DATA+AI SUMMIT BY databricks

Data Modeling Made Simple: A Non-Technical Beginner's Guide

Jason Zippro
June x 2024

Jason Zippro

HubSpot

Framework Data Group
Sr. Product Manager

Data Storage & Query Layer for all of HubSpot's CRM Data

Data modeling support for internal teams

GROW

Full-Stack BI Software
Head of Product & UX

Work with customers to build a single-source of truth for their business data

Copy of this Deck

Agenda

- Introduction
- The Data Model Components
- Additional Considerations
- Wrap Up

Translate your Business Model Into an actionable Data Model

Type of Data Models

Our focus is a blend of Conceptual + Logical

Conceptual	Logical	Physical
Identifies & Describes at a high-level the Entities & Relationships that a business uses in day-to-day operations	More fully defines the Entities & Relationships of the business by adding details such as Actions and Attributes and structure to the conceptual model	This is the actual implementation of the Data Model into the data platform(s) of choice of the business

What are the Benefits of doing this?

For You

Quickest way to make you the most knowledgeable person about the business and how it works

→ More Valuable Contributions

What are the Benefits of doing this?

For the Business

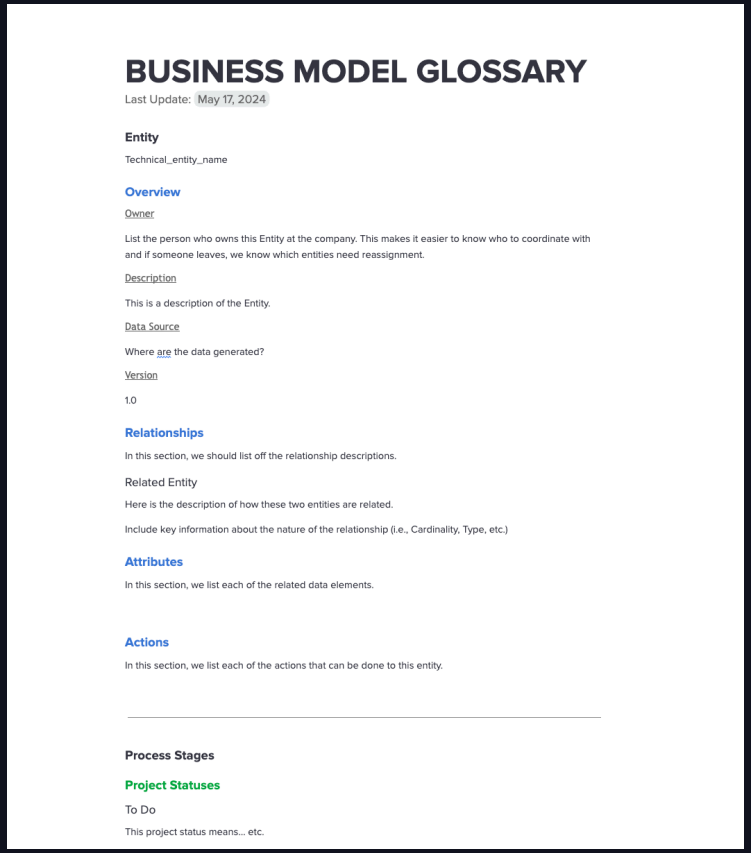
- Better allocation of resources
- Lowers risks and predicts problems before they arise
- Increases efficiency by strengthening cross-functional communication
- Guarantees underlying data quality & security/accessibility
- Makes the business more adaptable to change

Business Glossary

Key Artifact to Communicate

Captures the Details of your Conceptual Data Model

- Business Concepts
- Business Logic
- Data to track and how it's structured
- Changes to the Business Model



BUSINESS MODEL GLOSSARY

Last Update: May 17, 2024

Entity

Technical_entity_name

Overview

Owner

List the person who owns this Entity at the company. This makes it easier to know who to coordinate with and if someone leaves, we know which entities need reassignment.

