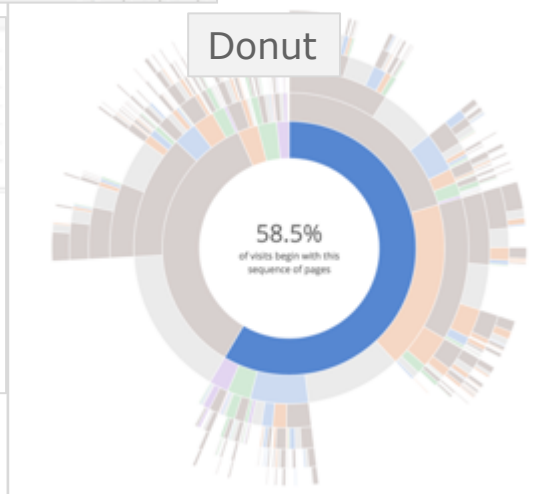
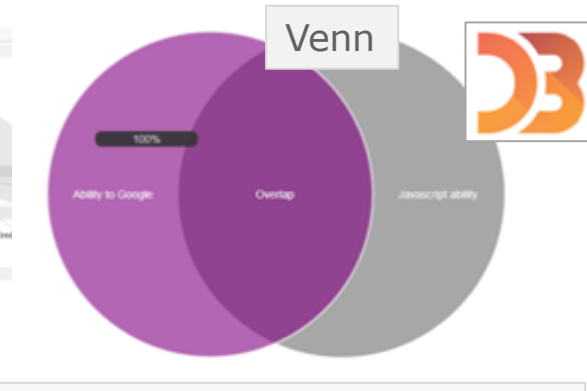
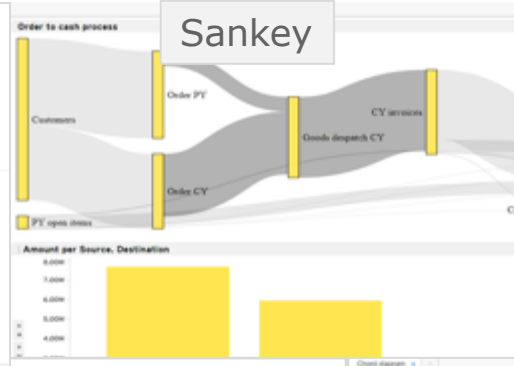
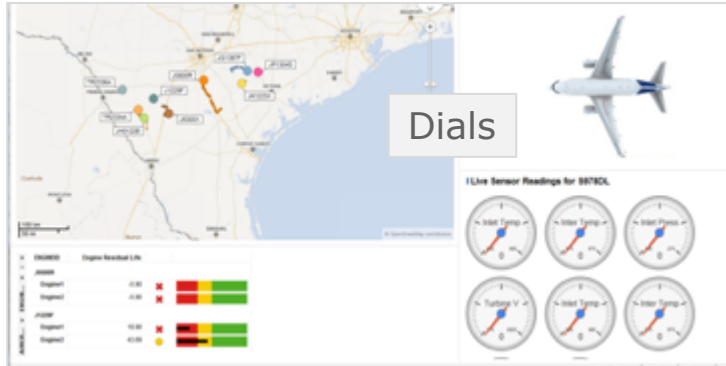
Smart Visual Analytics

Be first to insight, first to action



Visual analytics is like a **bicycle** for your **business mind**.





Connect Seamlessly with your Data Sources

- Hassle-free connectivity to your data sources
- Database drivers included with Spotfire software download



Amazon **Redshift**

salesforce

Spark  SQL

Recommended visualizations

E N Cen

Type to search columns

123 NUMBERS

- DeliveryTime
- Discount
- ForecastQuarter
- GrossProfit
- Lat
- Lon
- SalesTotal

TIME

- DeliveryDate
- ForecastMonth
- ForecastYear
- OrderDate

CATEGORIES

- BusinessUnit
- CustomerName
- InventoryStatus

Clear selections

DeliveryTime – DeliveryDate

BusinessUnit

DeliveryTime per BusinessUnit, DeliveryDate

DeliveryTime DeliveryTime

Office Supplies		Technology		Furni...
2003	2002		2002	2003
	2004		2003	2002
	2001	2001	2001	20 04 20 01

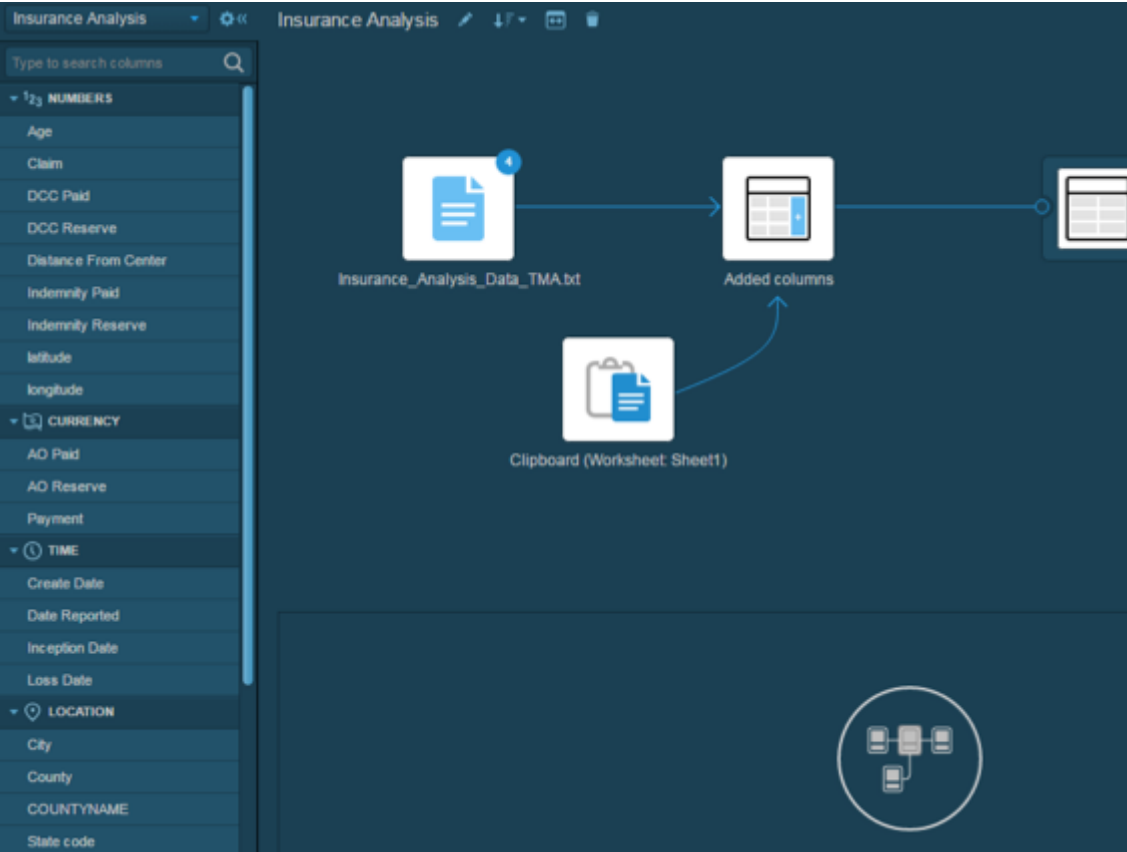
DeliveryTime per DeliveryDate

BusinessUnit

Jump Start Your Analysis

- Visualization recommendations based on your selected data
- Build one visualization – or entire dashboard – in just a couple of clicks
- Data previews accelerate discovery & insights
- Powered by Spotfire Analytics Recommendation Engine™





Visual Overview of data table sources & operations

Gain a better understanding of the data you're trying to visualize.

Speed up visualization design.

8/27/2007	\$12769.00	Hall
9/26/2007	\$3080.00	Hall
10/15/2007	\$5450.00	Hall
12/2007		Hall
1/1/2008	\$10000.00	Hall
1/30/2008	\$7853.00	Hall
2/5/2008	\$11200.00	Hall

The screenshot shows the 'Data Panel' in TIBCO Spotfire for a visualization titled 'Foreign Aid Analysis'. The panel is divided into several sections:

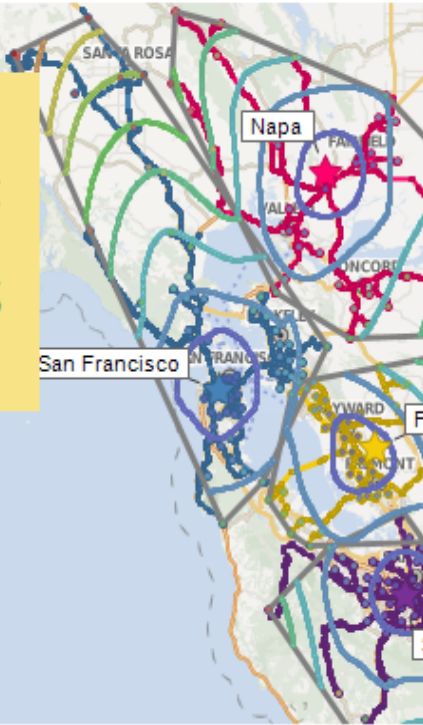
- Column List:** A list of columns on the left, with 'dac_purpose_name' selected and highlighted in blue.
- Summary:** A circular gauge shows a 'Unique count' of 186.
- Data Type:** A dropdown menu is set to 'Categories', and the 'Data type' is set to 'String'.
- Sort Order:** A dropdown menu is set to 'Standard alphabetical'.
- Most Common:** Lists 'Multisector aid', 'Administrative costs', 'STD control including HIV...', and 'Material relief assistance a...'.

Most common	Least common
Multisector aid	Rescheduling and refinancing
Administrative costs	Refugees in donor countries
STD control including HIV...	Relief of multilateral debt
Material relief assistance a...	Oil-fired power plants
- Least Common:** Lists 'Rescheduling and refinancing', 'Refugees in donor countries', 'Relief of multilateral debt', and 'Oil-fired power plants'.
- Click to sort:** A list of categories including 'Security system management ...', 'Public sector policy and admini...', 'Multisector aid', 'Sectors not specified', 'STD control including HIV/AIDS', 'Transport policy and administr...', 'Electrical transmission/ distribu...', 'Narcotics control', 'Legal and judicial development', 'General budget support', 'Financial policy and administrat...', 'Material relief assistance and s...', 'Energy manufacturing', 'Energy policy and administrativ...', 'Oil and gas', 'Food aid/Food security progra...', and 'Nuclear power plants'.

