Syllabus

Course Overview

IT4320 concentrates on a specific technical process known as extraction, transformation, and loading (ETL). These are critical steps in preparing and providing data that best supports the work of data analytics. The work of ETL is to identify, clean, transform, and serve the data that is the most effective for use in a specific data analytics project. You will apply tools and technologies specific to these activities in order to better understand the alternatives and application appropriate within a particular context.

Course Competencies

(Read Only)

To successfully complete this course, you will be expected to:

1. Define the role of data structure and quality as part of the Extraction, Transformation and Load (ETL) processes and workflows
2. Apply data quality and cleansing concepts to a specific ETL project.
3. Define the role that ETL plays in a specific data analytics project.
4. Apply the tools and technologies that are available to support ETL as a component of a data analytics project.
5. Communicate effectively.

Course Prerequisites

IT2230.

Syllabus >> Course Materials

Required

The materials listed below are required to complete the learning activities in this course.

Library

The following required readings are provided in the Capella University Library or linked directly in this course. To find specific readings by journal or book title, use Journal and Book Locator. Refer to the Journal and Book Locator library guide to learn how to use this tool.

- Data cleansers. (2011). *B to B, 96*(12), 15.
- Data migration requires process, tools, and expertise. (2008). *Computer Economics Report, 30*(6), 15–20.
- Ehrenmann, M., Pieringer, R., & Stockinger, K. (2012). Is there a cure-all for business analytics? *Business Intelligence Journal, 17*(3), 28–39.
- Eslambolchi, H. (2012). Making the most of big data. *CIO Insight,* (121), 10.
- George, G., Haas, M. R., & Pentland, A. (2014, April). Big data and management. *Academy of Management Journal*, 321–326.
- Henschen, D. (2012). Big data talent war: 10 analytics job trends. *Informationweek*.
- Jaspersoft delivers big data into Jaspersoft ETL. (2012). *Productivity Software, 25*(3), 7–8.
- Karacsony, K. (2006, January). ETL is a symptom of the problem, not the solution!. *DM Review,* 26–28.
- Kolfschoten, G. L., Niederman, F., Briggs, R. O., & de Vreede, G. (2012). Facilitation roles and responsibilities for sustained collaboration support in organizations. *Journal of Management Information Systems, 28*(4), 129–162.
- Krudop, M. (2005). Maximizing your ETL tool investment. *DM Review, 15*(3), 26–28.
- Kurukunda, P. (2013). Planning the move. *Best's Review, 113*(11), 71–73.
- Levy, P. (2009). Keep it clean. *Marketing News, 43*(20), 8–9.
- Longbottom, C. (2007, October 23). Clean data for an efficient enterprise. *Computer Weekly*, 44.
- Models of collaboration as the foundation for collaboration technologies. (2010). *Journal of Management Information Systems, 27*(1), 97–122.
- Peterson, T. (2003). Data scrubbing. *Computerworld, 37*(6), 32.
- Politano, T. (2009). Right-time ETL and integration. *Information Management, 19*(1), 28–31.
- Purdue employees FCU deploys high-tech scrubbing tools. (2009). *Credit Union Times, 20*(44), 35.
- Sallam, R., Beyer, M., & Heudecker, N. (2013). Key trends in big data technologies.
- Santaferraro, J. (2012). Offloading analytics. *Business Intelligence Journal, 17*(4), 43–48.
- Sherman, R. (2007, April). Trial-and-error method of ETL. *DM Review,* 27.
- Sherman, R. (2009). Beyond ETL & data. *Information Management* (*1521-2912*), 2–7.
- Smith, H. A., & McKeen, J. D. (2011). Enabling collaboration with IT. *Communications of the Association For Information Systems, 28,* 243–254.
- Songini, M. L. (2004). ETL. *Computerworld, 38*(5), 23.
- Stonebraker, M. (2013). What does 'big data' mean? *Communications of the ACM, 56*(9), 10–11.
- Zhang, F., Xue, H., Xu, D., Zhang, Y., & You, F. (2013). Big data cleaning algorithms in cloud computing. *International Journal of Interactive Mobile Technologies, 7*(3), 77–81.