Description

This is a description of the Entity.

Data Source

Where are the data generated?

Version

1.0

Relationships

In this section, we should list off the relationship descriptions.

Related Entity

Here is the description of how these two entities are related.

Include key information about the nature of the relationship (i.e., Cardinality, Type, etc.)

Attributes

In this section, we list each of the related data elements.

Actions

In this section, we list each of the actions that can be done to this entity.

Process Stages

Project Statuses

To Do

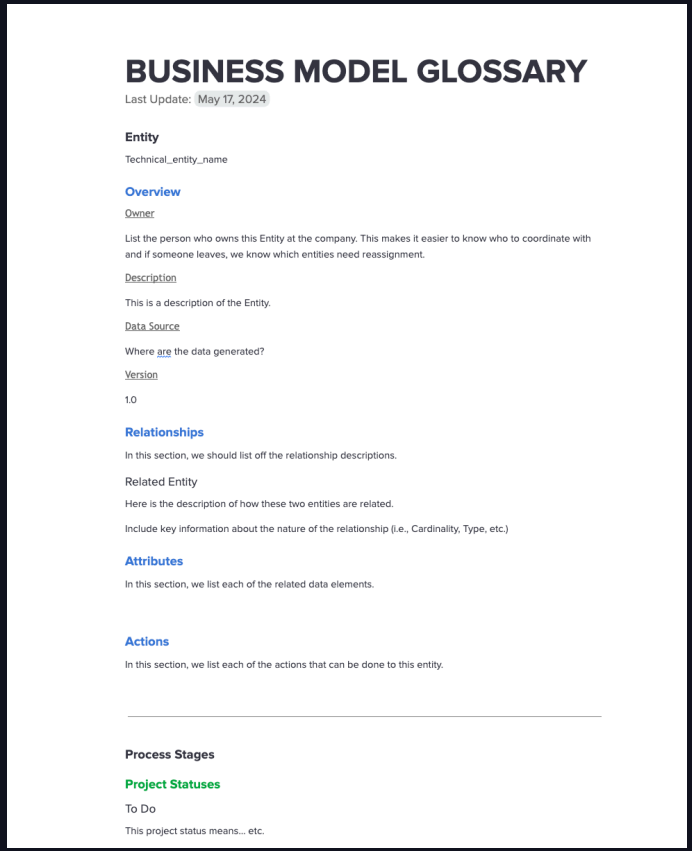
This project status means... etc.

Workshop

Activities to try as we go

Business Model Glossary

- Scan QR code
- Make a copy of this Template



BUSINESS MODEL GLOSSARY

Last Update: May 17, 2024

Entity

Technical_entity_name

Overview

Owner

List the person who owns this Entity at the company. This makes it easier to know who to coordinate with and if someone leaves, we know which entities need reassignment.

Description

This is a description of the Entity.

Data Source

Where are the data generated?

Version

1.0

Relationships

In this section, we should list off the relationship descriptions.

Related Entity

Here is the description of how these two entities are related.

Include key information about the nature of the relationship (i.e., Cardinality, Type, etc.)

Attributes

In this section, we list each of the related data elements.

Actions

In this section, we list each of the actions that can be done to this entity.

