

**Solution Design Primer**

A process primer for your organisation ro assist deisgning a solution using Eightwire

# Document Control

Document Information

|  |  |
| --- | --- |
| **Date:** | 11/05/2018 |
| **Owner:** | Development Team, Eightwire,  |
| **Approved by** | Jason Gleason, CEO |

Revision History

|  |  |  |  |
| --- | --- | --- | --- |
| **Version** | **Key Deliverables** | **Author** | **History** |
| 1.0 | 11 May 2018 | Kevin Rumble | First Draft |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

Table of Contents

Document Control 2

Document Information 2

Revision History 2

*Table of Contents* 3

Table of terms used in this document 4

Introduction 5

Purpose 5

Audiences 5

Scope 6

High Level Architecture 7

Eightwire Platform 7

Security Certification 13

Availability 13

Systems, Infrastructure, and Data 14

Connecting to Eightwire 14

Data preparation and consumption 15

Data staging 16

Agent Server 16

Network security 17

Processes 18

Operations 18

Monitoring 18

Support 19

Roles 20

Further Information 22

Table of terms used in this document

Terms used in this document have a specific meaning in a security context. This table provides clarity for those words where they may differ with the normalised language of your organisation.

| Description | Definition |
| --- | --- |
| Eightwire | The platform that is used to transfer data |
| Responsible | The role that ensures the necessary activities for a particular task gets done. |
| Eightwire Agent | A small piece of software that is installed on a Windows machine at an organisation that make secure HTTPS web service calls to the Eightwire services. |
| MOU | Memorandum of Understanding covering the exchange of data between two organisations. |
| Agent server orAgent hosting server | The terms used to refer to the Windows machine (physical or virtual) that hosts the Agent. |
| Data staging point or Data staging area | The terms used to refer to the location that data is presented to the Agent and written to the Agent. |
| Data preparation | The term used to refer to the components and processes that deliver prepared data to the data staging point. |
| Data consumption | The term used to refer to the components and processes that copy data from the data staging point to elsewhere in an organisation. |

Introduction

Purpose

This document presents the material for a detailed solution design for implementing connectivity to Eightwire Platform. Information in this document is intended to be drawn on and added to design documents for organisations connecting to Eightwire Platform.

Sufficient information should be included to allow suitably skilled and experienced people to build and deploy connectivity to Eightwire Platform. Whilst this document is considered comprehensive it is not exhaustive.

If additional information is sought please contact the Data Exchange Architect or the Data Exchange Technical Owner.

Audiences

This document is intended for the following audiences:

* ICT Management
* Security Architects
* Infrastructure Architects
* Solution Architects
* Data management practitioners
* Project managers

Scope

**In Scope**

The following componentry is in scope for this design primer:

1. Installation and configuration of the Windows Server virtual machine to host the Eightwire Agent.
2. Installation and configuration of the Eightwire Agent.
3. Installation and configuration of database drivers for data access by the Agent.
4. Design options for database storage of transferred data.
5. Design options for file system storage of transferred data.
6. Generation and configuring of an Active Directory user account to act as the Eightwire Agent service account.

**Not In Scope**

The following componentry is out of scope for this design primer:

1. The structure, format, and specific data to be transferred by the Eightwire.
2. Configuration of the Eightwire services located within the Revera Data Centre and Microsoft Azure.
3. Firewall / Proxy / Load Balancing configurations within the Revera Data Centre and Microsoft Azure.

High Level Architecture

Eightwire Platform

Eightwire branding of a product named Conductor, provided by Eight Wire Limited.

Eightwire is a data transfer platform hosted in the cloud, between two data centre providers. The platform consists of a management service that orchestrates data transfers (hosted solely at Microsoft Azure in Australia), and a data processing service that transfers data (that can be hosted anywhere if required).

Communication with the platform takes place using REST-style APIs. To make the platform more accessible (by not requiring bespoke software development) a small piece of software named the “Agent” is installed on a Windows machine at an organisation wanting to transfer (send or receive) data. The Agent must be able to access the data being read or the location data is to be written to. The data storage location is recommended to be a specific data staging point rather than the Agent directly connecting to operational systems. The Agent makes API calls to the platform and passes data to and from the platform. No bespoke development is required in this mode.

