
HHS |Talend   
ASPE Final Proposal

28 September, 2017

Matt Spencer

Federal Account Executive

(203) 984-3147

[mspencer@talend.com](mailto:mspencer@talend.com)

Abe Ansary

Solutions Engineer

(516) 238-3882  
[aansary@talend.com](mailto:aansary@talend.com)

**Table of Contents**

About Talend

* Founded 2006
* 800+ employees across 11 countries
* 3,000+ Customers
* 15 Offices Worldwide
* HQ in Redwood City, CA
* Over 3 million downloads of Open Source Products
* Recognized by Gartner and Forrester as a Leader in the Integration Space 2 years in a row
* Only Open Source ETL Company recognized as a Leader

At Talend, it’s our mission to connect the data-driven enterprise, so our customers can operate in real-time with new insight about their customers, markets and business.

**Cover Page………………………………................................………............1**

**about Talend………………………………................................….............2  
ASPE requirements: "BRIDG"-ing the gap…………….……………….3**

**Proposed Solution: Platform for Data Management…….4**

**Education & knowledge transfer…………................….............7 Professional Services & Engagement...…………………………….…8**

**Infrastructure & architecture overview ………....………….9  
WHY TALEND: beyond this use case ………………….………………….10**

**Talend Pricing Model & Costs ……………..………........……………..11**

**“BRIDG”-ing the gap: Project Requirements**

In a relatively unique move within the Federal government, five parts of HHS: the FDA, NCI, ONC, NLM, and NCATS are working together to solve a major challenge for medical researchers: different real world medical data can use as many as four different coding systems, depending on the storage location. This different data affects thousands of real-world systems with limited connectivity to FDA research systems, severely limiting the ability to conduct large scale, real-world evidence research.

This resulting PCORTF CDM Harmonization project seeks to build a data infrastructure for conducting exactly that kind of research. A successful implementation would allow researchers to query significantly larger volumes of data than currently possible, leveraging open standards and controlled terminologies to advance Patient-Centered Outcomes Research and supporting evidence generation for regulatory and clinical decision making.

HHS will take the existing four systems - Sentinel, PCORNET, OMOP, and I2B2 – and integrate them as part of a unified model, known as “BRIDG”. This system will enable researchers to submit a single query to the BRIDG database, which can then distribute components of the query in the proper form to the appropriate source system(s) and retrieve any relevant data.

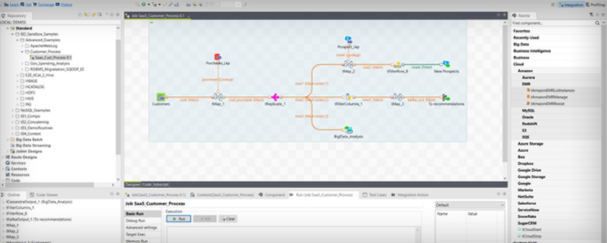
This objective is a complex integration and data quality problem, but it is not unprecedented, and Talend can help. As reviewed during the final demonstration, critical components would include:

* **Data Validation**: Sentinel, PCORNET, OMOP, and I2B2 grew and developed largely independent from each other; as such, they possess varying degrees of data quality. The selected solution must assist in ensuring the quality of any data received by researchers, through rules and validation scripts, including aiding the correction of errors in the data.
* **Data Mapping & Transformation**: As researchers search for and collect data today, they frequently struggle with the need to learn multiple languages to query the correct data source. The chosen solution must help map the four existing systems to BRIDG and to each other, to enable researchers to submit a single query and retrieve all relevant data. Importantly, the tool must enable the creation of these mappings through a drag-and-drop graphical user interface (“GUI”), rather than through writing code manually.
* **Metadata Usage**: To ease the burden on those constructing these mappings among the various models, the ideal tool will be able to interface with importable, machine-readable metadata to automate some of the mappings, particularly if changes are made in the future. Additionally, some mappings already exist in existing files, and the ideal tool will be able to save customization efforts by using those existing mappings, rather than require the recreation of each one.
* **Query Transformation**: Medical researchers do not always know the specific system in which their ideal data exists. As such, the ability to take a query which a researcher used one system to submit and automatically convert that query into another system’s format will save valuable time and effort. In particular, the ideal tool will enable two-way query transformation.