Process Stages

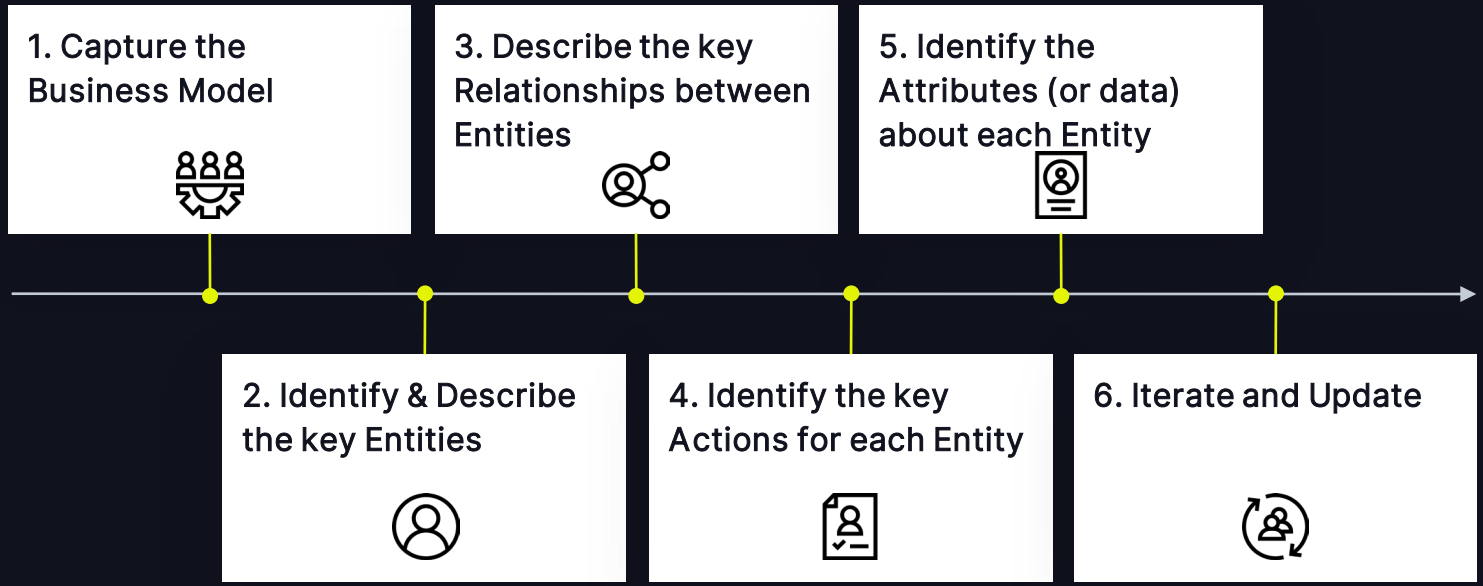
Project Statuses

To Do

This project status means... etc.

Let’s Begin!

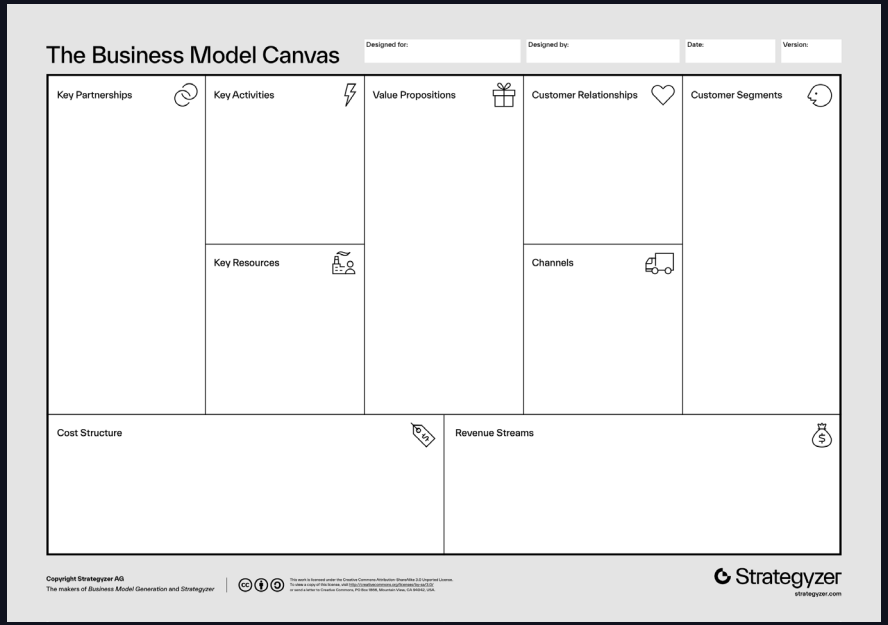
Data Modeling Process



Overview: Entities > Relationships > Actions > Attributes > Iterations

Capture Your Business Model

What does your business do? Write it out. Diagram it.