Key Information At Your Fingertips

- See all dimensions and measure available to visualize.
- Expanded view gives extra insights without going to other screens.
- Change data type, category, sort order all without dropping what you're doing.
- Data filters built-in.
- Views based on type of data in column.

We opened the Fremont Distribution Center this past September, which reduced our longest drive time by 23 minutes.



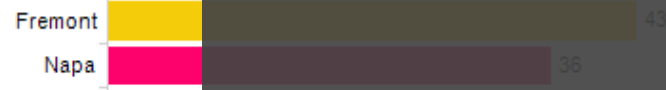
Add Commentary To Your Analysis

- Overlay annotations on any part of your analysis.
- Select size, color, and font to match the look and feel of the rest of your visualizations.

Distance



Drive Time



Connect Your Data To The Conversation

- Have conversations about specific issues and insights found in your data.
- Save your views and discoveries on each message in the conversation.

The screenshot displays the TIBCO Spotfire interface. On the left, a map shows several data points (blue and yellow circles) plotted on a street grid. A chat window is open over the map, showing a conversation between users 'jflint' and 'amckay'. The chat messages are:

- jflint 2:15 PM:** I'm on it. I've already alerted Adam McKay. It seems like most of the damage was in this particular neighborhood. We'll deploy people today.
- amckay 3:29 PM:** Thanks John, I'll get on this right away!

Below the chat messages, there is a text input field containing "Can we also alert" and a checked checkbox labeled "Include analysis state in the reply". At the bottom of the chat window are "Reply" and "Cancel" buttons.

On the right side of the interface, a table titled "Claims" is visible. The table has three columns: "Claim", "Loss Date", and "Reported Date".

Claim	Loss Date	Reported Date
600212	4/1/2014	5/19/2014
700205	7/19/2015	7/20/2015
700207	7/19/2015	7/20/2015
	7/19/2015	7/24/2015
	7/19/2015	7/27/2015
	7/17/2015	7/27/2015
	7/19/2015	7/30/2015
	7/17/2015	7/30/2015
	7/19/2015	7/30/2015
	7/19/2015	7/31/2015
	7/19/2015	8/1/2015

TIBCO GeoAnalytics

WHAT IS IT?

- High-accuracy, global base maps designed for data exploration & location analytics
- Visualize, explore, analyze data in the context of location
- Modern, interactive and easy map navigation
- Mash-in new data sources quickly and provide accurate enterprise geo-coding
- Use multi-layered analysis to understand geographic correlations in data and expand understanding

QUICK DETAILS

- Automatic, offline geo-coding
- Combine multiple layers of data on a same map
- Import & save geographic files for future use
- Cloud based base map always available, always up-to-date
- Drill-down to a deeper level of detail

FOR WHOM

- Analysts
- Data Scientist
- Business Users

MATURITY CURVE STEP
Diagnose (foundation for
Predict & Optimize)