The Eightwire can be illustrated as the following interacting components:



* Cloud-based data transfer application stack.
* Organisation-hosted Agents (a MS Windows service) that read and transfer data from and to secured data staging locations in each connected organisation.
* Legal agreements between data transferring organisations that are the basis of permitting data to be accessible to the Eightwire.

The following points are important to take note of:

* Data will be transferred from and to secured landing locations in each organisation.
* Transferred data will generally be about people, including personally identifying data, however initial deployment of Eightwire during Prod-1 only transferred anonymised data about people in order to embed essential technology controls and business discipline.
* Transferred data will be used for a broad variety of purposes, including research, policy development, and performance measurement. Following Prod-1 data may be used operationally for service delivery.
* The data containers can be either files or databases.
* Each organisation has complete control of the data staging location, the schedule of the data being transferred, and the actual data being transferred.
* Data that is to be transferred is encrypted in transit by the Eightwire agent using industry standard SSL (TLSv1.2).
* Data is retained at the Data Exchange only for the duration of the transfer between organisations.

Below is a diagram of the types of inter-process communications that take place:



A detailed walk through of inter process communications:

* Agents are light weight clients that make secure HTTPS web service calls to the Eightwire services. Agents are not required to transfer data, they are readily manageable end points for interfacing to data stores without involving bespoke development by connecting organisations.
* Agents initiate all communications.
* When the Agent is not involved in any task it polls the Management Service every 10 seconds, requesting instructions. The Management Service may respond with instructions for the Agent to update itself (manageable), retrieve data and deliver it to a specific Data Processing Service instance, or retrieve data from a specific Data Processing Service instance and write the data to a nominated data store.
* When the Agent becomes involved in the orchestration of a data transfer it polls the Management Service as frequently as every 1 second.
* The Management Service periodically polls the Data Processing Service instances to ascertain their availability and processing capacity. This information is used to direct Agents to use specific Data Processing Service instances, or instantiate a new Data Processing Service instance. In effect, load balancing management.
* The Agent and the services mutually authenticate with each other using embedded certificates, trusted by each party. This occurs while establishing the TLS 1.2 connection.
* Agents encrypt all connections to Processing Services via TLS 1.2 and from the Data Processing Service using the same encryption before transferring data to Agents.
* When an Agent receives an instruction to retrieve and transfer data:
	+ A schedule associated with the receiving Agent triggers the Management Service to inform both the receiving and sending Agents to reduce their polling interval from 60 seconds to 5 seconds.
	+ The Management Service directs both Agents to use the same Data Processing Service instance.
	+ Depending on the type of data container, a query string may be delivered to the sending Agent and this is executed by the Agent using either a built-in data format library or using an external data format library.
	+ The service account for the sending Agent is the identity the Agent will present to any data hosting process.  This approach mitigates the need to store credentials for data access.  An exception to this is where a data hosting service requires credentials to be provided, such as SQL Server when not using Windows integrated authentication.
	+ Data is retrieved by the sending Agent in batches of 1,000 rows, transformed to a proprietary format, compressed, encrypted via SSL, and transferred to the Data Processing Service.
	+ Multiple read threads can be executed in parallel and the extent of parallelism can be managed on a per-Data Store basis to manage load on the machines involved in retrieving the data.
	+ The sending Agent transfers the encrypted data to a Data Processing Service instance as directed by the Management Service.
* When data arrives at the Data Processing Service:
	+ Upon arriving at the Data Processing Service the data is decrypted and transforms are applied, if configured.
	+ The data is re-encrypted and buffered to disk. Buffered data is immediately deleted after use. This buffering takes place in the form of encrypted BLOBs and does not utilise native SQL Server cryptography.
	+ Encrypted data remains with the Data Processing Service until it can be delivered to the receiving Agent.
* When an Agent receives an instruction to transfer and receive data:
	+ The Management Service instructs the receiving Agent to contact a specific Data Processing instance to start receiving data.
	+ The Data Processing Service packages data into a proprietary streaming format, compresses the data, encrypts the data, and deliver the data to the receiving Agent.
	+ Upon arriving at the Agent the data is decrypted, decompressed, re-formatted for the destination container, and written to the destination container.
	+ Once all data has been streamed successfully the receiving Agent notifies the Management Service of successful transfer and the Processing Service discards the data just transferred.
* Eightwire supports several modes of data maintenance:
	+ Append: Data is appended to the end of any existing data in the container.
	+ Overwrite: Any existing data is first deleted before new data is written to the container.
	+ Partial Merge: New data is appended, matching existing data is updated. Non-matched existing data is unaffected.
	+ Full Merge: New data is appended, matching existing data is updated. Non-matched existing data is deleted.
	+ Change Set: The user nominates a scalar field in the source and destination data stores. Where this field contains greater values in the source than any in the destination, this row is appended to the destination. No other rows are affected.
* Data transfers are atomic tasks and are unidirectional.  Composing many transfers and of different types between several Agents can result in novel capabilities.  This orchestration is executed by the Management Service.
* Eightwire maintains serialisation of delivered data as it is read by the sending Agent.
* On platforms that support transactions, Eightwire endeavours to partially rollback uncompleted transfers at the data block level rather than at the batch level. This means that within any given 1000-row block, if an error occurs, that block will not be written. The best way to recover from write issues is to simply re-run the process a second time.
* Agents do not call or respond to calls to or from processes other than Eightwire services.  A corollary to this is that Agents do not handshake with any processes within customer sites.  This lack of handshaking results in uncoupled task execution between data preparation/presentation to the Agent and the Agent reading the data.  As the data receiving party controls the data transfer schedule, both sending and receiving parties must agree a schedule that coincides data availability and data transfer. Eightwire have committed to considering handshaking options to enable more robust synchronisation of tasks.

