



Capability White Paper

---

## ***Straight-Through-Processing (STP)***

Drag-and-drop to create automated, repeatable, flexible and powerful data flow and application logic orchestration without programming to support smarter digital enterprise and transformation

## Business Challenges

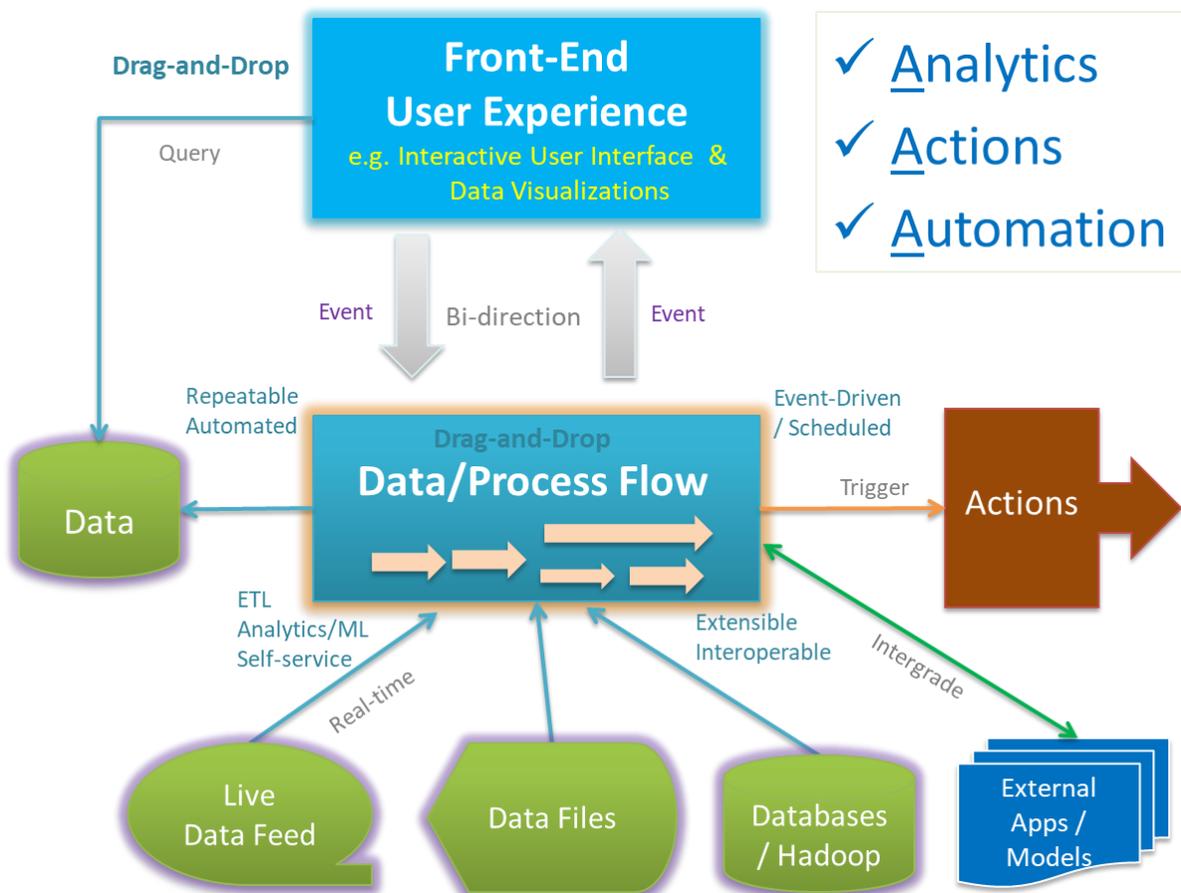
Today's businesses are facing growing challenges of extracting meaningful information from ever growing disparate data sources and silo applications, and integrating them together end-to-end - for timely and holistically visibility; also to embed the analytics seamlessly into new digital applications and processes rapidly for date-driven decision making and digital transformation.

While Interactive Data Visualization and Visual Analytics is a powerful way to enable analysts and non-technical business user to see and understand their data more quickly, a key *prerequisite* of that is to be able to (1) easily create *repeatable and automated process flows* to quickly and effectively *prepare and transform* data from the disparate sources or applications into data sets that are easy and ready for the data visualization and data analytics engine to use; (2) can also apply *user-defined or software-defined advanced data analytics, machine learning* or can plug in *existing* or 3<sup>rd</sup>-party *predictive analytical models* and *machine learning (ML)* easily into the workflows and end-to-end flexible process integration and orchestration.

	Challenges	Requirements
1	How to enable business people to <i>define process flows themselves</i> based on own situational data preparation, integration and analytical requirements?	Self-service No-Programming
2	How to enable business people to visually capture and define repeatable <i>process flows</i> on data ingestion, preparation, transformation, fusion, integration, analytics, monitoring, visualization activities so it can be fully automated for an operational environment?	Process-Oriented Repeatable
3	How to closely <i>integrate</i> this kind of repeatable and flexible processing <i>with the front-end</i> Interactive Data Visualization and Data Discovery, as well as support <i>Operational Intelligence (OI)</i> requirements such as <i>real-time</i> data feeds, real-time monitoring and alerting?	Front-End Integration + Operational intelligence
4	How to <i>ingest, clean, transform, fuse and prepare</i> the data from disparate data sources with a wide variety of formats and data quality into data that is more appropriate for integration, data analytics or data visualization?	Data Preparation ETL, Governance
5	How to enable user to insert <i>own data analytics</i> into the data processing flows, as well as to <i>plug-and-play more advanced data analytics, statistics, existing 3rd-party models and Machine Learning algorithms</i> easily in an <i>extensible and interoperable</i> way?	Analytical Extensible Interoperable

