alteryx

Alteryx Technical Overview

v 1.0, August 2014

Contents

System Overview 3
Alteryx Designer 3
Alteryx Engine 3
Alteryx Service 5
  Alteryx Scheduler 5
  Alteryx Service Controller 5
  Alteryx Service Worker 5
  Alteryx Service Persistence 6
  Alteryx Service Client API and Direct Connect 6
Alteryx Gallery 7
  Alteryx Analytics Gallery 7
  Alteryx Private Gallery 7
Deployment 8
  Deploying Alteryx Designer 8
  Deploying Alteryx Server 8
Scaling Out Alteryx 9
  Handling Increases in App Execution Requests 9
  Allowing More Users to Create Apps 10
  Handling Increased Load on the Gallery 10
  Data Backup and Reliability 11
Security 12
Data Connectivity and Sources 12
Appendix 13
Communication with the Engine (SDK and API) 13
  Alteryx SDK 13
  Alteryx API 13
Integration with Other Systems (Email and Run Command) 14
  Email 14
  Run Command 14

System Overview

The Alteryx Analytics platform offers two deployment options: Alteryx Designer and Alteryx Server.

Alteryx Designer provides business analysts the ability to create repeatable workflow processes for accessing and blending data and performing advanced analytics (such as predictive, spatial, and statistical). Alteryx Designer includes Designer and Engine components and a local version of the Service.

Alteryx Server provides a comprehensive and scalable server-based analytics solution that provides end users and business decision makers the ability to share and run analytic applications in a web-based environment. It also gives users of Alteryx Designer the ability to schedule their workflows to run at specific times. In addition to Designer and Engine components, the Alteryx Server includes Service and Gallery components.

Alteryx Designer

Alteryx Designer is a Windows software application that provides an intuitive drag-and-drop user interface for users to create repeatable workflow processes. Users can drag tools from a toolbox onto a canvas, connect them together, and edit their properties to create Alteryx modules and apps. Users can use these workflows to blend and enrich data from a range of sources, perform advanced analytics, and quickly produce results that can be easily shared with others.

The unique drag-and-drop workspace within Alteryx allows users to run workflows as they are being developed in order to instantly access and process data. This allows for continuous enhancements and refinements to the workflow and data results. Additionally, users can view data anywhere within a created workflow stream as a table, map, graph, or other visual representation.

Alteryx Designer is written mostly in C# and executes the workflow modules and apps through a local instance of the Alteryx Engine. In an Alteryx Server deployment, the Scheduler interface component within Alteryx Designer allows users to schedule modules and apps to be executed at predetermined times or specific recurring intervals. The Scheduler then communicates with the Alteryx Service where jobs are queued for execution at the appropriate time. Additionally, users may use Alteryx Designer to publish their workflows to the Alteryx Analytics Gallery as apps where others can have access to running them.

