## We consider splitting entities with optional arrows and merging those with similar relationships.

indicate solutions that are likely, but not absolutely certain. They warn us to investigate further.

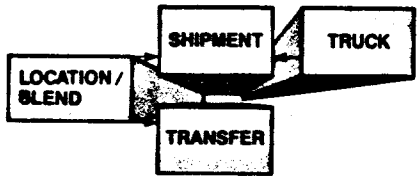
Consider the following database design for a pipe-tobacco wholesaler. Cut tobacco is stored in warehouses and distributed in trucks.


The location/blend record tells how many pounds of one tobacco blend there are at a specific warehouse. If there are 20 warehouses and 50 blends in all, we could have up to 1,000 occurrences of this record (perhaps not every blend will be stored at each location). Any blend may be shipped or received many times. Similarly, each truck may be involved in many movements. But, by definition, each movement represents an occasion where a single location's inventory of a blend was either incremented because tobacco arrived or decremented because some was sent off. Also, each movement involves only one vehicle.

The situation seems straightforward. The fields in the location/blend record include location number, blend number, inventory balance, and the like. Truck record holds vehicle number, cargo capacity, miles-since-maintenance, etc. And movement contains date, quantity shipped, and an in-or-out flag. But doesn't the truck-to-movement arrow mean each movement must involve a truck?

Think about transferring tobacco within a location. Many pipe tobaccos are produced by blending others—that's why they're called blends. The process is one of simply shoveling a measured amount from one hopper to another and, though it certainly affects the inventory balances, no vehicle is involved. The truck-movement relationship is then optional in some sense. Only one truck is involved in any shipments that go by truck, but some shipments do not involve trucks at all.

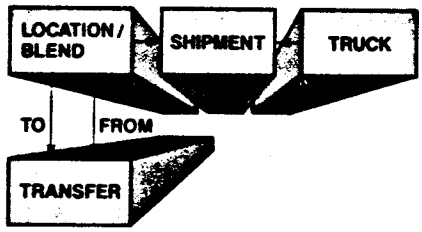
The optional arrow is a warning that the movement box, in an information modeling sense, could represent two fundamentally different entities. Redrawing it as two boxes, we have:




Shipment and transfer tell about different real-world events. As we refine the design further, their record layouts continue to diverge until, eventually, we find the only data common to both are date and quantity.

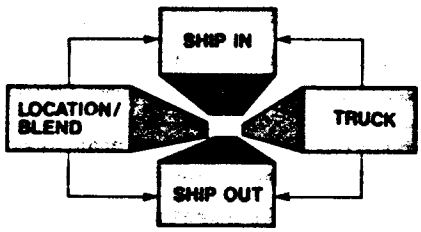
Our first indication of their duality was that optional arrow.

Now look at the arrow between location/blend and transfer. We can't say each transfer is associated with only one location/blend because the transfer event affects two different balances. It decrements the inventory balances and increments the location/blend balance it goes to.




The diagram does not mean that each transfer is related twice to the same location/blend, by the way. On the contrary, each transfer has two different relationships with two different location/blend record occurrences—a "from" relationship with one and a "to" with the other.

Merging records with similar relationships is basically the reverse of the process just described. For example:




Both intersection records, inbound shipment and outbound shipment, hold the same data elements and are related to everything around them in precisely the same way. This situation warns us to look at them more closely and see if they're not really modeling the same class of real-world event.

Next time, we'll look at hard sets, soft sets, and summary fields. ◉

Frank Sweet is corporate manager of data administation for the Charter Co., Jacksonville, Fla.