

CLAIMS

1. A method for providing statistical data from a data warehouse (410) characterized in that it comprises the following steps performed with at least one data processor:
- 5
- defining a plurality of index fields (114) each index field accepting a plurality of index field values;
 - creating a plurality of files (432) and hierarchically indexing the files as trees of indexes (300) which includes, for each tree, the steps of:
- 10
- hierarchically ordering the index fields (201);
 - defining bins, each bin being associated to an index field and gathering one or more index field values for that index field;
 - creating indexes by concatenating one or more bins in compliance with the hierarchy of hierarchically ordered index fields to form a sequence of bins,
- 15
- an index comprising only one bin per index field;
 - hierarchically indexing the files as trees of indexes (300), each index having zero or more children indexes or and having at most one parent index, so that each child index comprises the same sequence of bins as the index of its parent index plus at least an additional bin associated to an additional
- 20
- index field;
 - providing each index with a data container (325, 335) configured to store statistical data, so that each data container is indexed and is directly addressable from within the files hierarchically indexed;
 - receiving (436) one or more input files (434) comprised of raw data and
- 25
- updating the data containers with the input files (434), which includes for each input file using one or more processor arranged for performing the steps of:
 - identifying and extracting from the raw data at least one attribute to be analyzed through statistics and one or more input file parameters characterizing the attribute;

- creating at least an individual record (620) from the input file (434), each individual record (620) comprising at least an attribute and the one or more input file parameters characterizing said attribute;
 - associating each input file parameter with at least an index field;
 - 5 – establishing a correspondence between each input file parameter and a bin of the at least one index field associated to that input file parameter;
 - identifying data containers indexed with the one or more bins that all correspond to an input file parameter of said individual record (620);
 - 10 – incrementally updating the identified data containers (325, 335) with the at least one attribute of said individual records (620) to obtain statistical data describing the attribute.
2. The method according to claim 1 comprising an additional step of receiving a query and for each query using one or more processors programmed for:
- 15 – identifying in the query one or more query field and at least one query field value associated to the query field;
 - establishing a correspondence between each query field and an index field;
 - establishing a correspondence between each query field value and a bin of each index field corresponding to said one or more identified query field defining thereby bins corresponding to the query;
 - 20 – searching and identifying a relevant data container indexed with an index that comprises the bins corresponding to the query;
 - retrieving and sending to the user the statistical data of the identified relevant container.
3. The method according to any one of the preceding claims wherein statistical
- 25 data relate to a frequency distribution of the attribute or measures of a central tendency of the distribution of the attribute or measures of the dispersion of the distribution of the attribute.
4. The method according to the preceding claim wherein statistical data are based on statistical data related to prices actually paid by travelers for transportation services.

5. The method according to any one of the preceding claims wherein index fields are at least one of the following: city of origin, country of origin, geographical area of origin, city of destination, country of destination, geographical area of destination, cultural destination, sport activity, gastronomy, wildlife observation, entertainment, departure date by exact departure date, departure period by month, departure period by week, return date by exact date, return period by month, return period by week, duration of the trip after arrival, advance purchase category.
6. The method according to any one of the preceding claims wherein at least an index having no parent index is created by concatenating at least two bins.
7. The method according to the preceding claim wherein an index that has no parent index includes a bin associated to an index field related to the city of origin, a bin associated to an index field related to the duration of the trip before departure and a bin associated to an index field related to the week of departure of a flight.
8. The method according to the preceding claim wherein a bin gathers some index field values related to one of the following: cities of origin, cities of departure, departure dates, departure weeks, return dates, return weeks, number of days for trip duration, number of days for advance purchase.
9. The method according to any one of claims 1 to 5 wherein at least an index having no parent index comprises only one bin.
10. The method according to any one of the preceding claims wherein an input file is an electronic ticket of a transportation service and comprises at least a coupon (434) embodying all information for a segment of the transportation.
11. The method according to the preceding claim wherein the attribute is the price actually paid for the ticket or the coupon and the input file parameters describe at least one of the following fields: city or origin, city of departure, departure date, return date, departure date, return date, trip duration, number or days between booking and departure.
12. The method according to the preceding claim wherein a query field relates to at last one of the following fields: city of origin, country of origin, geographical area of origin, city of destination, country of destination, geographical area of destination, cultural destination, sport destination, gastronomy, wildlife observation, entertainment, departure date by exact departure date, departure period by month, departure period

by week, return date by exact date, return period by month, return period by week, duration of the trip after arrival, number of days between booking and departure.