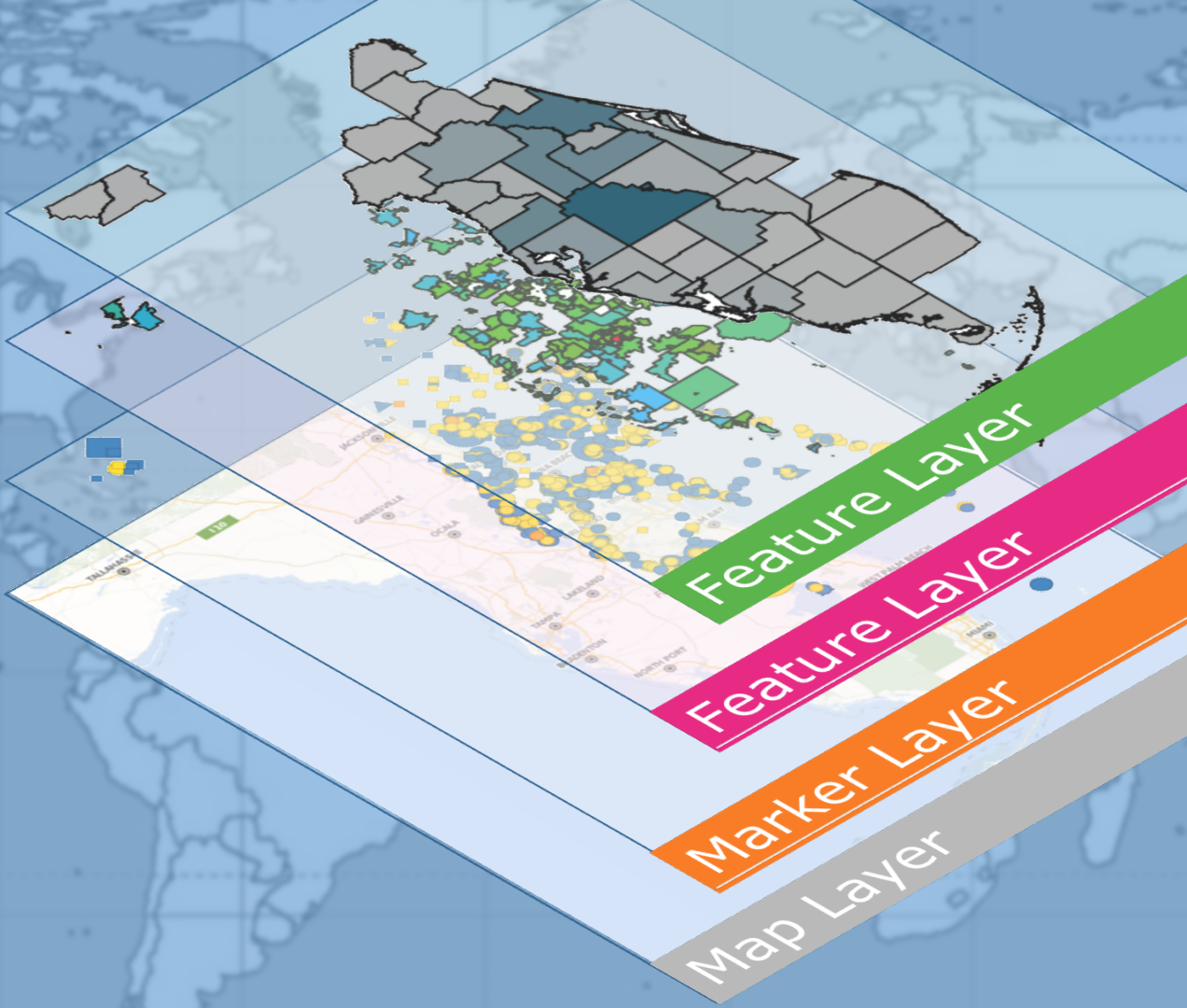
@@Data Layers

- Stores locations
- Market performance
- Catchment areas

@@Map Layers

- Maporama Layer



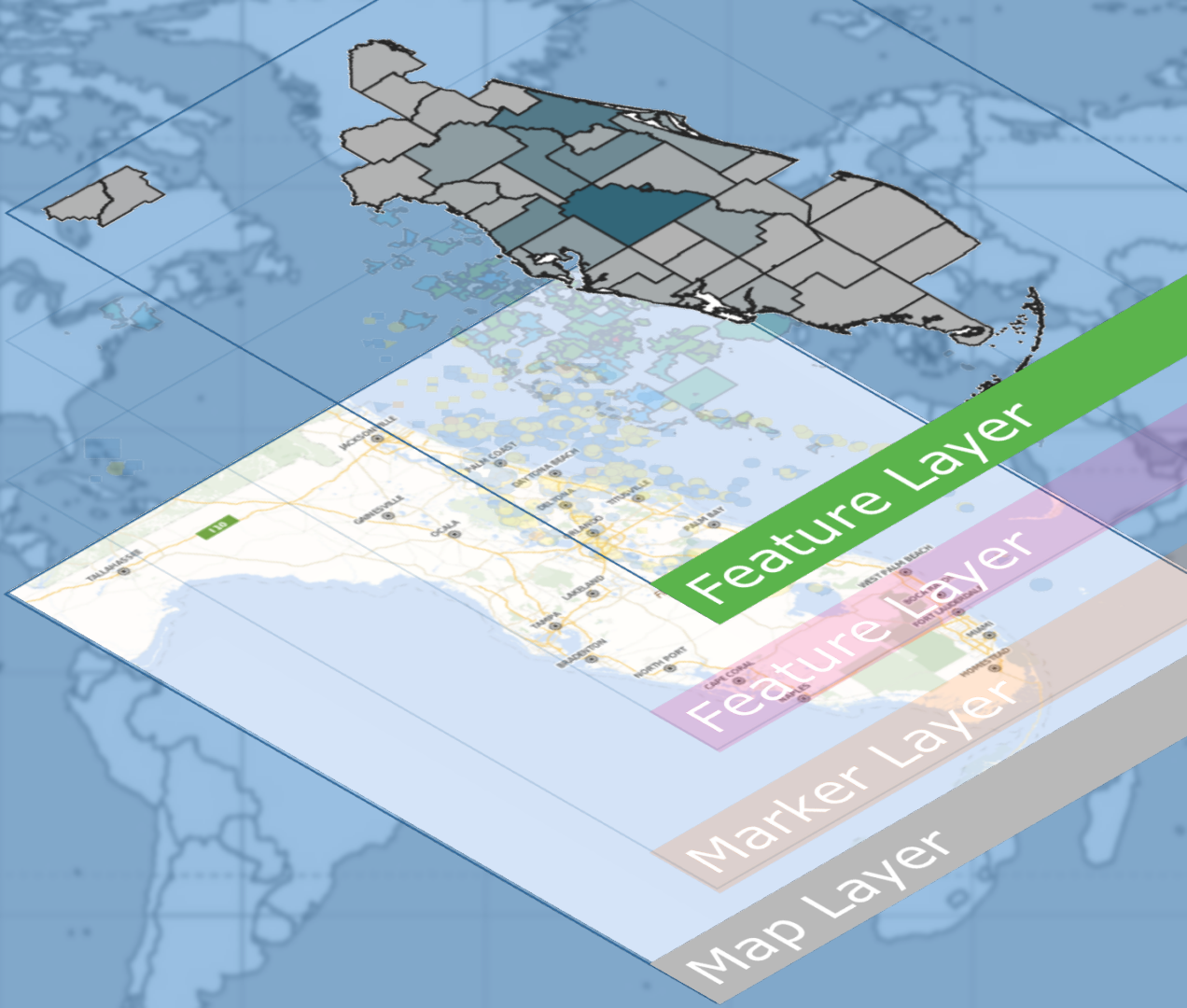


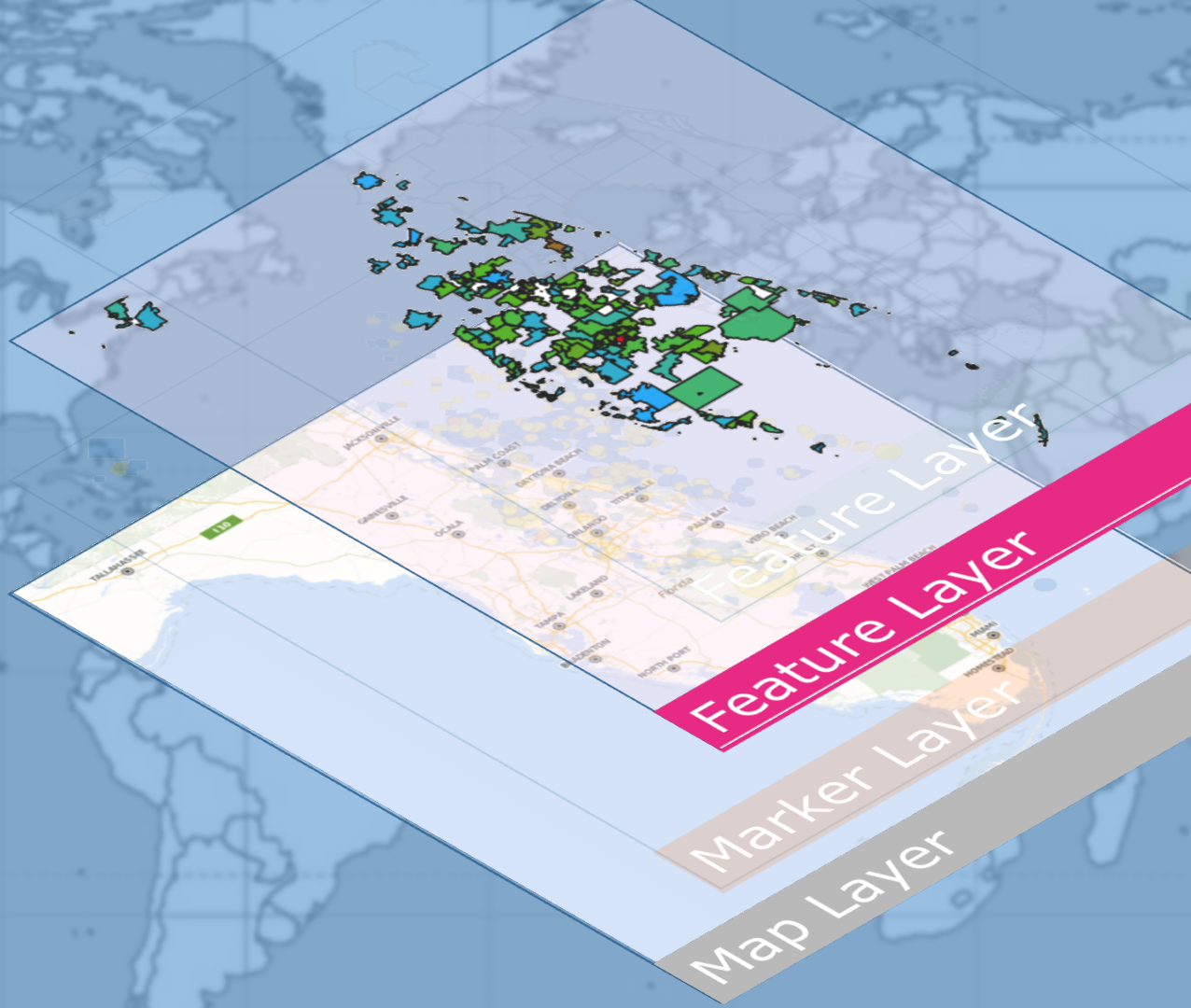
Feature Layer

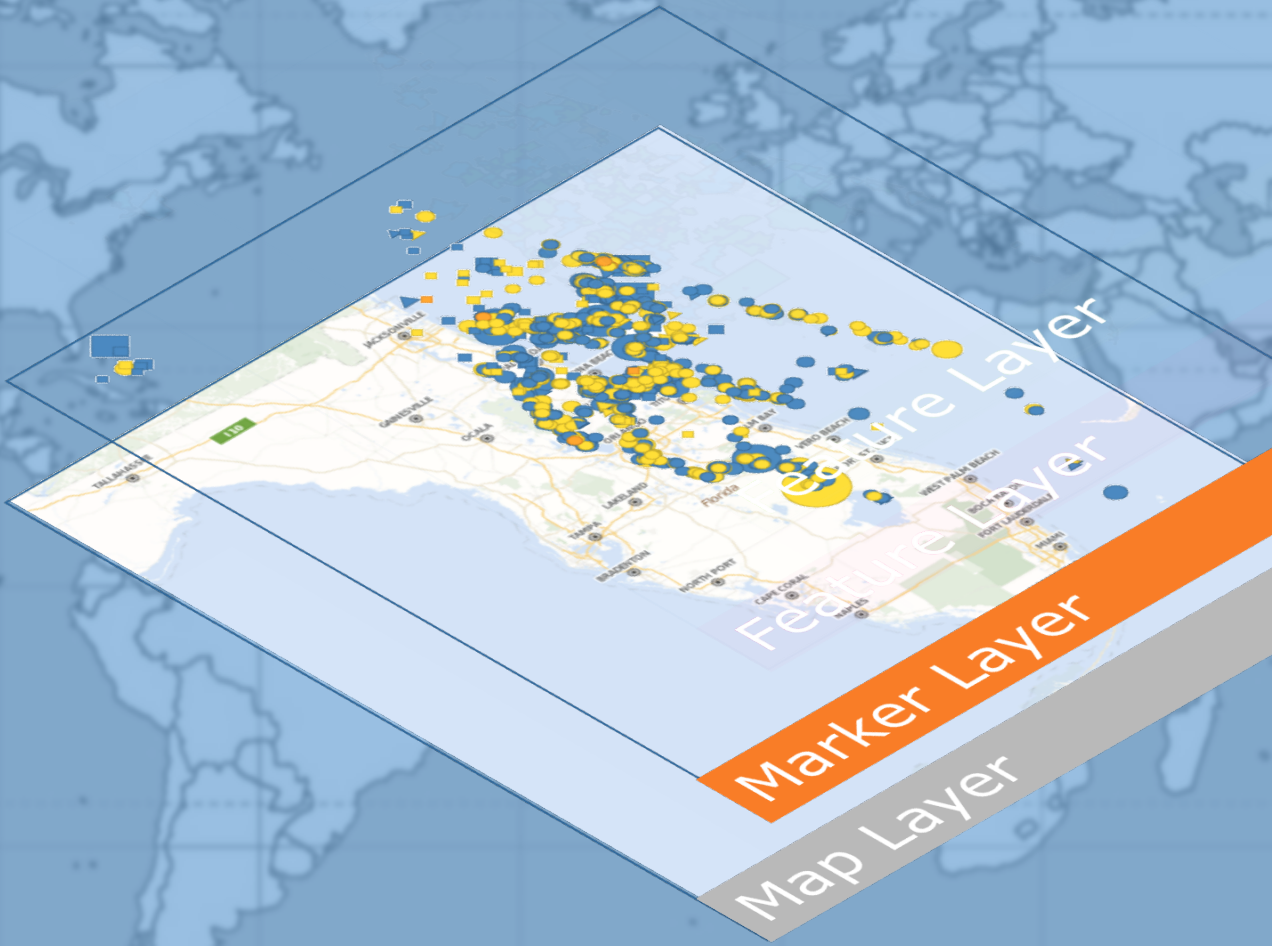
Feature Layer

Marker Layer

Map Layer





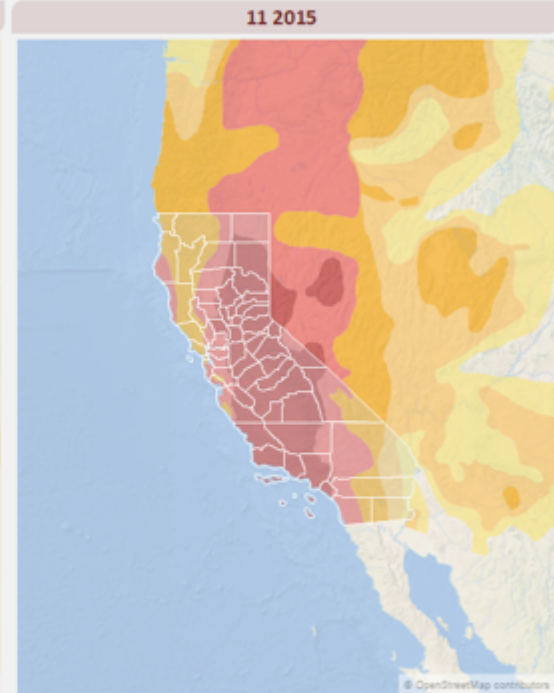
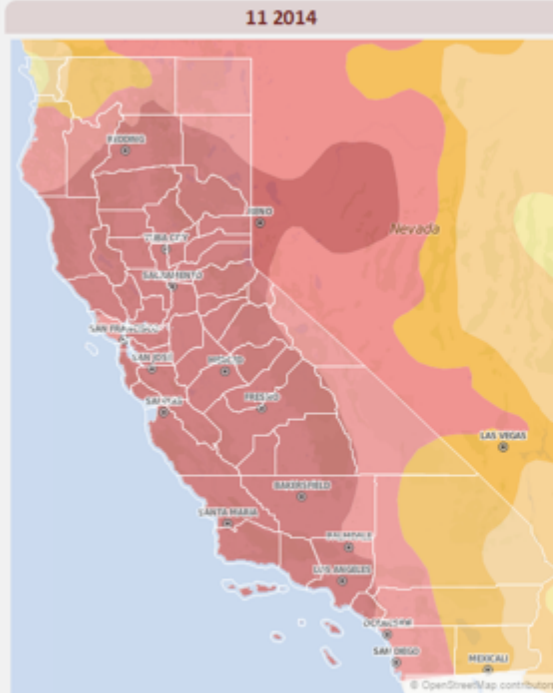
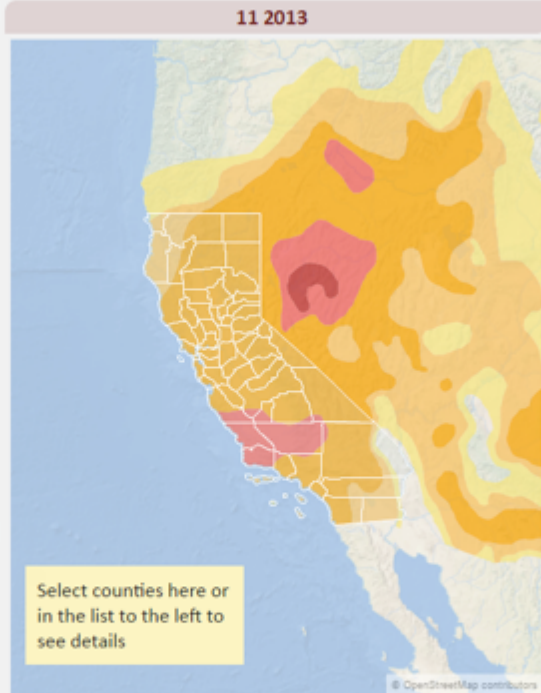
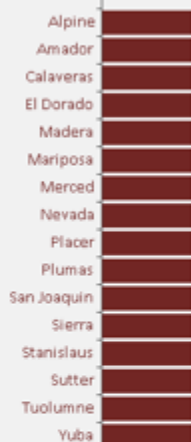