Security Certification

Eightwire has been certified by the New Zealand Government – Ministry of Social Development, Chief Information Security Office.

Security certification was at IN CONFIDENCE and was raised to SENSITIVE in November 2017. A copy of the Security Risk Assessment (SRA) and Security Certificate are available on request.

Also available from Eightwire is a Security Risk Assessment primer to aid organisations

connecting to Eightwire to assess their security risks.

Availability

Eightwire has 99.95% uptime of Eightwire Platoform across Azure’s infrastructure. Failover of Eightwire services is automatic. The availability requirements of the connectivity to Eightwire will vary for each organisation based on the role and business dependence on Eightwire as a data transfer channel. If an organisation is predominately a data sharer then the availability expectations of the organisations receiving the shared data will higher than organisations just receiving data.

Organisations should ensure they have fallback data transfer channels in place in the event Eightwire becomes unavailable or their connectivity to Eightwire fails.

Systems, Infrastructure, and Data

Connecting to Eightwire

Eightwire is a REST-style API communicating collection of services. Bespoke

development to use Eightwire Services is certainly possible but for the majority of

data transfer scenarios this is unnecessary due to the existence of Eightwire

Agent.

The Agent is a light weight (less than 2 MB) Windows service that communicates outward to

Eighwtire services using web services (an interface specific to the Agent) and

inward to data stores using native or locally installed drivers. The Agent effectively shifts the

boundary of Eightwire from being Eightwire provider’s cloud to being

within another organisation’s environments.

This approach results in no bespoke development required by a connecting organisation,

merely a suitable secured data store (file system or database) the Agent can read from and

write to, and permitting network configuration to allow the Agent to conduct HTTPS

communications on TCP 443.

Some new language:

* The term “Agent server” or “Agent hosting server” is used to refer to the Windows machine (physical or virtual) that hosts the Agent.
* The term “Data staging point” or “Data staging area” is used to refer to the location that data is presented to the Agent and written to the Agent.
* The term “data preparation” is used to refer to the components and processes that deliver prepared data to the data staging point.
* The term “data consumption” is used to refer to the components and processes that copy data from the data staging point to elsewhere in an organisation.

The general functional domains of the infrastructure architecture for connectivity to Eightwire are (the arrows indicate only flows of data):



Data preparation and consumption

When organisations are forming a data sharing arrangement a specification for the format and content of the data should be defined.