Overview: Entities > Relationships > Actions > Attributes > Iterations

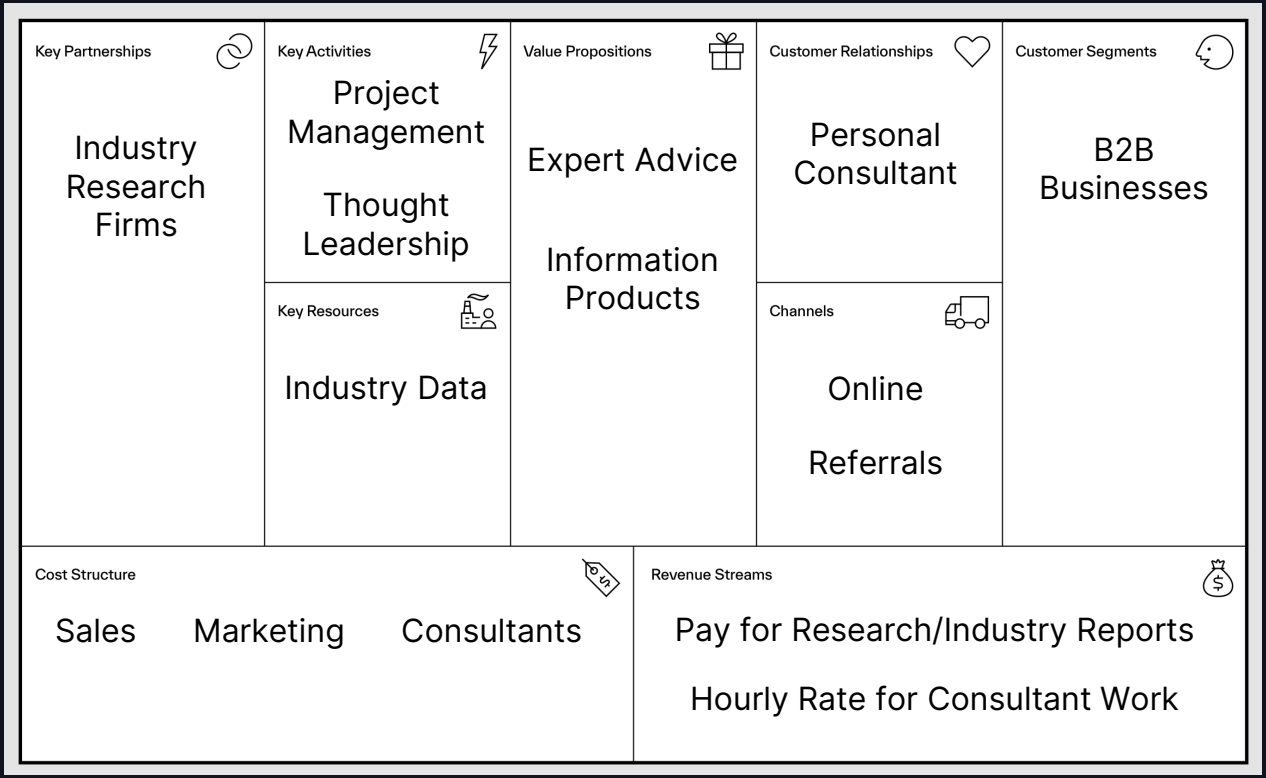
Example Business Model

Business Consulting Firm Example

The firm offers expert advice and information products to businesses to help them refine their strategy, increase profits/lower costs, add value, and resolve issues.

