(54) **Title:** DATA PROCESSING AND ANALYSIS BY COMPONENT ANALYSIS CONFIGURATOR
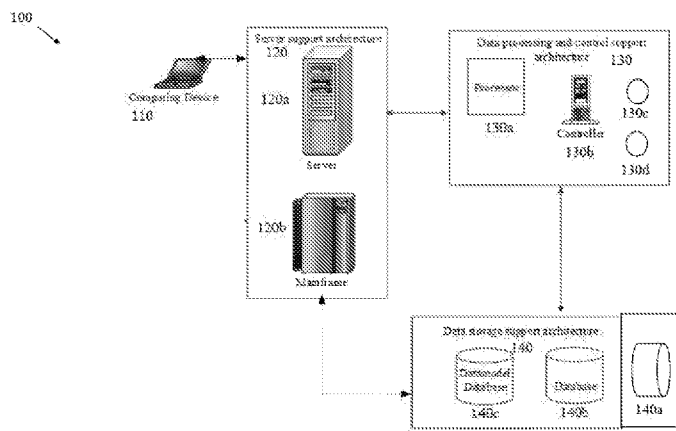

Fig. 1

(57) **Abstract:** The present invention provides a system for data processing and analysis by Component Analysis Configurator (CAC) for report generation. The system includes an electronic user interface configured for receiving a set of information from a user, a data lake storing a plurality of data attributes, an abstraction base data set layer embedded over an underlining data structure such that the plurality of data attributes and the data structure are identified using functional objects thereby enabling the user to define a plurality of components wherein the base data set is identified from the data lake based on the information, a processor configured for processing the information using at least one decision tree data model to extract the plurality of data attributes from the data lake. The system also includes a controller coupled to the processor and encoded with instructions enabling the controller to function as a bot for controlling multiple components of the system for data processing and analysis; a definition engine configured for capturing a metadata associated with the plurality of attributes and storing the metadata in a metadata database; and an executing engine configured for generating at least one query based on the metadata and extracting a plurality of reporting data based on the query from the data lake, wherein at least on report is generated based on the extracted data.

AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:
— *of inventorship (Rule 4.17(iv))*

Published:
— *with international search report (Art. 21(3))*
— *in black and white; the international application as filed contained color or greyscale and is available for download from PATENTSCOPE*

DATA PROCESSING AND ANALYSIS BY COMPONENT ANALYSIS CONFIGURATOR

FIELD OF THE INVENTION

The present invention relates to data processing. More particularly, the invention relates to system and method for data processing and analysis for reporting in financial institutions.

BACKGROUND

The reporting capability of any database management system is extremely essential. Reports generally present results in user friendly formats, such as graphs, tables, crosstabs, or forms. There has been a continuing need for analytical applications capable of providing asset and liability management capabilities through reporting. Finance, treasury, and risk groups of a bank requires to have a group wide view of a financial measure and then drill down across different levels of hierarchies to view contributions. For example, the treasury user would like to view a net interest income for the bank. The user would then like to contribution of each line of business and then for each country within the line of business and then by customer segment within each country. Another group of users within treasury department may want to view the same net interest income for a banking book of records for a country. Requirements from risk team may also be similar in nature. They may like to view the Risk Weighted Assets for a line of business.

However, the existing reporting tools expect the users to have good knowledge of SQL (Structured Query Language) and underlying data structures. Hence, users with no programming skills have to rely on developers to configure reports before the reports can be used. Entire development process has to be followed even if there is minor change to existing report or when ad hoc reports have to be generated. Further, SQL used to extract and visualize data can have high latency and there is no provision to pre-run the report and make it available for analysis.

In view of the above, there exists a need of improved systems and methods that overcome the shortcomings associated with existing technologies and prior arts.

SUMMARY OF THE INVENTION

Accordingly, the present invention provides a method for data processing and analysis by Component Analysis Configurator (CAC) for report generation. The method includes receiving a set of information from a user through an electronic user interface, identifying a

5   base data set from a data lake based on the information wherein the data set is an abstraction layer embedded over an underlining data structure such that a plurality of data attributes and the data structure are identified using functional objects thereby enabling the user to define a plurality of components, processing the information using at least one decision tree data model to extract the plurality of data attributes from the data lake, capturing by a definition

10   engine, a metadata associated with the plurality of attributes and storing the metadata in a metadata database, generating at least one query based on the metadata and extracting a plurality of reporting data based on the query from the data lake; and generating at least on report based on the extracted data.

15   In an embodiment, the present invention provides a system for data processing and analysis by Component Analysis Configurator (CAC) for report generation. The data processing and analysis occur in real-time. The system includes an electronic user interface configured for receiving a set of information from a user, a data lake storing a plurality of data attributes, an abstraction base data set layer embedded over an underlining data structure such that the

20   plurality of data attributes and the data structure are identified using functional objects thereby enabling the user to define a plurality of components wherein the base data set is identified from the data lake based on the information, a processor configured for processing the information using at least one decision tree data model to extract the plurality of data attributes from the data lake. The system also includes a controller coupled to the processor

25   and encoded with instructions enabling the controller to function as a bot for controlling multiple components of the system for data processing and analysis; a definition engine configured for capturing a metadata associated with the plurality of attributes and storing the metadata in a metadata database; and an executing engine configured for generating at least one query based on the metadata and extracting a plurality of reporting data based on the

30   query from the data lake, wherein at least on report is generated based on the extracted data.