External Resource

Please note that URLs change frequently. While the URLs were current when this course was designed, some may no longer be valid. If you cannot access a specific link, contact your instructor for an alternative URL. Permissions for the following links have been either granted or deemed appropriate for educational use at the time of course publication.

- Alteryx. (Producer). (2013). *Data cleansing – Cleansing data and enriching it for customer analytics* [Video]. Retrieved from http://www.youtube.com/watch?v=BOlyyg64Hmc
- Hillam, J. (Producer). (2011). What is an ETL tool? [Video]. Retrieved from http://www.youtube.com/watch?v=K_FCHYWGGug
- LearnItFirst (Producer). (2013). *The importance of ETL process* [Video]. Retrieved from http://www.youtube.com/watch?v=FzayiGi97bc
- Mir, H. (Producer). (2013). *Business intelligence component – Extract, transform, and load (ETL)* [Video]. Retrieved from http://www.youtube.com/watch?v=pXxRtOlDSXw
- Musick, B. (Producer). (2010). *Data quality – Data cleaning* [Video]. Retrieved from http://www.youtube.com/watch?v=hUY3skSprvo
- Passionned Group. (2014). List of ETL tools. Retrieved from http://www.etltool.com/list-of-etl-tools/
- Song, X., & Liu, X. (2011). An approach for designing, modeling, and realizing ETL processes based on unified views model. *International Journal of Software Engineering & Knowledge Engineering*, *21*(4), 543–570. Retrieved from https://www.researchgate.net/profile/Xiaobing_Liu4/publication/220344306_An_Approach_for_Designing_Modeling_and_Realizing_ETL_Processes_Based_on_Unifie
  Approach-for-Designing-Modeling-and-Realizing-ETL-Processes-Based-on-Unified-Views-Model.pdf
- Survis, G. (Producer). (2014). *Taming big data* [Video]. Retrieved from http://www.youtube.com/watch?v=hu8x2iBw6V4&list=PL2F3CA233083B6DEC

- Unique Solutions. (2013). *What is ETL process?* [Video]. Retrieved from http://www.youtube.com/watch?v=nkVzN62LisA

Suggested

The following materials are recommended to provide you with a better understanding of the topics in this course. These materials are not required to complete the course, but they are aligned to course activities and assessments and are highly recommended for your use.

Optional

The following optional materials are offered to provide you with a better understanding of the topics in this course. These materials are not required to complete the course.

Projects

Project >> Presentation and ETL Tools

Project Overview

Data cleansing is necessary for data quality and the transformation process. In this project, you will research and select a type of data cleansing technology, software, or product. The purpose is to identify, cleanse, and correct data discrepancies. Therefore, your chosen software or product should support the ETL process. In your research, consider how the selection can impact an organization's ability to effectively manage the data transformation process.

You will also select a specific organization or type of organization for which the software or product could be implemented. A data cleansing software or product designed for smaller data projects is not going to be the appropriate solution for a multinational or multi-facility organization. By selecting a specific organization engaged in a particular industry and of a particular size you can demonstrate your ability to apply the knowledge you gained through the research projects and study materials. You are welcome to select an organization and then research an applicable software or product or you may select a software or product and identify an organization that would be an appropriate environment in which to implement it.

The project can be thought of as having three parts, described as follows:

- In the first part, you will identify a data cleansing technology, software, or product according to the guidelines above. You will describe it and address the impact that it will have on the ETL process within the particular organization you have selected. You will also describe the specific and direct correlation to data quality within the organization.
- The second part will be focused on the implementation plan, including specific strategies and the impact on data analytics.
- In the third part of the project, you will compile all previous components into a presentation.

- **Written communication:** Written communication is free of errors that detract from the overall message.
- **APA formatting:** Resources and citations are formatted according to APA (6th Edition) style and formatting.
- **Number of resources:** Minimum of 10 resources.
- **Length of paper:** 8–10 pages typed double-spaced pages.
- **Font and font size:** Times, 12 point.

Unit 1 >> Data Quality and ETL Process

Introduction

This unit discusses the role of data structure and quality as it relates to the extraction-transformation-load (ETL) process, which is a concept related to data warehousing and the storage of data.

