eLearningCurve

Data Integration Education

Online Education • Certification • Enterprise Solutions

- Data Integration Fundamentals and Best Practices
- Data Warehousing Fundamentals
- Data Virtualization
- Ensuring Data Quality in Data Integration
- Big Data Fundamentals
- MDM Fundamentals: Architecture and Implementation
- Data Profiling
- Data Integration Techniques for Designing an ODS
- Data Parsing, Matching, and De-Duplication
- Metadata Management Fundamentals
- DW & BI Data Modeling
- Streaming Data: Concepts, Applications, and Technologies
- Introduction to Graph Databases

March 2021

TABLE OF CONTENTS

Data Integration (DI) is an essential information management capability that provides the foundation for enterprise-wide views of consistent, connected, and trusted business information. The alternative to data integration is data disparity which leads to miscommunication, misunderstanding, confusion, uncertainty, and misinformed business decisions and actions.

Data integration is important, but it is complex and challenging. The variety of reasons for data integration (data warehousing, master data management, data migration, etc.), the increasing scope of data sources (enterprise data, external data, big data, web data, etc.), and the growth in data integration techniques and technologies (ETL, ELT, federation, virtualization, etc.) all contribute to data integration complexities.

From defining integration requirements to acquiring and unifying data the integration choices are abundant. Effective and sustainable data integration systems depend on skilled and educated data professionals from architects to implementers.

WHAT PEOPLE ARE SAYING ABOUT ELC

I found the courses to be very well designed; they were intellectually stimulating and challenging, but the workload was manageable.

-- Marina Severinovskaya, CIMP Ex - Data Governance, CIMP - Data Quality, MDM, IM Foundations, USA

CURRICULA AT-A-GLANCE

Full course descriptions begin on page 7.

Data Integration Fundamentals and Best Practices

Instructor: Dave Wells

This 5-hour course discusses architectures, requirements, methods, roles and activities of data integration that can be applied to achieve successful data integration projects for a variety of applications and circumstances.

Data Warehousing Fundamentals

Instructor: Mark Peco

This 4.5 hour online course re-defines the scope of the "modern" data warehouse. We will discuss the challenges of gathering useful information requirements, design approaches, development, testing and quality management techniques.

Data Virtualization

Instructor: Dave Wells

This 3-hour online training course will introduce you to the concepts, techniques, and capabilities of data virtualization.

Ensuring Data Quality in Data Integration

Instructor: Arkady Maydanchik

. This 5-hour course discusses various practices that can maintain high data quality through data integration.

Big Data Fundamentals

Instructor: William McKnight, Jake Dolezal

This 3.5 hour course will help you make the most of big data and make the best choices to ensure information remains an unparalleled corporate asset.

MDM Fundamentals: Architecture & Implementation

Instructor: William McKnight

This 4.5 hour course provides a comprehensive look at the elements of an MDM program and the key success factors for MDM.

Data Profiling

Instructor: Arkady Maydanchik

Data profiling is one of the most common and important activities in information management. This 5-hour course teaches all practical skills necessary to succeed in a data profiling initiative.

Data Integration Techniques for Designing an ODS

Instructor: Angelo Bobak

Business data integration is a complex problem that must be solved when organizations change or enhance their internal structures. This 3-hour online course presents a simple yet thorough process that describes the challenges of building an Operational Data Store (ODS) and the solutions to these challenges.

Data Parsing, Matching, and De-Duplication

Instructors: Kathy Hunter, William McKnight, Henrik Sørensen

To take advantage of the worldwide marketplace, businesses need to manage data globally. In this 4.5-hour course you will learn to identify and avoid the pitfalls of global information.

Metadata Management Fundamentals

Instructors: Arkady Maydanchik and Dave Wells

This 4-hour course covers foundational metadata knowledge needed by anyone who has data management roles and responsibilities, from metadata basics to modeling, profiling, and cataloging data.

DW and BI Data Modeling

Instructor: Rick Sherman

This 4-hour course includes a mix of data modeling concepts, best practices, applications and practical examples that will help you build effective data warehouse and business intelligence applications.

Streaming Data: Concepts, Applications, and Technologies

Instructors: Kevin Petrie & Dave Wells

This 3-hour course covers the concepts, applications, and business and technical drivers for streaming data adoption, including an in-depth discussion of Apache Kafka.

Introduction to Graph Databases

Instructor: Johns Singer

This 4-hour online course will provide an overview of property graph database technology and teach the student how to translate business requirements to a property graph database design that can be implemented on any modern property graph database.

CERTIFICATION PROGRAM

CIMP: Demonstrate Mastery. Achieve Success.

Certification is an important tool for job seekers and for employers seeking to hire the most qualified people. eLearningCurve offers a robust certification program, Certified Information Management Professional (CIMP) that builds upon education to certify knowledge and understanding of information management.

The CIMP Data Integration designation will make a clear statement that you have learned from the industry leaders and have demonstrated thorough understanding of data integration by passing several challenging exams.

For the true experts and standard bearers in the industry we offer the second level of CIMP certification - CIMP Ex. To earn the CIMP Ex designation you must demonstrate a combination of great Expertise, Experience, and Excellence.

What Sets CIMP Apart?

Rigorous exam system: We go beyond the basics. Rather than testing for knowledge that any industry professional should know, CIMP exams test an in-depth knowledge, comprehensive understanding, and ability to apply various concepts to a problem. You can be proud of your achievement of the CIMP designation, and hiring managers can be sure they are getting a highly knowledgeable employee.

Education to support certification: We believe that the best way to ensure success is to combine meaningful industry experience with thorough academic study. To that end, CIMP exams are aligned with our courses, developed and taught by top industry educators and professionals.

Designed with busy, working professionals in mind: No time-consuming or costly travel is required to complete coursework or to take your CIMP examinations. All courses and exams are available online. All that's required of candidates is an internet connection and the desire to demonstrate mastery of data integration topics and achieve success.

How Do I Enroll?