The format of the data being transferred should in general suit the party receiving that data - this is due to the distribution of data transfers roles that are envisaged across the industry.

Following mutual agreement on the business and legal basis for sharing data, and a data format specification has been agreed, a technical process for generating data in the agreed format and ready for the agreed transfer schedule is required. This design primer does not define this data preparation process as that is the responsibility of the data sharing organisation. The prepared data will need to be delivered to the data staging point before the agreed transfer schedule.

Once data is transferred by Eightwire it is the responsibility of the data sharing organisation to remove any no longer required data from the data staging point. At this time Eightwire does not signal that transfer is complete and correct to the data sharing Agent. However, it is readily possibly to orchestrate the primary data transfer to be followed by a secondary data transfer back to the data sharing Agent that equates to a record of success and this could be used to signal a completed and correct data transfer.

Data consuming organisations have similar responsibilities as data sharing organisations: data management between the data staging point and the data consumption point is the responsibility of the data receiving organisation.

Data staging

The data staging point can be a file system, a database, or both. Most organisations are used to sharing data in file containers rather than reading and writing from and to database objects. Eightwire supports both types of containers, with slight differences in functionality: in general, the entire contents of a file will be read, whereas subsets of the contents of a database object can be read. The Agent will create files to store transferred data if it has suitable permissions, and it will create database objects to store data if it has suitable permissions.

The data staging points can reside on the Agent server but this is not a requirement - it is an economic and security decision. Whether the data staging points are on the Agent server or not is not material to the Agent so long as the Agent has the necessary permissions and data drivers in place to read from and write to the data staging points.

Agent Server

The Agent server hosts the Agent. The Agent can be installed manually and once installed updates can be conducted manually or automatically by Eightwire. If an Agent is not found during initial use of Eighwtire via the web browser management user interface then a prompt to install the Agent will be presented.

The Agent is in the order of 2 MB in size and does not have esoteric operating system rights to execute. More than one Agent can be installed if data management requires this - such a configuration is beyond the scope of this design primer - please contact Eightwire if you have questions on this front.

The Agent has relatively low resource and operating system requirements:

1. CPU: 4 cores
2. RAM: 8 GB
3. Storage: 10 GB
4. Network: 1 GB+
5. Operating system: Windows 8.1 +, Windows Server 2012 +
6. .Net Framework 4.5 +

The Agent logs its starts, stops, and updates in a local clear text log file. No other information is logged locally. No transferred data is buffered locally. No information about data transfers is logged locally.

##

Network security

The Agent only initiates communications with Eightwire - it never responds to incoming requests. All communications take place using HTTPS (TLS 1.2) on TCP 443. Please note that operation of Eightwire by role holders at an organisation is conducted using a web browser and those communications are independent of this discussion on Agent communications.

Agents and Eightwire mutually authenticate using externally issued certificates. The traffic established within the SSL session is encrypted in transit between the Agent and Eightwire services. The data itself is not separately encrypted from the SSL channel.

Any network configuration will need to meet the following requirements:

1. MUST NOT interfere with the application protocol communication with the cloud service. This does not prohibit lawful traffic inspection so long as the integrity of the communications is not impacted.
2. MUST support DNS resolved network addresses. Eightwire services are addressed using DNS names and this is unchangeable.
3. MUST NOT require static network addresses. Eightwire IP addresses will change as scale out and failover occurs.
4. MUST support outbound connections. The Agent initiates all communications with Eighwire.
5. MUST support inbound and outbound data flows. Data may be required to flow in or out of an organisation.
6. MUST support hosting the Agent on Microsoft Windows.

Processes

Operations

Eightwire related operations are based on the following foundations:

1. Eighwtire core technical solution is provided by Eight Wire Limited and is a cloud hosted service.
2. Eight Wire Ltd is the owner of Eightwire services.
3. Whilst the physical topology of Eightwire is a hub, the data transfer and legal arrangement relationships remain “organisation to organisation”.

