IBM i2 Intelligence Analysis Platform

Data Acquisition Overview

Introduction
The IBM® i2® Intelligence Analysis Platform is an extensible, scalable, service oriented analytical environment that is designed to provide organizations with access to intelligence when and where they need it, so they can make faster, more informed decisions. For the analysis of data and creation of actionable intelligence users need the ability to acquire data from a myriad of data sources.

Built upon 20 years experience of delivering data acquisition solutions as part of i2 intelligence analysis products, the Data Acquisition Framework within the i2 Intelligence Analysis Platform offers flexibility in approach and depth of capability to support data acquisition in the intelligence analysis process.

The Data Acquisition Framework helps to achieve this by:

- Providing developers with options and tools to build data acquisition services that match user requirements while meeting the challenges of working within the constraints of the data sources accessed
- Giving users access to multiple sources of data in a common and familiar way, using i2 discovery capabilities that provide fast, effective and targeted methods to find the data they need
- Reducing the effort required for users to acquire the data they work with, freeing up more time for analyzing, enriching and turning data into intelligence products
Data Acquisition—The Need

Data is the lifeblood for IBM i2 customers in the commercial, defense and national security and law enforcement industries. The availability of data from multiple sources supports the intelligence-led approach, which has become the default methodology to stay ahead of criminals, terrorists and fraudsters. Organizations have the challenge of improving levels of service while reducing operational costs. They face more complex challenges, with an increasingly high expectation from governments, regulators, customers and tax payers for a correct response delivered with speed.

Data is available in multiple sources and multiple formats. It is available internally and externally, in legacy data systems and/or in open source systems, in structured or unstructured formats. There is a need to support fast and accurate methods for acquiring this data that feature automated routes to mitigate the slow pace and errors introduced by re-keying of data. In addition, the available data varies in terms of volumes and quality, which calls for processes to cleanse and process the data as it is acquired.

Access to disparate sources of data is critical for users as analysis of those data sets provides baseline information. Even more analytical insight can be drawn when combining data from multiple sources. This will increase the likelihood of uncovering previously unknown links and will minimize risks of overlooking crucial information. Once acquired however, the data needs to be analyzed in a manner which enables users to share their intelligence products. In addition, intelligence should be stored in a repository so that it can be shared, re-used and referenced at a later date. The stored intelligence and data should maintain links back to the source and ensure data provenance is retained. This will facilitate future analysis of that stored intelligence and data.

Data Acquisition—i2 Intelligence Analysis Platform

Acquisition and management of data are capabilities inherent to the i2 portfolio and continue to be delivered in the i2 Intelligence Analysis Platform. The user has at their disposal the capabilities they need to acquire the data in a single common format at which point it can be combined with data from other sources to help build a single cohesive intelligence picture.

Options to connect to data sources directly means that i2 Intelligence Analysis Platform users have access to up-to-date data. The Data Acquisition Framework provides an extensible, customizable environment to develop specific data acquisition solutions to satisfy both user’s requirements and constraints around the use of and connectivity to the data source. Upon acquisition the data is transformed into i2 entity, link, property formats which means the data can be queried and analyzed using i2 discovery and analysis capabilities. Analysis can be carried out against data from a single source or combined with data from any other source within the i2 Intelligence Analysis Platform and i2 Analyst’s Notebook Premium should more advanced multi-dimensional visual analysis capabilities be required.

Once created, the intelligence can be stored in the Analysis Repository, which serves as a security rich collaborative environment addressing both ‘need to know’ and ‘need to share’ directives, giving users access to the high value intelligence that they and others have created. Every item created in the i2 Intelligence Analysis Platform has provenance data recorded against it. The provenance data includes a retrieval block which details the data source and reference to the item within the data source, providing both a link back to the origin data and a description of where it came from. This enables merge of items from multiple sources without data loss and conveys the quality and reliability of each item of data.
Who should read this white paper?
This white paper is intended for:

• All members of the team who wish to learn about data acquisition approaches and capabilities that the i2 Intelligence Analysis Platform offers and how to choose the capabilities to deploy
• Managers and Team Leads who need to understand how their teams could use data acquisition capabilities

This white paper aims to give the reader an overview of the capabilities the i2 Intelligence Analysis Platform provides to help organizations overcome the challenges associated with acquiring data by detailing the options and the potential benefits.