6	How to enable data-driven <b>Actions</b> triggered by the data analytics plugged-in or embedded in, for use cases such as <i>anomaly detection, risk mitigation security management and resource optimization</i> ?	Actionable
7	How to seamlessly <i>integrate</i> the user-defined application integration ( <b>Application Domain</b> ), data preparation, data analytics, rapid-fire data visualization ( <b>Data Domain</b> ) and human <i>user interface</i> and <b>collaboration</b> ( <b>Human Domain</b> ) <i>into one single integrated and holistic visual and modular configurable environment</i> for agile decision making?	Integrate Data + Application + Human Domains
8	How to quickly enable agile and relevant digital applications, more than just visualization and analytics, or giant complex enterprise software, so can hide all the complexity away from end-users using simple and most relevant user interfaces, yet magically integrate and orchestrating various backend silo-ed data and systems together?	Agile Digital Applications Assembly Capability - "Digital Enterprise Enablement Platform" ( <b>DEEP</b> ) in SitScape

The following diagram illustrates the high-level idea of the requirements out of the table above.



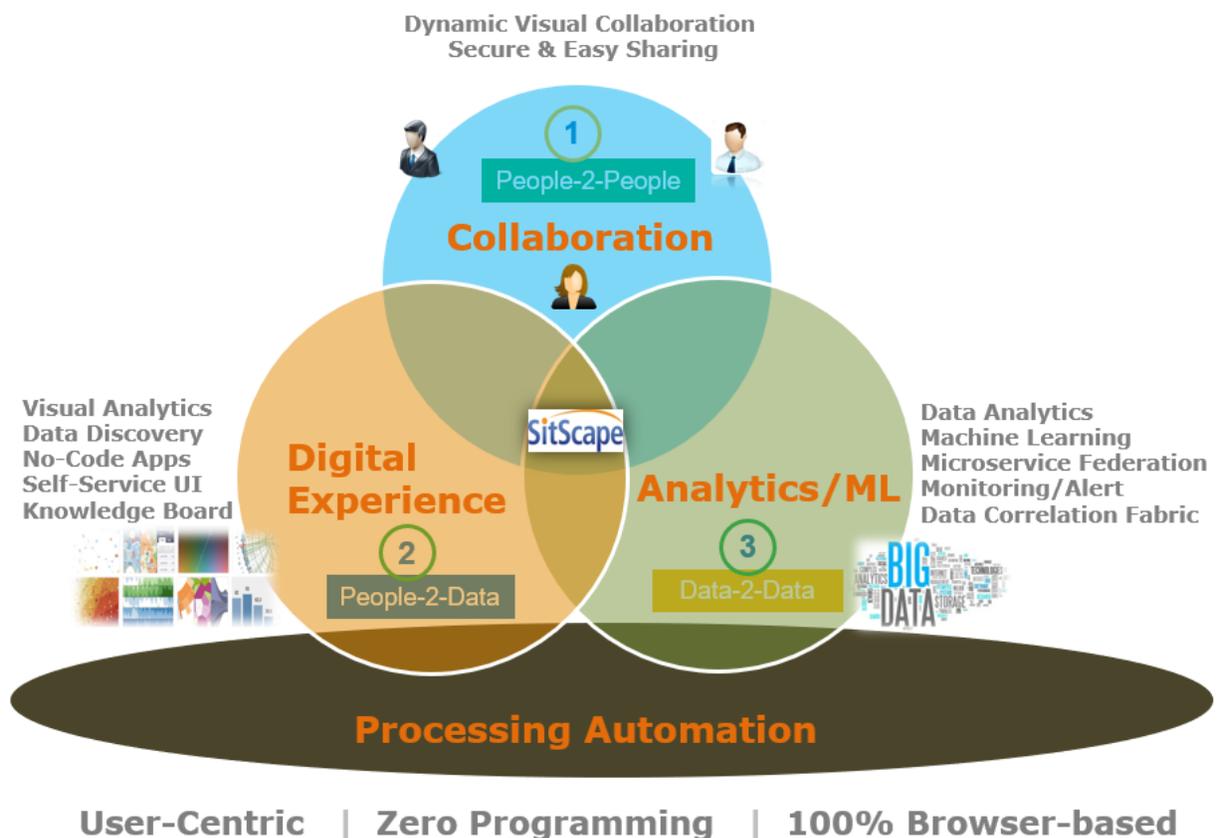
## Current Software Tools Available on the Market

While there are a number of tools on the market, each has some major limitations, and is lacking some key capabilities to add those challenges above holistically.