Data Source: Drought Monitor, National Drought Mitigation Center (NDMC), the U.S. Department of Agriculture (USDA) and the National Oceanic and Atmospheric Administration (NOAA)

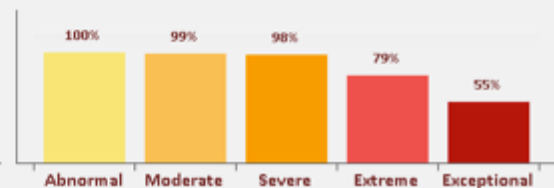
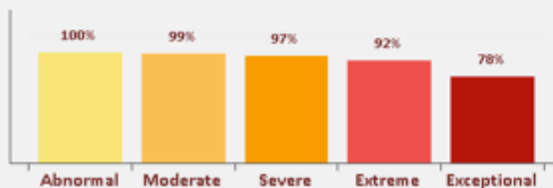
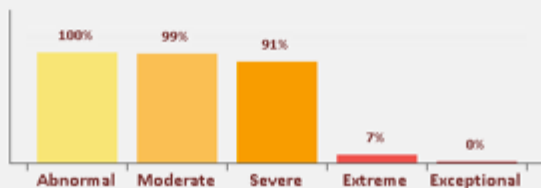
California Drought Conditions 2013 - 2015

November

Top Worsening Conditions



Top Improving



#2. Numerical Models

Analytics Apps

Build and broadcast smart analytics

$$\int dG(x) \geq \frac{1}{2} \sum_{h \rightarrow \infty} e^{-\frac{h^2}{2}} = H(h)$$

$$\int_0^1 f_n(u) f_n(t-u) du = \frac{\lambda^{n+1} t^n e^{-\lambda t}}{n!}$$

$$\log \varphi(t) = i\gamma t - c|t|^\alpha \left[1 + i\beta \frac{t}{|t|} \omega(t, \alpha) \right]$$

$$B(u) = \sum_{r=1}^{\infty} \Psi^*(b_r u)$$

$$g = e$$

$$\prod_{h \leq b} ; \bigcup_{i=1}^{\infty} M_i ; \prod_{n=0}^{\infty} X_n$$

$$C_{iv} = \sum_{j=1}^n a_{ij} b_{jv}$$

$$\lim_{t \rightarrow \infty} \frac{f(u)}{u} = P_e$$

$$\lim_{n \rightarrow \infty} P \left(\frac{\sum_{r=1}^n a_{r1} - n c_1}{\sqrt{\sum_{r=1}^n a_{r1}^2 - n c_1^2}} \right) = \frac{1}{\sqrt{2\pi}}$$

$$f(y|x) = \sup_{y' < y, y' < 2} P(\eta < y' | \xi = x)$$

$$y = \sqrt{\frac{\lambda u}{2n}} \left(\frac{\eta_{2n}}{\sqrt{2n}} + \frac{\eta_{2n} - \eta_{2n}}{\sqrt{2n}} \right)$$

$$f_n(t) = \frac{\lambda^n t^{n-1} e^{-\lambda t}}{(n-1)!}$$

$$f(t|y) = \frac{2e^{\frac{\lambda t^2}{2}}}{\sqrt{2\pi}} \int_{\frac{y}{\sqrt{t}}}^{\infty} \frac{e^{-\frac{u^2}{2}} du}{\left(1 - \frac{y^2}{u^2}\right)^{\frac{3}{2}}}$$

$$H_r(x) = \frac{G_r(x)}{1 + G_r(x)}$$

$$R = \int_{-\infty}^{\infty} \varphi(t) dt$$

$$C_n(\alpha) \geq \frac{n!}{\Gamma(\dots)}$$

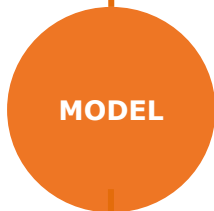
$$U_{n^{(+)}} = (2n)$$

$$\frac{\sinh t}{t} [\varphi(t)]$$

$$\frac{t}{u} \varphi(t) = \varphi$$



Insight to Action : the Model

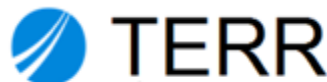


- PMML
 - R PMML (CRAN)
 - R2PMML (JPMML)
 - KNIME PMML
 - SAS PMML
 - ... many

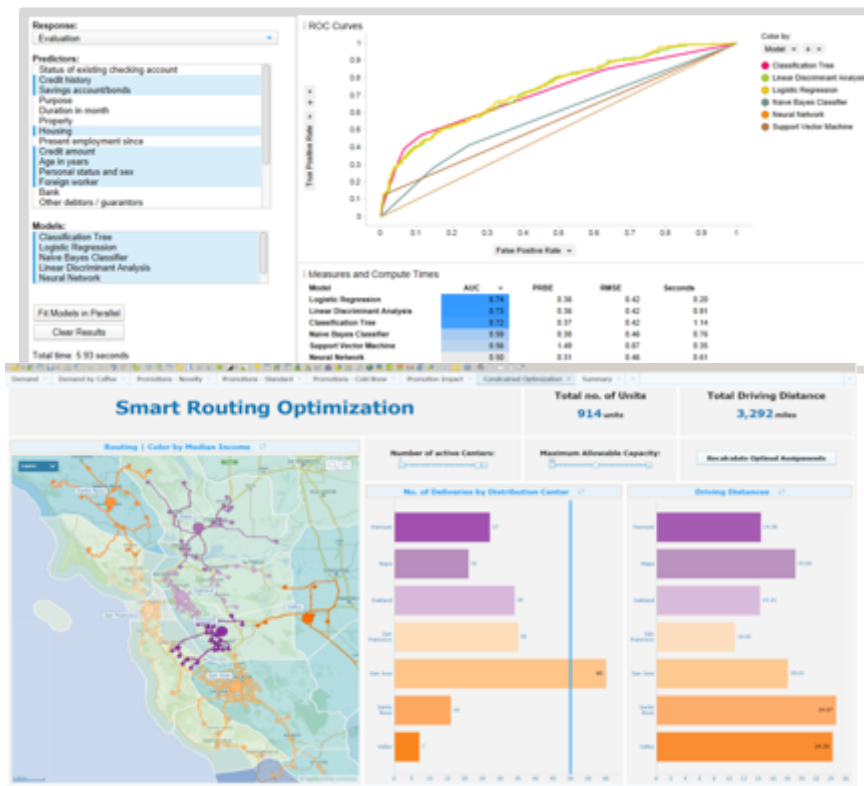
- H2O
 - POJO
 - Model object

- R Model Object
 - *.rds
 - *.RData
 - *.mdl (TERR)

- Matlab Model Object
 - MDS
 - PMML



- Declarative & Heuristic Rules
- SPC and Anomaly Detection
- Machine Learning
 - Supervised
 - Unsupervised
 - Gradient Boosting Machines
 - Random Forests
 - Deep Learning
- Optimization
 - Linear & Quadratic Programming
 - Genetic Algorithms
 - Process optimization
 - Capacity constraints

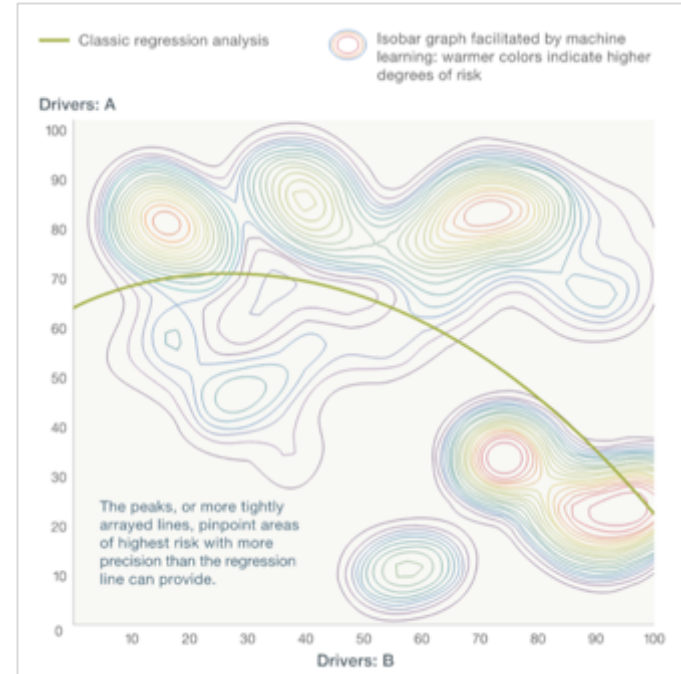


Machine Learning



Machine Learning finds predictive models in data without being told where to look

- Supervised – Solve known problems: $y=f(X)$
 - Build a model that predicts a condition (failure, success, ..)
 - What factors are driving customer network issues?
 - Decision Trees, Random Forests, Gradient Boosting Machines, Deep Learning
- Unsupervised – Identify new patterns, Detect anomalies X only
 - Are there new patterns or failure modes emerging?
 - Clustering, Principle Components, Deep Learning Autoencoder
- Optimization – Support Decision-making
 - Find best solution even when there are constraints on the process
 - What is the optimum allocation of resources for equipment maintenance?
 - Genetic Algorithm, Linear/Quadratic Programming



Anomaly Detection

Sensor Data / Anomaly Detection

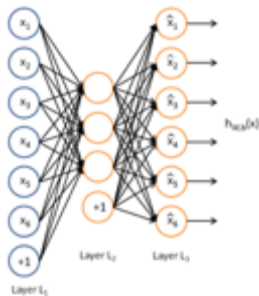
Existence of known anomaly ?