Given the technologies in place at the five HHS agencies today, Talend can complete these requirements and provide an enterprise-grade platform upon which the project can grow and mature – while limiting the burden placed on HHS resources to learn several platforms themselves.

**Talend’s Solution: Platform for Data Management**

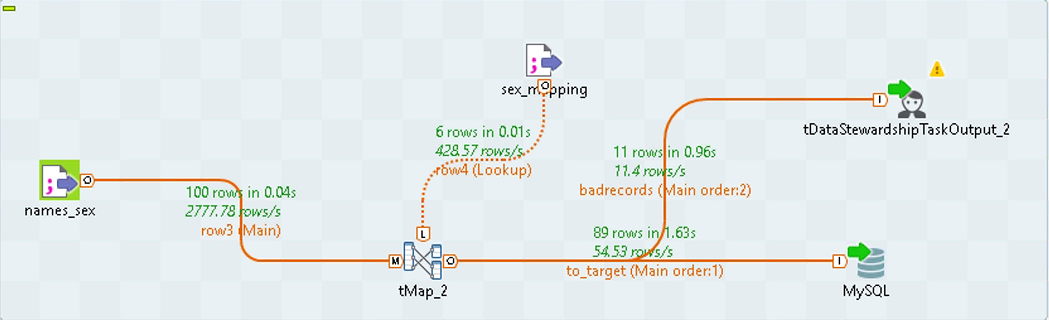
Talend’s Platform for Data Management encompasses Data Integration and Data Quality functionality, accessible via a unified, eclipse-based developer tool known as the Talend Studio. Data experts without development expertise can use the web-based Data Preparation and Data Stewardship interfaces to clean their data, then provide reusable “recipes” to the development team for integration into a data workflow.



**Data Validation**

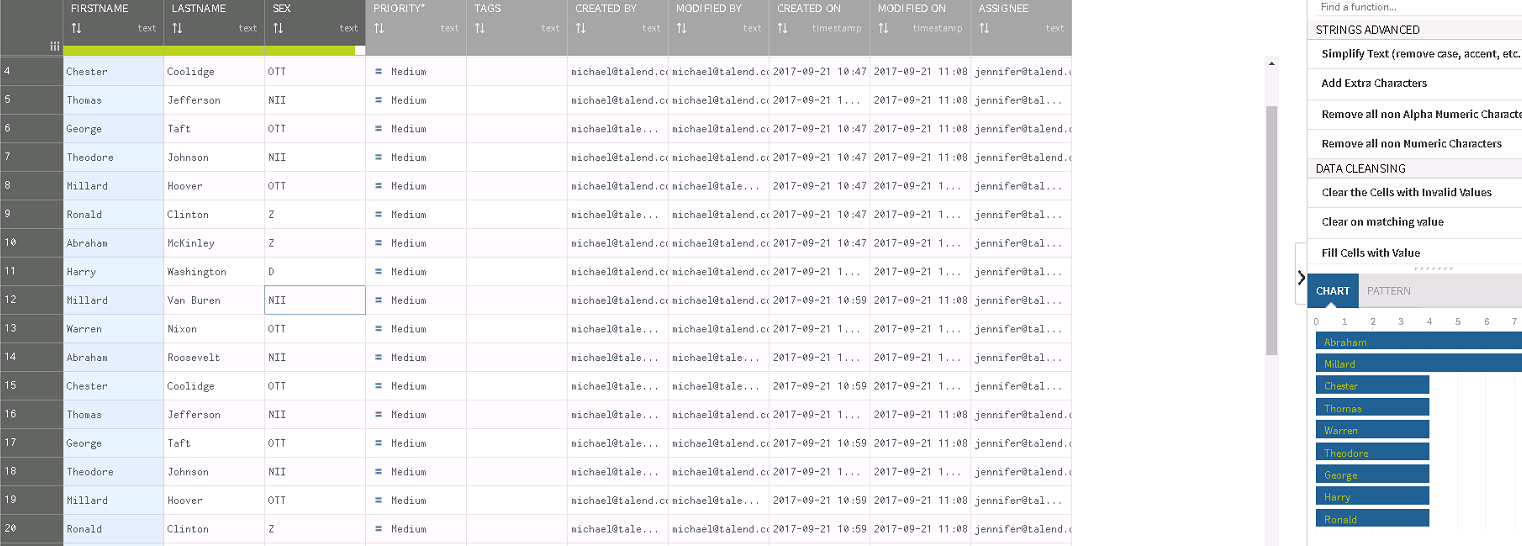
Talend includes a component called “tMap” which enables the creation of mappings among several data sources, via a drag-and-drop interface. As part of an integration job, a simple version of which we showed during the demo, the tMap component can also help validate expected joins and identify bad records. A job can then split the imported records accordingly – exporting them into separate tables for data ready for analysis and data that returned an error.