1. Pure-play Data visualization tools such as Tableau, Qlik are initially desktop-based for individual analysts' use, and focus on creating visualizations and supporting visual data discovery by data analytics type of users. Those tools *assume* data is well prepared and integrated in ready-to-use format, which is usually not the case in the real-world. They do not have the *process flow* concept in their environments at all. Plus they are not designed for, so do not work well for application and operational environment to support monitoring, alert, actions based analytics, nor designed for data governance.
2. There are tools focusing on desktop-based data blending, ETL and analytics such as Alteryx and Lavastorm. Those tools allow analysts to clean and prepare data sets into a format used by the separate data visualization tools. The limitation is they are desktop-based for single user on data preparation, not for enterprise-scale usage with governance. Their visualization capability is very limited and operation intelligence capability, no Common-Operating-Picture (COP), User-Defined-Operating-Picture (UDOP) and collaboration, no true application integration capability.
3. There are visual ETL-centric tools such as Trifacta that are mainly focused on data cleansing and data quality, not on process definition and automation, not to mention analytics, integration, real-time operational monitoring and front-end visualization and dynamic user collaborations.
4. There are new data curation tools such as Tamr that is focusing primarily on enterprise's disparate data's collaborative curation and data quality control with some human input.
5. There are traditional IT-centric ETL tools such as Informatica that have heavy footprint and less business user friendly.
6. There are application integration vendors such as DELL Boomi that is focused on integration of multiple cloud applications, not on preparing data for analytics, visualization and live monitoring and collaboration, nor application-logic automation.
7. Specialized tools such as Splunk is more focused on machine log files indexing and search, and related visualizations, not a general-purpose tool for data transformation, fusion, and process automation with any data format and sources.

## SitScape’s Solution

The **Straight-Through-Processing (STP)** module, as a key part of the SitScape’s overall software platform, is SitScape’s innovative answer to the above business challenges. The below diagram illustrates *where STP fits* in the overall capability picture of SitScape’s **Digital Enterprise Enablement Platform (DEEP)**.



SitScape’s STP capability provides a simple, light weight, elegant, holistic and fully integrated solution to address those challenges.

1. It is a 100% browser-based environment
2. It is fully integrated with SitScape’s User-Defined-Operating-Picture (UDOP) environment in one single software platform
3. It’s user-centric and visual. Business user can define repeatable, automated data and process flows in a visual drag-and-drop environment with a set of Nodes, and a Canvas to drag those Nodes into, and can connect Nodes into process flow.

- It supports some proven best practices of [Flow-Based-Programming \(FBP\)](#) paradigm, which uses a 2-dimensional graph with Nodes and Edges to represent the flow logic and supports the best practice of [Configurable Modularity](#)
  - It can support rapid creation, testing, simulation and iterative refinement similar to an Integrated Development Environment (IDE), yet without programming required when assemble a flow-based application with maximum productivity, reusability and maintainability
4. It is **modular**. A business user can simply drag modular Nodes on to the visual Canvas, and define and test the workflows quickly.
  5. It is **extensible**. Additional STP Nodes can be added if necessary into the available ones without impacting the existing Nodes and STP Flows.
  6. It is **interoperable**. With connection and integration Nodes such as REST API, Databases, Files and Web Services etc., it can re-use and is interoperable with existing internal or external applications and data sources, and can re-use their functionalities and integrate and glue them together into the flows by serving as the middleware in between.
    - One key use case of this is to integrate and re-use other or 3<sup>rd</sup>-party analytical models or engines (statistical or machine learning) and make them part of the STP flows.
  7. It is highly **parallel**, and can run multiple paths of a STP in a parallel fashion to take advantage of multi-core/multi-CPU and distributed computing environment.
  8. It supports **near-real-time operational intelligence**. It is not just for analysts to do a few ETL processing, and then hand the output to a separate data visualization tool. It can support repeatable and automated processes for dynamic data updates in a 24x7 operational environment.
  9. It supports various data calculation and data transformation functions such as *filtering, replacement, sorting, column re-ordering/rename, encoding, decoding,*

*various string manipulation, split, token extraction, regular expression, math calculation, aggregation, fusion, merge, transpose* etc.

10. It also supports various flow controls and advanced logic such as *looping, decision routing, advanced IF-ELSE logic* and *advanced expressions* and enables rapid Application Building
11. It supports STP process run based on scheduler, as well as triggered by events.
12. It can do the default full data-set-based processing of many records in STP flow, or can also *iterate* through incoming data *record-by-record* during a sub-processing of the flow for per-record-based processing logic and flow control.
13. It can calculate key metrics and KPIs, and can supports monitoring of those metrics and auto generate alerts based on threshold in an operational environment. Alerts can manifest as email, SMS, video-wall big screen animation.
14. It can trigger and generate advanced and sophisticated multiple page Reports based on the STP processing condition at run-time
15. It is fully integrated with the front-end user-interface and interactive data visualization layer
  - It can automatically trigger on-screen refresh, animation, dashboard and visualization updates and visual alert from the STP side
  - It can also subscribe to and listen on events generated from the UDOP front-end user interface such as Web forms, data changes, user selection, or user-actions, and then trigger the corresponding user-defined STP process flow on-demand, and back to the front-end user interface again. This can all be done with simple drag-and-drop configuration without any programming or coding for *both Front-End User Interface Design and backend Business Logic*
  - This kind of *2-way bi-directional* communication and full-integration between the front-end User Interface layer and the STP back-end middleware and process flows makes it extremely powerful and agile to support dynamic business requirements to support agile digital applications with significantly improved efficiency and effectiveness.