The first part of an ETL process involves extracting the data. Data extraction can occur from multiple source systems and sets the stage for how subsequent processes occur. The transformation stage is where a series of rules or functions of the data are determined in order to become meaningful information for the organization. This is

also the phase where data manipulation occurs, in order to meet the business needs of the organization. The last phase, load, involves the process of loading or storing the data. Data can be store in a variety of ways; the purpose is to have access to and maintain historical data for reporting and other business processes.

The ETL process is an essential component to establishing data structure and quality. The ETL process can help to resolve challenges of extracting data from various sources in addition to cleaning and transforming the data. It is directly related to data quality, which refers to processes that result in the completeness and validity of data. ETL impacts the way in which data quality is defined as a process to essentially gather, enter, transform, and extract data.

The readings this week will provide you with an understanding of key concepts of the ETL process as well a data quality and structure.

Learning Activities

u01s1 - Required Readings

- Kurukunda, P. (2013). Planning the move. *Best's Review, 113*(11), 71–73. The article focuses on data migration by using extract, transform and load (ETL) technology. It also discusses other considerations that impact data integration and migration, such as project management and software development.
- Song, X., & Liu, X. (2011). An approach for designing, modeling, and realizing ETL processes based on unified views model. *International Journal of Software Engineering & Knowledge Engineering, 21*(4), 543–570. This article discusses Extraction-Transformation-Loading (ETL) tools as pieces of software responsible for the extraction and cleaning of data from several sources, into data warehouses.
- Sherman, R. (2009). Beyond ETL & data. *Information Management (1521-2912)*, 2–7. The article discusses Extract, Transform, Load (ETL) processes and data warehousing.
- Politano, T. (2009). Right-time ETL and integration. *Information Management, 19*(1), 28–31. This article describes the best practices used to fully leverage developments in data loading and integration.

Resources

u01s2 - Web Resources

Videos

The objective of these videos is to enhance the required reading, to provide context on data integration tools, and to support completion of the Unit 1 assignment and discussion.

- Unique Solutions. (2013). *What is ETL process?* [Video]. Retrieved from http://www.youtube.com/watch?v=nkVzN62LisA – 5 minutes.
- Mir, H. (Producer). (2013). *Business intelligence component – Extract, transform, and load* (ETL) [Video]. Retrieved from http://www.youtube.com/watch?v=pXxRtOlDSXw – 8 minutes.
- LearnItFirst (Producer). (2013). *The importance of ETL process* [Video]. Retrieved from http://www.youtube.com/watch?v=FzayiGi97bc – 15 minutes.

Internet and Library Research

Use the following keywords to engage in research on the topic of data integration tools and techniques, which will support completion of the Unit 1 assignment and discussion.

- Extraction-Transformation-Load (ETL) process.
- Data warehousing.
- Key aspects of data quality.
- ETL tools.
- ETL limitations.

Resources

u01s3 - SAS E-Learning and Software

Accessing Your SAS E-Learning Materials and SAS Certification Preparation Courses

As a business intelligence or data analytics learner at Capella University, you have access to a comprehensive list of SAS e-learning materials and SAS certification preparation courses. Whether you are interested in becoming SAS certified or you are just looking for additional resources to help advance your understanding of a particular SAS tool, simply activate your Capella SAS account to gain access to resources that will help you advance in the field of data analytics.

- How to Activate Your Capella SAS Account.
- Your E-Learning Material and Certification Preparation Courses.

Access SAS Analytics U Community and Software Tools

Once you create your Capella SAS profile account, you become a member of the SAS Analytics U community. This is an online interactive community where you will find valuable resources and free software tools.

SAS University Edition

As a member of the SAS Analytics U, you can download your own free version of SAS University Edition or gain Web access to more robust and powerful SAS tools by registering for SAS OnDemand. Use this link to learn more about SAS University Edition and download your own copy of SAS University Edition.

- SAS University Edition.

SAS OnDemand Tools

To obtain cloud access to more powerful and robust SAS Tools, you must first register with SAS OnDemand.

Once you have registered for SAS OnDemand and have obtained your user ID, simply click the link below that corresponds with the tool that you would like to access.