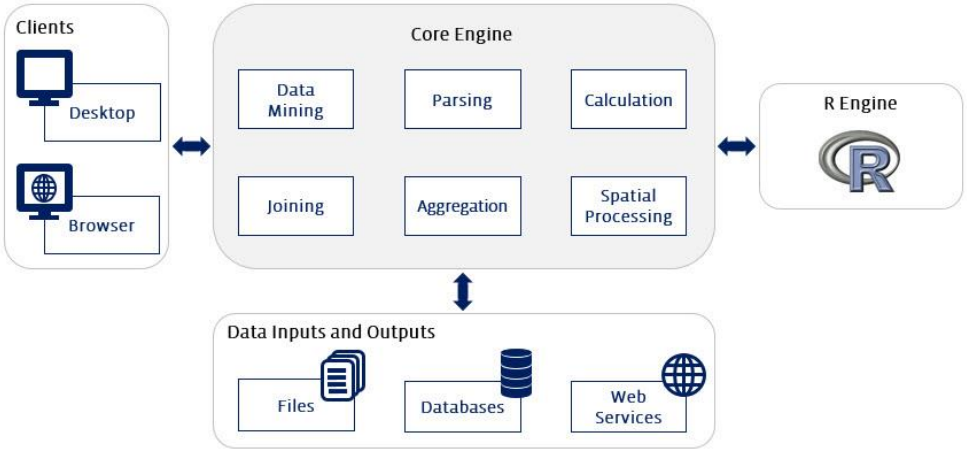
Alteryx Engine

The Alteryx Engine, which is written in C++, executes the workflows that are built as modules or apps in the Alteryx Designer and produces the output. The Engine supports direct connections to various data sources for accessing the data and then processes it in-memory during the execution of the workflow. Processing that exceeds memory limitations is written to temp files on disk, which are deleted once the processing is complete. The Engine can be entirely self-contained in an Alteryx Designer deployment,

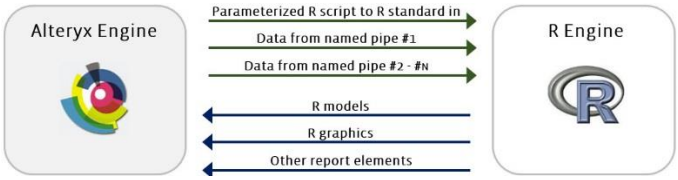
scaled across an organization via the Alteryx Server, or deployed in the cloud via the Alteryx Analytics Gallery.

Depending on the workflow module and process, the Alteryx Engine may:

- Read or write to input/output files.
- Read or write to one or more databases.
- Upload or download data from the web.
- Submit email to an email server through SMTP protocols.
- Execute external runtime commands.



The Alteryx Engine seamlessly integrates with R. When users choose to install the suite of R-based tools and macros used for predictive analysis, Alteryx also installs the R program and the additional packages that provide connectivity between Alteryx and R and are required for the tools and macros to work. Users can use R scripts or the R-based macros to process data directly within a workflow. When the workflow is run, the Alteryx Engine communicates with the R engine via a command line executable to process and send back the data.



Developers can write applications that call directly into the Alteryx Engine using the Application Programming Interfaces (APIs) and the Alteryx Software Development Kit (SDK) gives developers an easy way to add their own programs to the Alteryx toolbox. Both the Alteryx API and SDK are available with an Alteryx Server deployment. *For more information refer to the Appendix.*

Whether the Alteryx Engine is running on the desktop, through the API, or via the Alteryx Service, modules may leverage tools that communicate with databases, output files, web services, email servers, or run commands. *For a list of data connectivity options refer to the Alteryx website:* http://www.alteryx.com/products/technical-specifications.

Alteryx Service

The Alteryx Service allows the Alteryx Engine to be deployed across multiple servers, providing a highly scalable architecture for the scheduling, management, and execution of analytic modules and apps.

The Alteryx Service, which is written in C++ with some C# wrappers, can be deployed across multiple servers by using a Controller-Worker architecture. This means one server acts as the Controller and manages the job queue, and the others act as the Workers and perform the work. The Service relies upon the Service Persistence tier to store information critical to the functioning of the Service, such as Alteryx application files, the job queue, and result data. The Service also serves content and information to the Gallery when it requests it.

Alteryx Scheduler

The Alteryx Scheduler is the component that provides the ability for a user to schedule the execution of workflows developed within the Designer, either on a recurring basis or at a specific time in the future. The Scheduler interface component, which is written in C#, is accessed via the Designer and is used to specify the date, time, and frequency for running a module or app. The Scheduler then communicates with the Alteryx Service Controller, which manages the job queue. In an Alteryx Server deployment, the Scheduler interface can be set up to communicate with the Controller on the same machine, or it can be configured to connect to a remote Controller so that modules and apps are processed on a different machine.

