
Certificate in Master Data Migration

Master Data Implementation and Deployment

Attribute

Concept: A single data element that describes a characteristic of an entity. **Related Terms:** field, property, column. **Explanation:** In master data, an attribute holds the value for a specific aspect such as "Customer Name" or "Product Code." Attributes are the building blocks of a data model and are defined in the data dictionary. **Example:** The attribute "Email Address" stores the email contact for a customer record. **Practical Application:** During migration, attributes are mapped from source to target schemas, ensuring that each piece of information is transferred accurately. **Challenges:** Inconsistent naming conventions, missing attribute definitions, and varied data types across source systems can cause mapping errors and require extensive profiling.

Attribute Domain

Concept: The set of permissible values that an attribute can contain. **Related Terms:** value set, lookup table, validation rule. **Explanation:** Domains enforce data integrity by restricting inputs to defined ranges or lists, such as "Country Code" limited to ISO-3166 codes. Domains can be static (enumerated list) or dynamic (referencing another entity). **Example:** The "Status" attribute may have a domain of {Active, Inactive, Pending}. **Practical Application:** When loading master data, the migration engine validates each attribute against its domain to prevent invalid entries. **Challenges:** Divergent domain definitions between source and target systems often require domain harmonization, and legacy data may contain values outside the accepted domain.

Attribute Mapping

Concept: The process of linking source attributes to target attributes. **Related Terms:** field mapping, data transformation, mapping matrix. **Explanation:** Mapping defines how each source attribute is transferred, transformed, or combined to populate the target attribute. It includes rules for data type conversion, concatenation, or splitting. **Example:** Mapping "First_Name" and "Last_Name" from the legacy system to a single "Full_Name" attribute in the target system using a concatenation rule. **Practical Application:** Mapping documents serve as blueprints for ETL developers and are used to generate automated migration scripts. **Challenges:** Complex transformations, missing source attributes, and ambiguous business semantics often lead to iterative refinement of the mapping specification.

Business Rule

Concept: A condition or constraint that governs data behavior based on organizational policies. **Related Terms:** validation rule, logic, policy. **Explanation:** Business rules dictate how data should be processed, such as "A customer must have a valid tax identification number before becoming active." They are implemented during migration to enforce consistency. **Example:** A rule that prevents duplicate product SKUs by checking existing records before insertion. **Practical Application:** Rules are encoded in the migration engine to automatically cleanse and enrich data, reducing manual intervention. **Challenges:** Capturing implicit rules from legacy systems, reconciling conflicting rules across business units, and maintaining rule performance

during large-scale loads.

Canonical Data Model (CDM)

Concept: A unified, neutral data representation that facilitates interchange between heterogeneous systems. **Related Terms:** enterprise data model, standard model, reference architecture. **Explanation:** The CDM abstracts source and target structures into a common format, enabling consistent mapping and transformation logic. It reduces the number of direct source-to-target mappings by serving as an intermediary. **Example:** Defining a generic “Customer” entity with standard attributes (ID, Name, Address) that all systems map to. **Practical Application:** In multi-system migrations, the CDM acts as a hub, simplifying integration and supporting future data exchanges. **Challenges:** Designing a CDM that captures all necessary nuances without becoming overly complex, and keeping it aligned with evolving business requirements.

Change Data Capture (CDC)

Concept: A technique for identifying and capturing changes made to data in source systems. **Related Terms:** incremental load, log mining, delta extraction. **Explanation:** CDC enables the migration process to extract only new or modified records since the last extraction, minimizing disruption and reducing load times. **Methods include** database triggers, transaction log reading, or timestamp columns. **Example:** Capturing all customer record updates that occurred after the initial full migration to synchronize the target system. **Practical Application:** CDC is essential for “big-bang” migrations followed by a cut-over period, ensuring data consistency between source and target until go-live. **Challenges:** Ensuring reliable capture across heterogeneous databases, handling out-of-order events, and managing the increased complexity of incremental reconciliation.