- SAS Enterprise Minor.
- SAS Studio.
- SAS Enterprise Guide.
- SAS Forecast Studio. The name of the Environment is C867a0717de594199827634ad02887e5b.
- SAS JMP.

As you work your way through this course and through your program, lean on the SAS Analytics U Community to connect with fellow SAS users. Reach out other community members to obtain peer support, get questions answered, share ideas and best practices, and collaborate on projects.

Another available resource is the SAS OnDemand for Academics User's Guide.

Resources

u01a1 - The Extraction-Transformation-Load (ETL) Process

You have recently been hired as a consultant for a local high school, to develop a data model and process to better retain and maintain student records. The school's process will involve data extraction from various sources.

Create a diagram that illustrates the concept of Extraction-Transformation-Load (ETL) process. Using your model of the ETL diagram, write a paper to:

- Outline the ETL process.
- Describe importance of data quality in the ETL process.
- Describe any assumptions or limitations with data collection, extraction, and storage.
- Compare and contrast the key aspects data quality aspects.
- Analyze the school's requirements and review of required data elements.
- Overview of types of reports that can be used for reviewing student data.

Assignment Requirements

Written communication: Written communication is free of errors that detract from the overall message.
APA formatting: Resources and citations are formatted according to APA (6th edition) style and formatting.
Length of paper: 2–3 pages, excluding the references page.
Font and font size: Times New Roman, 12 point.

Resources

- Capella Undergraduate Resources: APA Citing and Referencing.

Course Resources

APA Style and Format

u01d1 - ELT Limitations

Discuss potential limitations within the ETL process in relation to data quality. Provide examples of how limitations can occur and how they can be resolved through the use of quality data.

Response Guidelines

After reviewing several of your peers' posts, compare your findings to at least two of your peers.

Resources

Course Resources
Undergraduate Discussion Participation Scoring Guide

Unit 2 >> Data Quality and Cleansing Concepts

Introduction

This unit discusses the importance of data quality and concepts of data cleansing. Data cleansing, sometimes referred to data scrubbing, is the process of identifying and correcting data discrepancies and inaccurate information. The goal of data cleansing is not only for "cleaning" purposes, but to ensure that data is consistent, relevant, and capable of integrating with other databases and systems. All these characteristics are reflective of data quality.

Data is often used for making business decisions; therefore, it is very important to have reliable, high-quality data. As discussed in Unit 1, data quality is also critical to the extraction, transformation, and load workflows.

The readings this week will provide you with an understanding of data cleansing and its relationship to data quality. You will have an opportunity to review how data cleansing techniques are implemented in addition to exploring specific data cleansing products. You will also complete the first part of the course project, which focuses on the selection and implementation of a data cleansing process as part of an ETL project.

Learning Activities

u02s1 - Required Readings

- Zhang, F., Xue, H., Xu, D., Zhang, Y., & You, F. (2013). Big data cleaning algorithms in cloud computing. *International Journal of Interactive Mobile Technologies, 7*(3), 77–81. This article discussed big data cleaning as an important research issues.
- Henschen, D. (2012). Big data talent war: 10 analytics job trends. *Informationweek*. The article discusses salary trends for business intelligence and information management professionals.
- Longbottom, C. (2007, October 23). Clean data for an efficient enterprise. *Computer Weekly, 44*. The article presents the author's comments on the importance of cleansing corporate data in ensuring efficiency and compliance.
- Peterson, T. (2003). Data scrubbing. *Computerworld, 37*(6), 32. Discusses the importance of data hygiene to computers, in addition to causes of dirty data and the difference between data hygiene and data quality.
- Purdue employees FCU deploys high-tech scrubbing tools. (2009). *Credit Union Times, 20*(44), 35.The article reports on the data scrubbing tools used by Purdue Employees Federal Credit Union (PEFCU), in a case study.
- Levy, P. (2009). Keep it clean. *Marketing News, 43*(20), 8–9. The article focuses on maintaining accurate information contained in databases. Benefits of vendors and how they correct issues are discussed, as well as the benefits of database cleaning software. Insets include 5 Tips for Clean Data Entry and 3 Ways to Clean Your Database.
- Data cleansers. (2011). *B to B, 96*(12), 15. This document presents a chart that lists the data cleanser companies.