Alteryx Service Controller

The Alteryx Service Controller acts as an orchestrator of app executions and is responsible for the management of the service settings and the delegation of work to the Alteryx Service Workers. The Controller receives jobs from the Scheduler, looks at them within the persistence layer where it maintains all of the queued jobs, and delegate jobs to the Workers. If Alteryx is deployed across multiple servers, only one machine may be enabled as a Controller.

Alteryx Service Worker

The Alteryx Service Worker contains an instance of the Engine and is responsible for executing the modules and apps. Once the Controller delegates a job to the Worker, the Worker runs it and produces the output. There must be at least one machine enabled as a Worker in order to execute applications through the Service. You may configure the same machine to be both the Controller and a Worker or they can be configured on separate machines.

The actual number of Workers needed will depend upon the required performance for the system. If the Worker is not the same machine as the Controller, then the connection information for the Controller, which includes the Controller's name or IP address and the unique security token for that Controller, must be specified in order for the Worker and Controller to communicate with each other.

Quality of Service (QoS)

The Quality of Service (QoS) setting is used to manage resource allocation in an environment where multiple Workers are deployed by restricting which jobs will be run by each Worker. For normal operation with one machine configured as a worker, the QoS value should be set to 0. The QoS value for a Worker can be set higher for specific types of jobs to allow the Worker resource to be reserved for higher priority requests.

For example, the default QoS settings in the Alteryx Gallery are:

- Normal application execution: QoS = 0
- Chained application execution: QoS = 4 (all apps in the chain aside from the last)
- Application validation requests: QoS = 6

When a job request is handled by a Worker, it will compare the QoS value of the job to the Minimum QoS Priority for the Worker. Only jobs that have a value greater than or equal to the Minimum QoS Priority for the Worker will be handled by that Worker. For instance, if a Worker has a Minimum QoS of 0 (the default value) that worker will handle any requests if it is available. However, a Worker with a Minimum QoS of 5 will only handle jobs that have a QoS of 5 or higher. This allows resources to be reserved for higher priority requests.

Alteryx Service Persistence

The Alteryx Service includes a persistence layer that it uses to store information critical to the functioning of the service, such as Alteryx application files, the job queue, and result data. The Service supports two different mechanisms for persistence: SQLite and MongoDB. For lightweight and local deployments, SQLite is adequate for most scheduling needs. For heavier usage, or if the Alteryx Gallery is deployed, MongoDB must be used.

If MongoDB is used, you can elect to either use the embedded MongoDB support native to the Alteryx Service or connect the Service to your own implementation of MongoDB. The embedded MongoDB supported is limited to one server and does not support replica sets or automated backups. It is highly recommended that you provide an automated backup system for whatever persistence mechanism you choose.

Alteryx Service Client API and Direct Connect

The Alteryx Service includes a client API through which it communicates with the Alteryx Gallery and the Alteryx Designer Scheduler to do things such as schedule modules, view schedules, etc. The client API consists of a low-level C++ library (AlteryxService_Client.dll) as well as a higher-level .NET assembly (AlteryxService_Client.Net.dll).

For performance reasons, the client API can communicate directly with the Alteryx Service Persistence tier for the majority of the requests it makes. This “Direct Connect” functionality is enabled by default but can be overridden if necessary. If Direct Connect is turned off, all communications from the client API will be funneled through the Service Controller.

At startup, the client API will communicate with the Service Controller to determine if the “Direct Connect” functionality is enabled and to obtain the MongoDb connection

information if it is. Then, for requests that can be fulfilled directly from the database, the client API will talk directly to the persistence tier.

The Alteryx Service Client API is currently only consumed internally within the product. External use or consumption of the API is not yet available.

Alteryx Gallery

The Alteryx Gallery is a cloud- or self-hosted application for publishing, sharing, and executing apps. Alteryx offers the Alteryx Analytics Gallery (https://gallery.alteryx.com/) where users can sign up and share apps publicly or with selected users. An Alteryx Server deployment allows companies the ability to offer a private Gallery to their internal users hosted on their own server infrastructure.