In an embodiment, the present invention provides a computer-readable non-transitory storage medium storing executable program instructions for data processing and analysis by

Component Analysis Configurator (CAC) which when executed by a computer cause the computer to perform operations as described above.

In an advantageous aspect, the system and method of the present invention functions with roll-up hierarchy i.e., the system has the ability to look at a set of data starting at one level and further drilling through multiple lower levels of dimensions. The system is configured to use attributes of some dimensions to further filter the data and select measures from the base data set for plotting in a graph or in a report. Further, the invention includes the capability to set a lot of filters to the base data where the filters could be value filters or date filters. The invention includes ability to do point-in-time reporting, forecasting or past date range analysis.

In another advantageous aspect, the system of the present invention computes complicated results from the base set of measures available in the base data set and also pre-execute some queries and store the results in a form that is easily reportable on demand. Also, the system provides the output of the analysis in various formats.

DESCRIPTION OF THE DRAWINGS

Fig. 1 shows an architecture diagram of the system configured for data processing and analysis by Component Analysis Configurator (CAC) in accordance with an embodiment of the present invention.

Fig. 1a shows a structural block diagram of the system with constituting components in accordance with an embodiment of the present invention.

Fig. 1b shows an application architecture block diagram depicting high level flow of information in accordance with an embodiment of the present invention.

Fig. 2 shows a flow diagram of a method of data processing and analysis in accordance with an embodiment of the present invention.

Fig. 2a shows a high-level process flow diagram of the data processing and analysis of the present invention in accordance with an embodiment of the present invention.

DESCRIPTION OF THE INVENTION

Various embodiments of the present invention provide system and method for data processing and analysis by Component Analysis Configurator (CAC). The following description provides specific details of certain embodiments of the invention illustrated in the drawings to provide a thorough understanding of those embodiments. It should be recognized, however, that the present invention can be reflected in additional embodiments and the invention may be practiced without some of the details in the following description.

The various embodiments including the example embodiments are described more fully with reference to the accompanying drawings, in which the various embodiments of the invention are shown. The invention may, however, be embodied in different forms and should not be construed as limited to the embodiments set forth herein. Rather, these embodiments are provided so that this disclosure is thorough and complete, and fully conveys the scope of the invention to those skilled in the art. In the drawings, the sizes of components may be exaggerated for clarity.

It should be understood that when an element or layer is referred to as being "on" "connected to" or "coupled to" another element or layer, it can be directly on, connected to, or coupled to the other element or layer or intervening elements or layers that may be present. As used herein, the term "and/or" includes any and all combinations of one or more of the associated listed items.

Spatially relative terms, such as "data structure," "data attributes," "base data set layer" and the like, may be used herein for ease of description to describe one element or feature's relationship to another element(s) or feature(s) as illustrated in the figures. It should be understood that the spatially relative terms are intended to encompass different orientations of the structure in use or operation in addition to the orientation depicted in the figures.

It should be noted that the term "dimensions" as used herein means a collection of related reference data; the term "filters" as used herein means restriction(s) applied on data set to

retrieve a sub set of data; the term "measures" as used herein means combination of mathematical operation(s) applied on one or more attributes to generate a financial formula; and the term "hierarchy" as used herein means a level at which data is aggregated for visualization.

Embodiments described herein refer to plan views and/or cross-sectional views by way of ideal schematic views. Accordingly, the views may be modified depending on simplistic assembling or manufacturing technologies and/or tolerances. Therefore, example embodiments are not limited to those shown in the views but include modifications in configurations formed on basis of assembling process. Therefore, regions or regions of elements exemplified in the figures have schematic properties and shapes, and do not limit the various embodiments including the example embodiments.

The subject matter of example embodiments, as disclosed herein, is described with specificity to meet statutory requirements. However, the description itself is not intended to limit the scope of this patent. Rather, the inventors have contemplated that the claimed subject matter might also be embodied in other ways, to include different features or combinations of features similar to the ones described in this document, in conjunction with other technologies. Generally, the various embodiments including the example embodiments relate to system and method for Component Analysis Configurator (CAC).

Referring to Fig. 1, a system architecture 100 for data processing and analysis by Component Analysis Configurator (CAC) for report generation is shown in accordance with an embodiment of the present invention. The system 100 include at least one computing device 110, a server support architecture 120, a data processing and control support architecture/mechanism 130, a data storage support architecture 140. The server support architecture may include server 120a and mainframe 120b. The data processing and control support architecture/mechanism 130 may include a processor 130a, a controller 130b, a definition engine 130c and an execution engine 130d. The data storage support architecture 140 may include a data lake 140a, a database 140b and a data model database 140c.

