CLAIMS

1. A computer-implemented method for unambiguously identifying entities in a database system, the method comprising:

storing data items in a table of a database, wherein the data items are stored as records comprising a plurality of attributes;

storing naming rules for selected combinations of the attributes of the data items, wherein the naming rules are based on the plurality of attributes including at least, in combination, a customer name and the customer's related employer identification number (EIN);

prioritizing the naming rules by defining a sequence of applying the naming rules or defining an order of the naming rules, depending on an importance of the naming rules for entity detection, wherein the combination of the customer name and the customer's related EIN have a comparably high priority as compared to the customer name alone;

determining a hash value for each of the selected combinations of the attributes of the data items,

identifying duplicate data items using the determined hash values and the prioritized naming rules; and

merging the identified duplicate data items into a merged data item, with the naming rules with higher priority defining where the attributes for the merged data item come from.

2. The method according to claim 1, wherein the database system is a relational database system, and wherein an entry into the database system automatically triggers a database engine to create the naming rules associated with the entry.

3. The method according to claim 1, wherein the database system is a configuration management database that underlies a specific internal organization and is used to manage a plurality of technical devices and applications in a plurality of data centers.

4. The method according to claim 1, further comprising merging the identified duplicate data items by maintaining the determined hash values as a multi-valued key for the merged data item.

5. The method according to claim 4, further comprising merging other data items that are in composite relationship with the identified data items.

6. The method according to claim 4, further comprising maintaining a pointer to a same row identifier of one of the merged data items for the determined hash values.

7. The method according to claim 1, further comprising:

maintaining an index of the table; and

maintaining a pointer in a search tree related to the index, such that the pointer points to the same record identifiers of a combined data item, wherein the combined data item is determined based on splitting a superior alias into two or more individual aliases and each of the two or more individual aliases is used as a single value in the index.

8. The method according to claim 1, further comprising:

using a create SQL statement adapted for a creating of the naming rule and its related priority.

9. The method according to claim 1, further comprising:

using a multi-value primary key for sorting records in the table of the database, wherein the multi-value primary key is a unique row identifier.

10. The method according to claim 1, wherein a multi-value primary key is used for clustering cluster data on multi-node database engines.

11. The method according to claim 9, wherein a multi-value primary key is comparable to a single value column data item.

12. The method according to claim 1, further comprising

collecting statistical database data for data blocks for single-valued primary keys and multi-valued primary keys.

13. A computer system for unambiguously identifying entities in the database system, the computer system comprising:

one or more computer processors, one or more computer-readable storage media, and program instructions stored on the one or more of the computer-readable storage media for execution by at least one of the one or more processors capable of performing a method, the method comprising:

storing data items in a table of a database, wherein the data items are stored as records comprising a plurality of attributes;

storing naming rules for selected combinations of the attributes of the data items, wherein the naming rules are based on the plurality of attributes including at least, in combination, a customer name and the customer's related employer identification number (EIN);

prioritizing the naming rules by defining a sequence of applying the naming rules or defining an order of the naming rules, depending on an importance of the naming rules for entity detection, wherein the combination of the customer name and the customer's related EIN have a comparably high priority as compared to the customer name alone;

determining a hash value for each of the selected combinations of the attributes of the data items,

identifying duplicate data items using the determined hash values and the prioritized naming rules; and

merging the identified duplicate data items into a merged data item, with the naming rules with higher priority defining where the attributes for the merged data item come from.

14. The computer system according to claim 13, wherein the database system is a relational database system.

15. The computer system according to claim 13, wherein the database system is a configuration management database.

16. The computer system according to claim 13, further comprising merging the identified duplicate data items by maintaining the determined hash values as a multi-valued key for a merged data item.

17. The computer system according to claim 16, further comprising merging other data items that are in composite relationship with the identified data items.

18. The computer system according to claim 16, further comprising maintaining a pointer to a same row identifier of one of the merged data items for the determined hash values.

19. The computer system according to claim 13, further comprising:

maintaining an index of the table; and

maintaining a pointer in a search tree related to the index, such that the pointer points to the same record identifiers of a combined data item, wherein the combined data item is determined based on splitting a superior alias into two or more individual aliases and each of the two or more individual aliases is used as a single value in the index.

20. A computer program product for unambiguously identifying entities in a database system, the computer program product comprising:

one or more non-transitory computer-readable storage media and program instructions stored on the one or more non-transitory computer-readable storage media capable of performing a method, the method comprising:

storing naming rules for selected combinations of the attributes of the data items, wherein the naming rules are based on the plurality of attributes including at least, in combination, a customer name and the customer's related employer identification number (EIN);

prioritizing the naming rules by defining a sequence of applying the naming rules or defining an order of the naming rules, depending on an importance of the naming rules for entity detection, wherein the combination of the customer name and the customer's related EIN have a comparably high priority as compared to the customer name alone;

determining a hash value for each of the selected combinations of the attributes of the data items,

identifying duplicate data items using the determined hash values and the prioritized naming rules; and

merging the identified duplicate data items into a merged data item, with the naming rules with higher priority defining where the attributes for the merged data item come from.