The Gallery user interface, which is written in Javascript, HTML5, and CSS3, communicates with the Gallery server component via HTTP requests using a REST API, which is a JSON transport protocol. The Gallery web server is written in C# and WCF (Windows Communication Foundation) and handles all the backend logic for what is displayed within the Gallery. WCF receives the HTTP requests and sends them to the server to handle the logic. The server then communicates directly with the MongoDB for handling its persistence, including information such as: the users who have registered with the system, the collections that exist, and the apps that are in those collections.

The Gallery server also communicates with the Alteryx Service for the management and execution of analytic apps. When someone requests an app to be executed the Gallery server communicates with the service layer to execute those apps and then retrieve the output.

Alteryx Analytics Gallery

The Alteryx Analytics Gallery (https://gallery.alteryx.com/), where users can sign up and share apps, is hosted on an Amazon EC2 deployment and consists of several nodes. Web requests are filtered through the elastic load balancer available within Amazon Web Services. The load balancer sends the requests for the web pages to multiple web servers.

The web servers communicate with the Controller when there is work to be processed. If someone wants to run an app that has been published to the Gallery, the web server will communicate with the Controller to schedule the job to be run immediately. The Controller will send the work to the Workers, the Workers will process it, and the results will be sent back to the web server for the end user to see.

Alteryx Private Gallery

A private Gallery allows organizations to deploy the same architecture used in the Alteryx Analytics Gallery to their internal users within their own IT server infrastructure. This provides a greater level of control over the system as well as allowing internal data to remain inside an organization's firewall.

By default, a private Gallery will be configured to run on a single server, but it can be configured to work across multiple servers in the same way as the Alteryx Analytics Gallery. Each layer of the Gallery architecture is independently scalable, allowing a private Gallery deployment to be customized to each individual organization's needs.

Deployment

The Alteryx Analytics platform offers two deployment options: Alteryx Designer and Alteryx Server. Decisions around selecting a deployment option should take the following into consideration: analytics processing goals, machine specifications, number of users and their needs, automation required to service jobs, access to APIs and SDKs, and volume of data to be processed. *For system requirements refer to the Alteryx website:* http://www.alteryx.com/products/technical-specifications.

Deploying Alteryx Designer

Alteryx Designer provides business analysts the ability to create repeatable workflow processes for accessing and blending data and performing advanced analytics (such as predictive, spatial, and statistical). Alteryx Designer includes Designer and Engine components and a local version of the Service.

A deployment of Alteryx Designer is typically done on a desktop and all data is processed in real time on that machine. Although a Scheduler interface is available via the Designer for scheduling work to be processed at a predetermined time or frequency, it can only be used if there is a connection to a machine that has Alteryx Server installed.

Alteryx Designer may use around 600 MB of space when installed on a single desktop. The Engine is designed to make use of the resources available on the machine on which it is deployed and its performance is dependent upon the hardware specifications. The Engine consumes more memory, processor, and disk space when workflow modules and apps are executed.

The Engine may run modules that use tools that communicate with databases, output files, web services, email servers, or run commands. The Engine performance is dependent on the ability to access data and bring it into the workflow, as well as the amount of temp space available for storing files.

Alteryx provides a wide assortment of datasets for analysts who integrate data into their analysis. Depending on an analyst's needs, an install of the data may use up to 180 GB of space on a machine. *For information about available datasets refer to the Alteryx website.*

Deploying Alteryx Server

Alteryx Server provides a comprehensive and scalable server-based analytics solution that provides end users and business decision makers the ability to share and run analytic applications in a web-based environment. It also gives users of Alteryx Designer the ability to schedule their workflows to run at specific times. In addition to Designer and Engine components, the Alteryx Server includes Service and Gallery components.