For further information regarding products and capabilities referenced in this white paper please read the following:

• IBM i2 Intelligence Analysis Platform Product Overview Whitepaper
• IBM i2 Analyst’s Notebook Premium Product Overview Whitepaper

Potential Benefits of the i2 Intelligence Analysis Platform Data Acquisition Framework
Smarter, more informed decision making across a centralized aggregated view of data from disparate sources

• Obtain greater insight and produce richer intelligence products by fusing data from multiple data sources
• Improve both speed and accuracy of data acquisition with targeted approaches to user requirements and data source constraints
• Increase focus on analysis by reducing time spent on discovering pertinent results from disparate data sources
• Retain and track the provenance of data gathered from various sources for a more in-depth understanding of items in an aggregated view
• Leverage valuable data stored across the organization at the point of need
• Enable collaboration and sharing through load of data into shared repositories
Increase efficiency and effectiveness by bridging knowledge gaps through accessing data from a single, security-rich collaborative environment

- Reduce educational ramp up time, costs, and skill fade with a single user interface to access data sources across the organization
- Minimize the need for post processing activities by using targeted queries to discover pertinent data
- Efficiently pin-point, save, and share relevant findings to support operational needs
- Mitigate against mistakes introduced by manual acquisition methods such as re-keying of data
- Gain access to the latest data with a live connection to the supplying source
- Quickly identify and understand the network of related data and see the ‘big picture’ by visually exploring interconnected items from multiple sources

Help minimize cost of ownership by integrating with existing business infrastructure

- Use the correct approach to suit data source size, availability and update frequency
- Create data acquisition services to suit deployment requirements through the Developer Essentials Toolkit
- Reduced cost of ownership with a scalable deployment architecture that can grow alongside the organization’s data acquisition needs

Data Acquisition — Key Concepts

The data model configuration for an i2 Intelligence Analysis Platform deployment is described in a schema which details the entity, links and property types that are valid for that deployment.

Data of interest will be available in the supplying data sources in different forms such as rows or tables of data. The Data Acquisition Framework provides routes to acquire this data and present it in entity, link and property format. This enables organizations to:

- Combine all available data to help build a single cohesive intelligence picture
- Create insight and understanding of complex networks
- Detect non-obvious relationships or patterns
- Increase the depth of intelligence products for more effective resource utilization

Data Model — Entities, Links and Properties

In the IBM i2 Intelligence Analysis Portfolio work with data in entity, link and property formats which are also the basic concepts of the i2 Intelligence Analysis Platform data model.

The elements that make up this model are as follows:

- **An entity** represents a real world object — such as a person, phone, vehicle or an event such as crime
- **A link** associates entities with each other usually representing a relationship between entities — such as subscriber between a person and a phone or owner between a person and a vehicle
- **A property** stores data that characterizes an entity or a link — such as date of birth for a person, telephone number for a phone, vehicle identification for a vehicle.

Entities and links can have multiple properties.
Data acquisition—data discovery capabilities

Acquiring data and transforming it into 'analysis-ready' information is key to the success of every i2 Intelligence Analysis Platform deployment. The Data Acquisition Framework can be used to provide connectivity to sources of data for the purpose of searching and retrieving data of interest. The discovery capabilities provided for this include:

- **Text search** — a broad search for item the same or similar to the search term
- **Property based search** — a more targeted search for items with specific property values
- **Visual query** — a targeted search for items connected to one or more other items with certain property values where the user can 'draw' a query
- **Show context** — a request to show items that are in some way related to a particular item from the data source it originated from
- **Expand** — a request to show items that are linked to a particular item from the i2 entity, link and property based data source it originated from

The data discovered can then be surfaced in i2 Intelligence Analysis Portfolio products to be analyzed standalone or alongside data from other sources. The resulting information (intelligence) can be stored in the i2 Intelligence Analysis Platform data store known as the Analysis Repository. It can be also saved as a visualization in IBM® i2® Analyst's Notebook® charts.

**Intelligence Portal and i2 Analyst's Notebook Premium**

There are two client user interfaces that can be used with the i2 Intelligence Analysis Platform and work with data provided by the Data Acquisition Framework:

**Intelligence Portal**

This is a web based application that provides the operational team with on-demand access to the intelligence stored within the i2 Intelligence Analysis Platform and to data stored in third party sources. This intuitive thin client presents a single interface to explore and analyze data across disparate sources, reducing the time and cost associated acquiring data. The Intelligence Portal can be used to discover data using text search, property based search and visual query. The results can then be sorted in a results grid and visualized within the portal where further discovery can be performed using show context and expand capabilities which display the results directly to the visualization.

