IDERA® | Data Modeling Whitepaper

IMPROVE YOUR DATA MODELING SKILLS

INTRODUCTION

Data professionals should know that improving their data modeling skills increases productivity and efficiency. Certifications can demonstrate these skills, which also improves your marketability. Learning the skills discussed here will help achieve the success you want by allowing you to become a sought-after data modeler. Data modeling skills include a basic understanding of how to create and read such a model. Fundamental data design techniques are also essential skills for data modelers.

OVERVIEW

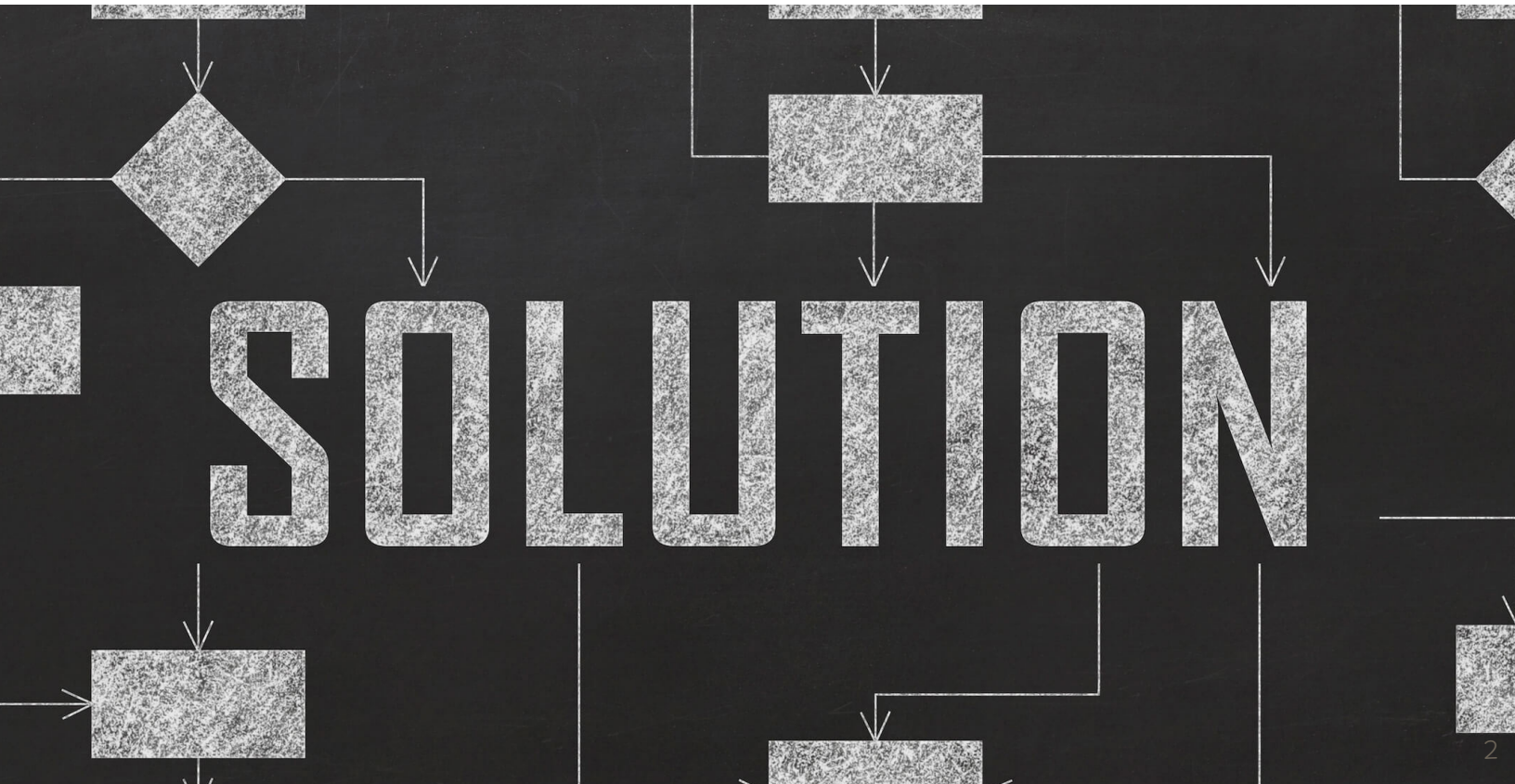
Data modeling is the process of analyzing data-oriented structures. These models include a variety of specific types ranging from models for physical data to models that illustrate high-level concepts. Data modeling is similar to class modeling from the perspective of object-oriented (OO) design except that you're modeling entity types instead of classes. A data modeler assigns attributes to these entity types, just as an OO developer assigns attributes and operations to classes. The associations between entities in data modeling are also similar to those between classes in OO design.