Resources

u02s2 - Web Resources

Videos

The objective of these videos is to enhance the required reading, to provide context on data integration tools, and to support completion of the Unit 2 assignment and discussion.

- Alteryx. (Producer). (2013). *Data cleansing – Cleansing data and enriching it for customer analytics* [Video]. Retrieved from http://www.youtube.com/watch?v=BOlyyg64Hmc – 25 minutes.

- Musick, B. (Producer). (2010). *Data quality – Data cleaning* [Video]. Retrieved from http://www.youtube.com/watch?v=hUY3skSprvo – 10 minutes.

Internet and Library Research

Use the following keywords to engage in research on the topic of data integration tools and techniques, which will support completion of the Unit 2 assignment and discussion.

- Data cleansing.
- Data quality.
- Data quality versus data cleansing.
- Data cleansing product.

Resources

u02a1 - Selection for Data Cleansing

The purpose of this part of the project is to select and summarize your data cleansing software or product.

Write a 2–3 page paper with the following sections:

Project Definition

- Choose a data cleansing software or product.
- Provide a description and describe how it relates to data quality.
- Discuss how it may contribute to organizational effectiveness, in relation to ETL.
- Identify potential advantages and disadvantages of using the software or product.

Type of Organization

- Define specific organization or type of organization for which the data cleansing software or product will/can be implemented.
- Discuss reasons for implementation (i.e., in support of specific issues with data management, poor-quality data, data transfer, data integration, etc.)
- Discuss the importance and/or need.

Assignment Requirements

Written communication: Written communication is free of errors that detract from the overall message.
APA formatting: Resources and citations are formatted according to APA (6th edition) style and formatting.
Length of paper: 2–3 pages, excluding the references page.
Font and font size: Times New Roman, 12 point.

Resources

- Capella Undergraduate Resources: APA Citing and Referencing.

Course Resources

APA Style and Format

u02d1 - Data Quality Versus Data Cleansing

Discuss the difference between data cleansing and data quality. Then, describe how they work collectively to contribute to an effective ETL process. Feel free to be resourceful and find resources in addition to the required readings for the week.

Response Guidelines

After reviewing several of your peers' posts, compare your findings to at least two of your peers.

Resources

Course Resources

Undergraduate Discussion Participation Scoring Guide

u02d2 - Choosing a Data Cleaning Product

What are some common practices for data cleansing? What factors are considered when an organization selects a data cleansing or scrubbing product? Compare the information to the software or product you have chose based on organizational needs.

Response Guidelines

Respond to at least one peer.

Resources

Course Resources

Undergraduate Discussion Participation Scoring Guide

Unit 3 >> ETL and Data Analytics

Introduction

At some point in time, we have all been faced with the dilemma of forgetting an online password. Take online banking, for example. In order to reset login credentials, you often have to go through a series of security questions, then a process of e-mails or responses. There is an enormous amount of information captured in this process and with customer accounts and e-mails, this is a prime example of the Big Data concept.

In today's world, almost every aspect of our lives is impacted by the collection of data; from online banking, social media, and even logging into your course. The concept of Big Data consists, in great part, of the extract, transform, and load process. The ETL is the process of taking data from sources and combining them in a format to perform data analysis. Further, data analytics is only as good as the data in the ETL process.

The readings this week will provide more information on the concept of Big Data, its relationship to ETL, and the impact on data analytics. You will also continue with your course project by exploring the implementation of your data cleansing software or product in addition to its relation to data analytics.