The most convenient and cost-efficient method to enroll in the CIMP program is with one of our Education Packages. Each package includes all courses and exams necessary to earn CIMP or CIMP Ex. Alternatively you can enroll in courses one at a time.

ENTERPRISE SOLUTIONS

Today more than ever companies are watching expenses and looking for ways to streamline processes, make training convenient, and create a consistent, scalable learning environment.

eLearningCurve Enterprise is a flexible, convenient, and cost-effective way to train your employees and ensure that all team members have access to information management training they need when they need it. Whether your team or department work in the same office, or are on the other side of the world from each other, you can train them on time and on budget with eLearningCurve Enterprise.

Why eLearningCurve Enterprise?

- Comprehensive educational solution from a single provider
- Employees can take the courses they need when they need them
- Ensure all team members are trained to the same high standard
- Train employees no matter what their geographic location
- Employ a fully scalable education solution
- Minimize disruption to the business
- Maximize your employee training ROI
- Achieve 100% information comprehension
- Get "live" time with our instructors
- Stretch your training budget
- Get solutions for your specific needs

When you become an Enterprise customer:

We'll work with you to develop educational programs for different roles, positions, teams, departments, and manage and track enrollment of all students in online classes and CIMP exams. We'll rack and report educational progress for each student and work with you to meet any specific educational needs including:

- Organize question and answer meetings (via Webinar) with course instructors for groups of students who complete online courses
- Organize onsite sessions when appropriate, often for senior management.
- Prioritize new course development, or customize existing courses, per customer needs
- Create custom instances of our Learning Management System to reflect customer branding
- Mount our online courses on the customer's Learning Management System

ELEARNINGCURVE ENTERPRISE BENEFITS

PARTNERSHIP: *Comprehensive educational solution from a single provider.* We'll be your educational "partner-for-life" providing employees with continuous information management education they need over the course of their careers.

FLEXIBILITY: *Employees can take the courses they need when they need them.* Our flexible program allows employees to take the courses they need when they need them to best suit their role, projects, backgrounds or interests.

CONSISTENCY: *Ensure all team members are trained to the same high standard.* Train your existing team, and set up courses for new hires and transfers. Consider CIMP exams to verify that your employees utilize the same methodology, techniques, and terminology.

SCALABLITY: Select an Education Partner who truly understands scalability. Roll out to a few employees, or your entire organization. Our solution can quickly and effortlessly accommodate groups of all sizes, even if they are geographically dispersed.

BREADTH: Acquire comprehensive education and certification. We offer a full information management education. We have you covered with a comprehensive set of courses, exams, and certifications designed to impart knowledge, test understanding, and validate learning.

LOCATION: *Train employees no matter what their geographic location.* Overcome geographical barriers to training. You can train your entire team whether they are in the same office, or on the opposite sides of the world. Everyone can access our online courses from any place at any time.

LOGISTICS: *Minimize disruption to the business.* Our online format allows employees to study from their office or home, allocate full training days, or study an hour a day during lunch breaks.

ROI: *Maximize your employee training ROI.* No need to worry about paying for flights, hotels and other travel expenses. 100% of what you spend goes towards learning, thus achieving top quality education at a fraction of the cost of in-person training.

RESULTS: *Achieve 100% information comprehension.* Learn from top industry experts in information management topics. Study at your own pace, listen to the material many times, and test your knowledge through CIMP certification exams.

SAVINGS: *Stretch your training budget.* We offer various pricing options including volume discounts, pay-as-you-go model with increasing discounts, and other alternatives.

"LIVE" INTERACTION: *Spend time with our instructors.* Organize question and answer meetings (via Webinar) with course instructors for groups of students who complete online courses.

INFORMATION MANAGEMENT 101 *Mini-classes.* As a benefit to our enterprise customers we offer a certain number of complimentary licenses for our 101 mini-classes.

CUSTOME COURSEWARE: *Get solutions for your specific needs.* Tell us which courses your organization needs the most. We'll work with top instructors in the industry to meet your needs in the most expedient manner.

Data Integration Fundamentals & Best Practices

Instructor: Dave Wells
Duration: 5 hours

Integrated data is at the heart of many business and technical disciplines today. Data warehousing, operational data integration, and master data management focus on integration as a key part of managing data as an asset. Business intelligence, performance management, and business analytics depend on integrated data to meet business requirements for management and decision-making information. Legacy system replacement, ERP implementation, and application integration all have integrated data dependencies. Integration is important, but it is challenging to understand data sources, select and apply integration techniques, and design and deliver integrated databases.

This online training course discusses architectures, requirements, methods, roles and activities of data integration that can be applied to achieve successful data integration projects for a variety of applications and circumstances.

You will learn:

- Fundamental concepts, principles, and terminology of data integration
- Common methods of data integration with attention to techniques, timing, and integration process automation
- How to perform the essential steps of data integration including requirements definition, data capture, data transformation, and data delivery
- Data integration techniques and technologies including ETL, ELT, virtualization, and federation
- Techniques for source-to-target mapping and data transformation
- Roles, purpose, and variations of data integration architecture including architectural constructs for data warehousing, master data management, and operational data integration
- Business and technical roles, responsibilities, knowledge, and skills that are central to data integration projects and processes

This course is geared towards:

- Data integration architects, designers, and developers
- Business intelligence practitioners, project managers, and architects
- Data warehousing practitioners, project managers, and architects
- MDM practitioners, project managers, and architects
- ERP implementers and system integrators
- Business subject experts and data subject experts with roles in data integration projects and processes

Course Outline

About the Course (5min)

Data Integration Concepts (44 min)

- Data Integration Defined
- Data Integration Dependent Programs
- Data Integration Projects

Data Integration Methods (49 min)

- Data Integration Techniques
- Data Integration Frequency
- Data Integration Systems
- Data Integration Challenges
- Data Integration Activities

Understanding Data (61 min)

- Identifying Data Sources
- Profiling Data
- Qualifying Data Sources
- Documenting Data Sources

Integrating Data (78 min)

- Integration Requirements
- Data Capture
- Data Transformation
- Data Delivery

Data Integration Architecture (37 min)

- Architecture Concepts
- Data Warehousing Architecture
- MDM Architecture
- Operational Data Integration Architecture

Roles & Responsibilities (28 min)

- Knowledge and Skills
- Understanding the Data
- Getting the Data
- Transforming the Data
- Delivering the Data
- Using the Data

Data Warehousing Fundamentals

Instructors: Mark Peco
Duration: 4.5 hours

Due to rapid growth of non-traditional data sources, availability of new technologies and growing expectations of managers to compete on analytics, the traditional data warehouse is re-defined and presented within a broader modern context. A corporate data ecosystem is evolving and presents new opportunities for creating business capabilities that were not previously possible. Amidst these changes, the data warehouse continues to play foundational and integral roles within the expanding data landscape.