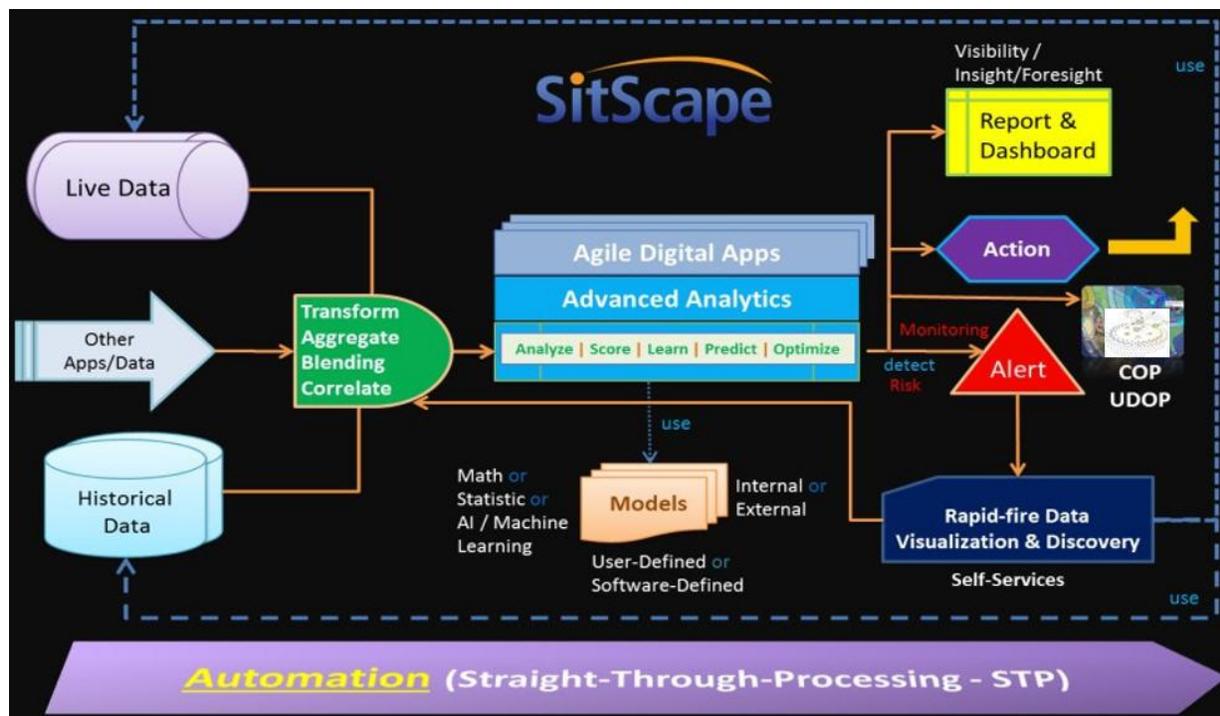
## Applications of STP in Real-World

STP can be used to support a wide spectrum of business and technical *use cases* such data preparation, data integration, no-code application assembling by enabling backend application logic without coding, application integration, operational intelligence, business intelligence, business visibility and enterprise performance management.

It clearly can support traditional data Extraction, Transform, Loading (ETL) very well in a very visual, scalable and managed environment by non-technical users.

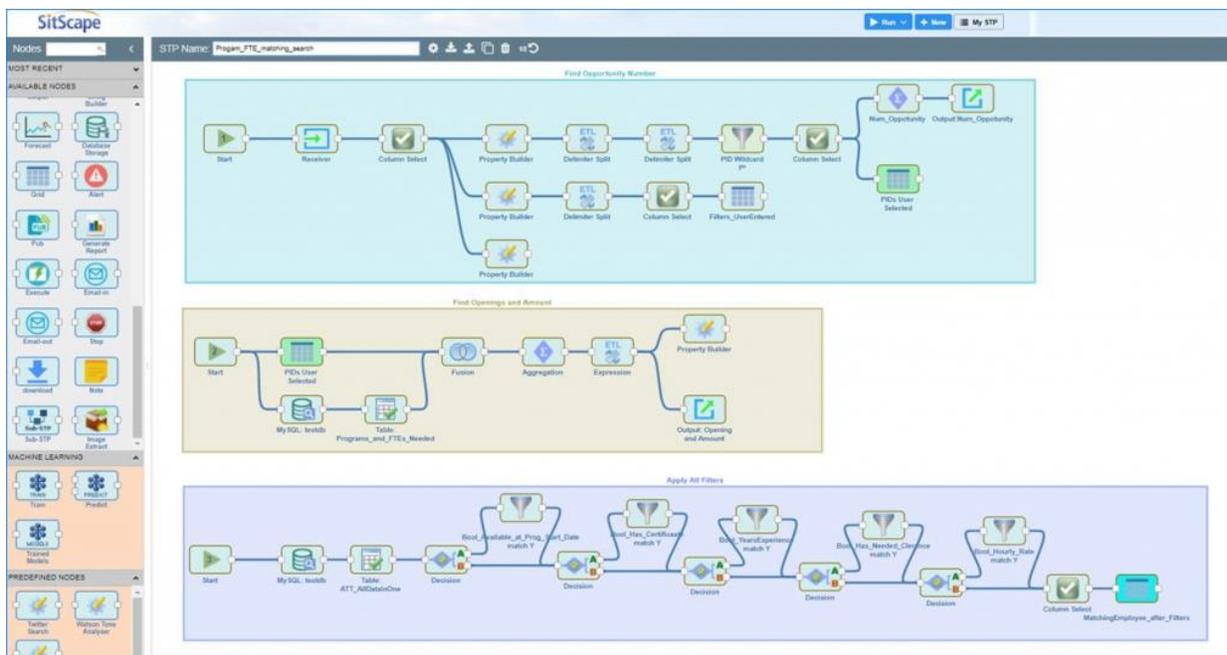
It can also plug in analytics and integrate with other applications and analytical engines. As it can trigger actions with *Execution Node*, it makes analytics not only visible, but also actionable.