Learning Activities

u03s1 - Required Readings

- Songini, M. L. (2004). ETL. *Computerworld, 38*(5), 23. This article discusses extract, transform and load (ETL), the processes that enable companies to move data from multiple sources, reformat and cleanse it, and load it into another database or on another operational system.
- George, G., Haas, M. R., & Pentland, A. (2014, April). Big data and management. *Academy of Management Journal*, 321–326.
- Karacsony, K. (2006, January). ETL is a symptom of the problem, not the solution!. *DM Review*, 26–28. The article presents the author's views on extract, transform and load (ETL) processes and direct correlation between the extent of redundant data and the amount of ETL processes.
- Sallam, R., Beyer, M., & Heudecker, N. (2013). Key trends in big data technologies. The article discusses the trends in information technology related to Big Data.
- Santaferraro, J. (2012). Offloading analytics. *Business Intelligence Journal, 17*(4), 43–48. The article focuses on the importance of data analytics and challenges faced by old technologies. It states that by creating a performance-based data solution can help to improve data warehouse infrastructure.
- Stonebraker, M. (2013). What does 'big data' mean?. *Communications of the ACM, 56*(9), 10–11.
- Ehrenmann, M., Pieringer, R., & Stockinger, K. (2012). Is there a cure-all for business analytics?. *Business Intelligence Journal, 17*(3), 28–39.This article presents a study of three brands representing business analytics in the fields of banking, telecommunications, and retail. Analytical questions are discussed of the respective industries and provide an overview of the analytical requirements. The biggest impact of specific data integration requirements are also discussed.
- Eslambolchi, H. (2012). Making the most of big data. *CIO Insight*, (121), 10. This article may be helpful in identify potential implementation issues, such as costs. The author discusses the best practices in information technology (IT) in businesses industries in the U.S. He mentions that good results happen after balancing capital and operational costs constraints. Also, the technologies to be considered for chief information officers (CIO) who are looking to capitalize big data are explored.
- Jaspersoft delivers big data into Jaspersoft ETL. (2012). *Productivity Software, 25*(3), 7–8. This article provides an example of implementing ETL processes. It reports on Jaspersoft Corp.'s original equipment manufacturing (OEM) agreement with Talend Inc. towards the inclusion of native connectors to Apache Hadoop big data software in its extract, transform and load (ETL) operations in the U.

Resources

u03s2 - Web Resources

Videos

The objective of these videos is to enhance the required reading, to provide context on data integration tools, and to support completion of the Unit 3 assignment and discussion.

- Survis, G. (Producer). (2014). *Taming big data* [Video]. Retrieved from http://www.youtube.com/watch?v=hu8x2iBw6V4&list=PL2F3CA233083B6DEC – 15 minutes.

Internet and Library Research

Use the following keywords to engage in research on the topic of data integration tools and techniques, which will support completion of the Unit 3 assignment and discussion.

- Data analytics.
- Big Data.
- Implementing ETL tool.
- Implementing data cleansing tool.
- Role of ETL in data analytics.

u03a1 - Implementation

This part of your course project will be focused on the implementation plan, specifically factors that impact data analytics. You will also identify potential issues in implementing your chosen IT tool (data cleansing software or product) based on your case or organization.

Write a 3–4 page paper with the following sections:

Implementation and Collaboration Plan

- This section should discuss the approach for implementing the data cleansing software or product. You can either use a case study or actual organization for which the product will be implemented.
- Specifically examine how the software or product will support data management and analytics.
- Discuss the relationship between concept of the software or product and ETL.
- Discuss the impact of software or product on the data analytics process.
- Explain the role that ETL plays in relation to data analytics project.

Potential Issues

- Provide a description of potential issues as a result of the implementation.
- Develop an outline of action items to address issues.

Implementation Schedule

- Establish a baseline schedule for the project.
- Discuss ways in which progress will be monitored.
- Outline responsible persons and their roles.
- Discuss a contingency plan for schedule variances.
- Include milestones with an estimated completion timeframe.
- Discuss time constraints and objectives.

Assignment Requirements

Written communication: Written communication is free of errors that detract from the overall message.
APA formatting: Resources and citations are formatted according to APA (6th edition) style and formatting.
Length of paper: 3–4 pages, excluding the references page.
Font and font size: Times New Roman, 12 point.

Resources

- Capella Undergraduate Resources: APA Citing and Referencing.

Course Resources

APA Style and Format

u03d1 - Implementation Plan

An implementation plan can serve as the "roadmap" to the implementation process. What elements do you think should be included in an implementation plan for an IT tool, such as a data cleansing software or product? How might the implementation of your chosen data cleansing software or product differ from other IT tools?

Response Guidelines

After reviewing several of your peers' posts, compare your findings to at least two of your peers.