This course re-defines the scope of the "modern" data warehouse. The need for planning and the role of architecture are described and clarified, followed by a discussion about the challenges related to gathering useful information requirements. This is followed by a discussion of design approaches, development, testing and quality management techniques.

The true value delivered to the organization by a data warehouse depends on operational and service activities that leverage the data components previously implemented. It is through this combination of data and technology assets with managed operations and services that desired analytical and business capabilities are created. The design and Implementation approaches of operations and services are provided to highlight this key requirement. The course material presents a full life cycle of the data warehouse including business context, scope, requirements, design, implementation and operations.

You will learn:

- The components that define a data warehouse platform
- Trends that are impacting the modern data warehouse
- To position the data warehouse platform in the big data era
- Architectural options and considerations
- Development options and approaches
- The requirements gathering process
- Necessary design activities
- How operations and service processes enable business capabilities

This course is geared towards:

- Data warehousing program and project managers
- Data warehouse architects
- Data scientists and analytics professionals
- Big Data practitioners
- Data warehouse designers and developers

Course Outline

About the Course (8 min)

Introductory Concepts (71 min)

- Overview
- Data and Information
- The Modern Data Landscape
- Generating Information
- The Need for Metadata
- Defining the Data Warehouse
- Implementation Approaches

Planning and Architecture (72 min)

- Overview
- Implementation Planning
- Architecture Overview
- Requirements Analysis
- Information Requirements

Design and Development (73 min)

- Overview
- Design Activities
- Design Decisions
- Design Example
- Development

Operations and Service Delivery (39 min)

- Overview
- Services
- Categories of Services
- Managing a Service Catalog
- Managing Performance

Data Virtualization

Instructor: Dave Wells
Duration: 3 hours

The work of data integration has become increasingly complex in recent years. Business needs for real-time and low latency data, expanded uses of unstructured data, and accelerated interest in big data analytics are but a few of the trends that change the data integration landscape. Extract-transform-load (ETL) processing was sufficient for the once relatively simple task of combining data from multiple transactional databases was to build a data warehouse, operational data store, or master data hub. Today's data integration challenges go well beyond the capabilities of ETL technologies with needs to integrate enterprise data with external data, Web data, clickstream data, end-user data, big data, cloud data, and more. To meet these new requirements, data integrators need more tools in the integration toolbox. Data virtualization doesn't replace ETL; it complements ETL and offers new tools to meet new integration needs.

Data virtualization is a core component of next-generation data integration architectures, techniques, and technology. This online training course will introduce you to the concepts, techniques, and capabilities of data virtualization. It will prepare you to expand your data integration capabilities, deliver business-speed information, and make the most of recent advances in data integration technology.

You will learn:

- Data virtualization definitions, concepts, and terminology
- Business case and technical rationale for data virtualization
- Foundational principles of virtualization – abstraction, views, and services
- How to extend the data warehouse with virtualization
- How virtualization is applied for unstructured data, big data, and cloud data challenges
- How to mix and match virtualization with ETL technology to optimize data integration architectures and processes

This course is geared towards:

- BI, MDM, and data warehousing program and project managers
- Data integration architects, designers, and developers
- Data and technology architects

Course Outline

About the Course (6 min)

Data Virtualization Concepts and Principles (29 min)

- Overview
- Data Virtualization Basics
- Why Data Virtualization
- The Data Virtualization Foundation
- Review

Data Integration Architecture (19 min)

- Overview
- Integration Architecture Concepts
- Reference Architectures
- Integration Architecture Examples
- Review

Data Virtualization in Integration Architecture (49 min)

- Overview
- Virtualization in Data Integration Projects
- Data Virtualization Use Cases
- Data Warehousing Use Cases
- Data Federation Use Cases
- MDM and EIM Use Cases
- More Data Virtualization Applications
- Practical Data Virtualization
- Review

Data Virtualization Platforms (20 min)

- Overview
- Platform Requirements
- Platform Capabilities
- Platform Variations
- Some Platform Vendors
- Review
-

Implementing Data Virtualization (16 min)

- Overview
- Analysis
- Design and Modeling
- Development
- Deployment and Operation
- Review

Getting Started with Data Virtualization (28 min)

- Overview
- Skills, Competencies, and Human Factors
- Goal and Expectations
- Best Practices
- Case Studies
- Review

Ensuring Data Quality in Data Integration

Instructor: Arkady Maydanchik
Duration: 5 hours

Corporate data universe consists of numerous databases connected by countless real-time and batch data interfaces. The data continuously move about and change. The databases are endlessly redesigned and upgraded, as are the programs responsible for the data integration. The typical result of these dynamics is that information systems get better, while data quality deteriorates. Without a comprehensive data quality monitoring program bad data spread like viruses.

This online training course discusses various practices that can be put in place to mitigate the problem and maintain high data quality through data integration.