What if we don't have this ?

Unsupervised – no y variable

Can still model ! 😊

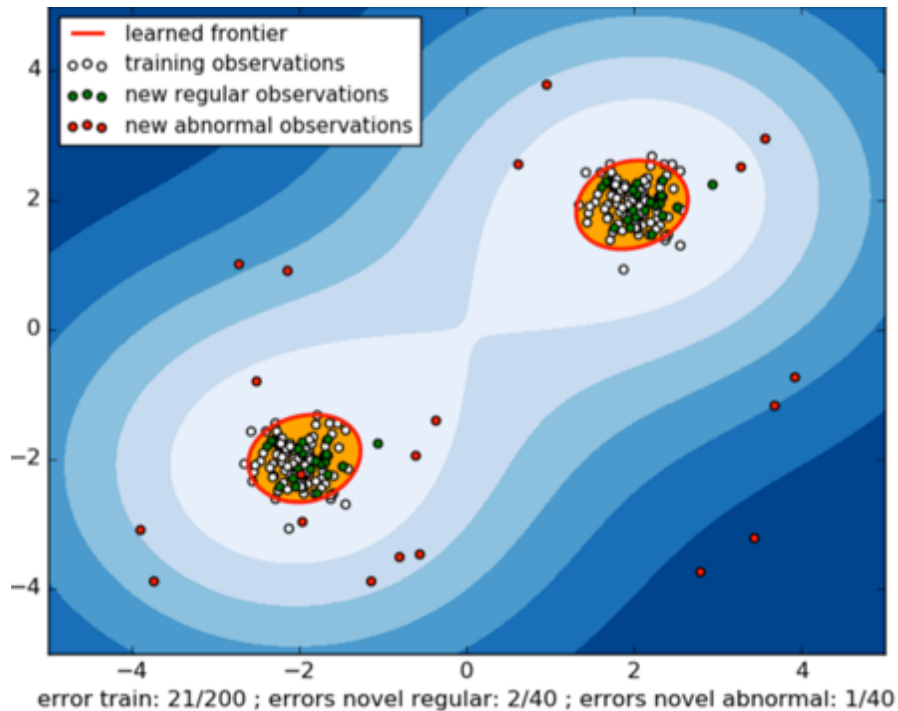
- SPC and Western Electric Rules (ESP)
- Single class Support Vector Machine
- Principal Components: PC Score
- Deep Learning: Auto encoding



Computing and visualizing

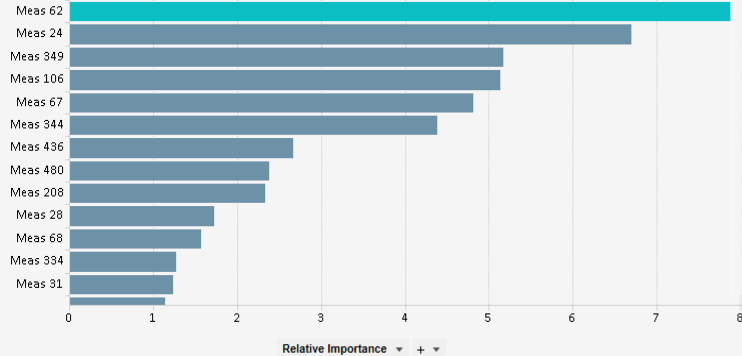
PCA in R

November 26, 2013
By thiago



GBM Results

Predictor Importance - Effect on Yield



Predictor Interactions Summary Table

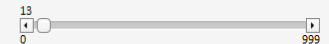
var1.name	var2.name	index
Meas 62	Meas 24	0.20
Meas 349	Meas 67	0.05
Meas 24	Meas 106	0.03
Meas 24	Meas 349	0.03
Meas 24	Meas 67	0.03
Meas 106	Meas 67	0.02
Meas 62	Meas 67	0.02
Meas 349	Meas 106	0.02
Meas 62	Meas 349	0.01
Meas 62	Meas 106	0.00

Heat Map Setup

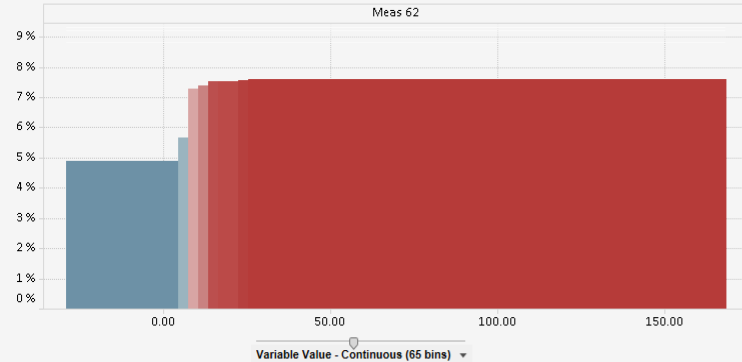
Bin Continuous variables into N groups for display:



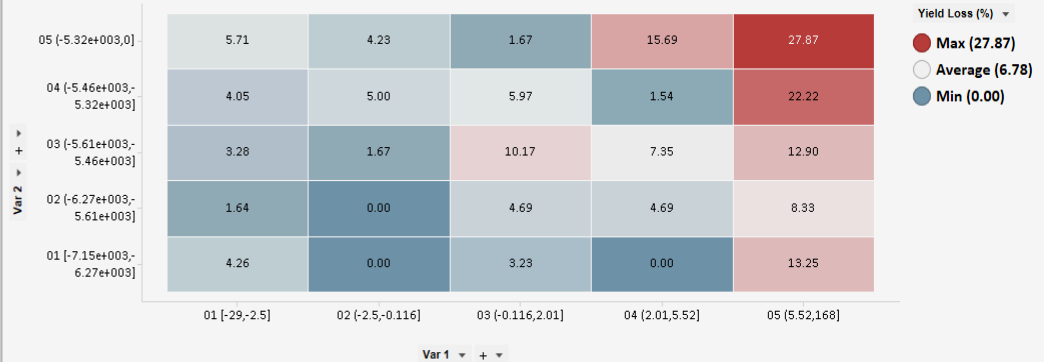
Eliminate small cells with unstable target average.
Only show cells with n rows greater or equal to:



Predictor Effect on Yield Detail



Predictor Interactions Detail



Model: Gradient Boosting Machine

Package 'gbm'

March 11, 2015

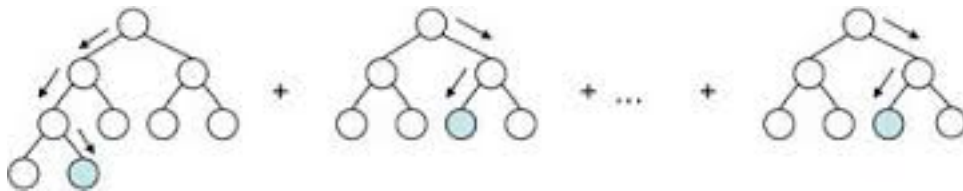
Version 2.1.1

Date 2015-03-10

Title Generalized Boosted Regression Models

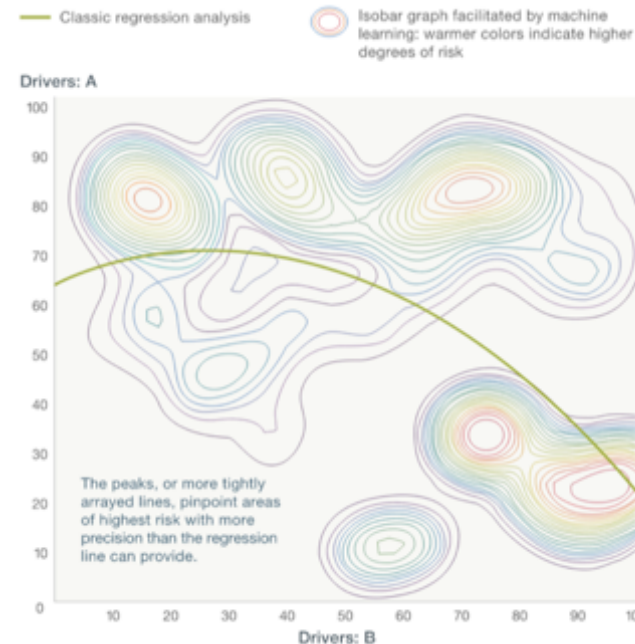
Author Greg Ridgeway <greg.ridgeway@gmail.com> with contributions from others