13. The method according to claim 1 wherein a single index field value can be associated to a plurality of index fields.

5 14. The method according to the preceding claim wherein an index field value that describes a city of destination is associated to at least two of the following index fields: city of origin, country of origin, geographical area of origin, cultural destination, sport activity, gastronomy, wildlife observation, entertainment.

10 15. The method according to any one of the preceding claims wherein creating individual records (620) includes the steps of: assigning a date or week or month or semester or year to each individual record, the date corresponding to the reception of the input file; storing the individual record (620) in data storage means.

15 16. The method according to the preceding claim wherein updating the data containers with the individual records (620) includes the steps of: creating batches of individual records (620) having the same dates or having same date periods; updating the data containers by batch of individual records (620).

20 17. The method according to any one of the two preceding claims comprising the following steps for maintaining accurate statistical data: reading the dates of stored individual records (620); identifying prior individual records (620) assigned with a date older than a given date; locating through the input file parameters of these identified prior individual records (620) the data containers that were updated with these identified prior individual records (620); updating the located data container by deleting these identified prior individual records (620).

25 18. The method according to any one of the preceding claims wherein incrementally updating the identified data containers (325, 335) comprises the step of incrementally computing updated statistical data from previously computed statistical data and from the at least one attribute of said individual records (620).

30 19. The method according to claim 1 wherein a data container is identified for an update by an input file in case said data container is indexed with a number of bins that is equal or inferior to the number of input file parameters of said input file and provided each of these one or more bins correspond to an input file parameter of said input file.

20. The method according to claim 1 wherein in case an input file parameter of an individual record (620) is associated to an existing index field but does not correspond to any bin of this associated existing index field, then creating an additional bin for said associated existing index field; creating an additional data container indexed with said additional bin; updating the additional data container with said individual record.

21. The method according to claim 1 wherein in case each input file parameter of an individual record (620) corresponds to a bin but there is a combination of one or more bins corresponding to said input file parameters that does not correspond to an index, then creating an additional data container indexed with this combination of one or more corresponding bins; updating the additional data container with said individual record.

22. The method according to any one of the preceding claims wherein all data stored in data containers are stored under the form of flat files.

23. The method according to claim 1 comprising the step of creating indexes through forming a sequence of one or more bins, at least one bin of this sequence gathering all possible index field values accepted by the index field associated to said at least one bin.

24. A system for providing statistical data from a data warehouse (410) comprising processing means characterized in that the processing means are arranged to perform the following steps of:

- defining a plurality of index fields (114) each index field accepting a plurality of index field values;
- creating a plurality of files (432) and hierarchically indexing the files as trees of indexes (300) which includes, for each tree, the steps of:
 - hierarchically ordering the index fields (201);
 - defining bins, each bin being associated to an index field and gathering one or more index field values for that index field;
 - creating indexes by concatenating one or more bins in compliance with the hierarchy of hierarchically ordered index fields to form a sequence of bins, an index comprising only one bin per index field;

- 5 – hierarchically indexing the files as trees of indexes (300), each index having zero or more children indexes or and having at most one parent index, so that each child index comprises the same sequence of bins as the index of its parent index plus at least an additional bin associated to an additional index field;
- providing each index with a data container (325, 335) configured to store statistical data, so that each data container is indexed and is directly addressable from within the files hierarchically indexed;
- 10 • receiving (436) one or more input files (434) comprised of raw data and updating the data containers with the input files (434), which includes for each input file using one or more processor arranged for performing the steps of:
- identifying and extracting from the raw data at least one attribute to be analyzed through statistics and one or more input file parameters characterizing the attribute;
- 15 – creating at least an individual record (620) from the input file (434), each individual record (620) comprising at least an attribute and the one or more input file parameters characterizing said attribute;
- associating each input file parameter with at least an index field;
- establishing a correspondence between each input file parameter and a bin of the at least one index field associated to that input file parameter;
- 20 – identifying data containers indexed with the one or more bins that all correspond to an input file parameter of said individual record (620);
- incrementally updating the identified data containers (325, 335) with the at least one attribute of said individual records (620) to obtain statistical data
- 25 describing the attribute.
25. A computer program product stored on a non-transitory computer readable storage medium, comprising computer readable code means instructed for causing at least one microprocessor to operate the method for providing statistical data according to any one of claims 1 to 23.