Overview: Entities > Relationships > Actions > Attributes > Iterations

Example Business Model



Overview: Entities > Relationships > Actions > Attributes > Iterations

Entities

Objects or Nouns of your Business Model

Overview: Entities > Relationships > Actions > Attributes > Iterations

Entities

Objects or Nouns of your Business Model

Who is your Customer?

What is your Product / Service?

Where does your Customer buy?

What or Who delivers your Product / Service?

👍 Rule of Thumb: Is it Countable?

Overview: Entities > Relationships > Actions > Attributes > Iterations

Entities

Objects or Nouns of your Business Model

External People	Things	Engagement	Internal People
Prospect	Content	Campaign	Marketer
Client	Project	Meeting	Sales Rep
	Product	Task	Consultant

Overview: Entities > Relationships > Actions > Attributes > Iterations

Entities

Objects or Nouns of our Business Model

External People	Things	Engagement	Internal People
Prospect	Content	Campaign	Marketer
Client	Project	Meeting	Sales Rep
	Product	Task	Consultant

Overview: Entities > Relationships > Actions > Attributes > Iterations

Entities

Objects or Nouns of our Business Model

External People	Things	Engagement	Internal People
Prospect	Content	Campaign	Marketer
Client	Project	Meeting	Sales Rep
	Product	Task	Consultant

Overview: Entities > Relationships > Actions > Attributes > Iterations

Entities

Objects or Nouns of our Business Model

External People	Things	Engagement	Internal People
Prospect	Content	Campaign	Marketer
Client	Project	Meeting	Sales Rep
	Product	Task	Consultant

ACTIVITY

Write down a few Entities for your business model

Overview: Entities > Relationships > Actions > Attributes > Iterations

Entities

Objects or Nouns of your Business Model

Now write a 2-3 sentence description of each entity.

Overview: Entities > Relationships > Actions > Attributes > Iterations

Entities

Objects or Nouns of your Business Model

Projects: consulting engagements aimed at achieving the client's objectives. They are composed of various tasks that need to be completed by consultants.

Consultants: professionals who carry out the tasks within projects. They have various skills and expertise relevant to the client's needs.

Clients: the businesses that hire the consulting firm. They have specific objectives, such as improving efficiency, entering new markets, or enhancing technology infrastructure.

Overview: Entities > Relationships > Actions > Attributes > Iterations

Relationships

How your Entities are related or interact with each other

Overview: Entities > Relationships > Actions > Attributes > Iterations

Relationships

How your Entities are related or interact with each other.