Data Governance

Concept: The framework of policies, processes, and responsibilities that ensure data is managed as a strategic asset. **Related Terms:** stewardship, policy, compliance. **Explanation:** Governance establishes ownership, data quality standards, and lifecycle management, providing the control needed for successful master data implementation. It defines who can create, modify, or delete master records. **Example:** A governance charter that assigns the “Product Data Owner” role to oversee product master data quality. **Practical Application:** Governance structures guide the migration team on approval workflows, data validation checkpoints, and post-migration monitoring. **Challenges:** Aligning governance across multiple business units, gaining executive sponsorship, and balancing control with agility during rapid migration phases.

Data Integration

Concept: The process of combining data from disparate sources into a unified view. **Related Terms:** ETL, ELT, data federation. **Explanation:** Integration involves extracting data, transforming it to meet target standards, and loading it into the master data repository. It may include real-time streaming or batch processing. **Example:** Merging customer records from CRM, ERP, and legacy billing systems into a single master data hub. **Practical Application:** Integration pipelines are built using integration platforms or custom scripts to automate the migration workflow. **Challenges:** Handling schema mismatches, reconciling duplicate records, and ensuring performance when processing large volumes.

Data Migration

Concept: The systematic transfer of data from one environment to another. **Related Terms:** data conversion, load, cut-over. **Explanation:** Migration encompasses planning, extraction, transformation, loading, validation, and post-migration support. It is a core activity in master data implementation projects. **Example:** Moving all product master data from an on-premise ERP to a cloud-based MDM solution. **Practical Application:** Migration plans outline timelines, resource allocation, risk mitigation, and success criteria for each phase. **Challenges:** Data loss, downtime, incompatibility, and insufficient testing can jeopardize project goals.

Data Quality

Concept: The degree to which data meets the requirements for accuracy, completeness, consistency, and timeliness. **Related Terms:** profiling, cleansing, validation. **Explanation:** High data quality is essential for reliable master data; quality issues are identified through profiling and remediated via cleansing rules. **Example:** Detecting and correcting malformed phone numbers in the customer dataset. **Practical Application:** Quality dashboards monitor key metrics such as “% of records with missing mandatory fields” throughout the migration lifecycle. **Challenges:** Legacy systems often contain entrenched errors, and enforcing uniform quality standards across diverse sources can be resource-intensive.

Data Stewardship

Concept: The responsibility for managing and safeguarding data assets. **Related Terms:** ownership, custodianship, accountability. **Explanation:** Data stewards define data standards, approve changes, and resolve data issues. They act as the bridge between business users and technical teams. **Example:** A steward reviewing and approving new supplier records before they are loaded into the master repository. **Practical Application:** Stewardship workflows are embedded in migration tools to route data change requests for approval. **Challenges:** Securing sufficient time from subject-matter experts, clarifying authority boundaries, and maintaining stewardship continuity after project completion.

Data Validation

Concept: The process of verifying that data conforms to defined rules before it is accepted. **Related Terms:** checks, assertions, error handling. **Explanation:** Validation occurs at multiple stages—pre-extraction, during transformation, and post-load—to ensure integrity. It includes format checks, referential integrity, and business rule enforcement. **Example:** Ensuring that every “Country Code” matches a valid entry in the ISO list before loading. **Practical Application:** Automated validation scripts generate error reports that guide remediation activities. **Challenges:** Designing comprehensive validation suites without causing performance bottlenecks, and handling large volumes of validation failures efficiently.

Data Warehouse

Concept: A centralized repository designed for analytical reporting and decision support. **Related Terms:** OLAP, dimensional model, ETL. **Explanation:** While not the primary target of master data migration, the warehouse often consumes master data as reference dimensions. Aligning master data structures with warehouse schemas is crucial for consistent reporting. **Example:** Loading the “Customer” master as a dimension table in the data warehouse for sales analytics. **Practical Application:** Synchronization jobs keep the warehouse dimensions up-to-date with the master data repository. **Challenges:** Managing divergent refresh cycles, handling slowly changing dimensions, and ensuring data lineage traceability.

ETL (Extract, Transform, Load)

Concept: A classic pattern for moving data from source to target systems. Related Terms: ELT, pipeline, batch processing. Explanation: ETL extracts raw data, applies business logic and cleansing during transformation, and loads the refined data into the target. In master data migration, ETL tools automate many repetitive tasks. Example: Using an ETL tool to extract product records, standardize naming conventions, and load them into the MDM hub. Practical Application: ETL job scheduling aligns data loads with maintenance windows to minimize impact on operational systems. Challenges: Complex transformations may require custom coding, and performance tuning is needed to handle high-volume loads within acceptable timeframes.