Resources

Course Resources
Undergraduate Discussion Participation Scoring Guide

u03d2 - What Would You Do?

As the IT compliance manager for Bank of CCA, you are responsible for ensuring that the review of all customers' information follows proper procedures. After a review of the monthly access report, you noticed that an employee has been reviewing customers' bank information for long periods of time and without reason. What would you do? What policies or procedures would you put in place to ensure that proper protocol is followed for reviewing customer information? How would your recommendations relate to ETL processes and future use of data?

Response Guidelines

Respond to at least one peer.

Resources

Course Resources
Undergraduate Discussion Participation Scoring Guide

Unit 4 >> Tools and Technologies

Introduction

As discussed in previous units, the ETL process is used to transfer data from one place to another or to convert data from one format to another. This is often accomplished through the use of ETL tools.

Some of the most common tools are Oracle and SQL products. But the selection of an ETL tool is dependent upon organizational needs and current data processes and systems. ETL tools provide the ability to take data in different formats and systems and transform into quality information that is used for business reporting and analytics.

The readings this week will provide additional information on ETL tools and technologies. Specific ETL products will be explored as well as examples of organizations that have implemented ETL tools. You will continue with your course project by identifying and describing the potential use of ETL tools.

Learning Activities

u04s1 - Required Readings

- Song, X., & Liu, X. (2011). An approach for designing, modeling, and realizing ETL processes based on unified views model. *International Journal of Software Engineering & Knowledge Engineering, 21*(4), 543–570. Extraction-Transformation-Loading (ETL) tools are pieces of software responsible for the extraction of data from several sources, their cleaning, customization and insertion into Data Warehouses (DWs). Complexity, usability and maintainability are the primary problems concerning ETL processes. This paper discusses those problems and provide a dynamic approach for designing, modeling and realizing ETL processes.
- Krudop, M. (2005). Maximizing your ETL tool investment. *DM Review, 15*(3), 26–28. This article explores that ETL tools serve two very specific purposes. First, they provide a development environment that is easy to manage. Second, ETL tools provide increased throughput over scripting and database coding. The increased throughput is achieved by separating data management from data access.
- Sherman, R. (2007, April). Trial-and-error method of ETL. *DM Review*, 27
- Data migration requires process, tools, and expertise. (2008). *Computer Economics Report, 30*(6), 15–20. The article focuses on the report regarding data migration process and its open source tools. It states that successful migration process includes mapping, extracting, cleansing, loading, and verifying. It also mentions that data migration decisions should consider costs and benefits, partial migration, data compatibility, outsourcing, documentation, and automation.

Resources

u04s2 - Web Resources

Videos

The objective of these videos is to enhance the required reading, to provide context on data integration tools, and to support completion of the Unit 4 assignment and discussion.

- Hillam, J. (Producer). (2011). *What is an ETL tool?* [Video]. Retrieved from http://www.youtube.com/watch?v=K_FCHYWGGug – 15 minutes.

Internet and Library Research

Use the following Web site and keywords to engage in research on the topic of data integration tools and techniques, which will support completion of the Unit 4 assignment and discussion.

- List of ETL Tools that are used widely used for extracting, cleaning, transforming, and loading data from different systems.
- Keywords:
  - ETL tool.
  - Use of ETL tools.
  - Benefits of ETL tools.
  - Alternative ETL tools.
  - ETL techniques.

Resources

u04a1 - Presentation and ETL Tools

This part of your course project will compile all components into a comprehensive presentation and include an additional section on the use of ETL tools. Your presentation should include an overview of requirements in previous units.

Submit your 8–10 page written report as well as a presentation, with the following sections:

Project Definition

- Type of organization.
- Implementation and collaboration plan.
- Potential issues.
- Implementation schedule.

Use of ETL tools

- Discuss the definition of an ETL tool.
- Describe the use of your chose product or software as an ETL tool.
- Describe ways in which an ETL tool can be used.
- Describe why the ETL tool is beneficial to organization.

- Discuss alternative ETL tools that can be used or in addition to chosen tool.

