

**Data Warehouse  
Assessment  
Recommendations  
Summary  
for  
University of Maine System**



April 2018  
Version 7

## Executive Summary

Athena IT Solutions is pleased to present this Data Warehouse (DW) Assessment to the University of Maine System. Our detailed findings and recommendations are documented in the accompanying PowerPoint file, which outlines the DW assessment process, participants, what was reviewed, our findings and detailed recommendations in the areas of architecture, people, projects, products and processes.

The purpose of this document is to highlight keys findings and recommendations.

### ***DW Complete Rebuild***

We strongly recommend that the UMS completely rebuild its DW and BI architectures. This rebuild requires:

- **Architecture** - a new data architecture
- **Best practices** - the use of modern best and pragmatic practices
- **Commitment** - a commitment to a DW program, especially in terms of business participation
- **Training** - an investment in training for technical and business people
- **Technology** - selecting products on a best fit basis so that people (IT and business) can, and will, use them

A complete rebuild is necessary because the DW does not meet the UMS business needs and is not built in a way that enables the UMS staff to sustain it. The current DW's limited use is a clear indication that it does not meet the needs of the UMS community.

### ***Architecture and Best Practices***

There are five fundamental flaws in the current DW design and deployment:

First, the current DW embeds specific business and operational process rules as it initially integrates data from systems of record (SOR), such as PeopleSoft. This severely limits its applicability to the broader UMS community and, even for its narrower audience, other shortcomings, most notably a lack of transparency, further limit its utility. The UMS needs to build an analytical data architecture (ADA) that supports the dual purposes of data integration and business intelligence (BI) for the UMS communities. An ADA has an integration and a BI schema for these distinct purposes. Business rules are applied in loading schemas built for specific business purposes, and not in the integration schema.

Second, the current DW integration processes do not effectively support detecting changes in the source systems, referred to as Change Data Capture (CDC), nor track and record changes to data over time, referred to as Slowly Changing Dimensions (SCD). This means that DW users do not know when information about people, places or things have changed, inhibiting analysis. Often

campuses and departments keep copies of historical data to maintain a record of historical changes.

Third, the DW weekly update cycle is not frequent enough for many BI uses.

Fourth, the DW integration processes do not have robust error handling, recovery and restart, nor do they track data about its process such as when started, finished, if successful, how many records were successfully processed and what failures occurred, if any.

Finally, there are many architectural, schema and integration processes that are incorrect (for details, please refer to the slides).

### **BI Program and Projects**

The UMS should commit to a long-term BI program that incrementally and iteratively builds out the DW and BI capabilities for the UMS communities such as UMS systems administration, campuses and individual departments across the system and campuses. A BI Roadmap should be agreed to for the DW program with each of the communities incrementally obtaining analytical capacities and integrated data to enhance what they provide to their stakeholders. These capabilities will enable the communities to shift from gathering data and building spreadsheets to leveraging actionable analytics.

We recommend three categories of BI projects serving each of the UMS communities:

- Develop analytics to assist in system-wide goals such as student success
- Shifting the burden of data integration from each campus to the DW
- Replace departmental data shadow systems that are a burden to these departments.

### ***Commitment***

The UMS community needs to be involved in the full lifecycle of DW design, deployment and operations in addition to providing requirements. This includes active involvement in BI application design, assuming responsibility for self-service BI, and UMS-wide data governance. UMS communities need to commit to the use of the BI applications and provide feedback when these applications fall short of expectations rather than just retreating to spreadsheets.

### ***Training***

The UMS needs to invest in training for UMS communities for both IT and businesses. Much of this investment is not in vendor training, but rather in the investment of people's time. IT should leverage online training, particularly vendors' functional tutorials, which often provide videos and hands-on exercises. Fundamentals and concepts training for DW, data integration and BI are just as critical as product-specific training.

For the business community people who are BI “power users,” there are benefits in formal vendor training or online tutorials like the approach mentioned for IT.

### ***Technology***

While it is tempting to simply select technology products from the largest IT providers, or products that have the most technical features, this is not the best choice for an enterprise. The UMS needs to select BI and data integration tools that are the best fit for the intended use cases and UMS users from both IT and business perspectives. If a BI tool is used infrequently or not at all by the business, or if a data integration tool is inadequately used or improperly implemented, then it does not matter what features it offers when compared to other tools. Many feature-rich tools, in both the BI or data integration categories, are overly complex and are best suited to staff who are fully dedicated to full-time use of these tools.