You will learn to:

- The data quality challenges that are inherent in data integration
- The critical role of data quality monitoring in data integration
- Specific techniques to monitor and manage quality for batch data integration
- Use of Statistical Process Control (SPC) methods in monitoring data quality
- The impacts of change on data quality and techniques to address those impacts
- How an enterprise integration hub can be applied to managing data quality

This course is geared towards:

- Data integration practitioners
- Data quality practitioners
- Data warehousing practitioners
- MDM practitioners
- Others in the trenches involved in design, development, and maintenance of data integration systems

Course Outline

About the Course (8 min)

Introduction (60 min)

- Data Integration Basics
- Data Quality in Data Federation
- Data Quality in Data Consolidation
- Data Quality in Real-Time Interfaces
- Data Quality in Batch Interfaces
- Beyond Monitoring

Data Consolidation and Cleansing (60 min)

- Data Quality in Data Conversion
- Data Cleansing
- Data Quality in Data Consolidation

Error Monitors in Batch Interfaces (77 min)

- Monitoring Techniques
- Batch Integrity Rules
- Master Data Integrity Rules
- Dynamic Integrity Rules
- Monitor Management
- Error Correction
- Root Cause Analysis
- Statistical Process Control

Change Monitors in Batch Interfaces (95 min)

- Examples
- Change Monitor Defined
- Which Metrics to Monitor
- Implementing the Monitors
- Heuristic Monitors
- Basic Statistical Monitor
- Handling Non-Stationary Data

Big Data Fundamentals

Instructor: William McKnight & Jake Dolezal
Duration: 3.5 hours

Big data has gone main stream. It reaches well beyond the initial group of Silicon Valley "new economy" tech companies and the new media companies that helped launch the industry. The big data adoption landscape has expanded to include automakers, big finance, big insurance companies, telecommunications, healthcare companies and big retailers. Big data is past the hype phase and adoption is accelerating, but success is not a given and challenges remain.

This informative technical general session is full of the "need to know" for anyone involved in an enterprise data landscape. Learn from experienced enterprise information strategists with real project experience about the path that big data is on, the obstacles along the path, and how to confidently join the big data revolution. Learn the players in the technology landscape and the ideal workloads for big data in enterprises. Learn where big data adds value to an existing enterprise information strategy and how to get the projects started and dropping the "not in production" label.

This 3.5-hour online course addresses the technical community as well as the user community, providing guidance on how to penetrate and benefit the enterprise. This practical session will help you make the most of big data and make the best choices to ensure information remains an unparalleled corporate asset.

You will learn:

- A workable definition of big data so you know it when you see it
- Drivers for big data
- Big data in the enterprise
- The Hadoop framework for analytical big data
- NoSQL and operational big data
- An overall information architecture with big data

This course is geared towards:

- Business and Data Analysts
- BI Architects and BI Developers
- Data Architects
- Data Integrators
- Analytics Developers and Consumers
- Anyone who needs to understand the business and technical implications of Big Data

Course Outline

About the Course (8 min)

Big Data Definition (34 mins)

- Big Data Introduction
- Big Data Technology
- Enablers for Big Data

Big Data Drivers (28 mins)

- Value Density of Data
- Before Data was Big…
- Once Big Data Grew, Value was Realized
- Data is too Valuable to Discard
- Data is too Valuable to Ignore
- Focus Before Big Data
- Focus After Big Data
- Performance/Workload Optimization
- Cost of Storage
- Other Cost Drivers
- Analytic Need
- Implication for IT Skills

Big Data in the Enterprise (21 mins)

- The Great Database Thaw
- Data Access in the Modern Enterprise
- Marz's Lambda Architecture
- Row vs. Columnar Stores
- In-Memory
- Big Data & Analytics
- Leveraging Hadoop for Analytics

Hadoop Ecosystem (40 mins)

- Hadoop Overview
- Hadoop Distributions
- Hadoop Framework

NoSQL (31 mins)

- NoSQL "Schemaless" Data Modeling
- NoSQL Heartburn
- Key-Value Stores
- Document Oriented Database
- Graph Oriented Database
- Stream Processing Engines
- NewSQL

Enterprise Architecture with Big Data (45 mins)

- Modern Components of Information Architecture
- ETL with Big Data Systems
- Analytic Patterns with Hadoop
- Where Do We go from Here?

MDM Fundamentals: Architecture and Implementation

Instructor: William McKnight
Duration: 4 hours and 30 minutes

Proliferation of heterogeneous systems creates a pressing need for data sharing and data consistency. When many different systems collect data about master entities – customers, products, suppliers, employees, accounts, etc. – you can be certain that you'll find inconsistencies, conflicts, and confusion. At best, conflict and confusion leads to waste and inefficiency in business process. More severe consequences include damaged credibility and reputation when errors and inconsistencies are visible to customers, suppliers, and employees. Today's complex business and information systems must synchronize master data. That is the role and purpose of Master Data Management (MDM) systems.

MDM is not a casual endeavor. It is a complex data management challenge that requires a formal and well-managed program. The unique challenges of an MDM program are often not apparent even to seasoned data management professionals. The complexities of managing identities and resolving conflicts among disparate data sources make MDM an ambitious undertaking that must address business, architectural, people, process, project, and technology dimensions to succeed. This course provides a comprehensive look at the elements of an MDM program and the key success factors for MDM.

You will learn

- The what and why of Master Data Management (MDM)
- A variety of architectural approaches to MDM and how to determine which is the best fit for your MDM program
- The human dimension of MDM including roles and responsibilities of sponsors, managers, analysts, architects, designers, and developers
- The state of MDM technologies along with techniques and guidelines for tool selection
- The process dimension of MDM including impacts upon business processes and information management processes
- The project perspective of MDM including organizing and executing the activities of planning, requirements analysis, design, development, testing, data migration, and implementation.

This course is geared towards

- MDM Program and Project Managers
- MDM Analysts, Designers, and Developers
- Business Data Owners, Data Stewards, and Data Consumers
- Data Architects
- Information Systems Project Managers
- Data Integration Program and Project Managers
- Data Stewards, Data Governance Professionals, and Data Quality Practitioners

Course Outline

About the Course (7 min)

Introduction (54 min)

- MDM Overview
- MDM Justification and Outcomes
- Master Data Management for Customers

MDM Architecture (90 min)

- Architecture Approaches
- Conforming Dimensions for the Enterprise
- Data Quality
- Syndicated Data
- Additional Considerations

MDM Tool Selection (48 min)

- Process Considerations
- Proof of Concept and Final Selection

MDM Project Execution (77 min)

- MDM Project Management
- MDM Project Roles and Responsibilities
- Organizing and Planning for MDM Success
- Case Study

Data Profiling

Instructor: Arkady Maydanchik
Duration: 5 hours

Data profiling is the process of analyzing actual data and understanding its true structure and meaning. It is one of the most common and important activities in information management. Data profiling is the first critical step in many major IT initiatives, including implementing a data warehouse, building an MDM hub, populating metadata repository, as well as operational data migration and integration. It is also the key ingredient to successful data quality management.