Subject + Verb + Direct Object

Overview: Entities > Relationships > Actions > Attributes > Iterations

Relationships

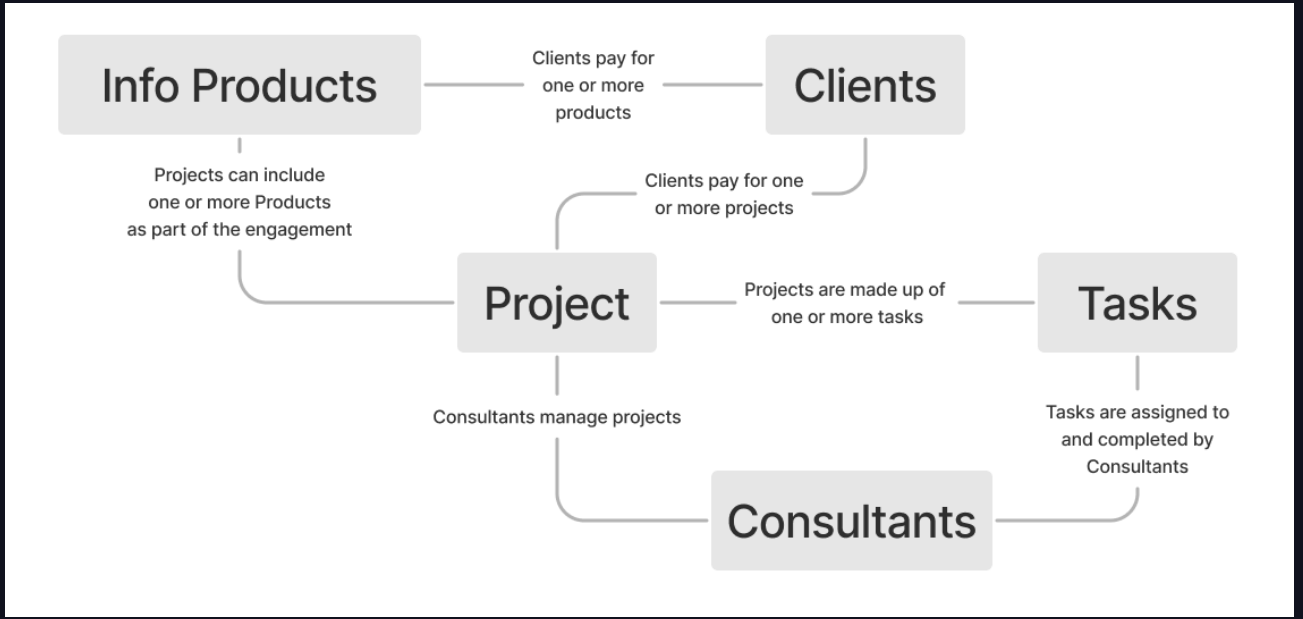
How your Entities are related or interact with each other.

- Clients pay for one or more projects.
- Clients purchase one or more information products.
- Projects can include one or more products as part of the engagement.
- Projects are made up of one or more tasks.
- Tasks are assigned to and completed by a consultant.
- Marketers create and run campaigns.

Overview: Entities > Relationships > Actions > Attributes > Iterations

Relationships

How your Entities are related or interact with each other.



Overview: Entities > Relationships > Actions > Attributes > Iterations

Relationships

How your Entities are related or interact with each other.

ACTIVITY

Describe the relationship between your entities

Overview: Entities > Relationships > Actions > Attributes > Iterations

Actions

What can be done to your entities

Overview: Entities > Relationships > Actions > Attributes > Iterations

Actions

What can be done to your entities

Events (Verbs) of your Business Model

Overview: Entities > Relationships > Actions > Attributes > Iterations

Actions

What can be done to your entities

Content

- Published
- Clicked
- Downloaded

Project

- Paid for
- Delivered
- Created
- Cancelled

Task

- Created
- Edited
- Deleted
- Assigned

Overview: Entities > Relationships > Actions > Attributes > Iterations

Actions

What can be done to your entities

ACTIVITY

Describe what can be done to your entities

Overview: Entities > Relationships > Actions > Attributes > Iterations

Attributes

The information that describe specific characteristics of an entity

Overview: Entities > Relationships > Actions >Attributes > Iterations

Attributes

The information that describe specific characteristics of an entity

Client

- Client ID
- Name
- Contact Info
- Industry

Project

- Project ID
- Name
- Start Date
- End Date
- Objective
- Status

Task

- Task ID
- Description
- Status
- Assigned To
- Due Date

Overview: Entities > Relationships > Actions >Attributes > Iterations

Attributes

The information that describe specific characteristics of an entity

- What are the most important data related to each Entity?
  Don’t try to capture everything.

Overview: Entities > Relationships > Actions > Attributes > Iterations

Attributes

The information that describe specific characteristics of an entity

ACTIVITY

List the top 10 Customer properties to track

Overview: Entities > Relationships > Actions > Attributes > Iterations

Iterations

Capturing changes in your business model

Overview: Entities > Relationships > Actions > Attributes > Iterations

Iterations

Capturing changes in your business model

- There is no correct way to model your data — You decide what is most helpful
- Business models change over time, ideally every quarter.
- Capture those changes and iterate your way into the best data model for your business, while minimizing model re-implementation as much as possible.