### **BI Tool Recommendations**

BI tools that are overly dependent on IT setup and maintenance have a history of driving business people to spreadsheets and Microsoft Access databases, creating data shadow systems. Currently, the UMS is using IBM Cognos, which falls in the category of an IT-centric tool that is complex to set up and administer, and typically requires dedicated IT resources to keep up with business demand.

We recommend that business people use three categories of tools: self-service BI tools, often referred to as data discovery tools, guided discovery BI tools where IT or other business users (power users) create dashboards for others, and spreadsheets. We are recommending that the use of Excel shift from gathering, integrating and reporting on data to reporting only with the DW (integration schema) and BI schemas being where sourcing and integration occurs. Any self-service BI and guided discovery BI tool(s) that are selected should work in conjunction with spreadsheets.

The top data discovery tools are Microsoft PowerBI, Tableau and Qlik Sense, with other vendors such as Information Builders and Oracle catching up. Tableau is viewed by many business people as the most robust if they use many data visualizations in their analysis, but other products are slowly catching up. Tableau Server will be required if the UMS wishes to broadly deploy and to date that has been expensive. We recommend Microsoft PowerBI for the UMS community who typically uses Microsoft Excel and does not need sophisticated data visualizations (and Microsoft PowerBI is the top of Gartner’s Magic Quadrant for BI tools and will likely catch up to Tableau at some point regarding visualizations.)

### **Data Integration Tool Recommendations**

Although Informatica and IBM DataStage have been in the top tier of Gartner’s Data Integration Quadrant for more than a decade, there are other tools, such as Talend and Oracle Data Integrator (ODI) that are also in that top quadrant, and that we recommend for UMS. These tools can meet UMS data integration requirements.

ODI and Talend are implemented with Extract, Load & Transform (ELT) architectures, while Informatica and DataStage use a legacy Extract, Transform & Load (ETL) architecture. ETL requires an intermediate integration server where data is transformed and then moved to the target databases while ELT move data directly to the target databases where the transformation takes place.

The benefits to the ELT architecture:

- a variety of integration processes are more effective because they leverage the underlying database, file system or API functionality
- easier to learn, develop and maintain, especially with an IT staff that may be working on a variety of tasks and not just dedicated to data integration
- more robust with best practices such as CDC, SCD, error handling, restart and recovery, and managing integration processes
- performance tuning can be done by database administrators or ELT developers
- avoids the expense of acquiring, maintaining and tuning the additional server

### Other Tool Recommendations

Both Oracle and Microsoft SQL Server would perform well as a DW. During the assessment, however, there was some discussion about what options UMS was licensed for on an enterprise level. If our recommendations for Oracle database features are not available, then Microsoft SQL Server should be used; it offers the advantage of being on a Microsoft platform. If the UMS is considering cloud deployments, then Azure offers a very robust platform and Oracle Cloud is catching up.

UMS should also use a data modeling tool such as Erwin, ER/Studio or Toad Data Modeler.

### *Concluding Remarks*

The UMS needs to invest in creating an enterprise DW and enabling BI that will be used to better serve its students, faculty, administration and other stakeholders. The UMS is in a situation shared by many other enterprises, where its initial DW effort was too operationally focused with a misfit of the architecture, data and tools that resulted in the DW being bypassed for custom-coded applications and spreadsheets. The UMS needs to use its initial DW as a learning experience, leverage current best practices and create a new DW built for its various communities and stakeholders.

The key design principles that will drive success for the UMS are:

- **Focus on architecture** - build an analytical data architecture that supports all the UMS communities in their need for data integration and BI. Allow the UMS team to detect changes in source systems and track and record changes to data over time, update the data warehouse more frequently, and enhance error-handling.

- **Use best practices** – including building incrementally and iteratively. Do not try to create an enterprise DW for everyone all at once, but select projects that will benefit targeted communities and stakeholders.
- **Get commitment** - obtain community participation and commitment in the design, development and use of the DW and BI solutions. People are the most important part of any project.
- **Invest in training** – train both the IT staff and the people who will use BI. A small portion of this is vendor-specific tool training, but the bulk of the training should be learning the best practices in analyzing and integrating data.
- **Choose technology** - select best-fit technology for the BI community and IT.

Although creating an enterprise DW and deploying new BI solutions often seems overwhelming, the UMS has the needs, resources and skills to be highly successful in its future efforts. Athena IT Solutions is here to help guide and implement that process if the UMS would like assistance in their future efforts.