Entity

Concept: A distinct real-world object represented in master data, such as a Customer, Product, or Supplier. Related Terms: record, object, instance. Explanation: Entities are defined by a set of attributes and have unique identifiers. They form the core of the master data model. Example: The "Customer" entity includes attributes like CustomerID, Name, and Address. Practical Application: Entity definitions guide the creation of database tables, API contracts, and user interfaces. Challenges: Over- or under-modeling entities can lead to redundancy or loss of critical information, and aligning entity definitions across business units is often difficult.

Master Data

Concept: The core, non-transactional data that is shared across an organization. Related Terms: reference data, core data, golden record. Explanation: Master data provides a single source of truth for critical business objects, enabling consistent processes and analytics. It is the primary focus of migration and MDM initiatives. Example: A consolidated "Product" master that contains the authoritative list of SKUs, descriptions, and pricing tiers. Practical Application: Master data is synchronized with downstream systems (e.g., ERP, CRM) to ensure uniformity of product information. Challenges: Identifying true master data, eliminating duplicates, and establishing governance for ongoing maintenance.

Master Data Management (MDM)

Concept: A discipline and technology suite for creating and maintaining a single, reliable view of master data. Related Terms: hub, registry, golden record. Explanation: MDM solutions provide data modeling, workflow, stewardship, and survivorship capabilities to manage master records throughout their lifecycle. Example: An MDM hub that consolidates customer data from multiple acquisition channels and resolves conflicts using predefined rules. Practical Application: MDM platforms enforce data quality standards, support real-time synchronization, and provide APIs for consumption by other applications. Challenges: Integration complexity, change-resistance from business units, and the need for robust governance to prevent "data silos" within the MDM solution itself.

Metadata

Concept: Data that describes other data, providing context such as structure, lineage, and semantics. Related Terms: data dictionary, catalog, schema. Explanation: Metadata enables understanding of source and target structures, supports impact analysis, and drives automated mapping. It includes technical details (data types) and business definitions (meaning). Example: A metadata entry that defines the "OrderDate"

field as a date type with the format "YYYY-MM-DD". Practical Application: Migration tools ingest metadata to auto-generate mapping suggestions and validate data type compatibility. Challenges: Incomplete or outdated metadata hampers accurate mapping and can cause runtime errors during migration.

Normalization

Concept: The process of organizing data to reduce redundancy and improve integrity. Related Terms: de-duplication, entity-relationship, first normal form. Explanation: Normalization restructures data into related tables, each representing a single entity, which simplifies maintenance and supports referential integrity. In master data, normalization helps create a clean, non-redundant model. Example: Splitting a flat "Customer" record that contains multiple address lines into separate "Customer" and "Address" tables linked by a foreign key. Practical Application: Normalized structures are used in MDM hubs to store core entities while allowing flexible extensions. Challenges: Over-normalization may degrade performance for read-heavy analytical workloads, requiring a balance between structure and accessibility.

Operational Data Store (ODS)

Concept: A temporary repository that consolidates data from multiple source systems for short-term use. Related Terms: staging area, integration layer, temporary store. Explanation: The ODS holds raw extracted data before it undergoes transformation and loading into the master data repository. It enables data profiling, cleansing, and validation in a controlled environment. Example: An ODS that receives daily extracts of supplier information from three legacy ERP systems. Practical Application: ETL processes read from the ODS, apply business rules, and write to the final master data store, ensuring that source systems remain unaffected during migration. Challenges: Managing storage capacity, ensuring data consistency across refresh cycles, and preventing the ODS from becoming a permanent data silo.

Reference Data

Concept: Static or slowly changing data that provides classification or categorization for master data. Related Terms: lookup, code list, enumeration. Explanation: Reference data includes codes such as country, currency, or industry classifications that master entities reference. Maintaining consistency of reference data across systems is essential for accurate reporting. Example: A "Currency Code" reference table containing ISO-4217 values like USD, EUR, and JPY. Practical Application: Migration scripts validate that all master records reference existing codes, and any missing codes are added to the reference table prior to load. Challenges: Divergent code sets between source and target, outdated reference values, and the need for periodic updates to reflect regulatory changes.