ENTITY TYPES

An understanding of entity types is one of the most fundamental skills that data models must develop. Entity types represent a collection of similar objects such as people, places and things, although they can also represent non-physical concepts like events. Customer, Order and Item are all common entity types in an order entry database. The main difference between an entity type and class is that an entity types only represents data whereas a class also describes an object's behavior.

ATTRIBUTES

Entity types have at least one attribute. For example, attributes for a Customer entity type typically include First Name and Last Name. Developers typically implement attributes as column in database tables, although achieving the optimum level of detail is often the most challenging aspect of this process. Expressing a single attribute with multiple columns can provide greater control over the data, but it also incurs development and maintenance costs. For example, a phone number in North America has three components, including the area code, prefix and line number. However, there's rarely a compelling reason to assign each component to a separate column.



NAMING CONVENTIONS

A data modeling project should include naming conventions that are typically maintained by enterprise administrators. These standards are essential for making code easy to understand and modify. Physical and logical data models typically have different naming conventions since they have different purposes. For example, the naming conventions for logical data models should give greater priority to human readability, while the naming conventions for physical models focus more on technical considerations.

RELATIONSHIPS

Learning about the relationships between entities is a key requirement for developing your data modeling skills. The relationships between entities are conceptually identical to the associations between objects in OO programming. For example, Customers place Orders, so placement is the typical relationship between customers and orders in an order entry system. Similarly, Customers live at an Address, and a Zip Code is part of an Address. Naming a relationship often becomes unnecessary if you specify the entities' role in that relationship with sufficient clarity.

KEY ASSIGNMENT

Data modeling uses two basic strategies to assign keys to tables. Assigning a natural key is usually the best option when a table has at least one attribute that's unique to the table's business concept. For tables where this isn't the case, a data modeler needs to add a new column known as a surrogate key. This column has no business meaning and merely serves to identify an entity type. For example, addresses don't have an obvious natural key because you need the entire address to identify it. In this case, a data modeler often uses a surrogate key called something like "AddressID" to identify the address.

NORMALIZATION

Normalization is the process of organizing data within a data model to make its entity types more cohesive. This generally involves reducing data redundancy, which is highly beneficial for application development. Storing objects in a relational database becomes much easier when information about those objects is maintained in only one place.
The first three levels of normalization are the most common, although higher levels are possible. Normalization is a progressive hierarchy such that the next level meets all the requirements of the previous level. For example, an entity type is in first normal form (1NF) if it doesn't contain repeating groups of data. An entity type is in second normal form (2NF) if it's in 1NF and its non-key attributes are fully dependent on its primary key. An entity type is in third normal form (3NF) if it's in 2NF and its attributes are directly dependent on the primary key.

However, normalization incurs a performance cost, so denormalization is also an important skill for data modelers. As a result, data models often bear little resemblance to their normalized schema.

SUMMARY

In addition to proper training, the key to improving data modeling skills is practice. Data modelers who follow the agile methodology should welcome the offer of assistance as part of the Model With Others practice. Answering questions about why you do things a certain way is also a great way to hone your data modeling skills.

IDERA'S SOLUTION

ER/Studio Data Architect

Document, design, and communicate data assets faster and easier.

Round-trip database support gives ER/Studio Data Architect users the power to easily reverse- and forward-engineer, compare and merge, and visually document data assets across multiple relational, NoSQL, and ETL platforms and data sources. Built-in facilities automate routine data modeling tasks so users can analyze and optimize database and data warehouse designs faster than ever.

Start a free, fully-functional, 14-day trial today!

Start for FREE