Overview: Entities > Relationships > Actions > Attributes > Iterations

Digging Deeper

Entities

Objects or Nouns of your Business Model

Dig Deeper: Entities > Relationships > Actions > Attributes > Iterations

Entities

Two main categories of Entities: People & Things

People: *Active: Do*

- Client
- Prospect
- Marketer
- Consultant

Things: *Passive: Done to*

- Project
- Product
- Task
- Content

- Automations*

Dig Deeper: Entities > Relationships > Actions > Attributes > Iterations

Entities

People split into External & Internal

External Roles

- Client
- Prospect
- Vendor
- Partner

Internal Roles

- Sales Rep
- Marketer
- Consultant
- Support Rep

Dig Deeper: Entities > Relationships > Actions > Attributes > Iterations

Entities

How to model “Activities”: Entity vs Action?

Should you model “Meeting” as an entity or “Meet Client” as an action?

Marketing, Sales, and Support activities are better captured as entities: Meeting, Call, Visit, etc.

Then have actions that make sense: Attended Meeting, Rescheduled Meeting, etc.

Dig Deeper: Entities > Relationships > Actions > Attributes > Iterations

Entities

Single or Multiple Entities?

Employee might just have a “Role” attribute: Marketer, SDR, Support, Consultant

OR

Entity for each role with a “Group” entity: Employee

If each *type* of entity needs a set of attributes that are unique to it, then multiple entities, otherwise just capture it as a single entity with an enumeration attribute.

Dig Deeper: Entities > Relationships > Actions > Attributes > Iterations

Entities

Single or Multiple Entities?

Single Entity: Attribute

- Partner Types
    - Integration Partner
    - Retail Partner
    - Channel Partner
    - Etc.

Multiple Entities

- Consultant
    - Expertise
    - Utilization Score
    - Hourly Rate

Note: Subtype entities inherit the attributes of their supertype.

Dig Deeper: Entities > Relationships > Actions > Attributes > Iterations

Relationships

How your Entities are related or interact with each other

Dig Deeper: Entities > Relationships > Actions > Attributes > Iterations

Relationships

Cardinality — Shows how often one thing is related to another

- 1:1 — Employee : Security Info
- 1:M — Project : Task
- M:1 — Meeting : Client
- M:M — Client : Product

Dig Deeper: Entities > Relationships > Actions > Attributes > Iterations

Relationships

Cardinality — Shows how often one thing is related to another

- Clients pay for one or more projects.
- Clients purchase one or more information products.
- Projects can include one or more products as part of the engagement.
- Projects are made up of one or more tasks.
- Tasks are assigned to and completed by a consultant.
- Marketers create and run campaigns.

Dig Deeper: Entities > Relationships > Actions > Attributes > Iterations

Relationships

Unique relationship types to consider

- Aggregation
  - A collection of an entity. An Order is a collection of Products.
- Inheritance
  - A parent/child relationship. Consultant and Marketer are both an Employee.
- Component
  - An entity dependency. A Review is a component of a Product.

Dig Deeper: Entities > Relationships > Actions > Attributes > Iterations

Actions

What can be done to your entities

Dig Deeper: Entities > Relationships >Actions > Attributes > Iterations

Actions

Put actions on the *Thing* not the *People*

- People & Processes are Action Triggers
- Smaller subset of actions
- You know what entity general actions are associated with: e.g., “Create” can be associated with most objects.