An Alteryx Server deployment provides for the scheduling and sharing of the workflow processes that have been built via the Designer. Additionally, once the Server is installed

it is scalable to accommodate for higher levels of usage by enabling certain components on additional machines.

With the Alteryx Service (which includes the Controller, Worker, and Persistence) jobs can be scheduled to be executed at predetermined times, or at specific recurring frequencies, by connecting to the Controller via the Scheduler interface in the Designer. The Service manages the scheduled job queue, how the job processes, and the results and logs associated with the job in a persistence layer that can be either SQLite or MongoDB. Both SQLite and MongoDB are included in the Server installation, or an existing MongoDB instance may be used.

The Alteryx Gallery component is a web-based application that allows users to publish, share, and execute apps in a browser. The Gallery web server communicates with the Service for the management and execution of apps and a MongoDB persistence layer for maintaining information about the users who have registered with the system, the collections that have been created, and the apps that are in those collections. When deploying Alteryx Server, companies can take advantage of all the functionality provided in the Alteryx Analytics Gallery while offering a private Gallery to internal users hosted on their own server infrastructure

The Alteryx Service is composed of several different processes that may be running on the machine to handle various aspects of the Alteryx Server functionality such as the Gallery web interface, execution and scheduling of the modules, and the database persistence layer. To ensure system stability, the Service continually monitors these processes and if a process is not running, the Alteryx Server will try to start it back up. If it is unsuccessful after multiple attempts, the Alteryx Server will stop all of the other processes, shut down, and log the problem.

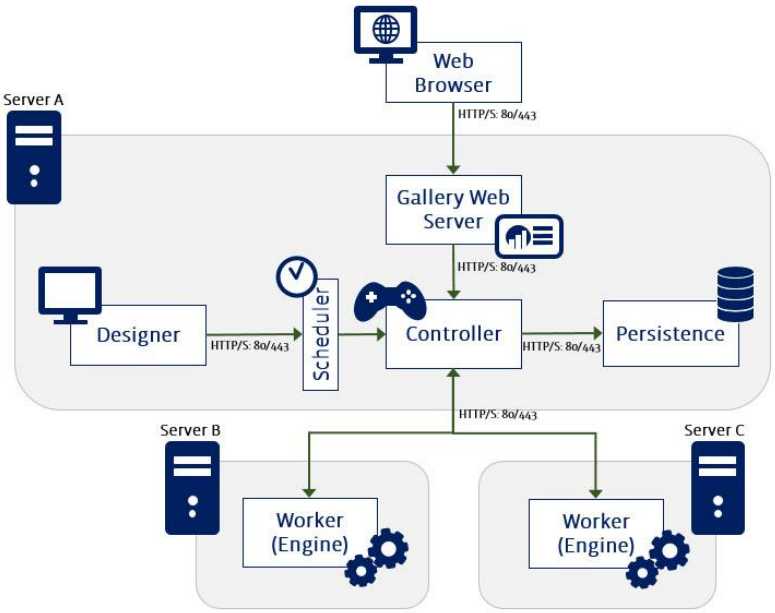
Scaling Out Alteryx

The complete Alteryx Server can be enabled on a single machine, selected components can be enabled on a machine, or certain components can be enabled on multiple machines to accommodate for higher levels of usage.

A typical deployment of Alteryx Server includes installing the Controller, Worker, and Galley web server component, on a single machine. There can only be one instance of the Controller, but the Worker, Designer, or web server components can be enabled on individual and multiple machines as the need to handle different workloads increases. Decisions around which Alteryx components to scale out are based on the type of work that is needed and who is doing it.