Additionally, the Data Profiling perspective within the Studio enables many different types of quick looks at a dataset, creating a PDF report with basic to advanced analyses of a given dataset.



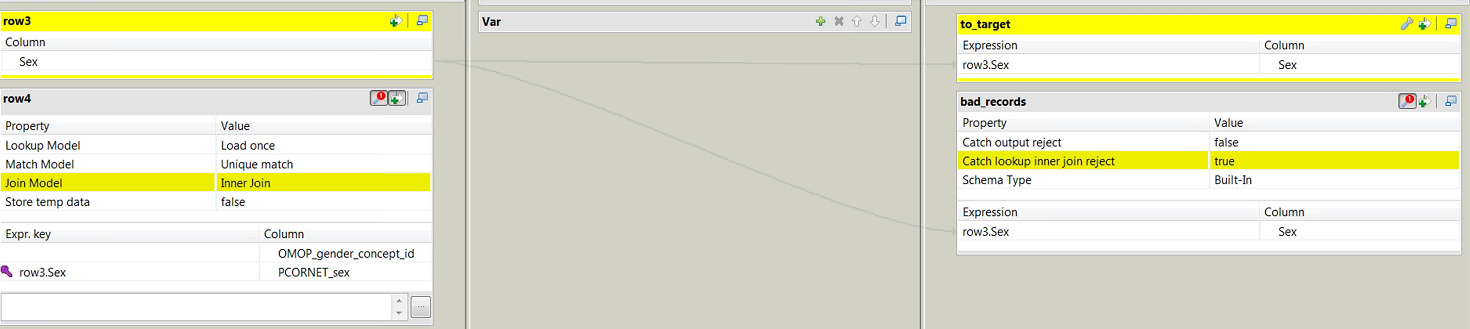
Once the job has identified bad records, a user with knowledge of the data can use the Talend Data Stewardship application to update fields with an error. These fixes can address issues across multiple records in the same column, create rules such as removal of whitespace, and overwrite specific text on an individual record.

This Data Stewardship tool is extremely powerful in the hands of a “business” or “mission” user, particularly one who understands the dataset well. As they move through the data and make their changes, the applications will generate a “recipe” of changes, allowing for easy reuse of their steps for the next time a job exports the data. They also create a log of all changes, which an administrator can review in the Talend Administration Center (“TAC”).



**Data Mapping & Transformation**

The tMap component provides the primary GUI to map fields across multiple datasets. A drag-and-drop interface, tMap enables developers to create a wide variety of mappings and joins to facilitate the movement and profiling of data across disparate data models.