Maintainer Harry Southworth <harry.southworth@gmail.com>

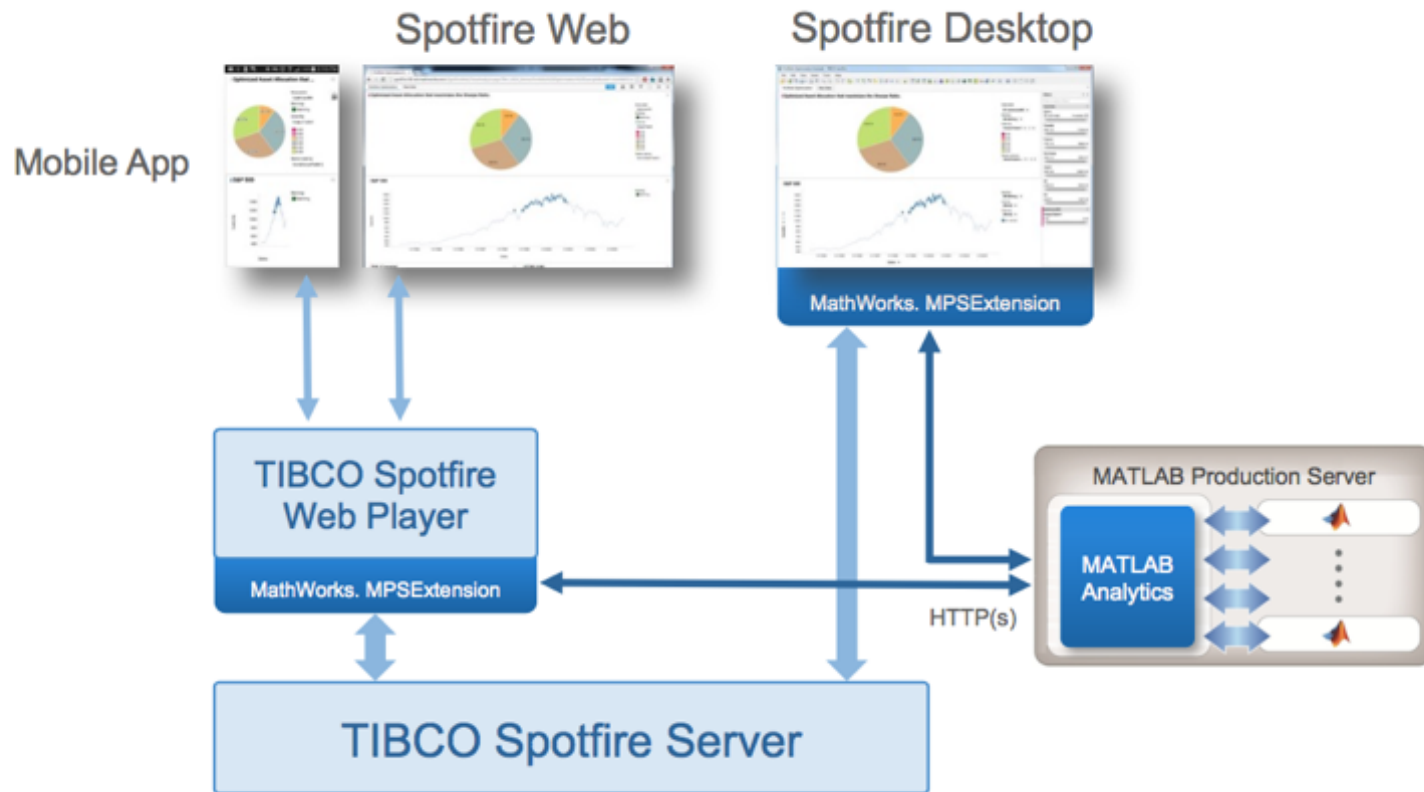


• Boosting Process

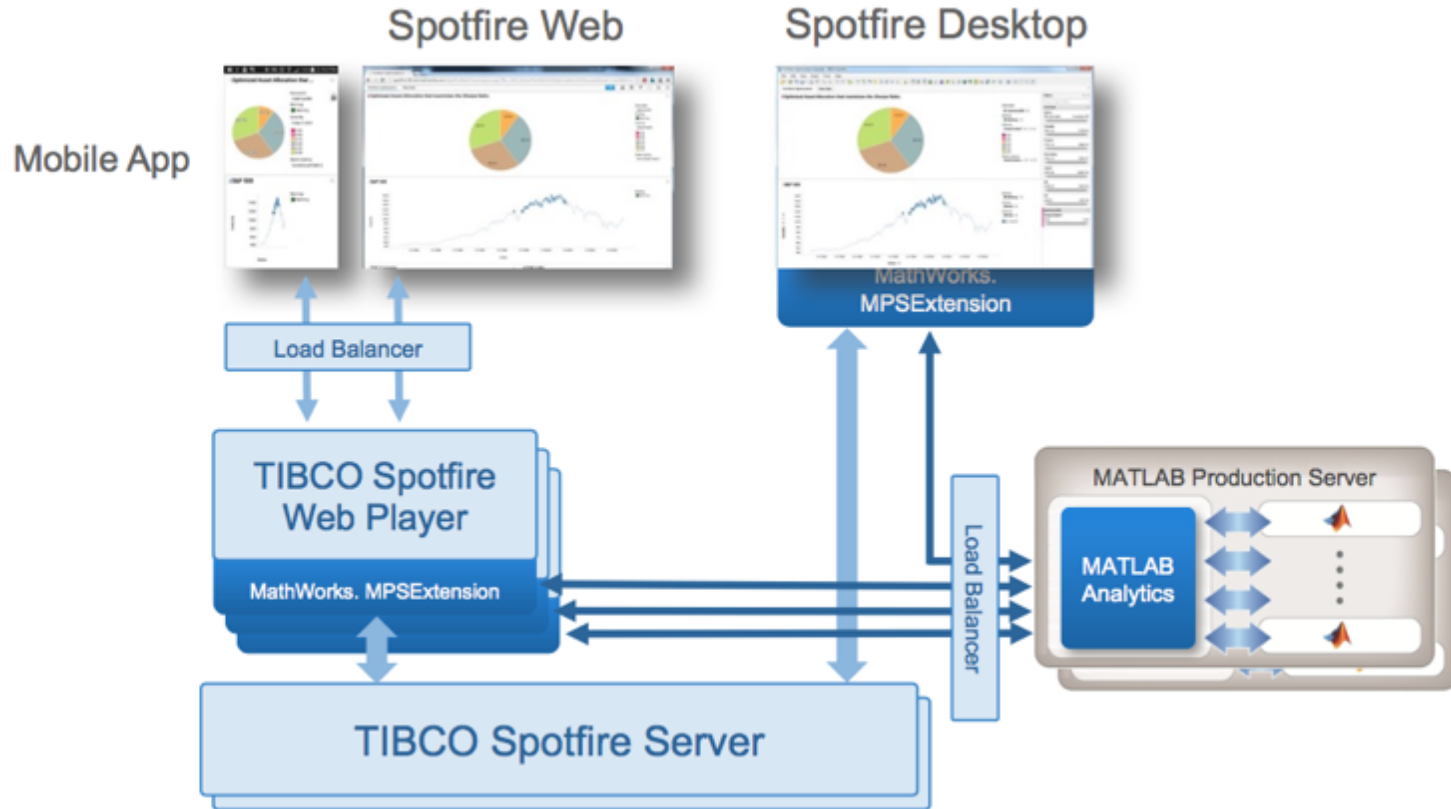
- Sample the data: fit a tree: recursive partitioning
- Drop the observations down the tree
- Re-sample the data; up-weighting the observations that weren't fitted well in previous model
- Save all the trees and average them
- Excellent fit + prediction



Reference Diagram



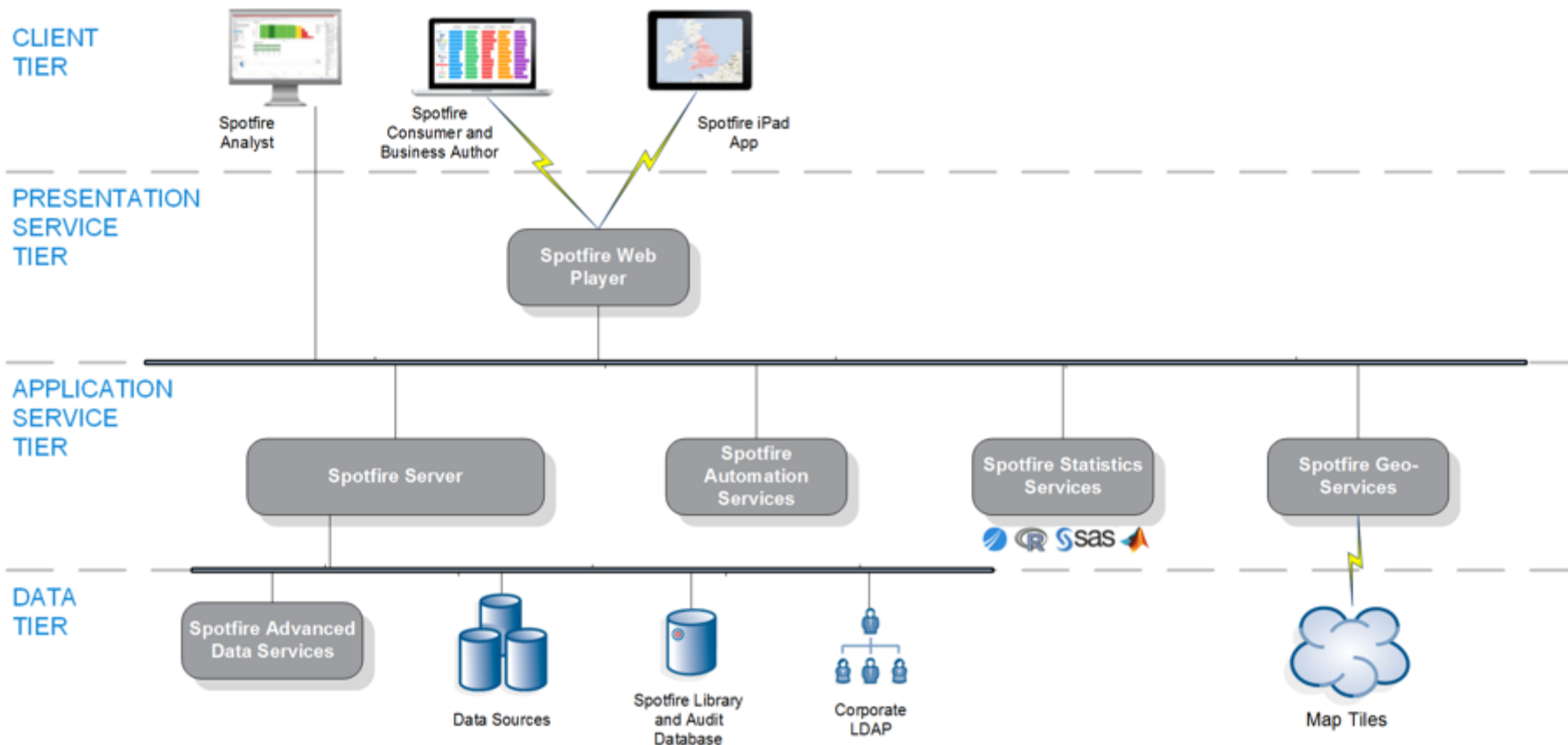
Increasing Capacity and Resilience



Spotfire & Matlab Demo



Spotfire Platform – Complete

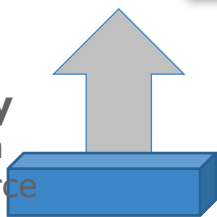


How does Spotfire Consume Data?



In-Memory

Load data from source in to memory



SQL
MDX

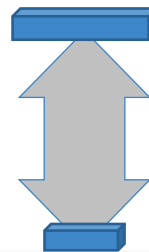
In-Database

Leave data in DB
Dynamically load and discard data to visualize



On-Demand

Dynamically swap data in and out of memory.