Source System

Concept: The originating application or database that holds the data to be migrated. Related Terms: origin, legacy system, upstream. Explanation: Source systems may be on-premise, cloud-based, or hybrid, each with its own data model, access method, and constraints. Understanding the source environment is critical for extraction planning. Example: An on-premise ERP that stores product master data in a relational database. Practical Application: Connectivity adapters are configured to read data from each source system, respecting security and performance considerations. Challenges: Limited documentation, proprietary data formats, and restrictive access controls can impede extraction efforts.

Target System

Concept: The destination platform where master data will reside after migration. **Related Terms:** destination, MDM hub, sink. **Explanation:** The target system defines the final data model, storage technology, and interfaces for consumption. It may be a dedicated MDM solution, a data warehouse, or a cloud service. **Example:** A SaaS-based MDM platform that stores consolidated customer records. **Practical Application:** Target schemas are provisioned, indexes created, and security roles assigned before loading begins. **Challenges:** Aligning target capabilities with business requirements, handling schema changes during migration, and ensuring that the target can support required volume and performance.

Transformation

Concept: The set of operations that convert source data into the format required by the target system. **Related Terms:** data mapping, logic, conversion. **Explanation:** Transformations may include data type conversion, standardization, enrichment, aggregation, and survivorship rules. They are defined in the mapping specification and executed by the migration engine. **Example:** Converting a "Date" field from "MM/DD/YYYY" to "YYYY-MM-DD" and applying timezone adjustments. **Practical Application:** Transformation scripts are version-controlled, allowing repeatable execution and rollback if needed. **Challenges:** Complex business logic can lead to performance bottlenecks, and ensuring that transformations are auditable and reversible is essential for compliance.

Unique Identifier

Concept: A value that uniquely distinguishes each master record within an entity. **Related Terms:** primary key, surrogate key, natural key. **Explanation:** Unique identifiers enable reliable linking, de-duplication, and referential integrity. They may be system-generated (surrogate) or derived from business data (natural). **Example:** A "CustomerID" generated by the MDM system that is independent of any source system's identifier. **Practical Application:** During migration, source identifiers are mapped to the new unique identifier, and cross-reference tables preserve the relationship for audit purposes. **Challenges:** Collisions when merging records from multiple sources, and ensuring that downstream systems adopt the new identifier without disruption.

Versioning

Concept: The practice of tracking changes to master records over time. **Related Terms:** history, audit trail, temporal data. **Explanation:** Versioning allows organizations to retain previous states of master data, supporting regulatory compliance and rollback capabilities. It can be implemented as row-level timestamps or separate version tables. **Example:** Maintaining a versioned record of a product's price changes, each with an effective start and end date. **Practical Application:** Migration tools copy both current and historical versions to preserve lineage and support downstream analytics. **Challenges:** Managing storage growth, ensuring correct sequencing of version records, and handling overlapping effective dates during consolidation.

Workflow

Concept: A defined sequence of tasks and approvals that govern the migration process. **Related Terms:** process, pipeline, orchestration. **Explanation:** Workflows automate activities such as data extraction, validation, approval, and loading, often using orchestration tools. They enforce consistency and provide

visibility into progress. Example: A workflow that routes newly identified duplicate records to a data steward for manual resolution before load. Practical Application: Monitoring dashboards display workflow status, enabling project managers to identify bottlenecks and intervene promptly. Challenges: Designing flexible workflows that accommodate exceptions, integrating with existing IT service management tools, and ensuring that automated steps do not bypass critical governance checks.

Xref (Cross-Reference)

Concept: A mapping that links identifiers from different source systems to a unified master identifier.

Related Terms: mapping table, correlation, linkage. Explanation: Xref tables preserve the relationship between legacy keys and the new unique identifier, supporting traceability and rollback. They are essential when merging duplicate records from multiple origins. Example: A table that maps "Legacy_Cust_ID_1" and "Legacy_Cust_ID_2" to the new "Master_Cust_ID". Practical Application: Post-migration, integration interfaces use the Xref table to translate incoming transaction IDs to the master identifier. Challenges: Maintaining Xref integrity as new sources are added, handling orphaned legacy IDs, and ensuring that the Xref does not become a performance bottleneck.