The diagram below illustrates how STP can be used to **drive dynamic data integration, analytics, modeling to calculate risk, detect anomaly, optimize performance, and trigger appropriate actions** with *actionable intelligence*, such as blocking an attaching IP address, suspend an account, increasing capacity of servers based on analytical forecasting, re-route a truck based on traffic condition, approve a loan application etc. in a dynamic environment.



SitScape’s STP module provides a very powerful way to support *end-to-end rapid application assembly or integration, orchestration, data ingestion, transformation, fusion, correlation, analytics, modeling, forecasting, KPI and performance metrics calculation, risk scoring, anomaly and outlier detection, alert, triggered actions, agile application assembly on-demand*, and can auto generate reports, and auto update dashboards, refresh monitoring console, and also on-demand alerting human operators and analysts so they can initiate and perform their rapid-fire visual data analytics and drill-down for root-cause analysis using SitScape, *all in a self-service visual environment*. All those fully integrated in the SitScape’s *comprehensive, holistic and easy to use software platform*.

Below is a sample screen of a real-world STP process, which is used to calculate the matching logic of supply-and-demand in a human capital management application:



## Sample Screens

The following screen is a simple example of STP process flow in action.

The screenshot displays the SitScape interface with a data pipeline and a console log. The pipeline starts with 'HistoricalData.csv' and 'MostRecentMinuteT.' feeding into a 'Fusion' node (30 rows, 17 cols). This is followed by 'Math Calculation' (ETL, 30 rows, 18 cols), 'Math Modeling' (ETL, 30 rows, 19 cols), 'Data Classification' (ETL, 30 rows, 20 cols), and 'Select Data Fields' (30 rows, 3 cols). The pipeline then branches into two paths: one leading to 'Update Dashboard' (30 rows, 3 cols) and another leading to 'Very High Risk Ones' (3 cols), which then feeds into a 'Count' node (1 row) and finally an 'Alert' node.

The console log shows the following data:

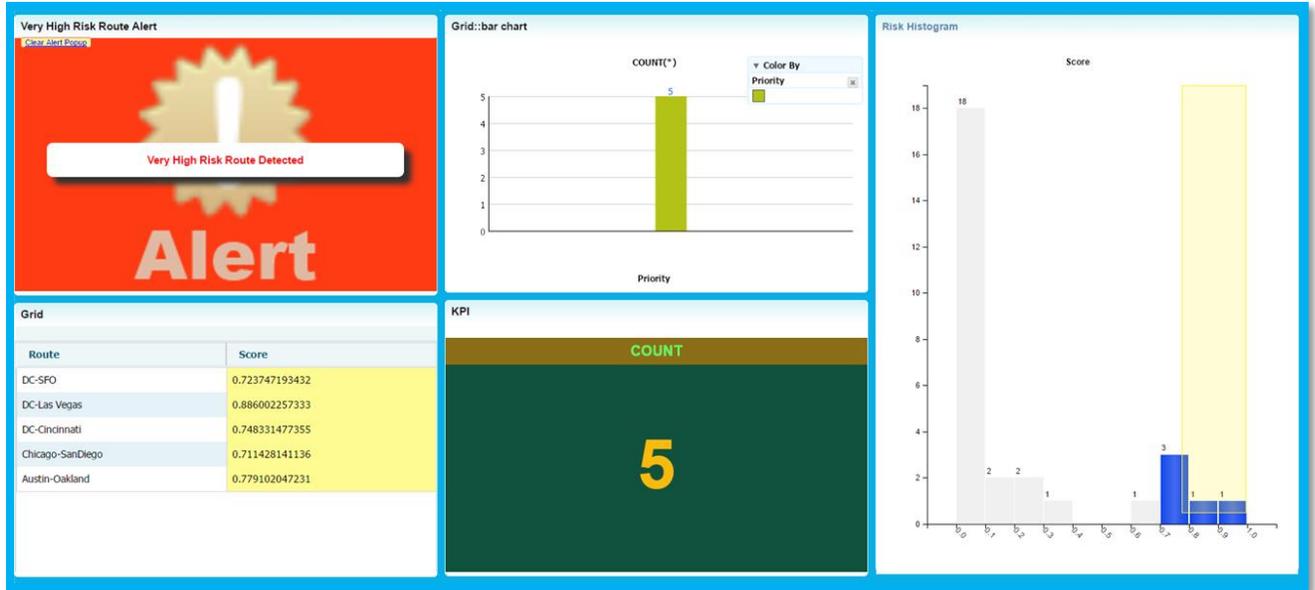
T1.DayOf	Ratio	T1.Hour0	T1.Minute	T1.Type	T1.Route	T1.Total Sales	T1.Ticket Sold	T1.Avg_P	T2.DayOf	T2.Hour0	T2.Minute	T2.Booker	T2.Route	T2.From	T2.To	T2.Price_	T2.Price_
WED	-0.07	0	1	Online	NYC-DC	0	400	410	0	0	0	1	Online	NYC-DC	4300	10	430
WED	-0.11	0	1	Online	NYC-Bosto	0	180	190	0	0	0	1	Online	NYC-Bosto	1200	6	200
WED	0.92	0	1	Online	NYC-Tokyo	0	1,300	1330	0	0	0	1	Online	NYC-Tokyo	42000	55	100
WED	0.05	0	1	Online	DC-Chicago	0	200	204	0	0	0	1	Online	DC-Chicago	950	5	190
WED	0.79	0	1	Online	DC-Las Veg	0	100	120	0	0	0	1	Online	DC-Las Veg	2150	100	21.5
WED	0.11	0	1	Online	DC-San Dk	0	700	666	0	0	0	1	Online	DC-San Dk	62300	100	623
WED	-0.02	0	1	Online	DC-Denver	0	440	444	0	0	0	1	Online	DC-Denver	3150	7	450
WED	0.56	0	1	Online	DC-Cincin	0	1,000	480	0	0	0	1	Online	DC-Cincin	1760	4	440
WED	-0.06	0	1	Online	Chicago-NI	0	330	345	0	0	0	1	Online	Chicago-NI	3500	10	350
WED	-0.09	0	1	Online	Chicago-Dk	0	220	219	0	0	0	1	Online	Chicago-Dk	5280	22	240