While proliferation of commercial tools made data profiling accessible for most information management professionals, successful profiling projects remain elusive. This is largely because the tools allow gathering large volumes of information about data, but offer limited means and guidelines for analysis of that information.

In this online training course you will learn all practical skills necessary to succeed in a data profiling initiative.

You will learn:

- The what, why, when, and how of data profiling
- Various data profiling techniques, from simple column profiling to advanced profiling methods for time-dependent and state-dependent data
- How to efficiently gather data profiles
- How to analyze the data profiling information and ask the right questions about your data
- How to organize data profiling results
- How to perform dynamic data profiling and identify changes in data structure and meaning

This course is geared towards:

- Data quality practitioners
- MDM practitioners
- Metadata management practitioners
- IT and business analysts involved in data management
- Those responsible for implementation and maintenance of various data management systems

Course Outline

About the Course (5 min)

Quality Basics (30 min)

- Quality Basics
- Quality Defined
- Quality and Defects
- Quality Economics
-

Quality Management (93 min)

- Quality Management Practices
- Quality Management Gurus
- Quality Management Methodologies
- Related Disciplines
- Measurement and Standards

Data Quality Basics (45 min)

- Data Quality Defined
- Data and Purpose
- Dimensions of Data Quality

Data Quality Management (74 min)

- Data Quality Processes
- Data Quality Techniques
- Data Quality Tools and Technology
- Data Quality Projects
- Building-In Data Quality
- Data Quality Organizations

Data Integration Techniques for Designing ODS

Instructors: Angelo Bobak
Duration: 3 hours

In today's modern business environment, corporate entities are constantly merging or splitting, internal divisions are sold to different companies, and new business lines are created in order to meet the challenges of difficult economic times. Business data integration is a complex problem that must be solved when organizations change or enhance their internal structures. New IT departments must be merged with old ones, and transactional, operational, and master data must be integrated in order to be managed efficiently, if the business is expected to grow and be profitable.

The goal of this course is to present a simple yet thorough process that describes the challenges of business data integration and the solutions to these challenges. It will show you how the application of a technique called "schema integration" addresses these challenges.

Schema integration is both a theory and process that was pioneered by experts in the field of data management. We will discuss the techniques of two of these pioneers, M. Tamer Ozsu and Patrick Valduriez in the design of an Operational Data Store (ODS) for a small business..

You will learn:

- The underlying architecture of the Operational Data Store (ODS)
- The different types of ODS Architectures
- The theory behind schema integration
- The schema integration process
- Identifying and resolving data conflicts when integrating data
- The importance of master data and data quality in schema integration

This course is geared towards:

- The Logical and Physical Data Modeler
- The Data Architect
- The Database Administrator
- Project Managers
- Data Warehouse Architects
- Anyone wishing to enter the field of database design and ODS implementation

Course Outline

About the Course (4 min)

Introduction to Operational Data Stores (45 min)

- Overview
- What is an ODS
- Master Data and The ODS
- Data Quality and the ODS
- Loading the ODS
- ODS and Data Warehouse Architectures

Theory of Scheme Integration (58 min)

- Overview
- Data Integration Pioneers
- Schema Integration Types
- Schema Integration Process
- Resolving Data Conflicts
- Profiling Data
- Defining the ETL Specifications to Merge Data
- Tracking Data Lineage
- Schema Integration ETL Tools

Maintenance (30 min)

- Overview
- Adding New Sources
- Adding New Destinations
- Modifying Existing Sources
- Modifying Existing Destinations
- Retiring Old Sources
- Retiring Existing Destinations
- Managing Security and Access
- Monitoring and Managing Storage Capacity
- Monitoring Performance
- Physical Design Techniques to Increase Performance
- Key Project Roles and Responsibilities

Case Study (38 min)

- Overview
- Databases to Integrate
- Data Dictionaries
- Tools You Will Need
- Performing the Integration
- Concluding Remarks

Data Parsing, Matching and De-duplication

Instructors: Kathy Hunter, William McKnight, Henrik Sørensen
Duration: 4 hours, 20 minutes

Data parsing, standardization, matching, and de-duplication are the cornerstones of successful Master Data Management (MDM). They are also critical parts of successful data quality programs, and are key steps in building data warehouses as well as any data integration and consolidation initiatives. You could say that today few organizations can function effectively without implementing data parsing and matching processes often in many data domains.

This need is further magnified if your company has gone global and plans to create databases that combine name- and address-related data from all corners of the world. Managing global information effectively takes specialist knowledge and the ability to show consideration for the differences that exist throughout the world. Worldwide there are more than 10,000 languages, 130 address formats, 36 personal and hundreds of business name formats. All of these variables are further complicated by the need to respect national and regional cultures. Failure to consider formats, styles, and cultures has huge impact on quality of data and quality of business relationships.

This online training course is aimed at data quality and master data management (MDM) professionals as well as those responsible to work with global information. The field is broad and the details are many. The purpose of this course is to provide a broad and in-depth review of data parsing, standardization, matching, and de-duplication techniques, as well as extensive overview of specific problems and solutions when dealing with global data.