**Figure 3: Data discovery capabilities**
**IBM® i2® Analyst’s Notebook® Premium**

This is a rich desktop application that provides the analyst with on-demand access to the intelligence stored within the i2 Intelligence Analysis Platform and data stored in other sources. It features a powerful set of visual analysis capabilities that are designed to help users quickly deliver high value intelligence. i2 Analyst’s Notebook Premium can help users discover data using text search, property based search and visual query using the embedded Intelligence Portal. The results can be sorted in a results grid and visualized on a chart from where further discovery can be performed using show context and expand capabilities which display the results directly on the chart.

This provides i2 Intelligence Analysis Platform users with a suite of data acquisition options:

- **Manual data entry** — simplified with configurable data entry forms
- **Ad hoc import** — using powerful and intuitive import capabilities
- **Data on Demand** — connecting to and questioning existing data sources in situ
- **Data Load ELP Stage** — the loading of ‘working data’ to pre-analysis staging areas
- **Data Load Direct** — bulk ingestion of vetted data directly to an ‘analysis-ready’ repository

These capabilities help ensure organizations maximize and leverage the value in their existing data sources, both internal and external to the enterprise, while supporting and enhancing existing workflow and practices.

**i2 Intelligence Analysis Platform Data Acquisition Framework**

The Data Acquisition Framework provides an Application Programming Interface (API) which developers write code against to create and deploy Data Acquisition services for i2 Intelligence Analysis Platform users.

**Data formats supported**

The Data Acquisition Framework supports any format of data that can be mapped to entities, links and properties. The Data Acquisition Framework API accepts an XML stream of data. Extensibility within the framework offers support for managing other types of data such as binary files.

**Data mapping implementation**

As part of the creation of the data acquisition service the developer maps data types from the supplying data source to the data types in the i2 Intelligence Analysis Platform.

**Pre-acquisition processing of data**

There are no pre-processing capabilities ‘built-in’. The Data Acquisition Framework provides flexibility for other applications, or code, to undertake pre-processing, if it is a requirement. This could include data cleansing, de-duping and entity extraction processes.
Security—user authentication
The Data Acquisition Framework offers flexible support for the i2 Intelligence Analysis Platform user authentication against the supplying data source, including the integration with existing enterprise user authentication services.

Security—items
The Data Acquisition Framework supports mapping of the security data for items in the data source to the security model in place for the i2 Intelligence Analysis Platform.

Provenance
The Data Acquisition Framework can provide provenance on items as they are created during acquisition. Provenance data includes a retrieval block which details the data source and reference to the item within the data source. This enables the Data Acquisition Framework to get back to the relevant data source and carry out further actions against the item of data such as view details or show context.

Data Discovery
Once a data acquisition service has been created and deployed i2 Intelligence Analysis Platform users can:

• Query the data using the search capabilities within the Intelligence Portal
  – Querying the data in an already familiar fashion and presenting the results in same format as searching the Analysis Repository
• Filter/refine results in the Intelligence Portal results grid
• Visualize chosen items in the Intelligence Portal or i2 Analyst’s Notebook Premium
• Choose to explore items that are connected to the selected item
  – ‘Expand’ which will query the i2 repository for items linked to the selected item.
  – ‘show context’ which will query the origin data source for more data

Data Acquisition Framework extensibility and customization
The Data Acquisition Framework is designed with extensibility and customization in mind. There are options to include support for metadata against items as they are created and also pass through other data such as files associated with items.

For example, an item may be related to a document in the data source; it is possible to pass that document through to be stored in the i2 Intelligence Analysis Platform document store.

Data Acquisition—Approaches
The i2 Intelligence Analysis Platform provides a range of data acquisition approaches. They are as follows:

• Data on Demand — connecting to and questioning existing data sources ‘in situ’
• Data Load — ingestion of all or part of the data source
  – Data Load ELP Stage — loading of ‘working data’ to pre-analysis staging areas
  – Data Load Direct — bulk ingestion of vetted data directly to an ‘intelligence’ repository
  – Data Load Ad hoc — import of file based data using i2 Analyst’s Notebook Premium Importer and i2 Analyst’s Notebook Premium Chart Item Uploader

These data acquisition options are intended to ensure that users have an acquisition model that supports their requirements while operating within the constraints of use in place for the supplying data source.

Examples of these constraints are below:

• Data governance policies — These are employed to ensure data integrity, quality and security. Organizations may have specific rules for how data can be moved around the organization, how long it can be kept and who can have access to it.
• Data source availability — It is unlikely the data source’s only function is to serve as a data source for the i2 Intelligence Analysis Platform deployment. The owner/administrator of the data source may limit access to the source as it may generate extra unpredictable load on the service.
• Data source connectivity — The data source connectivity profile determines the data acquisition approach. Some approaches require the data source to be up and running as a service and accessible via the network 100% of the time whereas others can be used when the data source is only running and connected for a limited time.
The Data on Demand approach provides the capability to search for data directly from the data source. This means the data is acquired at the point of need and the user can be sure it is the latest data at time of acquisition.