The use-case here is that it ingests incoming near-real-time online ticketing sales data, and dynamically fuses it with historical data's statistical output, and then applies user-defined mathematics, statistics or machine learning (ML) model to detect any *outlier* or *anomaly* in the incoming data comparing to the historical pattern.

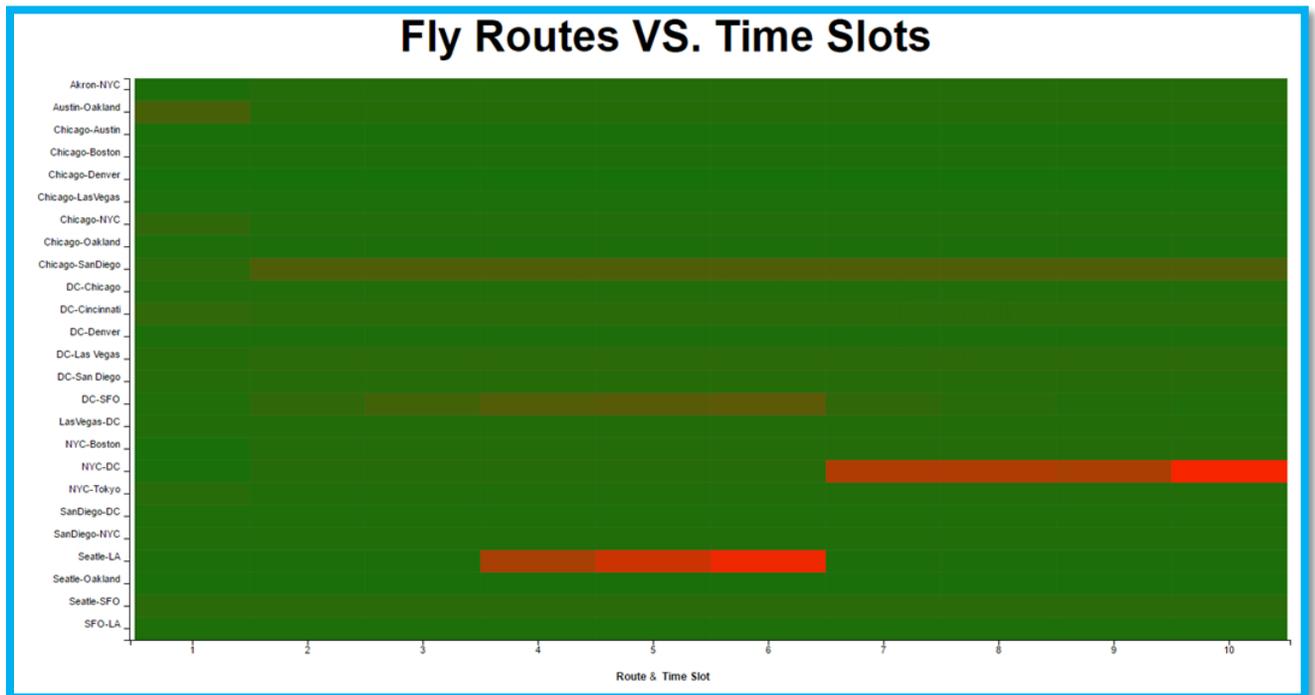
On the first sub-flow-route, it outputs the calculations to a front-end data grid, and auto update the corresponding dashboard on the big display monitor in the watch center using the latest processing and calculation outputs.

It also branches into second sub-flow-route where it focuses detecting on any very high-risk ticketing data set that is significant different from historical patterns (the anomaly), if there is a match, an alert with related info will be sent to the appropriate human operator so they can initiate further drill-down and interactive visual data analysis and discovery of the root cause.

Below is a sample screen output of the alert screen and related data visualizations on the big display monitoring console using a Web browser.



The below screen shows a visualization of the risk scores in timeline-based heat-map, which is updated on a continuously basis for an operational environment. The green color represents low risk, and the red color represents higher risk. It is easy to see where the high risk is happening and for how long at a glance in this way.