CLAIMS

1. A method for providing statistical data from a data warehouse characterized in that it comprises the following steps performed with at least one data processor:

- 5
 - defining a plurality of index fields each index field accepting a plurality of index field values;
 - creating a plurality of files and hierarchically indexing the files as trees of indexes which includes, for each tree, the steps of:
 - hierarchically ordering the index fields;
- 10
 - defining bins, each bin being associated to an index field and gathering one or more index field values for that index field;
 - creating indexes by concatenating one or more bins in compliance with the hierarchy of hierarchically ordered index fields to form a sequence of bins, an index comprising only one bin per index field;
- 15
 - hierarchically indexing the files as trees of indexes, each index having zero or more children indexes or and having at most one parent index, so that each child index comprises the same sequence of bins as the index of its parent index plus at least an additional bin associated to an additional index field;
- 20
 - providing each index with a data container configured to store statistical data, so that each data container is indexed and is directly addressable from within the files hierarchically indexed;
 - receiving one or more input files comprised of raw data and updating the data containers with the input files, which includes for each input file using
- 25
 - one or more processor arranged for performing the steps of:
 - identifying and extracting from the raw data at least one attribute to be analyzed through statistics and one or more input file parameters characterizing the attribute;

- creating at least an individual record from the input file, each individual record comprising at least an attribute and the one or more input file parameters characterizing said attribute;
- associating each input file parameter with at least an index field;
- 5 - establishing a correspondence between each input file parameter and a bin of the at least one index field associated to that input file parameter;
- identifying data containers indexed with the one or more bins that all correspond to an input file parameter of said individual record;
- 10 - incrementally updating the identified data containers with the at least one attribute of said individual records to obtain statistical data describing the attribute.

2. The method according to claim 1, further comprising an additional step of receiving a query and for each query using one or more processors programmed
15 for:

- identifying in the query one or more query field and at least one query field value associated to the query field;
- establishing a correspondence between each query field and an index field;
- 20 - establishing a correspondence between each query field value and a bin of each index field corresponding to said one or more identified query field defining thereby bins corresponding to the query;
- searching and identifying a relevant data container indexed with an index that comprises the bins corresponding to the query;
- 25 - retrieving and sending to the user the statistical data of the identified relevant container.

3. The method according to any one of the preceding claims, wherein statistical data relate to a frequency distribution of the attribute or measures of a

central tendency of the distribution of the attribute or measures of the dispersion of the distribution of the attribute.

4. The method according to the preceding claim, wherein statistical data are based on statistical data related to prices actually paid by travelers for transportation services.
5. The method according to any one of the preceding claims, wherein index fields are at least one of the following: city of origin, country of origin, geographical area of origin, city of destination, country of destination, geographical area of destination, cultural destination, sport activity, gastronomy, wildlife observation, entertainment, departure date by exact departure date, departure period by month, departure period by week, return date by exact date, return period by month, return period by week, duration of the trip after arrival, advance purchase category.
6. The method according to any one of the preceding claims, wherein at least an index having no parent index is created by concatenating at least two bins.
7. The method according to the preceding claim, wherein an index that has no parent index includes a bin associated to an index field related to the city of origin, a bin associated to an index field related to the duration of the trip before departure and a bin associated to an index field related to the week of departure of a flight.
8. The method according to the preceding claim, wherein a bin gathers some index field values related to one of the following: cities of origin, cities of departure, departure dates, departure weeks, return dates, return weeks, number of days for trip duration, number of days for advance purchase.
9. The method according to any one of claims 1 to 5, wherein at least an index having no parent index comprises only one bin.