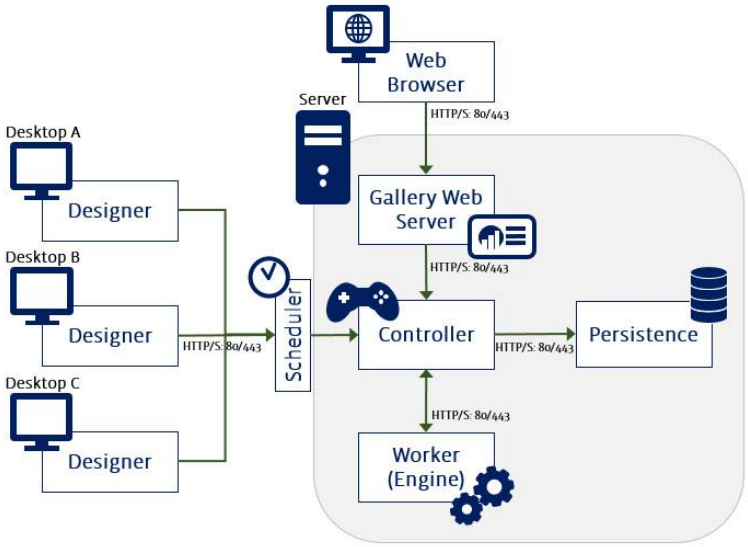
Handling Increases in App Execution Requests

If more work needs to be processed and there is not enough simultaneous processing to be able to handle it, the Engine processing capabilities can be increased by configuring multiple machines to act as Workers. Each additional Worker machine must be configured with the unique security token of the Controller to be able to communicate with it. Then, as app execution requests begin to queue up, each Worker machine will communicate with the Controller and the Controller will delegate a job for it to process.



Allowing More Users to Create Apps

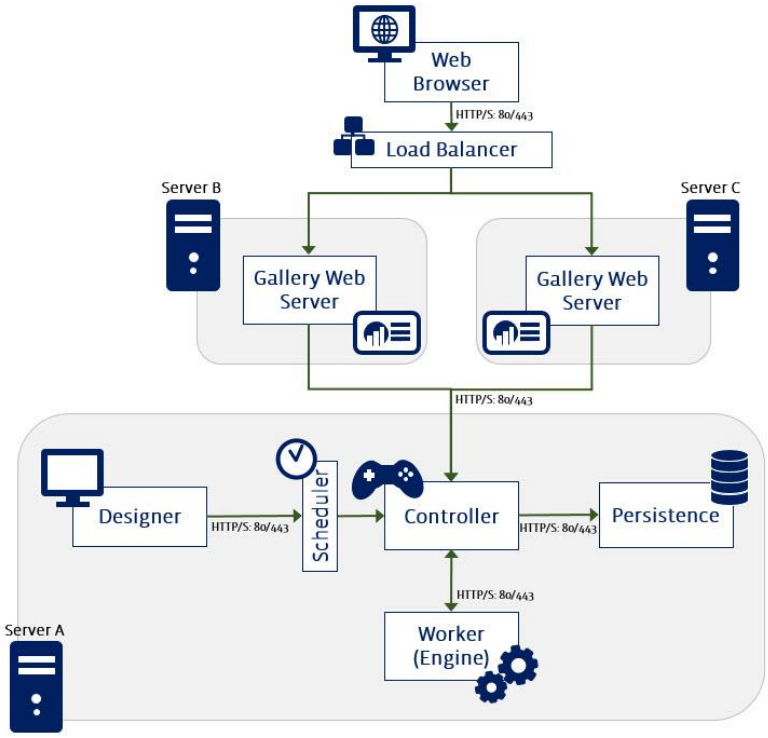
If there are multiple end users using more lightweight machines to create apps and modules in Designer within their own desktop, each machine can be configured to communicate to the same Controller so jobs can be scheduled to be processed on a more robust Worker machine. The Worker will execute the job and provide the output which the users can retrieve from their desktop.