You will learn:

- Data parsing, standardization, matching, and de-duplication techniques
- How to find and use external reference data
- How data parsing and matching contribute to improving data quality, MDM, and data warehousing
- Which data domains, entities and data elements may benefit from data parsing and matching
- Challenges of global data and ways to overcome these challenges

This course is geared towards:

- Master data management professionals
- Data quality professionals
- Information architects
- Developers of data warehousing systems
- Business professionals who work with global data

Course Outline

About the Course (8min)

Introduction (17 min)

Implementation Fundamentals (70 min)

- Parsing and Standardization
- Introduction to Data Matching
- Data Matching Techniques
- Data Matching Destinations
- Evaluating Data Matching Tools

External Reference Data (45 min)

- External Data Sources
- Syndicated Customer Data
- Syndicated Product Data
- Using the Web

Challenges of Global Data (58 min)

- Introduction to Global Information
- Global Data: What You Need to Know
- Variations by Country and Region
- Cultural and Legal Impacts
- Characters and Diacritics

Overcoming the Challenges of Global Data (59 min)

- Data Profiling
- Consistent Data Structures
- Preparing Global Data for Effective Use

Metadata Management Fundamentals

Instructor: Dave Wells and Arkady Maydanchik
Duration: 4 hours

Deriving value from data depends extensively on understanding the data and sharing knowledge among everyone who works with data. Sharing data knowledge is the core purpose of metadata. Just as you need financial data to manage financial resources, you need metadata to manage data resources. In today's data-driven world, the importance of managing data is certainly on par with that of managing finances.

This online training course is designed to provide the foundational metadata knowledge needed by anyone who has data management roles and responsibilities. It covers metadata basics such as the types and purposes of metadata, and explores core metadata disciplines of data modeling, data profiling, and data cataloging. Metadata roles in data governance, stewardship, security, quality, and analysis are explained.

You will learn:

- The scope and complexities of metadata management
- The roles of data models as metadata and the basics of data modeling
- The role of data profiling in metadata management and the basics of data profiling methods
- The roles of data catalogs in metadata management and the fundamentals of data curation and data cataloging
- Metadata dependencies of business processes, IT projects, data governance, data quality, business intelligence, self-service data, business analytics, and data science

This course is geared towards:

- Anyone with data management roles and responsibilities
- Data stewards and data governance practitioners and participants
- Data curators and data catalog administrators
- Data and database analysts and designers
- Data quality professionals and practitioners
- Aspiring data modelers who need to start with the basics
- Anyone with a role in information management who needs to understand data or help others to understand data

Course Outline

About the Course (8 min)

Understanding Data (15 min)

- Views of Data
- Projects Flow
- Describing the Data Meaning
- Describing the Data Constraints
- Describing the Data Relationships
- Describing the Data

Metadata Management (57 min)

- Metadata
- Metadata Management Processes
- Using Metadata
- Metadata Tools and Technologies

Data Modeling (24 min)

- Data Modeling Defined
- The Data Modeling Process
- Supplemental Models & Additional E-R Concepts
- Dimensional Data Modeling

Data Profiling (47 min)

- What is Data Profiling?
- Myth and Reality of Data Profiling
- Profiling Techniques
- Profiling Challenges
- Role of Profiling
- People and Technology

Data Curation and Cataloging (31 min)

- Data Curation
- Data Cataloging
- Metadata and the Catalog

Metadata Management for BI and Data Science (49 min)

- The Metadata Muddle
- Data Science and Metadata
- Data Provenance and Data Lineage
- Ontology and Taxonomy

DW and BI Data Modeling

Instructor: Rick Sherman
Duration: 4.25 hours

A well designed data model is the cornerstone to building business intelligence and data warehouse applications that provide significant business value.

Effective data modeling results in transforming data into an enterprise information asset that is consistent, comprehensive and current. Data is transformed from operational or source systems into a data warehouse and often data marts or OLAP cubes for analysis. This course provides the fundamental techniques to designing the data warehouse, data marts or cubes that enable business intelligence reporting and analytics.

This online training course discusses the two logical data modeling approaches of Entity-Relationship (ER) and dimensional modeling. ER modeling is used to establish the baseline data model while dimensional modeling is the cornerstone to Business Intelligence (BI) and Data Warehousing (DW) applications. These modeling techniques have expanded and matured as best practices have emerged from years of experience in data modeling in enterprises of all sizes and industries. These techniques improve the business value of the data, enhance project productivity and reduce the time to develop applications.

You will learn:

- The basics of Entity-Relationship (ER) and dimensional modeling
- The benefits and applicability of Dimensional Data Modeling
- How to create Dimensional Data Models for BI and DW applications
- How to learn more about Data Modeling

This course is geared towards:

- Beginning Data Modelers
- Business Analysts and Architects
- Database Administrators and Analysts
- Information Technology Managers, Project Managers
- Application Development Project Team Members
- People involved in design and maintenance of Data Warehousing and Business Intelligence applications
- People involved in data quality or data governance processes

Course Outline

About the Course (6 min)

Introduction to Data Modeling (75 min)

- Data Modeling Overview
- Entity-Relationship Modeling Overview
- Normalization

Dimensional Modeling Basics (88 min)

- What is Dimensional Modeling?
- Facts
- Dimensions
- Schemas
- Entity-Relationship vs. Dimensional Modeling
- Purpose of Dimensional Modeling
- Fact Tables
- Dimensional Modeling Vocabulary

Advanced Dimensional Modeling (84 min)

- Hierarchies
- Slowly Changing Dimensions
- Rapidly Changing Dimensions
- Casual Dimensions
- Multi-Valued Dimensions
- Snow flaking
- Junk Dimensions
- Value Brand Reporting
- Heterogeneous Products
- How Swappable Dimensions
- Too Few or Too Many Dimensions
- Benefits of Dimensional Modeling

Streaming Data: Concepts, Applications, and Technologies

Instructors: Dave Wells & Kevin Petrie
Duration: 3 hours

The analytics opportunities with IoT and application data streams are abundant, but the value of streaming technology is not limited to native data streams. In today's fast paced business world, the need for fast data is pervasive and tacit acceptance of high-latency data is rapidly diminishing. Streaming as an alternative to batch ETL is a practical way to meet the demand for fast data.

Change Data Capture (CDC) is a category of technology that captures data about changes made to a database – inserts, updates, and deletes – and makes that data available to downstream processing such as data pipelines that flow to data warehouses and data lakes. CDC can be combined with streaming to accelerate data flow and reduce data latency.