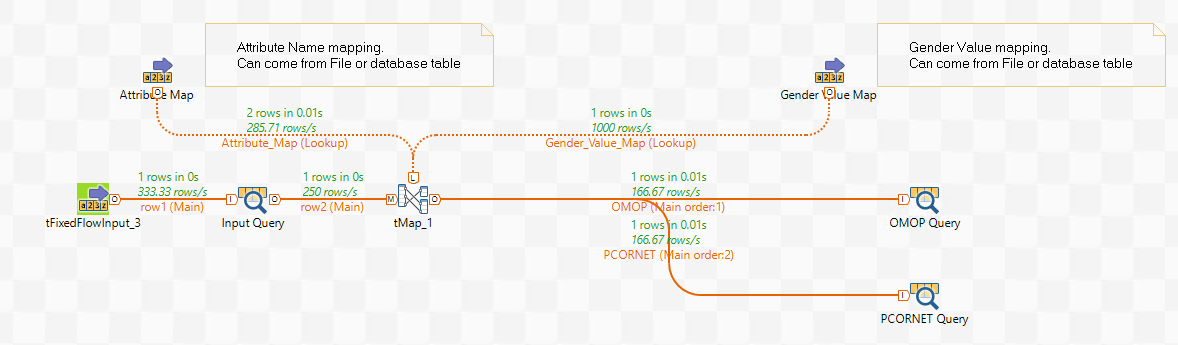


**Metadata Usage**

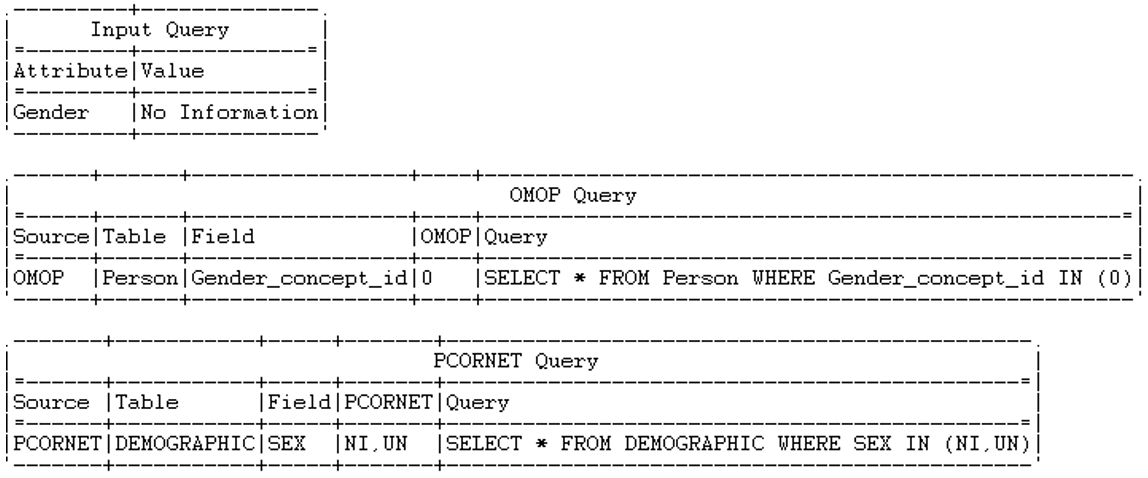
The Talend Studio enables direct import of metadata as a data source, and management tools within the Studio enable impact analysis and data lineage views to help keep manage any metadata used in a project. Importing metadata from a source can help easily define data mappings in the case of larger, more complex data mapping needs where mappings already exist in a file. As an example, ICD-9 to ICD-10 mappings could be imported from an existing file as metadata, rather than recreating them manually using the tMap component.

**Query Transformation**

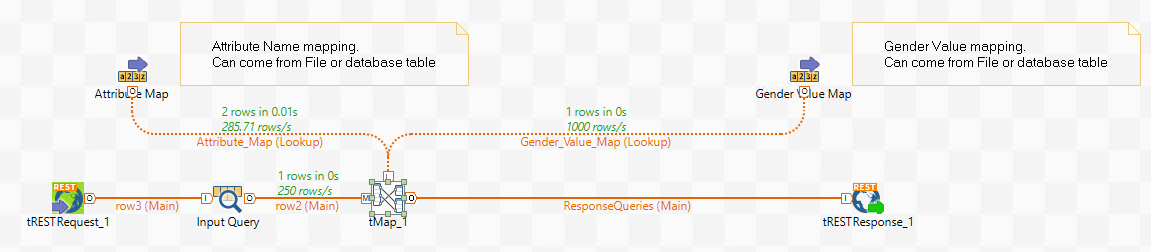
Talend would use a combination of several components to create a flexible, multiple-query job. This job would utilize an Attribute Map, to handle the mappings among the various data stores. The tMap component, referenced earlier, would help create a dynamic mapping, including the flexibility to compose a specific SQL statement as part of the query which could be customized for multiple data sources.