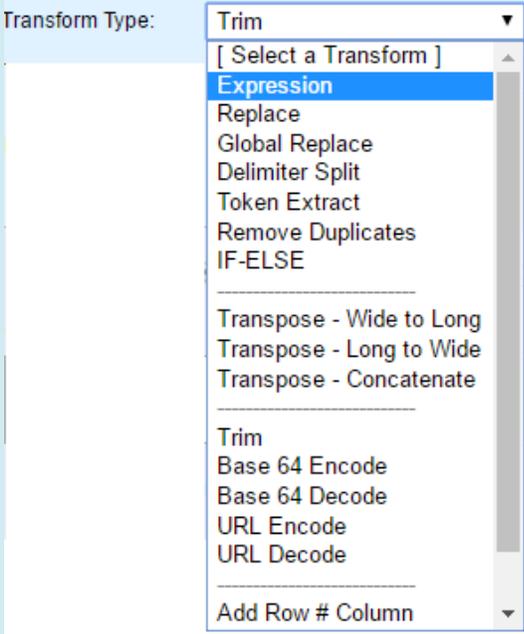


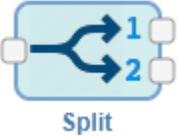
## Selected Nodes in STP

The following table lists some selected Nodes in SitScape's STP environment:

Node Icon	Node Name	Description
 Start	Start	This Node can trigger a process flow. It is fine to have multiple Start Nodes in one canvas, and the number in it will determine the starting <i>order</i> of each process flow on a canvas
 Scheduler	Schedule	This Node allow user to define schedules of process flow such as hourly, weekly, monthly etc.  A STP can be deployed to run automatically by the server per the schedule's configuration
 Database	Database Source	This Node can connect to database sources such as relational databases (Oracle, IBM DB2, MS SQL, MySQL, PostgreSQL), big data store (Hadoop) or Corporate Data Warehouse (Teradata) or MS Analytics Services (Cubes)
 SQL	SQL Node	Can run either static or dynamically generated SQL statement to targeted database or data sources that support SQL interfaces
 Table Select	Table Select	This Node allows user to pick table from a database source
 JSON / CSV	JSON Feed or CSV <u>Feed</u> Source	This Node allows user to connect to a JSON Feed or CSV Feed URL, and receive data from the feed. This can be an external Web services such as a RESTful microservice.

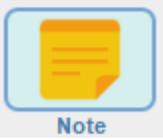
 File	File Data Source	This Node allows user to connecting to a file as the incoming data source, such as uploading a .csv data file
 Column Select	Column Select	This Node allows user to select columns or fields of a data table, change field names, change field type and subtype, re-order the fields in new order etc.
 Sort	Sort	This Node supports sorting of data in any combinations of multiple fields
 Aggregation	Aggregation	This Node support various math calculations based on Group-by over any number of category type of fields. It supports concatenation of strings of multiple rows too.
 Filter	Filter	This Nodes allow user to do various filtering operations on data based math values, string values, string patterns over regular expression, date and time values, GEO locations etc.
 ETL	Transformation Node	This Node has <i>extensive and broad</i> features. It actually contains many sub-Nodes such as

		<p>Transform Type:</p>  <p>For examples:</p> <ul style="list-style-type: none"> <li>• The <b>IF-ELSE</b> option is very powerful to do Boolean logic and advanced Expression to derive new values in new field or replace current field with sophisticated options</li> <li>• The <b>Expression</b> option supports math, text and logic <i>Expression</i> with various <i>Functions</i> to create new values from existing data values including <i>math</i> and <i>stats</i>.</li> <li>• The <b>Transpose</b> options support various pivot, un-pivot and concatenate transformation of data tables</li> </ul>
 <p>Fusion</p>	Fusion	This Nodes allow user to fuse multiple data sources into one table based on specified fusion options
 <p>Merge</p>	Merge	This Node allows user to combine data from multiple data tables with similar data structure into one data table
 <p>Decision</p>	Branching and Decision	This Node allows user to do dynamic <i>branching</i> based advanced logic, and can route process flow to the corresponding sub-branch based on run-time values and logical decision

	Split Node	Allow incoming data sets to be split into 2 subsets based on the splitting logic, and flow to each downstream sub-flow
	Looping	Looping Node allows user to do (1) <i>simple looping</i> of downstream process flows; or (2) support <i>iterating</i> upstream/incoming dataset itself record-by-record or per sub-records-set for downstream process flow's processing that similar to streaming
	Property Builder	<p>This Node allow user to create Name-Value-Pairs from upstream dataset, and make them into Property data structure, that can be used as parameters into many other Nodes.</p> <p>A property can also be optionally made as <i>global</i> so it is accessible from the whole canvas without explicit parameter passing.</p> <p>A property with it unique Name is dynamically updatable as well during the process flow.</p>
	String Builder	<p>This Node allows user to dynamically build any string at runtime with run-time values such as from the Property Builder Node in combination with String Template.</p> <p>It can be used to build <i>dynamic SQL Statement</i>, <i>JSON URL</i>, <i>RESTful API URLs</i>, <i>dynamic messages</i> or any other Strings on the fly.</p>
	Data Grid	<p>This Node basically allow data to be saved either in temporarily saved Data Table (can be local or global) or be presented as a visual Data Grid Table on the front-end user interface of a SitScape UDOP Page</p> <div data-bbox="678 1430 1409 1682" style="background-color: #e0e0e0; padding: 10px;"> <ul style="list-style-type: none"> <li><input checked="" type="radio"/> As a Temp Table</li> <li><input type="radio"/> Create a Data Grid on a SitScape Page</li> </ul> <p>Insertion Option:    <input checked="" type="radio"/> Replace Data    <input type="radio"/> Append Data</p> <p><input type="checkbox"/> Global</p> </div>
	Email Receive Node	This Node can connect to an email server via IMAP or POP3, and retrieve and download emails from the specified email server.