A data on demand service is a custom extension to the Intelligence Analysis Platform that:

- Maps data source data types to those used in the i2 Intelligence Analysis Platform schema in use
- Accepts queries from one or more of the i2 discovery commands and maps them to the relevant query on the data source
- Receives the data from the source and transforms to the relevant items in the i2 Intelligence Analysis Platform schema
- Feeds the data to the i2 intelligence Analysis Platform

**Data on Demand—Data sources supported**
The Data on Demand approach can be applied to a data source that:

- Runs as a service or a database that can be queried
  - Most applicable to a service that offers an Application Programming Interface (API)
- Is available all the time that i2 Intelligence Analysis Platform users require to run data discovery queries against the data source

**Data on Demand—Data Discovery capabilities**
As part of the creation of the Data on Demand connector the developer maps the i2 discovery capabilities to those available on the supplying source. It is possible to configure which i2 discovery capabilities are available per source. For example, if the text search query is the only query available on the source the connector can be configured to only offer that option to the i2 Intelligence Analysis Platform user.

**When to use Data on Demand**
Scenarios where Data on Demand may be more suited:

- Large Data Sources
  - Where the data source volume is too large to consider a Data Load approach
- Frequently updated sources, particularly when:
  - Updating existing items within the source
  - Users require up to date data
- Sources with discovery/query capabilities that map to i2 discovery capabilities
  - Consider i2 Intelligence Analysis Platform user requirements – Text Search and Get Context may be enough to satisfy
- Where data governance policies don’t allow bulk upload of data to another system

**Constraints for choosing Data on Demand**
- Data source must be online (connected and running) when the i2 Intelligence Analysis Platform user is online and requires access to the data source

**Data on Demand customization—subset retrieval**
The Data Acquisition Framework can support a scenario where a results set, or subset can be generated by a third party process as opposed to the i2 discovery capabilities provided by the Intelligence Portal. The third party process could be the discovery capabilities native to the data source or another mechanism that prepares the data.
The item types still need to be mapped to the deployed i2 Intelligence Analysis Platform schema and the data needs to be created in the format that can be consumed by the Data Acquisition Framework. Mapping of the discovery capabilities is no longer required, however, a reference for the supplied subset must be provided.

From the i2 Intelligence Analysis Platform users view, the subset will look like a data source they can access and search using the i2 discovery capabilities available to them. This customization is useful if the source has comprehensive discovery capabilities that do not readily map onto the i2 discovery capabilities. The Data Acquisition Framework supports combining this customization along side a ‘pure’ Data on Demand connector so that the user could explore the data further from the originating data source.

i2 Intelligence Analysis Platform Data Acquisition Framework—Data Load

The i2 Intelligence Analysis Platform Data Load approaches enable the ingestion of all, or part of, a source’s data to two kinds of data repositories:

- **Analysis Repository** — At the heart of the i2 Intelligence Analysis Platform the Analysis Repository is the storage area for high quality information and Intelligence in entity, link and property form.
- **ELP Stage** — separate data storage areas for working data. They are similar in structure to the Analysis Repository and store data in entity, link and property form. There can be multiple ELP Stages within the i2 Intelligence Analysis Platform and they can contain data from one or more source.

**Data Load - Service Creation**

A Data Load is deployed by creating a service that:

- Maps data source data types to those used in the i2 Intelligence Analysis Platform schema in use
- Extracts and transforms source items required to be loaded into the ELP Stage repository and pushes them to the required repository as XML stream
- Manages periodic updates to the repository — if required

**Data Load—Data Discovery capabilities**

The Analysis Repository and ELP Stage data repositories are native components of the Intelligence Analysis Platform. As the data has been transformed in to i2 entity, link and property format all i2 discovery capabilities can be used to query the data.
i2 Intelligence Analysis Platform Data Acquisition Framework—Data Load ELP Stage

Data Load ELP Stage provides an interface for pushing data into entity, link and property item stores (ELP Stage) within the i2 Intelligence Analysis Platform. The data is acquired ‘in bulk’ at a particular time so it is as up to date as the last load from the source.