Once the queries are created in tMap, multiple query components can create the specified queries to specific data sources. As this project matures, Talend’s MDM platform can help manage the reference data required to form the queries in a more automated form.



As a middle step between the manual process and one automated through MDM, Talend provides an Enterprise Service Bus component which would enable a user to call a Restful service to submit a query, and have a service automatically respond with a set of queries, customized to the appropriate data source(s). An administrator can start and end services and check the usage statistics of such a query builder via the Talend Administrative Center.



**Education and Knowledge Transfer for HHS Personnel**

We recognize that the best software in the world doesn’t mean much if its users lack an understanding of how best to use it. With that in mind, here is a recommended set of courses for HHS users.

Most courses can be delivered as private instructor-led (in person or remote), or as online self-guided instruction. Talend Data Preparation for Implementers is the only exception, as that course is only offered via online self-guided instruction.

**Talend Data Integration Basics & Advanced:**

These courses help students use Talend Studio for Data Integration as quickly as possible. The Basics course focuses on the basic capabilities of Studio and how you can use it to build reliable, maintainable data integration tasks that solve practical problems, including extracting data from common database and file formats, transforming it, and integrating it into targets.

The Advanced course enables students to use the more advanced features of Talend Data Integration as quickly as possible. Participants can work in teams on projects shared on a remote repository to monitor Jobs and database changes.

**Talend Data Quality Essentials**

After learning the basic and some advanced functions of Data Integration, we transition to Data Quality. Still working in the Talend Studio, this course is designed to help you immediately utilize its Data Quality functions. Students learn how to evaluate data quality according to a set of metrics and thresholds based on indicators, models, and rules for each data item to be analyzed or monitored.

**Talend Data Stewardship**

This course enables developers to build DI Jobs for Talend Data Stewardship in order to empower business users to quickly access and handle tasks. It covers creation of data models, semantic types, campaigns, tasks, and how to resolve several types of tasks in Talend Data Stewardship.

**Talend Data Preparation**

Talend Data Preparation is a self-service application that enables information workers to prepare data for analysis and other data-driven tasks. This course is designed to help you immediately utilize the Talend Data Preparation web interface.