You'll need to know the actions and responsibilities of data producers and of data consumers, as well as the capabilities for cluster management, data connections, and APIs. Integrating Kafka or other streaming technologies into your data ecosystem is an important consideration.

You will learn:

- The business and technical drivers for streaming data adoption
- Data pipeline processing patterns and the advanced patterns that are possible with streaming
- Use case patterns and a variety of use cases for streaming data
- Five kinds of Change Data Capture (CDC) and the strengths and weaknesses of each
- The concept and applications of streaming first architecture
- Kafka architecture and essential components
- Kafka data and process flow
- The roles and functions of Kafka broker, data producers, and data consumers
- Cluster management, data connections, and APIs with Kafka
- Integrating streaming into the data ecosystem

This course is geared towards:

1. Data and analytics leaders and managers
- Data and analytics architects
- Data scientists
- Data engineers
- Data governance professionals who need to understand the opportunities and implications of streaming data
- Anyone with a desire to know how streaming is changing the data management landscape

Course Outline

About the Course (5 min)

The Need for Speed (21 min)

- Batch Vs. Real Time
- The Speed of Business
- The Speed of Data Parts 1-3
- Fast Data Drives the Consumer World
- Fast Data Drives the Business World
- Fast Data Drives the Analytic World
- The Business Case for Fast Data

Moving Data through Pipelines (61 min)

- Data Pipeline Processing
- Data Stream Processing
- IoT and Edge Computing

Change Data Capture (10 min)

- Data Management – A New Streaming Use Case
- What is CDC?

Streaming First Architecture for Data Management (17 min)

- Modern Enterprise Dara Requirements
- Essential Characteristics of Streaming Technology
- Stream Processing Technologies
- Kafka Stream Processing Basics
- Replacing ETL with CDC and Streaming
- Streaming First Architecture
- Moving to Streaming First Architecture

Data Streaming with Apache Kafka (11 min)

- Introduction to Apache Kafka
- Kafka Predecessors

Kafka Architecture (19 min)

- Apache Kafka Architecture
- Essential Components Part 1-2
- Data and Process Flow Part 1-2
- Kafka Record – Key Components
- Kafka Broker Overview
- Cluster Management with Apache

Data Stream Ecosystems and Use Cases (40 min)

- Data Streaming Use Cases
- Streaming Integration with Data Ecosystems
- Kafka Alternatives from Cloud Service Providers
- Amazon Kinesis
- Azure Event Hub
- Google Cloud PUB/SUB

Full outline:
https://ecm.elearningcurve.com/Online_Data_Science

Introduction to Graph Databases

Instructor: John Singer
Duration: 4 hours

Entity Relationship modeling and relational databases have dominated the IT scene since the '80s, becoming the de facto standard approach for data persistence. However, the ubiquitous relational database has waned with the advent of NoSQL and big data technologies. Today's data architect must master a new database technology – graph database – that has emerged with a solid set of use cases based on mathematical graph theory and graph algorithms.

This online course will provide an overview of property graph database technology and teach the student how to translate business requirements to a property graph database design that can be implemented on any modern property graph database.

You will learn:

- The fundamental concepts and practices of Business Intelligence
- The fundamental concepts and practices of Business Analytics
- Concepts and principles of Predictive Analytics
- Concepts and principles of Data Mining
- Concepts and principles of Data Science
- The roles of data analytics in a data-driven organization

This course is geared towards:

- Data Architects that need to understand how graph database fits into the overall persistence architecture
- DBA's and Data modelers expanding into graph database
- Data Science and Data analysts interacting with graph databases
- Big Data Managers and decision makers

Course Outline

Module 0. About the Course (7 min)

Module 1. Introduction to the Graph Data Model (28 min)

- Graph Theory
- Anatomy of a Node
- Anatomy of a Relationship
- Properties and Paths

Module 2. The Graph Database Eco-System (65 min)

- Graph Database
- Graph Analytics
- Semantic Web Graph Database

Module 3. Agile Graph Data Modeling (39 min)

- Graph Data Modeling Overview
- Agile Graph Data Modeling
- Graph Data Modeling Process
- Working with Users

Module 4. Entity Modeling (42 min)

- Representing Things
- Describing Things
- Categorizing Things Part 1 & 2
- Entity Definition Best/Worst Practices

Module 5. Relationship Modeling (38 min)

- Representing Connections
- Naming Relationships
- Relationship Direction
- Describing Relationships
- Relationship Best/Worst Practices

Module 6. Complex Object Modeling (40 min)

- Complex Object Modeling
- Resolving Hypergraphs Part 1 & 2
- Complex Objects
- Data Structures-Linked Lists and Trees
- Managing Slowly Changing Dimensions

OUR INSTRUCTORS

Angelo Bobak

Angelo Bobak is a seasoned data architecture professional and published author with over 20 years' experience in Business Intelligence, Data Architecture, Data Modeling, Master Data Management, and Data Quality. Currently he is working at ATOS Origin NA as a Director/Senior Data Architect in the areas of Global Master Data Management, Data Integration and Data Quality. Past experience includes positions as an IT consultant, manager and data architect with companies such as Siemens, Praxair, Avaya, and Pepsi.

Jake Dolezal

Jake Dolezal has over 16-years' experience in the Information Management field with expertise in business intelligence, analytics, data warehousing, statistics, data modeling and integration, data visualization, master data management, and data quality. Jake has experience across a broad array of industries, including: healthcare, education, government, manufacturing, engineering, hospitality and gaming. He is also the author of the book I Survived Stats: A Less Frustrating Approach to Help Students Survive and Pass a Statistics Course (forthcoming).

Kathy Hunter

Kathy always says she has data in her blood. Joining Harte-Hanks in 2002, she built an information management practice and, with her highly skilled team, was responsible for instituting their highly successful Global Data Management solution set. From information quality and data governance through to providing global data solutions and guidance she attained a reputation for expert knowledge and successful delivery in global information management to her clients

William McKnight

William is president of McKnight Consulting Group, which includes service lines of Master Data Management, IT assessment, Big Data, Columnar Databases, Data Warehousing, and Business Intelligence. He functions as Strategist, Lead Enterprise Information Architect, and Program Manager for sites worldwide.