When to use Data Load ELP Stage
Scenarios where Data Load ELP Stage may be more suited:

- When i2 Intelligence Analysis Platform users need the power of the more complex i2 discovery capabilities over the source data.
- When the source has no data discovery capabilities or where the discovery capabilities do not easily map to the i2 discovery capabilities.
- Where the source is disconnected or unavailable for all or part of the time.
- When the administrator of the source will not allow unpredicted load but will offer a window of time to access data ‘in bulk’

Figure 8: Data Acquisition Framework—Data Load ELP Stage

i2 Intelligence Analysis Platform Data Acquisition Framework—Data Load Direct

Data Load Direct provides an interface for pushing high quality information and intelligence directly into the Analysis Repository in entity, link and property form. The data is acquired ‘in bulk’ at a particular time so it is only as up to date as the last load from the source.

When to use Data Load Direct
Scenarios where Data Load Direct may suit:

- When the data source is holding high quality information or intelligence.
- Static or Less frequently updated data sources.
- When the updates are additive, i.e. new items added rather than updates to existing ones.
- Data Load Direct is ideal for loading an existing intelligence database before it’s retirement.

Figure 9: Ad hoc Data Load
**i2 Intelligence Analysis Platform—Ad hoc Data Load**

Ad hoc Data Load or import approaches enable the ingestion of small volumes of data in file based data sources to the Analysis Repository. When using the i2 Intelligence Analysis Platform with i2 Analyst's Notebook Premium the user has the ability to import existing data from spreadsheet and tabular data files and intelligence from existing i2 Analyst’s Notebook Charts into the Analysis Repository. The data is imported using wizard based interfaces that allow resulting import specifications to be built, enabling subsequent imports of the same data structure to be run without user interaction.

**i2 Analyst's Notebook Premium Importer**

i2 Analyst's Notebook Premium includes a flexible and intuitive importer that enables users to import data from source files directly, as entities and links into the Analysis Repository, without set-up or assistance from their administrators.

i2 Analyst's Notebook Premium Importer is out of the box functionality of i2 Analyst's Notebook Premium. To import data the user:

- Maps data source columns to the types used in the i2 intelligence Analysis Platform schema in use
  - User can import character separated (CSV) and Microsoft Excel (XLS, XSLX) files
- Runs the import which extracts and transforms source items and loads them into the Analysis Repository
- Choose whether to find and update any duplicates that may already exist within the repository

**When to use i2 Analyst's Notebook Premium Importer**

Scenarios where i2 Analyst's Notebook Premium Importer is suitable:

- When data is made available to the i2 Intelligence Analysis Platform user and immediate load of the data is required
- When the i2 Intelligence Analysis Platform user is best placed to map the source data to i2 Intelligence Analysis Platform schema types
The Chart Item Uploader ingests data held in i2 charts into the Analysis Repository. Extracting individual entities from these charts enables a user to quickly build up detailed intelligence on individuals or events that may otherwise be hidden in numerous i2 Analyst’s Notebook charts.

Chart Item Uploader is out of the box functionality available with Analyst’s Notebook Premium. To import intelligence from charts the user:

- Configures the mapping of charts items to the types in the i2 Intelligence Analysis Platform schema in use
  - Chart Item uploader creates suggested mappings
- Runs the upload which extracts chart source items and loads them into the Analysis Repository

When to use i2 Analyst’s Notebook Premium Chart Item Uploader

Scenarios where i2 Analyst’s Notebook Premium Chart Item Uploader is more suited:

- When data is made available to the i2 Intelligence Analysis Platform user as i2 charts and immediate load of the data is required

Pre-requisites

The Data Acquisition Framework is an integral part of the i2 Intelligence Analysis Platform. Guidance on current minimum hardware, operating system, and third-party software requirements is provided on the following IBM website: ibm.com/support

Licensing

The Data Acquisition Framework application programming interface (API) and tools are included and licensed for use under the i2 Intelligence Analysis Platform license. IBM supplied data connectors may be licensed separately.

For more information

To learn more about IBM i2 Intelligence Analysis Platform, please contact your IBM representative, or visit: ibm.com/i2software

To learn more about all of the IBM Smarter Cities solutions, visit: ibm.com/smartercities

© Copyright IBM Corporation 2014
IBM Corporation
Software Group
Route 100
Somers, NY 10589

Produced in the United States of America
October 2014

i2, Analyst’s Notebook, COPLINK, IBM, the IBM logo and ibm.com are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at “Copyright and trademark information” at www.ibm.com/legal/copytrade.shtml.

The content in this document (including currency OR pricing references which exclude applicable taxes) is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

The performance data discussed herein is presented as derived under specific operating conditions. Actual results may vary. THE INFORMATION IN THIS DOCUMENT IS PROVIDED “AS IS” WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided.

Please Recycle

ZZW03285-USEN-01