The operational roles are:

1. Data Exchange Business Management
	1. Fulfilled by Connecting Organisation
2. Data Exchange Platform Management
	1. Fulfilled by Eight Wire Limited
3. Data Exchange Support
	1. Fulfilled by Eight Wire Limited
4. Data Exchange Agent Administrator
	1. Fulfilled by Connecting Organisation
	2. Installs and maintains the Agent and local data stores.
5. Data Exchange Data Manager
	1. Fulfilled by Connecting Organisation
	2. Configures data sets to be shared with partner organisation's and to receive data shared from partner organisations

Monitoring

Platform monitoring is provided by Eight Wire Limited. Reports on performance and usage will be provided by Eight Wire Limited to Connecting Organisation when requested.

Support

Support will be provided by different parties depending on the specific incident or problem.

Usability queries will be worked through in the following sequence:

1. Eightwire Operations Guide
2. Eightwire web site
3. Each organisation's Eightwire Data Manager
4. Eight Wire Limited

Business queries will be worked through in the following sequence:

1. Each organisation's Data Manager
2. Eightwire

Data Exchange incidents will be worked through in the following order:

1. Eightwure Support Guide
2. Each organisation's Data Manager
3. Each organisation's Eighwtire Administrator
4. Eight Wire Limited or Connecting Organisation personnel

Incidents and problem recording will be maintained within each organisation's support systems. Eight Wire Limited and Connecting Organisaton will also maintain records of their involvement.

Roles

Implementing Eightwire connectivity requires a Windows administrator, a network engineer, and a security engineer. Maintaining changes to connectivity typically requires a Windows administrator.

Once connected to Eightwire any person who has authority to reference the data being shared or being received requires access to a web browser with Internet connectivity and they can configure and operate their organisation’s use of Eightwire.

Between establishing connectivity and using Eightwire is data preparation and data consumption within the data transferring organisations.  The roles, processes, and tools required to prepare and consume transferred data are assumed to already exist and are referenced in this guide and not defined.

The phases of a data transfer are itemised below:

Eightwire manages its own login accounts and roles and does not integrate its authentication and authorisation with any third party regime.  This is important to understand and has the following implications:

1. Provisioning and de-provisioning of logins and role membership for Eightwire must be added to the organisation’s current processes.  Two factor authentication at login will reduce the likelihood of inappropriate logins from non-authorised people, particularly those no longer with authority to manage aspects of Eighwtire use, this alone will not disallow situations such as a past staff member with the same phone number they hold when employed being able to login and use Eightwire if their login has not been de-provisioned.
2. A connected organisation cannot assert its identity to another connected organisation.  A connected organisation is dependent on the following controls as proxies for assertion of identity:
	1. Prior and on-going inter-personal engagements between the data transferring organisations.
	2. At the time of sharing a data store for a particular organisation to access a PIN code is generated specific to that installed Agent, that data store being shared, and the Agent permitted access to the shared data store.  This PIN code is required during data transfer setup by the organisation permitted to receive the data.
	3. Integrating Eightwire with broader identity assertion regimes is being investigated but not on any release schedule at this point.

Eightwire provides pre-defined roles and an organisation is free to use as many or few of these as meets their needs:

* Viewer: This user is able to view data that has been shared with them.
* Author: This user is able to publish data to be shared with other Connecting Organisations.
* Administrator: This user is able to create new data sources, manage user accounts and privileges within a project, and establish and dis-establish new sharing arrangements with third parties for the project they administrate. This user cannot create new projects or control privileges outside of the project they administrate.
* Account Administrator: This user has full control over the account, including adding and removing users, adding and removing projects, and the full rights of an Administrator.

For simplicity it is proposed that two of these roles be utilised in general, at least at the outset of using Eightwire:

1. Account Administrator: manage logins, roles, and Agent configuration.
2. Administrator (referred to as Data Manager): manage data stores, projects, and schedules.

Further Information

A good deal of additional collateral is available. If you would like copies of the below documents or anything else please contact Eightwire Technical Owner.

* Eightwire Use Case Guide
* Eightwire Data Sharing MOU Template
* Eightwire Privacy Screening Template
* Eightwire Operations Guide
* Eightwire Installation Guide
* Eightwire Management Guide
* Eightwire Support Guide
* Eightwire Privacy Impact Assessment (PIA)