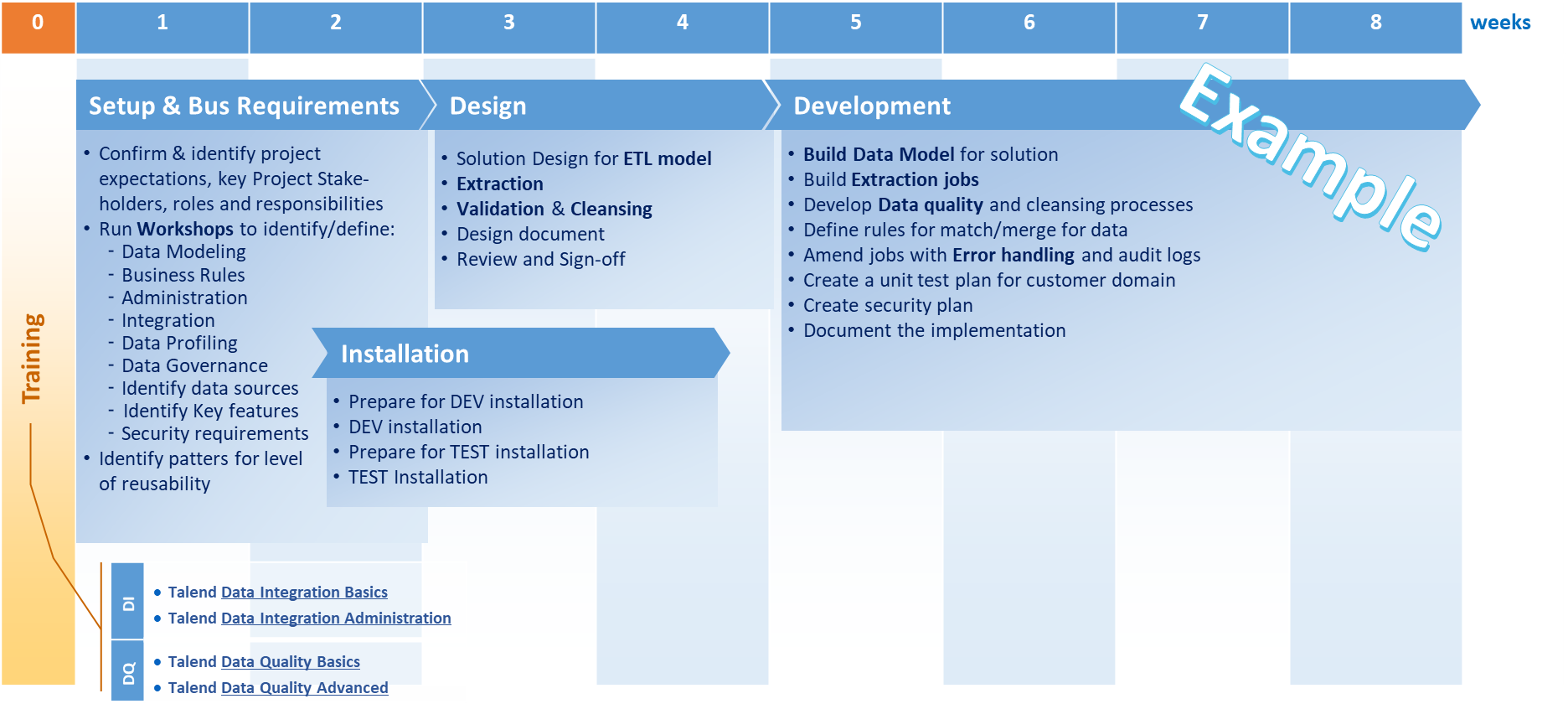
You learn how to create datasets and preparations to deliver cleansed, structured, enriched data to business users. You also learn how to use Talend Studio to execute preparations and create datasets in DI Jobs.

There are plenty of additional learning options available, and interested users can find a full catalog of Talend courses at <https://www.talend.com/services/training/catalog/>.

**Professional Services and Typical Engagement Scenarios**

**Recommended Engagement – Success Assurance**

The recommended engagement model is based on the principle of the Talend PS consultant being fully dedicated during the first 2 months. This will maximize the assurance of a professional installation, fastest customer team ramp up time and a qualitative initial design and development phase.

****

Consulting can be delivered onsite or remote, Talend’s recommendation is for at least the first 1/3 of the full time phase is delivered on site.

Following the 8 full time weeks the PS consultant will switch to a 2 months parttime remote mode to be of assistance to the team for ad-hoc problem solving and general advisory.

Further, with all of Talend’s PS offerings, customers are assigned a Project Delivery Manager. The DM oversees the engagement, from onboarding and weekly cadence calls to final engagement close review. This is a critical resource who ensures the team is meeting all expectations & timelines for a successful delivery.