 <p>Email-out</p>	<p>Email Send Node</p>	<p>This Node can send out email with text or rich text content, and also attachments.</p>
 <p>SQL</p>	<p>SQL</p>	<p>This Node can take a dynamically generate SQL String statement and run against a targeted database.</p>
 <p>Database Storage</p>	<p>Database Storage</p>	<p>Allows user to output, save and persist data from STP into a range of database storages in a structured way. It supports replacement or appending as well.</p>
 <p>Alert</p>	<p>Alert</p>	<p>This Node allows user to apply a logic from upstream data, and then generate dynamic alert messages and send out to users via email, text message, or update a monitoring console panel visually..</p>
 <p>Generate Report</p>	<p>Generate Report</p>	<p>This Node allows user to dynamically trigger the generation of any report using a front-end UDOP page the user specified, including pages with very nice looking Report layout and various dynamic content defined by the user on that Page.</p> <p>This dynamic and nice report can be delivered to users via email on-demand as PDF file.</p>
 <p>Execute</p>	<p>Execution and Action</p>	<p>This Node allows STP to intelligently or conditionally trigger external applications and pass the relevant data at run-time to the external application for processing. This makes it <i>actionable</i> and also <i>extensible</i> to integrate with any other applications or commands in whatever language they use. Those can also be custom application developed by the customers and be triggered and executed on-demand.</p> <p>The output of those external applications or commands can be the input into STP as well for further automation and process flow.</p>
 <p>Stop</p>	<p>Stop</p>	<p>This Node will stop the continuation of the current path of STP. It will not stop the processing of other paths or branches of STP on the same Canvas.</p>

	<p>Download</p>	<p>This is basically an output Node to allow STP to output data as CSV file, JSON file, or as a JSON feed URL from the upstream STP processing.</p> <p>The JSON feed output from can then serve as an input to another STP flow for further processing in a new flow.</p>
	<p>Publish Node</p>	<p>This Node allows STP to output data as event and payload to a messaging bus, so other listeners on the message bus can pick up the data and do further processing.</p>
	<p>Subscribe Node</p>	<p>This Node allows STP to listen on message bus and pick up the data automatically when new data of interest arrives.</p> <p>For example, if user clicks on a record in a table on a Front-end Dashboard, if that event is subscribed by a Sub Node on a STP, it will trigger the process flow and execution of the related STP using this Sub Node and incoming data. This effectively enables the on-demand communication of a front-end user interface with a backend customizable process flow without coding or programming.</p>
	<p>(Web) Receiver Node</p>	<p>This Node can receive form-type of data submitted from UDOP Web front-end Web page such as the data content from a text box, a pull-down, a radio list, or a row in a data table etc.</p>
	<p>(Web) Output Node</p>	<p>This Node allows a STP to send data back to a Web Front-end, i.e. a UDOP page, which can be used dynamically to populate content of a UDOP Page via dynamic data binding using the outputted data from a backend STP process and may display them in UI components such as data grid, text box, label, pull-down, tree etc.</p>

 <p>Forecast</p>	<p>Model and Forecast Node</p>	<p>This Node supports predictive analytics, forecasting, or additional Statistical or Machine Learning (ML) Models.</p>
 <p>Note</p>	<p>Note Node</p>	<p>This Node allow user to add sticky Notes to the STP Canvas for easily annotation and comments.</p>
 <p>Sub-STP</p>	<p>Sub-STP Node</p>	<p>This Node allow user to create re-usable sub-STP that can be either called or included into other STP processes. Think it is a sub-process or a sub-routine.</p>
 <p>Train</p>	<p>Machine Learning Training Node</p>	<p>Train Machine Learning Models with input data sets and user selectable Machine Learning algorithms</p>
 <p>Trained Models</p>	<p>Machine Learning Trained Modes Node</p>	<p>This Nodes contains the trained ML models that can be used as input to the Machine Learning Predict Node</p>

	<p>Machine Learning Prediction Node</p>	<p>This Nodes takes input of data, and user selected trained model to do ML-based prediction</p>
<p>Prdfined or User-Defined Extensible Nodes</p>	<p>Predefined Nodes</p>	<p>Can create template Nodes based on pre-configured REST web services sources such as:</p> <div style="display: flex; justify-content: space-around; align-items: center;"> <div style="text-align: center;">  <p>Twitter Search</p> </div> <div style="text-align: center;">  <p>Watson Tone Analyzer</p> </div> <div style="text-align: center;">  <p>Azure Income Predict</p> </div> <div style="text-align: center;">  <p>Image Extract</p> </div> </div>

## Learn More

To learn more about how SitScape’s technology solutions can help you increase your situational understanding, optimize your analytics-based decision making, improve your secure information sharing and real-time collaboration capabilities, please contact us at [info@sitscape.com](mailto:info@sitscape.com), or call us at 888-762-6562, or visit our web site at <http://www.sitscape.com>.

## About SitScape, Inc.

SitScape Inc., the recognized leader of software solutions for Intelligent Digital Operations, is trusted by the Federal government and various Fortune 500 organizations for digital enterprise enablement, real-time collaboration, agile data correlation, continuous monitoring, analytical visualizations and flexible straight-through-processing (STP) automation. Our solutions support critical decision making at real-time with our self-service, easy-to-use, highly collaborative User Defined Operating Pictures (UDOP) graphical user interface, and the underlying engines with unmatched data correlation, analytics, monitoring/alerting and process automation capability for the next-generation digital operations.



**Smart Operational Intelligence**

**Business Performance Analytics**

**Real-time Collaboration**

*Multi-Source Data Correlation Fabric*

*Continuous Monitoring | Alert | KPI*

*Straight-Through-Processing (STP)*

*Shared Situational Awareness*

*Common-Operating-Picture*

*User-Defined-Operating-Picture*

*User-Defined-Operating-Picture*

*Common-Operating-Picture*

*User-Defined-Operating-Picture*