Next Steps and Conclusion

Assignment Requirements

Written communication: Written communication is free of errors that detract from the overall message.
APA formatting: Resources and citations are formatted according to APA (6th edition) style and formatting.
Length of paper: 8–10 pages, excluding the references page.
Font and font size: Times New Roman, 12 point.

Resources

- Capella Undergraduate Resources: APA Citing and Referencing.

Course Resources
APA Style and Format

u04d1 - Common Mistakes

At a minimum, common mistakes are expected due to the effect of human error. But mistakes can also be avoided through careful planning and well-developed processes. What do you think are some common mistakes made by companies in their initial use of the ETL process? How would you plan to eradicate those mistakes?

Response Guidelines

After reviewing several of your peers' posts, compare your findings to at least two of your peers.

Resources

Course Resources
Undergraduate Discussion Participation Scoring Guide

Unit 5 >> Collaboration and IT Tools

Introduction

Collaboration is a critical component to any project. Getting everyone on the same page, at the same time, and with the same goal can be challenging.

The adoption of ETL and related tools requires an interrelated process that involves team management and organizational collaboration. Therefore, successful implementation generally cannot be achieved in isolation. It requires some level of collaboration and sometimes the use of collaboration tools.

This unit will explore the idea of collaboration as well as the use of collaboration methods to assist in the adoption of information technology. The readings this week will provide additional information on the importance of collaboration and team processes to address issues associated with IT.

Learning Activities

u05s1 - Required Readings

- Kolfschoten, G. L., Niederman, F., Briggs, R. O., & de Vreede, G. (2012). Facilitation roles and responsibilities for sustained collaboration support in organizations. *Journal of Management Information Systems, 28*(4), 129–162. This paper explores how IT tasks and associated roles can be anchored in organizations, and the relationship of task allocation patterns to the sustained use of collaboration technology in organizations.
- Models of collaboration as the foundation for collaboration technologies. (2010). *Journal of Management Information Systems, 27*(1), 97–122. This article discusses two basic models of collaborative teamwork, comparing their relative merits in terms of technology-based collaboration and other aspects of information systems.
- Smith, H. A., & McKeen, J. D. (2011). Enabling collaboration with IT. Communications of the Association For Information Systems, 28, 243–254. This article discusses the need to improve collaboration, while information technology (IT) is at the center of related trends. Additionally, different types of collaboration and why collaboration is becoming so important, is also discussed.

Resources

u05s2 - Web Resources

Internet and Library Research

Use the following keywords to engage in research on the topic of data integration tools and techniques, which will support completion of the Unit 5 assignment and discussion.

- Collaboration strategies.
- Use of teams and work groups for project implementations.
- Collaboration and communication strategies.
- ETL techniques.

Resources

u05a1 - Collaboration Strategies

Research and write a 2–3 page paper in APA format that covers the following topics:

- Importance of collaboration and teams in implementing IT systems and tools.
- Specific collaboration strategies and impact on data analytics.
- Collaboration strategies or a model that could be beneficial to your project.
- The use of teams and workgroups in managing implementations.

Assignment Requirements

Written communication: Written communication is free of errors that detract from the overall message.
APA formatting: Resources and citations are formatted according to APA (6th edition) style and formatting.
Length of paper: 2–3 pages, excluding the references page.
Font and font size: Times New Roman, 12 point.

Resources

- Capella Undergraduate Resources: APA Citing and Referencing.

Course Resources
APA Style and Format

u05d1 - Influence of Collaboration

Discuss the relationship between collaboration and impact (pros and cons) for data transformation processes. Also explain the potential influence on data transformation. Support your post with at least one reference to the literature.

Response Guidelines

After reviewing several of your peers' posts, compare your findings to at least two of your peers.

Resources

Course Resources
Undergraduate Discussion Participation Scoring Guide

u05d2 - Take-Aways

Explain to your peers and your instructor what you will be taking away from this course:

- What did you learn that surprised you?
- What did you find challenging to understand or grasp?
- What aspects of the course did you enjoy? Which did you not enjoy?
- What would you like to see added to the course for future sessions?

Response Guidelines

Read other learners' posts and reply as you see fit. It is not required to respond to your peers' postings, however,

Resources

Course Resources
Undergraduate Discussion Participation Scoring Guide