10. The method according to any one of the preceding claims, wherein an input file is an electronic ticket of a transportation service and comprises at least a coupon embodying all information for a segment of the transportation.
11. The method according to the preceding claim, wherein the attribute is the price actually paid for the ticket or the coupon and the input file parameters describe at least one of the following fields: city or origin, city of departure, departure date, return date, departure date, return date, trip duration, number or days between booking and departure.
12. The method according to the preceding claim, wherein a query field relates to at least one of the following fields: city of origin, country of origin, geographical area of origin, city of destination, country of destination, geographical area of destination, cultural destination, sport destination, gastronomy, wildlife observation, entertainment, departure date by exact departure date, departure period by month, departure period by week, return date by exact date, return period by month, return period by week, duration of the trip after arrival, number or days between booking and departure.
13. The method according to claim 1, wherein a single index field value can be associated to a plurality of index fields.
14. The method according to the preceding claim, wherein an index field value that describes a city of destination is associated to at least two of the following index fields: city of origin, country of origin, geographical area of origin, cultural destination, sport activity, gastronomy, wildlife observation, entertainment.
15. The method according to any one of the preceding claims, wherein creating individual records includes the steps of: assigning a date or week or month or semester or year to each individual record, the date corresponding to the reception of the input file; storing the individual record (620) in data storage means.

16. The method according to the preceding claim, wherein updating the data containers with the individual records includes the steps of: creating batches of individual records having the same dates or having same date periods; updating the data containers by batch of individual records.

5 17. The method according to any one of the two preceding claims, further comprising the following steps for maintaining accurate statistical data: reading the dates of stored individual records; identifying prior individual records assigned with a date older than a given date; locating through the input file parameters of these identified prior individual records the data containers that were updated with
10 these identified prior individual records; updating the located data container by deleting these identified prior individual records.

18. The method according to any one of the preceding claims, wherein incrementally updating the identified data containers comprises the step of incrementally computing updated statistical data from previously computed
15 statistical data and from the at least one attribute of said individual records.

19. The method according to claim 1, wherein a data container is identified for an update by an input file in case said data container is indexed with a number of bins that is equal or inferior to the number of input file parameters of said input file and provided each of these one or more bins correspond to an input file
20 parameter of said input file.

20. The method according to claim 1, wherein in case an input file parameter of an individual record is associated to an existing index field but does not correspond to any bin of this associated existing index field, then creating an additional bin for said associated existing index field; creating an additional data
25 container indexed with said additional bin; updating the additional data container with said individual record.

21. The method according to claim 1, wherein in case each input file parameter of an individual record corresponds to a bin but there is a combination of one or more bins corresponding to said input file parameters that does not

correspond to an index, then creating an additional data container indexed with this combination of one or more corresponding bins; updating the additional data container with said individual record.

22. The method according to any one of the preceding claims, wherein all data
5 stored in data containers are stored under the form of flat files.

23. The method according to claim 1, further comprising the step of creating indexes through forming a sequence of one or more bins, at least one bin of this sequence gathering all possible index field values accepted by the index field associated to said at least one bin.

10 24. A system for providing statistical data from a data warehouse comprising processing means characterized in that the processing means are arranged to perform the following steps of:

- defining a plurality of index fields each index field accepting a plurality of index field values;
- 15 • creating a plurality of files and hierarchically indexing the files as trees of indexes which includes, for each tree, the steps of:
 - hierarchically ordering the index fields;
 - defining bins, each bin being associated to an index field and gathering one or more index field values for that index field;
 - 20 - creating indexes by concatenating one or more bins in compliance with the hierarchy of hierarchically ordered index fields to form a sequence of bins, an index comprising only one bin per index field;
 - hierarchically indexing the files as trees of indexes, each index having zero or more children indexes or and having at most one parent index, so that each child index comprises the same
25 sequence of bins as the index of its parent index plus at least an additional bin associated to an additional index field;

- 5

 - providing each index with a data container configured to store statistical data, so that each data container is indexed and is directly addressable from within the files hierarchically indexed;
 - receiving one or more input files comprised of raw data and updating the data containers with the input files, which includes for each input file using one or more processor arranged for performing the steps of:
 - identifying and extracting from the raw data at least one attribute to be analyzed through statistics and one or more input file parameters characterizing the attribute;
 - 10 - creating at least an individual record from the input file, each individual record comprising at least an attribute and the one or more input file parameters characterizing said attribute;
 - associating each input file parameter with at least an index field;
 - establishing a correspondence between each input file parameter and a bin of the at least one index field associated to that input file parameter;
 - 15 - identifying data containers indexed with the one or more bins that all correspond to an input file parameter of said individual record;
 - incrementally updating the identified data containers with the at least one attribute of said individual records to obtain statistical data describing the attribute.
 - 20

25. A computer program product stored on a non-transitory computer readable storage medium, comprising computer readable code means instructed for causing at least one microprocessor to operate the method for providing statistical data according to any one of claims 1 to 23.

25