Arkady Maydanchik

For more than 20 years, Arkady Maydanchik has been a recognized leader and innovator in the fields of data quality and information integration. As a practitioner, author and educator he has been involved in some of the most challenging projects industry has seen. These projects were often the result of major corporate mergers and the need to consolidate and integrate databases of enormous variety and complexity. Arkady's client list includes such household names as Dun & Bradstreet, Hewitt Associates, Kimberly Clark, Raytheon, Sprint, Verizon, and Xerox.

Mark Peco

Mark Peco is an experienced consultant, educator, practitioner and manager in the fields of Business Intelligence and Process Improvement. He provides vision and leadership to projects operating and creating solutions at the intersection of Business and Technology.

Henrik Sørensen

Henrik Liliendahl Sørensen has over 30 years of experience in working with Master Data Management and Data Quality and is a charter member of the International Association of Information and Data Quality. Currently Henrik works with Master Data Management at Tata Consulting Services and as Practice Manager at Omikron Data Quality besides writing on a well trafficked blog about data quality, master data management and the art of data matching. Henrik is the founder of the Data Matching and the Multi-Domain MDM groups on LinkedIn.

Rick Sherman

Rick Sherman will admit that he's been doing data warehousing since before it was even called data warehousing. He is the founder of Athena IT Solutions, a Boston area business intelligence and data warehousing consulting firm that provides solutions for customers of all sizes and industries. Rick also teaches data warehousing, data integration and business intelligence for a masters' degree program at Northeastern University's graduate school of engineering. He is a published author of more than a hundred articles and a speaker at industry events.

Dave Wells

Dave Wells is a consultant, teacher, and practitioner in the field of information management. He brings to every endeavor a unique and balanced perspective about the relationships of business and technology. His career of more than thirty-five years has encompassed both business and technical roles. Dave is a frequent contributor to trade publications and is a co-author of the book BI Strategy: How to Create and Document. He also speaks at a variety of industry events.

John Singer

John Singer has 4 decades of experience in a variety of data architecture–related roles. John's past accomplishments include implementing metadata management solutions, data modeling processes and governance, master data management solutions, and an ITIL-based CMDB combining architecture, business, and IT metadata in a comprehensive solution. John is currently focusing on property graph technologies and is the Founder and CEO of NodeEra Software, makers of the NodeEra family of Property Graph Data Modeling tools for Neo4j.

OUR CUSTOMERS

eLearningCurve has students in almost every country in the world, including many enterprise customers.

- Variety of customers from small project teams to large enterprises
- Enterprise customers typically Fortune 500 and Global 1000 companies
- All major industries are represented

WHAT OUR CUSTOMERS ARE SAYING...

The courses are well laid out, build on each other, and are rich in practical content and advice.

-- Steve Lutter, CIMP Data Quality, DM and Metadata, IM Foundations, Business Intelligence, Data Governance, MDM, United States

It is evident that a thorough and considerable effort has gone into the preparation of this program.

-- Alfredo Parga O'Sullivan, CIMP Ex Data Quality, Ireland

The ability to take the courses at my own pace and at a time suitable for me was of great help.

-- Geeta Jegamathi, CIMP Data Quality, India

CONTACT US

Director, Enterprise Solutions
Arkady Maydanchik
Arkadym@elearningcurve.com

Director, Marketing
Michelle Johnson
Mjohnson@elearningcurve.com

Director, Education
Dave Wells
David.Wells@elearningcurve.com

Director, Technology
Varya Belyaevskaya
Admin@elearningcurve.com

Customer Support
Support@elearningcurve.com

Phone: 1-630-242-1659
Mailing Address: 305 Midwest Club Parkway, Oak Brook, IL 60523

RESELLERS

DENMARK
Right Training
Contact:
Jan Sørensen
+45 31 45 04 99
jms@righttraining.dk

SOUTH AFRICA & SUB-SAHARAN AFRICA
Master Data
Contact:
Gary Allemann
+27 11 485 4856
gary@masterdata.co.za

DATA INTEGRATION COURSE PRICING

Packages

CIMP Data Integration Package	$1,895.00
CIMP Ex Data Integration Package	$2,895.00
CIMP Ex Data Integration + MDM Package	$3,395.00
CIMP All Course Access License	$4,995.00
CIMP All Course Access License w/ Exam Package	$5,990.00

Individual Courses

Data Integration Fundamentals and Best Practices	$500.00
Data Warehousing Fundamentals	$445.00
Data Virtualization	$285.00
Ensuring Data Quality in Data Integration	$525.00
Big Data Fundamentals	$350.00
MDM Fundamentals: Architecture and Implementation	$460.00
Data Profiling	$520.00
Data Integration Techniques for Designing an ODS	$295.00
Data Parsing, Matching, and De-Duplication	$435.00
Metadata Management Fundamentals	$385.00
BI & DW Data Modeling	$420.00
Streaming Data Concepts, Applications, Technologies	$305.00
Introduction to Graph Databases	$435.00

Exams

Exam for each course	$80.00

Enterprise Discounts

We offer discounts to Enterprise customers who purchase in bulk. Please contact us for more information.

About eLearningCurve

eLearningCurve offers comprehensive online education programs in various disciplines of information management. With eLearningCurve, you can take the courses you need when you need them from any place at any time. Study at your own pace, listen to the material many times, and test your knowledge through online exams to ensure maximum information comprehension and retention.

eLearningCurve also offers two robust certification programs: CIMP & CDS. Certified Information Management Professional (CIMP) builds upon education to certify knowledge and understanding of information management. Certified Data Steward (CDS) is a role-based certification designed for the fast growing data stewardship profession.

Finally, eLearningCurve's Enterprise Program is a flexible, scalable, cost-effective solution for teams and enterprises.