

What is claimed is:

1. A computing system including:

a data storage system storing entity data representing a plurality of entities, with each entity having one or more attributes, at least some of the entities each having multiple instances, and at least some of the instances each having respective values for one or more of the attributes;

an input device or port for receiving input data that includes one or more mappings that each define a correspondence between one or more input attributes of an input entity and one or more output attributes of an output entity, where the input entity includes one or more key attributes identified as part of a unique key for the input entity, and the output entity includes one or more key attributes identified as part of a unique key for the output entity; and

at least one processor configured to process instances of the input entity to generate instances of the output entity according to the one or more mappings included in the input data, the processing including:

determining one or more mapped input attributes of the input entity that correspond to each of the one or more key attributes of the output entity, based on the one or more mappings; and

comparing the mapped input attributes with the one or more key attributes of the input entity to determine whether the mapped input attributes include: (1) all of the key attributes of the input entity, or (2) fewer than all of the key attributes of the input entity; and

generating the instances of the output entity based on: (1) a one-to-one correspondence between the instances of the output entity and instances of the input entity that have matching key attributes, in response to determining that the mapped input attributes include all of the key attributes of the input entity, or (2) an aggregation of multiple instances of the input entity that share the same values for the mapped input attributes, in response to determining that the mapped input attributes include fewer than all of the key attributes of the input entity.

2. The computing system of claim 1, wherein determining one or more mapped input attributes of the input entity that correspond to each of the one or more key attributes of the output entity includes determining whether the one or more mapped input attributes have a one-to-one correspondence with respective key attributes of the output entity.

3. The computing system of claim 1, wherein the entity data represent a plurality of output entities that are related according to a hierarchy, where at least one root output entity is at a highest level of the hierarchy and one or more output entities are at one or more levels below the highest level of the hierarchy, and each output entity at a level lower than the root entity is a sub-entity of a single output entity.

4. The computing system of claim 3, wherein the entity data represent a plurality of input entities that are related according to a hierarchy, where at least one root input entity is at a highest level of the hierarchy and one or more input entities are at one or more levels below the highest level of the hierarchy, and each input entity at a level lower than the root entity is a sub-entity of a single input entity.

5. The computing system of claim 3, wherein at least a first entity that is not related to the plurality of output entities that are related according to a hierarchy includes at least one attribute that is referenced as an output attribute by at least one of the mappings included in the input data.

6. The computing system of claim 5, wherein the first entity includes at least one attribute that is referenced as an input attribute by at least one of the mappings included in the input data.

7. The computing system of claim 1, wherein a plurality of instances of a first entity that is a sub-entity of a second entity each include a common value of a key attribute of the first entity that identifies a particular instance of the second entity.

8. The computing system of claim 7, wherein the first entity corresponds to a first set of records, the second entity corresponds to a second set of records, and the key attribute of the first entity corresponds to a foreign key field of the first set of records that identifies a value included in a primary key field of a particular record in the second set of records.

9. The computing system of claim 1, wherein a plurality of instances of a first entity that is a sub-entity of a second entity correspond to a plurality of elements of a vector that is included within a data structure of a particular instance of the second entity.

10. The computing system of claim 9, wherein the processing further includes generating the instances of the output entity using a dataflow graph to process the instances of the input entity to generate the instances of the output entity, the dataflow graph including nodes representing components configured to perform operations on instances of an entity, and links between nodes representing flows of instances between components.

11. The computing system of claim 10, wherein the dataflow graph includes at least one split component that is configured to extract one or more vectors of instances of a sub-entity from a data structure of an instance of another entity based on the input attributes of the one or more mappings, and at least one combine component that is configured to insert one or more vectors of instances of a sub-entity into a data structure of an instance of another entity based on the output attributes of the one or more mappings.

12. The computing system of claim 10, wherein the dataflow graph includes, for each mapping for which the mapped input attributes include fewer than all of the key attributes of the input entity, at least one component that performs an aggregation operation to aggregate multiple instances of the input entity that share the same values for the mapped input attributes.

13. The computing system of claim 1, further including at least one output device or port for displaying a user interface configured to receive the input data.

14. The computing system of claim 13, wherein the user interface is further configured to display result information characterizing a result of generating the instances of the output entity according to the one or more mappings included in the input data.

15. The computing system of claim 13, wherein the result information includes a total number of instances of the output entity that were generated.

16. A method for processing data in a computing system, the method including:
storing, in a data storage system, entity data representing a plurality of entities, with each entity having one or more attributes, at least some of the entities each having multiple instances, and at least some of the instances each having respective values for one or more of the attributes;

receiving, over an input device or port, input data that includes one or more mappings that each define a correspondence between one or more input attributes of an input entity and one or more output attributes of an output entity, where the input entity includes one or more key attributes identified as part of a unique key for the input entity, and the output entity includes one or more key attributes identified as part of a unique key for the output entity; and

processing, with at least one processor, instances of the input entity to generate instances of the output entity according to the one or more mappings included in the input data, the processing including:

determining one or more mapped input attributes of the input entity that correspond to each of the one or more key attributes of the output entity, based on the one or more mappings; and

comparing the mapped input attributes with the one or more key attributes of the input entity to determine whether the mapped input attributes include: (1) all of the key attributes of the input entity, or (2) fewer than all of the key attributes of the input entity; and

generating the instances of the output entity based on: (1) a one-to-one correspondence between the instances of the output entity and instances of the input entity that have matching key attributes, in response to determining that the mapped input attributes include all of the key attributes of the input entity, or (2) an aggregation of multiple instances of the input entity that share the same values for the mapped input attributes, in response to determining that the mapped input attributes include fewer than all of the key attributes of the input entity.

17. Software stored in non-transitory form on a computer-readable medium, the software including instructions for causing a computing system to:

store, in a data storage system, entity data representing a plurality of entities, with each entity having one or more attributes, at least some of the entities each having multiple instances, and at least some of the instances each having respective values for one or more of the attributes;

receive, over an input device or port, input data that includes one or more mappings that each define a correspondence between one or more input attributes of an input entity and one or more output attributes of an output entity, where the input entity includes one or more key attributes identified as part of a unique key for the input entity, and the output entity includes one or more key attributes identified as part of a unique key for the output entity; and

process, with at least one processor, instances of the input entity to generate instances of the output entity according to the one or more mappings included in the input data, the processing including:

determining one or more mapped input attributes of the input entity that correspond to each of the one or more key attributes of the output entity, based on the one or more mappings; and

comparing the mapped input attributes with the one or more key attributes of the input entity to determine whether the mapped input attributes include: (1) all of the key attributes of the input entity, or (2) fewer than all of the key attributes of the input entity; and

generating the instances of the output entity based on: (1) a one-to-one correspondence between the instances of the output entity and instances of the input entity that have matching key attributes, in response to determining that the mapped input attributes include all of the key attributes of the input entity, or (2) an aggregation of multiple instances of the input entity that share the same values for the mapped input attributes, in response to determining that the mapped input attributes include fewer than all of the key attributes of the input entity.