**Costs**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Project Description** | **Hours** | **List Price**  **$/h** | **Discount** | **Price**  **$/h** | **Price** |
| Fulltime phase 8 weeks | 320 hrs | $275 | 10% | $247.50 | $79,200 |
| Parttime phase 8 weeks | 128 hrs | $275 | 10% | $247.50 | $31,680 |
| **Total USD** |  |  |  |  | **$110,880** |

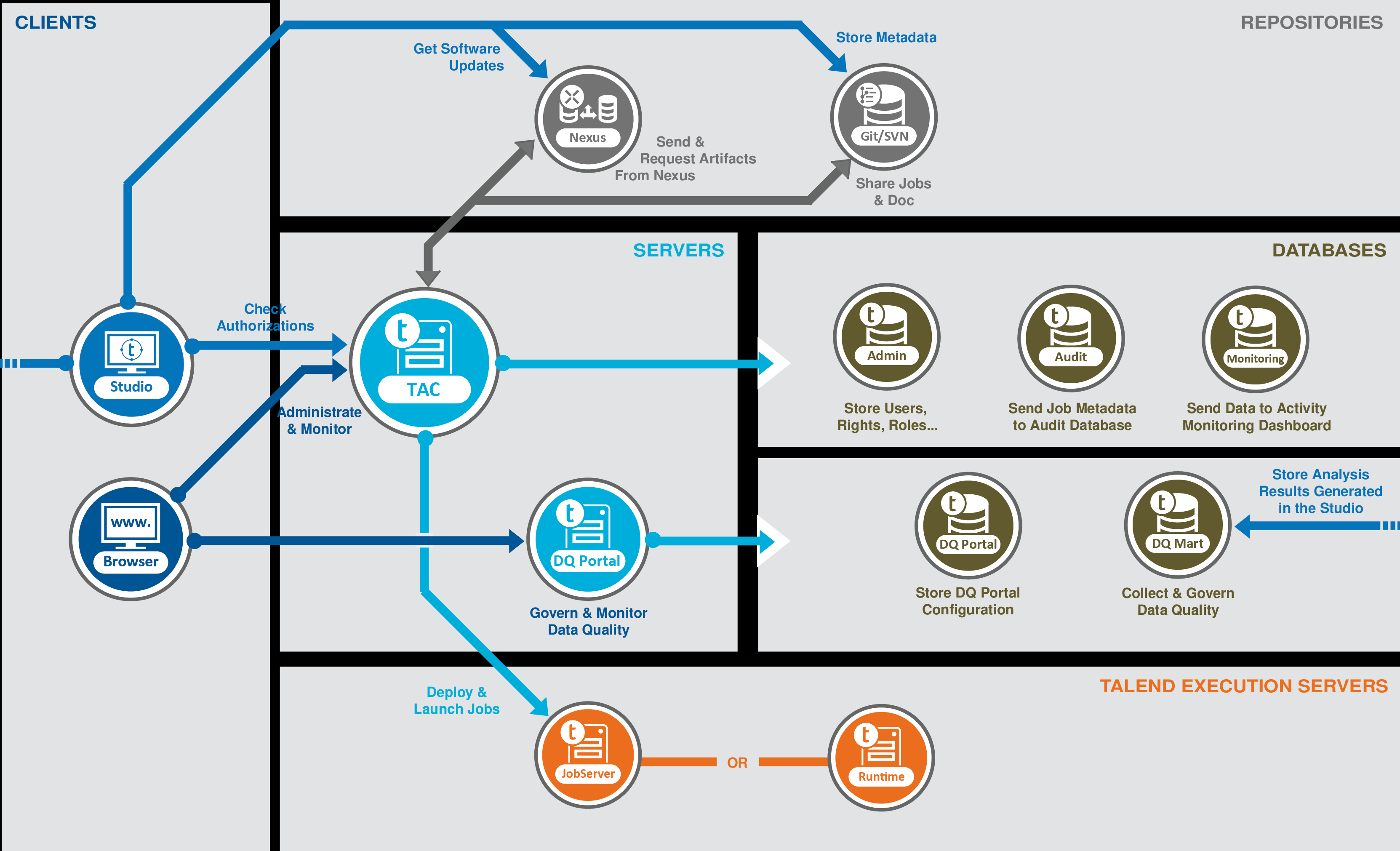
**Optional Engagement – Installation and Assistance – 4 weeks**

Talend PS consultant focus on baseline installation and them enablement of the team in combination with availability for ad-hoc problem solving.

