

ETL Tool Implementation Phase 1

Account/Department

ITS Enterprise Data Management

Divisional VP:

Dr. Alex Enyedi - Provost Office

Estimated annual initial costs:

\$40,000 - \$113,000

Estimated annual on-going costs:

\$10,000 - \$13,000

Funding Availability?:

Partially Funded - only initial OR on-going but not both

Need to Know:

Bethany Rizzardi, Hana Wills, Jeff Stebbins, Holly Aitken

Portfolio Type:

Infrastructure & InfoSec: Activities involving the admin, operation and maint. of campus IT sys.

Portfolio Type Alignment:

The tool selected by this project will be responsible for generating all reporting structures contained within Humboldt's Enterprise Data Warehouse. Reporting structures include General Ledger, Labor Cost, Human Resources, Application Statistics, Registration Statistics and Degrees Awarded.

Requestor name:

Christian Robbie

Project Champion:

Anna Kircher

Primary Customer(s):

Campus-wide

What problem are you looking to solve?:

One of the key responsibilities of Humboldt's data warehouse is to Extract data from source applications, Transform the data so that it can be easily queried and Load the data into OBI so the reports can be viewed (ETL). To do this the data warehouse team currently uses two ETL solutions, Oracle Warehouse Builder and IBM DataStage.

Oracle Warehouse Builder only supports older versions of Oracle's database software. Oracle has stated that no new versions of Oracle Warehouse Builder will be released and users should move to an alternate solution going

forward. Since this product is already at End of Life there is an increasing urgency to replace it and resolve the risk of continuing to use an unsupported tool. The second ETL tool used by HSU, IBM DataStage, was deemed not a viable sole solution given the historically problematic nature of the tool. We currently used a free open source version of IBM DataStage and even after evaluation of the premium version of DataStage, the features and functionality did not compare with other products on the market.

Which departments or processes are affected?:

All users of data warehouse data are affected (staff, faculty and students).

Implementing a new ETL Solution will enable existing reporting to be more reliable as well as allow new report requests to be implemented quicker.

What is the consequence of not doing this project?:

Eventually Warehouse Builder will need to be replaced, even if nothing is currently done this project cycle.

Additionally a higher proportion of the ITS Enterprise Data Management team's time will need to be spent on maintenance which will impact implementation time of new customer requests.

What would a successful solution look like?:

To be successful, this project will require the purchase, installation and configuration of a new ETL Solution identified by the ETL Tool Replacement Investigation project (Spring 2016). Also included in this scope is the migration of all transformations currently being performed by Oracle Warehouse Builder to the new solution. Once completed, and the Oracle Warehouse Builder product retired, the risk to HSU of continuing to use an unsupported program will be eliminated.

This then sets the stage for a Phase 2 project proposal to migrate transformations from the IBM DataStage's ETL, allowing retirement of that tool and consolidation of all data warehouse ETL processes to a single solution. This will enable quicker transformations, faster troubleshooting, less support time and a more stable data warehouse.

This ETL Tool Replacement Investigation project involved reviewing the market, identifying approximately 10 different tools for review, which was narrowed to a second level review of 5 of those, ultimately arriving at three potential solutions through an RFP process. Demonstrations were conducted by all three vendors, followed by the team test-driving each ETL tool, actually installing demonstration versions of the software and running the same test scenario through each tool during an evaluation period to identify the best solution to meet HSU's needs. While the evaluation is still in progress at time of proposal submission, the final selection and, thus, product costs, are on schedule to be completed by November 2016.

The IBM DataStage migration / consolidation is intentionally identified as a separate project, allowing the team to first learn from the phase one Oracle Warehouse Builder migration and apply that experience to the IBM DataStage migration.

How else might you solve this problem?:

The ETL Tool Replacement Investigation (SPRING 2016), is currently being implemented to allow the University the time to investigate solutions from other CSU campuses and evaluate multiple vendors' solutions to insure the correct selection for Humboldt.

Additional information about the outcomes of the ETL Tool Replacement Investigation is available upon request.

Student Success:

The ETL performed by the data warehouse supports many of the reporting capabilities on campus that are crucial to student success on campus. Examples include advisor roster and DARS (Degree Audit Report for Students).

Having a quick, reliable Data Warehouse will enable all departments to make data driven decisions ultimately improving all aspects of the student experience at Humboldt. This project will enable this by allowing OBI to be more dependable and to perform additional analysis to provide PeopleSoft reporting to Faculty and Advisors to support personalized and proactive advising to students. Data-driven decisioning is a hallmark of the campus Strategic Plan and WASC priorities.

There are many intangible improvements this project will have on student success, however ITS will be able to quantify improvements to the Enterprise Data Warehouse stability and speed.

What resources will this project require?:

The installation and migration of code will be performed by the ITS Enterprise Data Management team.

And is your deadline a hard deadline?:

No

Explanation of annual estimates:

Ranges based on tool selected from the ETL Tool Replacement Investigation Program (SPRING 2016). As noted elsewhere in this proposal, the evaluation is still in progress at time of proposal submission, with the final selection and, thus, product costs, on schedule to be completed by November 2016. This information can then be included in any final recommendations to Cabinet, including the extent to which either one-time or base can be covered by ITS budget. It is still being determined if initial costs can be covered out of ITS Innovations funds. Once initial licensing and the first year of maintenance have been covered, any additional hardware and training costs can be covered from the departmental budget. ITS may be able to fund annual maintenance. Additional details about this investigation are available upon request.

We do not expect to incur any additional costs when we move to the next phase of this project.

Type:

Replace

Affected or Related Systems/Modules:

OBI

Other

Other Affected or Related Systems/Modules:

Oracle Warehouse Builder (later project will include IBM DataStage)

Considerations / dependencies:

Completion of the ETL Tool Replacement Investigation Program (SPRING 2016) is required to complete before this project can begin. This project is on schedule for completion by November 2016.

Information Technology (IT) Feasibility Statement:

This project will require two new servers be installed and configured to host the new production and development ETL tool solution. The EDM team will then be responsible for the installation and configuration of the new tool and

migration of legacy code. Once the new solution is in place, the EDM team will be able to decommission the server instances used for Oracle Data Warehouse Builder and IBM DataStage. Training is rated as moderate as there will be some training required for the Enterprise Data Warehouse team (some of which they will have gained through the test-drive phase) and no training of end-users will be required.

This project breaks down into the following components:

1. Install new the ETL solution.

Analysis: Expected to require purchase and installation of a new server, involving ITS Admin Support and SysAdmin resources, in addition to Enterprise Data Warehouse department resources to install and configure the software. With dependencies across 1-2 teams, involvement of new technology and some team inexperience with the specific products this project section is sized a Medium.

2. Migrate code from Oracle Data Warehouse Builder into new environment.

Analysis: Similar to above, use of new technology and team inexperience with it, this project section is also sized a Medium. Depending on the extent of work distribution across the data warehouse team, it is possible that some work can be done concurrently but more likely consecutively.

3. Create Phase 2 project proposal for integrating data and processes from IBM DataStage into the new ETL solution, incorporating lessons learned from Phase 1 and retire DataStage.

Analysis: Extra Small.

Estimated project duration: 6 to 6 1/2 months.