Dig Deeper: Entities > Relationships > Actions > Attributes > Iterations

Actions

Example Actions from our Relationship Descriptions

- Clients pay for one or more projects.
  - Paid for Project
- Tasks are assigned to and completed by a consultant.
  - Assigned to Task
- Marketers create and run campaigns.
  - Created Campaign

Dig Deeper: Entities > Relationships >Actions > Attributes > Iterations

Attributes

The information that describe specific characteristics of an entity

Dig Deeper: Entities > Relationships > Actions >Attributes > Iterations

Attributes

Bulk of your time will be spent on attributes

Client

- Client ID
- Name
- Contact Info
- Industry
- Retainer

Project

- Project ID
- Name
- Description
- Start Date
- End Date
- Objective
- Status

Consultant

- Consultant ID
- Name
- Expertise
- Utilization Rate
- Hourly Rate

Dig Deeper: Entities > Relationships > Actions >Attributes > Iterations

Attributes

Data Types

- Text (String)
- Number
- Date / Time / Timestamp
- Boolean: T/F, On/Off, Yes/No, etc.
- Pick list (Enumeration)
  - e.g., Project statuses
  - e.g., Partner types
  - e.g., Hiring stages

Dig Deeper: Entities > Relationships > Actions >Attributes > Iterations

Attributes

Constraints

- Unique & Identifier (Primary key)
- Default Value
- Enumeration
- Null / Not Null
- Data Specific
  - Age: (number: 1-120)
  - SSN (Format: ###-##-####)

Dig Deeper: Entities > Relationships > Actions >Attributes > Iterations

Attributes

Data Governance

- Who in the business should have access to what data?
- PPI & Sensitive data constraints?

Dig Deeper: Entities > Relationships > Actions >Attributes > Iterations

Attributes

Additional Dimensions to Consider

- Date / Time
  - What timezone will you use?
- Currency
  - What currency will you use?
  - If you accept in other currencies, what is your logic for accounting for exchange rates?
- Location / Geo
  - Addresses or Geocodes

Dig Deeper: Entities > Relationships > Actions >Attributes > Iterations

Attributes

Naming Conventions

- _id for unique identifiers (primary key)
  - client_id, consultant_id, project_id
- _at for timestamp attributes
  - created_at, updated_at, completed_at
- snake_case
  - camelCase is alternative, but it's a lot harder for people to read and remember.

Dig Deeper: Entities > Relationships > Actions >Attributes > Iterations

Attributes

What about Aggregations & Calculated attributes?

- Aggregations & Calculations should be called out and treated differently because you need to treat them differently in your Physical Data Model.

- Business Glossary: Include the logic and/or calculations that make up this number. Use some symbol to signify that this data is derived.

Dig Deeper: Entities > Relationships > Actions > Attributes > Iterations

Attributes

Example Aggregation and Calculation

- Aggregations: Order Total
  - The sum (calculation) of each product_price multiplied by the product_quantity in the order.
- Calculation: Customer Acquisition Cost
  - Total marketing and sales expenses divided by the number of new customers acquired.

Dig Deeper: Entities > Relationships > Actions > Attributes > Iterations

Iterations

Capturing changes in your business model

Dig Deeper: Entities > Relationships > Actions > Attributes >Iterations

Iterations

There is no correct way to model your data

- Start simpler. Then get more sophisticated and complex, but try to eliminate complexity along the way.
  - It’s easier to add net new, then to fix existing old
  - Add nuance only when it will help with long-term decisions and tracking

- Make it Easily Accessible at first → e.g., a simple document anyone can access

Dig Deeper: Entities > Relationships > Actions > Attributes > Iterations

Iterations

Business Models change over time, plan for it

- Capture those changes in a Business Glossary Index
  - Log with versions: 2024, Q1 → 24.1
- Iterate your way into the best data model for your business
  - e.g., start with the enumeration instead of multiple entities and then evolve to multiple entities once it’s very clear you need to.
- Minimize model re-implementation as much as possible

Dig Deeper: Entities > Relationships > Actions > Attributes > Iterations

Iterations

Over time, enlist others in keeping it up-to-date

- Spend a day, once a quarter meeting with key stakeholders for updates
- Work with them to take ownership of their portion

Dig Deeper: Entities > Relationships > Actions > Attributes > Iterations

Thank you!