**Costs**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Project Description** | **Hours** | **List Price**  **$/h** | **Discount** | **Price**  **$/h** | **Price** |
| Fulltime 4 weeks | 160 hrs | $275 | 0% | $275 | $44,000 |
|  |  |  |  |  |  |
| **Total USD** |  |  |  |  | **$44,000** |

**Infrastructure & Architecture Overview**



**Why Talend**

Customer Success Program

As a Talend customer you will be assigned a Customer Success Manager (CSM), a non-billable Talend resource.

They will:

* Help you define, measure, and achieve your desired business results with your Talend implementation
* Connect you with the correct resources inside Talend (e.g. services scoping, ongoing strategy, and roadmap information).
* Help maximize the value you receive from your Talend investment
* Be your day-to-day point of contact for anything Talend related

**Open Source**

* No vendor lock in.
* No black box engine; customers can extend the solution   
   themselves.
* Talend’s commitment to open source and open standards,   
   including ongoing investment in new Apache community projects, helps streamline future technology transitions.

**Predictable Pricing**

* Talend provides a predictable pricing model based on the   
   number of users, paid via an annual subscription.
* Cost is NOT based on data volume, types, or throughput;   
   database adapters/connectors, or other “data taxes.”

**Native Code Generation**

* No hand-coding required. Talend developers create jobs using a drag-and-drop Graphical User Interface (GUI), while the product generates lines of code in the background.
* No proprietary code: Talend uses a code generation paradigm   
   to create Standards Based code (e.g. Java, SQL, MapReduce).   
   The code is all visible, and can be run on the environment of   
   your choosing.

**Unified Platform**

* Talend’s entire platform (ETL, Data Quality, Big Data ETL, Application Integration, and Master Data Management) use a single GUI. This platform, grown through development rather than acquisition, provides a consistent developer experience and facilitates training and expansion of functionality across the full technology stack.

**DATA  
 INTEGRATION**

These four pillars are the basis of Talend’s competitive   
 advantage in the market and make Talend especially well-  
 suited for HHS’s ASPE project needs.

**Pricing & Licensing: How it Works**

Talend prides itself on a common-sense, predictable pricing model. All licenses are sold as an annual subscription. Prices stem from the number of users, and in the case of HHS, these users fall into two buckets. The first bucket, the “Platform for Data Management”, focuses on the Developers using the Talend Studio; the second, “Data Preparation / Data Stewardship”, applies to the business users who will utilize the web-based interface.

Pricing includes support, maintenance, and upgrades. Talend can discount for multi-year commitments.

**Importantly, there are NO hidden/additional costs with Talend. We do not charge for:**

* Onsite vs Cloud deployment
* Multiple Environments (Dev, UAT, QA, Prod, etc.)
* Number of Data Sources/Targets
* Types of Sources
* Database Connectors/Components
* Number of Records
* Volume of Data Throughput, etc.

|  |
| --- |
| **Platform for Data Management Includes:  Data Integration & Data Quality** |
| NASA SEWP Standard Pricing:  10 Users: $171,600 25 Users: $429,000 | 10  Named Users | 25 Named Users |
| Term Length | **Offers for HHS** | |
| 1 Year Subscription | $163,020 | $407,550 |
| \*2 Year Subscription | $145,860 | $364,650 |
| \*3 Year Subscription | $121,550 | $286,000 |
| Data Preparation & Stewardship, per 5 users | $12,500 | |
| Training Option 1: 10 users / 5 Admins Training Option 2: Platinum TU Package | 120 Training Units | $24,000 250 Training Units | $48,000 | |
| Consulting: Option 1 Consulting: Option 2 | $110,880 $44,000 | |

This document was prepared for the U.S. Department of Health & Human Services. It contains CONFIDENTIAL information and should not be distributed without prior permission from Talend. Recommended procurement vehicle is NASA SEWP.

\* Multiyear options can be billed annually \*\*Pricing is valid through 31 October, 2017