Handling Increased Load on the Gallery

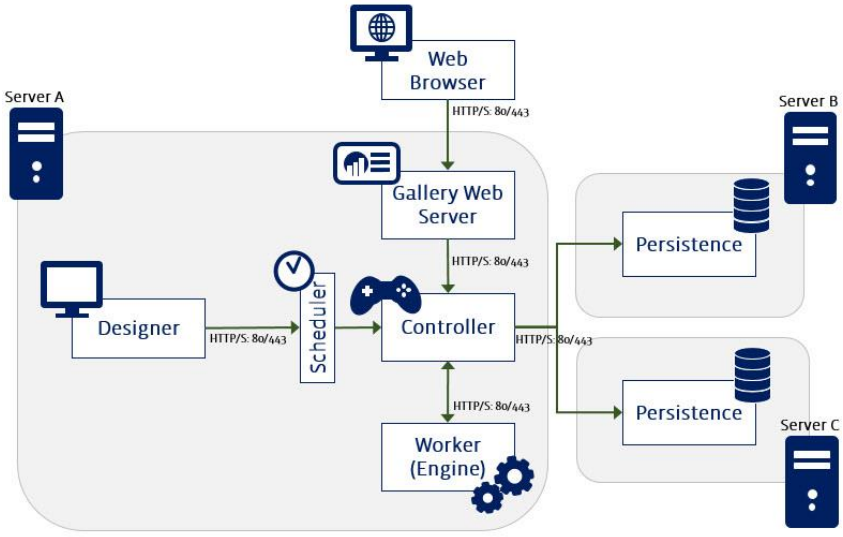
If web requests increase because there are many users who need to access or share the apps published to the Gallery, the Gallery web server can be scaled out by configuring additional machines to act as web server nodes. When setting up multiple web server nodes these nodes would generally be behind a load balancer to help distribute the

amount of web requests across all the web servers. Each individual web server will then communicate to the same Controller when there is work to be processed.



Data Backup and Reliability

If a machine is configured with only the Designer and Scheduler, SQLite can be used as the persistence layer. However, if the Gallery component is enabled on the machine as well, MongoDB must be used. Alteryx includes an embedded version of MongoDB for ease of setup and use but the embedded MongoDB is limited to one server and does not support replica sets or automated backups. If you deploy your own instance of MongoDB and manage it yourself, you can follow the MongoDB recommendations for scalability and redundancy.



Security

When implementing Alteryx Designer there are no security implications when it is installed locally on a desktop. Access to data assets is governed by the database or network permissions that are setup by a company's IT department at the local level drive.

For an Alteryx Server implementation, there are multiple security methods built in to the Gallery. The Gallery supports both built-in authentication (which requires email address and password for each user to sign in) and Windows authentication (which uses Windows active directory). For built-in authentication user registration information and log-in passwords are encrypted and not sent or stored in clear text. For Windows Authentication, instead of storing the password and network credentials of users, Alteryx refers to them via an SID within the active directory.

Users can be permissioned for roles that govern what they can and cannot do within the Gallery, such as publish apps or administer the site itself. Additionally, a user can only be logged in and have one active session in the Gallery at any time and the session will remain active for 1 hour after the user's last interaction before logging them out. The Gallery supports SSL encryption by using HTTPS for the Alteryx Analytics Gallery (https://gallery.alteryx.com). The Gallery web server communicates with the service layer through a method that encrypts the information as well.

Alteryx also supports SSL for communicating with the MongoDB persistence layer. MongoDB has its own REST HTTP interface for sending data and SSL can be enabled to encrypt the data that is being sent as well as the general user name and password authentication.

A machine using Designer, or a machine set up as a Worker, that needs to communicate with a different machine that has the Controller enabled must have the token for that Controller. It is that token that is added to the Schedule Module screen in Designer, or the System Settings window for the Alteryx Server, to establish communications between the two machines.

For information on how Alteryx integrates with other systems via Email and Run Command tools and events see the Appendix.

Data Connectivity and Sources

Alteryx provides data connectivity options to enable users to access and integrate data from many different sources. The various types of data sources Alteryx supports include relational databases, document stores, flat files, and web services.