#3. Streaming Analytics

Streaming Analytics

Continuous algorithmic awareness
and automation

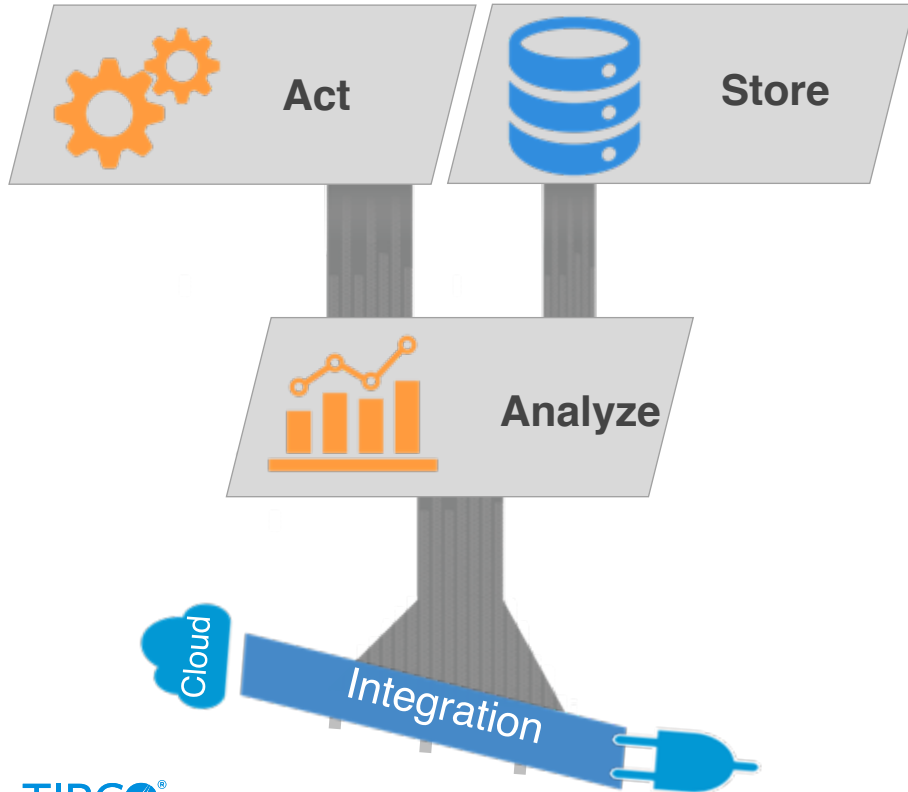


Issues with Traditional Data Processing



- Data collected from multiple sources and **periodically** placed in a persistent store.
- Analytical processes are executed against the **stored data**
- Introduces too much “**decision latency**”
- Responses are delivered “**after-the-fact**”.
- Decisions are made on **old and stale** data.
- Maximum **value is lost**.

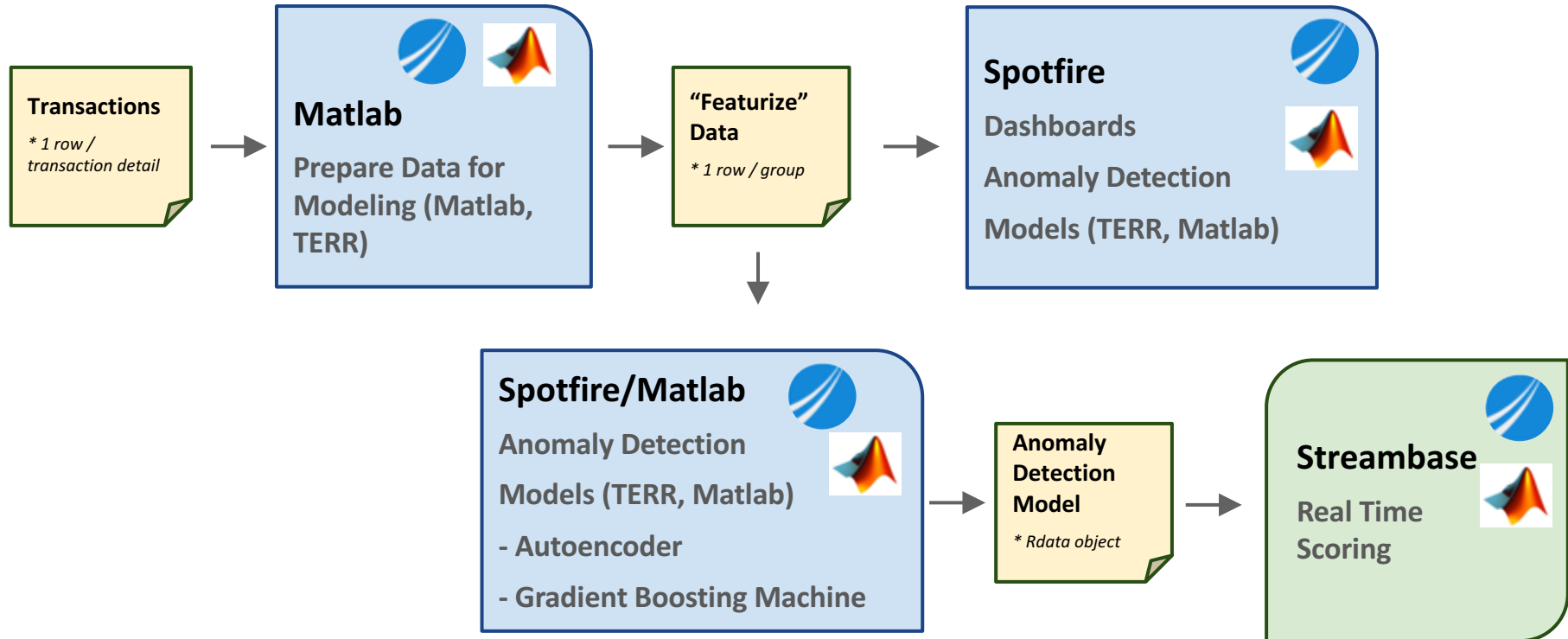
The New Era: Fast Data Processing



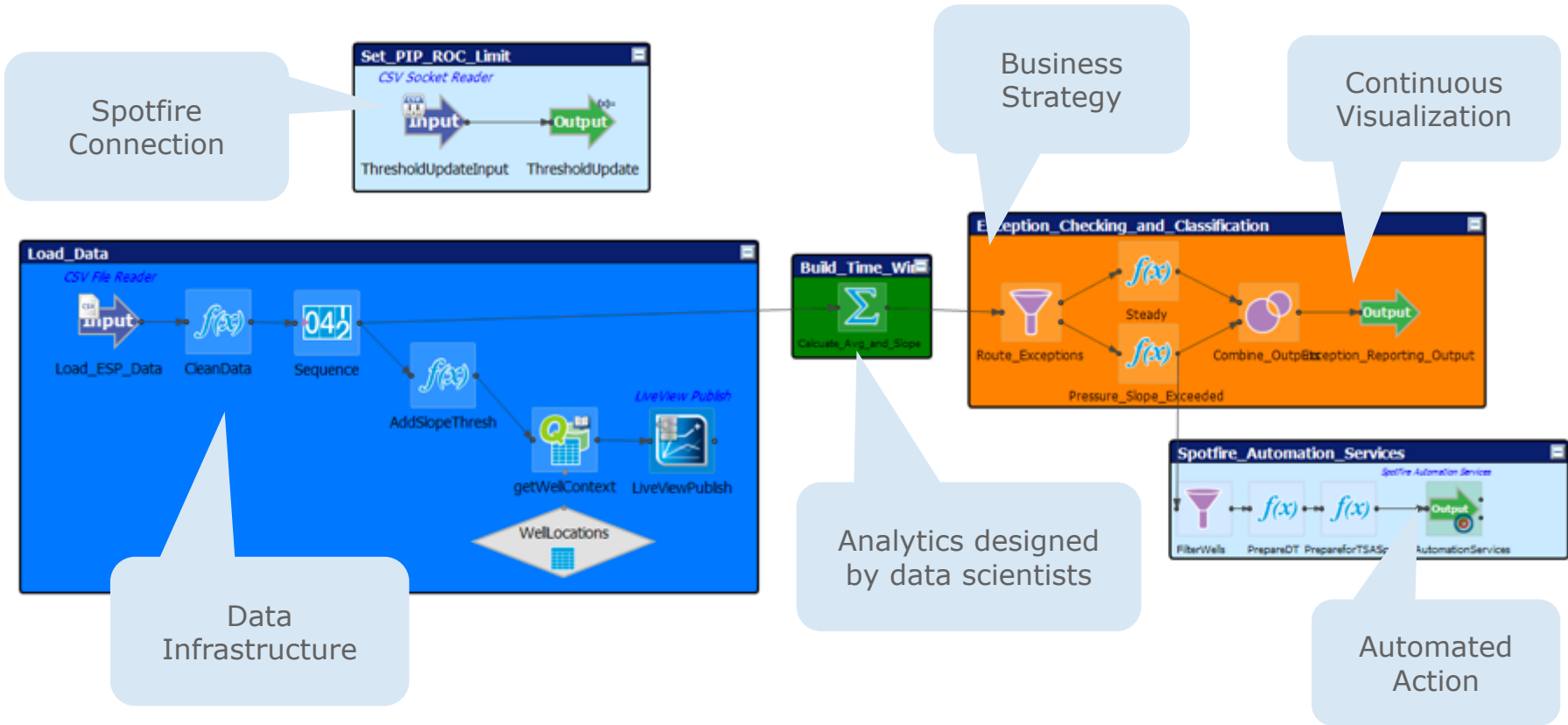
- Events are analyzed and processed in **real-time** as they arrive.
- Decisions are **timely**, **contextual**, and based on **fresh data**.
- **Decision latency is eliminated**, resulting in
 - Superior Customer Experience
 - Operational Excellence
 - Instant Awareness and Timely Decisions



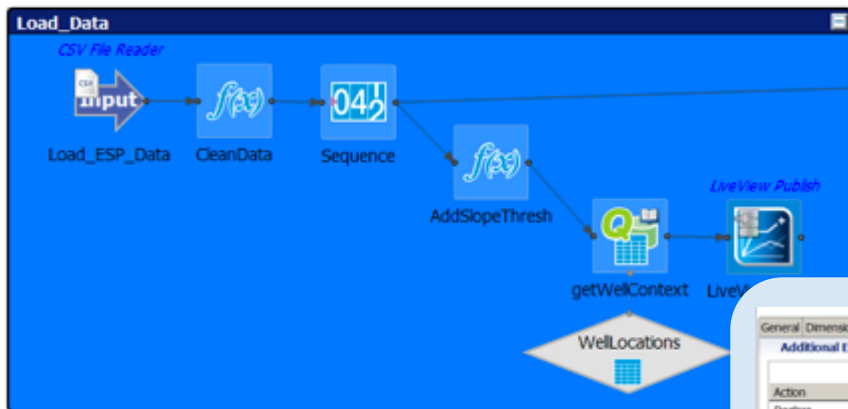
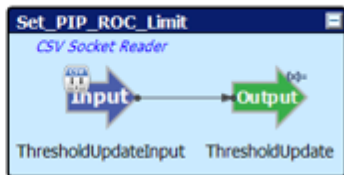
Anomaly Detection: Data Flow