Through partnerships with leading data providers, Alteryx also provides a wide assortment of datasets for analysts who require spatial or syndicated data to be integrated into their analysis. Some of this data is included when Alteryx Designer or Alteryx Server is installed and additional data can be added after installation.

Data is bundled into packages in a way that allows users to easily install all data at once or to install specific datasets within a package individually. The complete data package may be up to 180 GB in size when installed on a machine or network.

For a current list of data connectivity options and information about available datasets refer to the Alteryx website.

Appendix

Communication with the Engine (SDK and API)

Developers can write applications that call directly into the Alteryx Engine using the .NET or C++ Application Programming Interfaces (APIs). The Alteryx Software Development Kit (SDK) gives developers an easy way to add their own C and C++ programs to the Alteryx toolbox. The Alteryx API and SDK, which are available with an Alteryx Server deployment, give developers the freedom they need to deploy high-performance solutions that serve their user communities while maintaining control of the business logic.

Alteryx SDK

The Alteryx SDK allows developers to create their own tools for use within Alteryx modules. The SDK is available as a set of C++ libraries or through a .NET wrapper. All of the tools in the Desktop Designer were built with the Alteryx SDK.

Alteryx modules consist of a series of tools connected together as a workflow. Each of these tools falls into one of three categories:

- Input tools, which generate data either spontaneously (such as a random number generator) or by reading information from a data source (such as a file or database).
- Output tools, which consume data either displaying it (such as a table or map) or storing it to a data source.
- Computation tools, which consume data, process it, and then generate new or additional data.

All tools, regardless of their type, are implemented in a nearly identical fashion. Each tool has a user interface implementation, which handles the visual rendering of the tool object as well as its configuration properties dialog. Each tool also has an engine implementation, which performs the actual processing work. These two pieces together form the single tool.

When needed, Alteryx calls into a tool's user interface implementation to provide its Properties view. When the module runs in the Alteryx Engine, an instance of each tool's engine implementation is created, and the outputs of upstream tools are connected to the inputs of later tools. Alteryx then invokes each input tool in the module. During the run, tools push their output directly to one another, minimizing execution overhead.

User interface implementations can be written in any .NET language. Alteryx uses C# for this purpose, but others have been successful at using VB.NET. Engine implementations are written in any language which can produce DLLs with an unadorned C-style binding. Alteryx uses Visual C++ for its engine implementations, with a few exported C-style functions to provide the binding layer.

Alteryx API

The Alteryx API provides multi-level integration access to both manage the Alteryx Engine and also work directly with Alteryx supported data formats. The APIs consist of native

C++ dlls with.NET wrappers that allow you to access various portions of the Alteryx functionality.

Included with the APIs are two high-level .NET wrappers for running Alteryx modules.

- AlteryxAPI.Net.dll: This assembly includes classes for running modules and apps, accessing properties and settings for Alteryx, as well as high-level classes for accessing Allocate, Solocast and Calgary functionality.
- AlteryxDocument.dll: This assembly provides access to Alteryx module properties and settings, including runtime settings for wizards and macros.

In order to work with the API, simply add the appropriate assemblies to your .NET project. The Alteryx installation includes documentation with its APIs in order to help get you started. There are also direct APIs for specific functionality (Allocate, Solocast, Calgary, etc.).

Integration with Other Systems (Email and Run Command)

Alteryx can be expanded further and integrated into other systems by way of Email and Run Command tools and events.

Email

Using the Email tool or event, Alteryx can prepare outgoing emails and submit the emails through the SMTP protocol to any SMTP-enabled email server. The user can:

- Set the To, From, CC, and BCC fields.
- Create the subject line.
- Create the body of the message.
- Attach documents.

Run Command

Using the Run Command tool or event, Alteryx allows for the execution of command line functions through the “Run” shell. Common uses for this command include integration through:

- Batch commands to execute external processes.
- Complex file transfers.
- Custom scripts or other add-ons.