Streaming Analytics with Streambase



Streaming Analytics

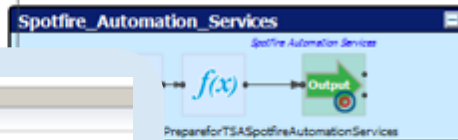
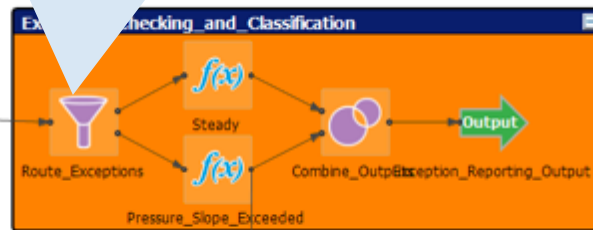


General | Predicate Settings | Concurrency

Create output port for non-matching tuples

Predicates:

Output Port	Predicate
1	MotorTemp_Max < 150
2	MotorTemp_Max > 150



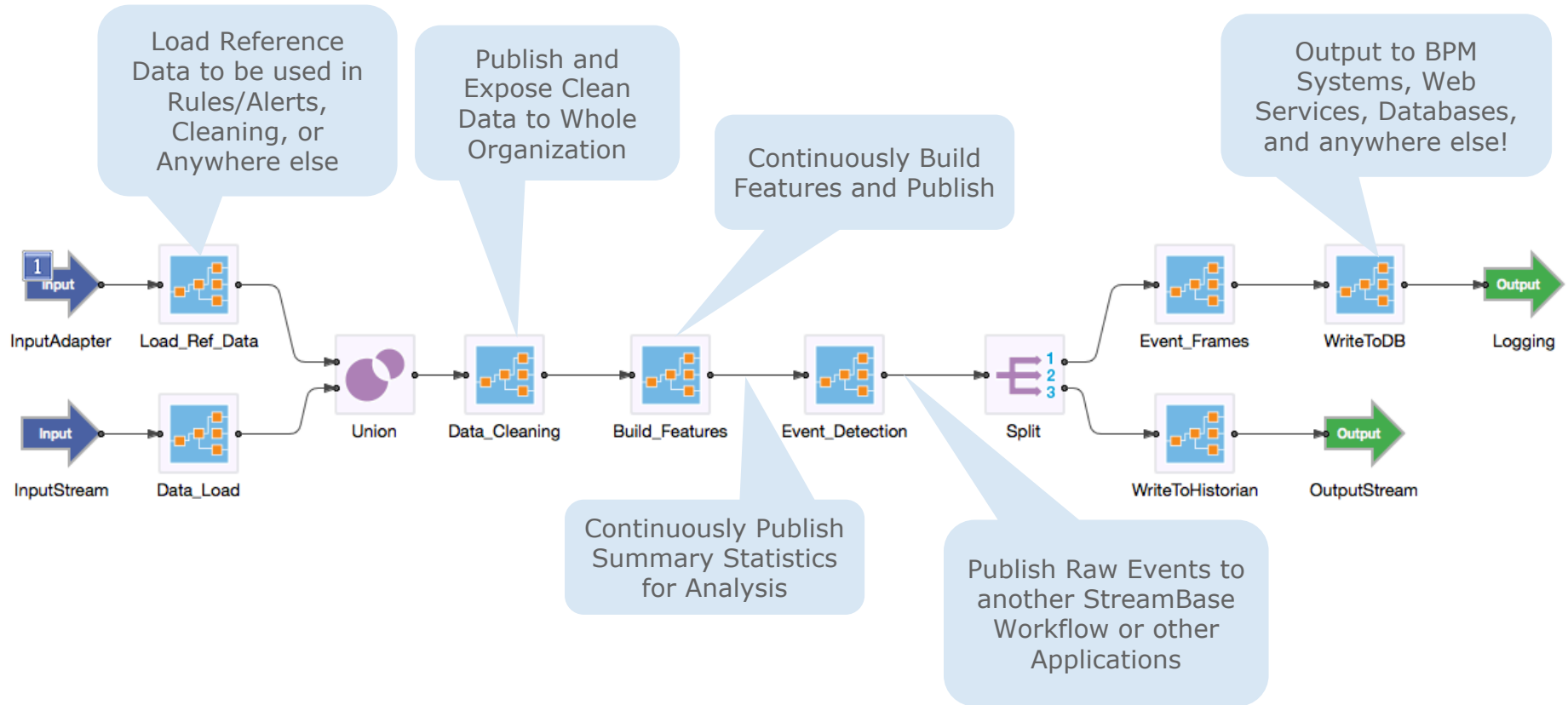
General | Dimensions | Aggregate Functions | Group Options | Concurrency

Additional Expressions

Action	Field Name	Expression
Declare	secs	to_seconds(DateTime)
Declare	hours	secs / 3600
Add	Pressure_Avg	avg(Pressure)
Add	Pressure_Max	max(Pressure)
Add	Pressure_ROC	slope(hours, Pressure)
Add	MotorTemp_Avg	avg(MotorTemp)
Add	MotorTemp_Max	max(MotorTemp)
Add	MotorTemp_ROC	slope(hours, MotorTemp)



Streaming Analytics : Equipment Monitoring



Financial



Databases



IoT: Integration



Messages



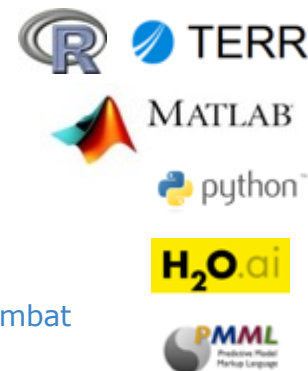
Big Data



Social



Compute



API's

C++, Java, .NET, JavaScript, Python

Messaging

TIBCO RV & EMS, JMS, MQ Series, RMDS, Solace, Tervela, Wombat

General Purpose Adapters

FIX, STAMP, JDBC, ActiveSpaces, POP3, IRC, HTTP, SMTP, IM, XML, RSS



Streambase & Matlab Demo



Insight Platform Demo

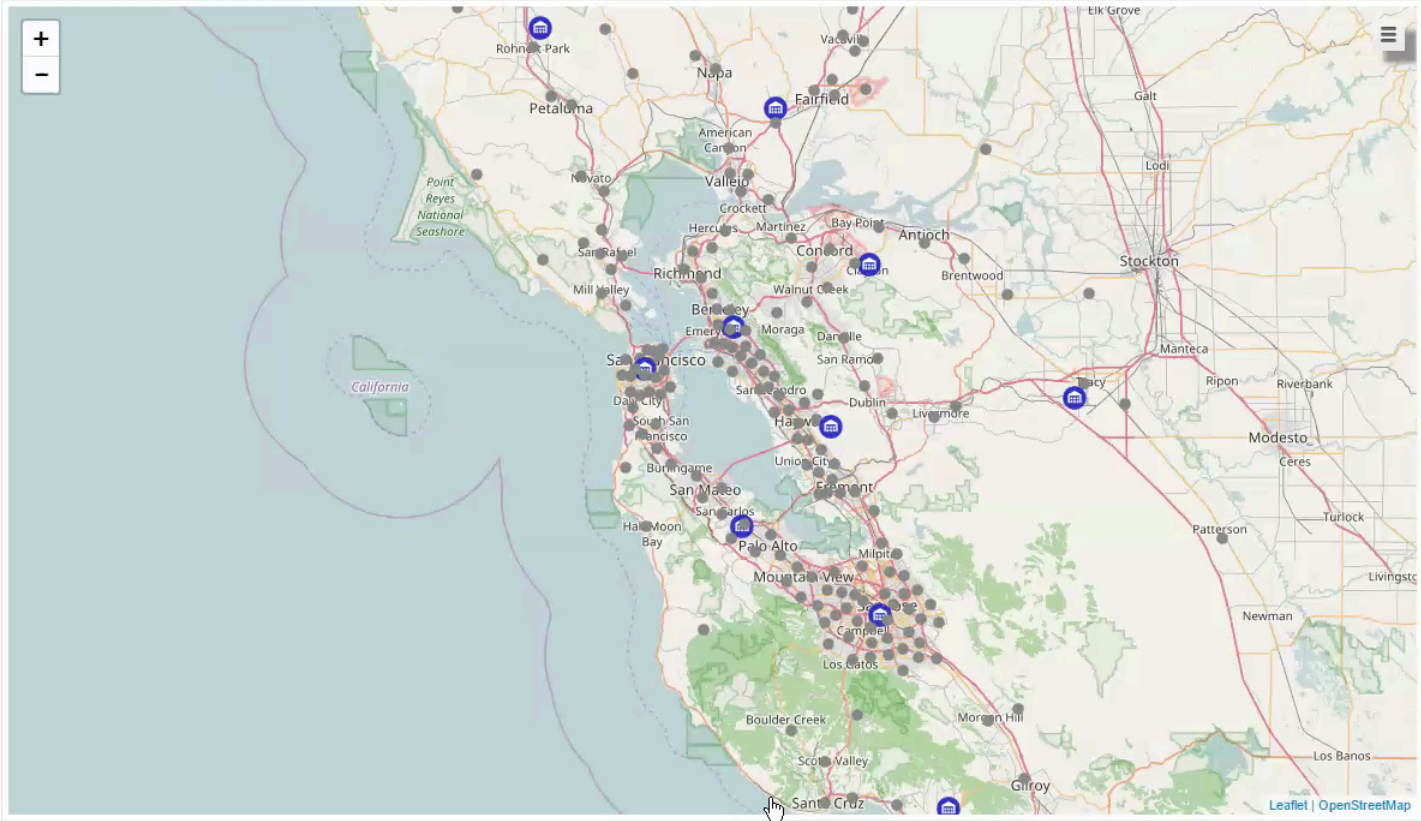


Trips

Search

Trip	Service	Upd	Status
No Data Found			

Headway



Learn more about TIBCO

- www.tibco.com
- Accelerators & Community
 - <https://community.tibco.com/>
 - <https://community.tibco.com/wiki/accelerators>
- Spotfire micro site
 - <http://spotfire.tibco.com/>
 - <http://spotfire.tibco.com/demos>



INTERCONNECT EVERYTHING