In an embodiment, the system includes an electronic user interface configured for receiving a set of information from a user. The data lake 140a is configured for storing a plurality of data

6

attributes. The data processing support architecture 130 includes an abstraction base data set layer embedded over an underlining data structure such that the plurality of data attributes and the data structure are identified using functional objects thereby enabling the user to define a plurality of components where the base data set is identified from the data lake 140a

5   based on the information. The processor 130a is configured for processing the information using at least one decision tree data model to extract the plurality of data attributes from the data lake 140a.

In an embodiment, the electronic user interface functions to create and maintain the plurality

10   of components, search and view existing attributes, setup chained relationship between components and/or attributes, and/or trigger on demand execution and viewing execution output / reports.

In an embodiment, the system 100 also includes a controller 130b coupled to the processor

15   130a and encoded with instructions enabling the controller to function as a bot for controlling multiple components of the system for data processing and analysis.

In an embodiment, the definition engine 130c is configured for capturing a metadata associated with the plurality of attributes and storing the metadata in a metadata database.

20

In an embodiment, an execution engine 130d is configured for generating at least one query based on the metadata and extracting a plurality of reporting data based on the query from the data lake, where at least on report is generated based on the extracted data.

25   The execution engine is a layer that consists of the core engine. It uses the definitions created by the user and extracts the data into format, which can then be used for reporting.

In an embodiment, the execution engine is a black box application which is invoked every time an execution request is triggered. The execution engine also functions as load

30   component definitions, generate SQL representation of component definitions, execute queries, and/or generate report based on a selected format.

In one embodiment, the system includes a reporting layer consisting of pre-defined reporting templates that combine with the extracted data to generate different types of visualization.

In another embodiment, the invention includes a scheduler configured to pre-execute components when required thereby enabling the report to be readily available. The scheduler is configured to schedule frequently used components and preprocess the execution in a batch mode thereby making it ready for analysis.

In an embodiment, the scheduler is configured to be invoked by an external scheduler.

In an example embodiment the server 120a may include electronic circuitry for enabling execution of various steps by the processor. The electronic circuity may have various elements including but not limited to a plurality of Arithmetic Logic Units (ALU) and Floating-Point Units (FPU), and/or the equivalents thereof. The ALU enables processing of binary integers to assist in generating a plurality of data models to be stored in the data model database 140c and associated with entity information to determine the data attributes from the data lake. In an example embodiment, the server electronic circuitry includes at least one arithmetic logic unit, floating point units (FPU), other processors, memory, storage devices, high-speed interfaces connected through buses for connecting to memory and high-speed expansion ports, and a low speed interface connecting to low speed bus and storage device. Each of the components of the electronic circuitry, are interconnected using various busses, and may be mounted on a common motherboard or in other manners as appropriate. The processor can process instructions for execution within the server 120a, including instructions stored in the memory or on the storage devices to display graphical information for a GUI on an external input/output device, such as display coupled to high speed interface. In other implementations, multiple processors and/or multiple busses may be used, as appropriate, along with multiple memories and types of memory. Also, multiple servers may be connected, with each server providing portions of the necessary operations (e.g., as a server bank, a group of blade servers, or a multi-processor system).

In an embodiment, the definition engine 130c is a layer that includes user interfaces used to identify the base data set, built components and trigger dynamic executions.

The processor 130a may communicate with a user through a control interface and display interface coupled to a display. The display may be, for example, a TFT LCD (Thin-Film-Transistor Liquid Crystal Display) or an OLED (Organic Light Emitting Diode) display, or other appropriate display technology. The display interface may comprise appropriate circuitry for driving the display to present graphical and other information to an entity/user. The control interface may receive commands from a user and convert them for submission to the processor. In addition, an external interface may be provided in communication with processor 130a, so as to enable near area communication of device with other devices. External interface may be suitable, for example, for wired communication in some implementations, or for wireless communication in other implementations, and multiple interfaces may also be used.

The data storage support architecture 140 may include memory units that may be a volatile, a non-volatile memory or memory may also be another form of computer-readable medium, such as a magnetic or optical disk.

The data storage 140 may also include storage device capable of providing mass storage. In one implementation, the storage device may be or contain a computer-readable medium, such as a floppy disk device, a hard disk device, an optical disk device, a tape device, a flash memory or other similar solid state memory device, or an array of devices, including devices in a storage area network or other configurations.

Referring to Fig. 1a, a structural architecture 100a of the system with components is shown in accordance with an embodiment of the present invention.

In an embodiment, the system components broadly include data storage 140 and processing layer 150. The processing layer includes web services interface 152, a user interface 154, the execution engine 130d, over the top of data access layer 156. The data storage 140 includes the data lake 140a and a processing layer data store 140d.

Referring to Fig. 1b, an application architecture block diagram 100b depicting high level flow of information is shown in accordance with an embodiment of the present invention. The application includes a maintenance layer 160, an engine layer 170, a reporting layer 180 and a batch layer 190. The information flow includes defining of component, identifying base data set and selecting measure and formula. The information flow also includes selecting filter, selecting dimension for roll up hierarchy, execution and viewing of reports where batch

triggers execution of CAC (Component Analysis Configurator) which is executed by the engine layer based on certain captured parameters for generation and viewing of reports. The batch layer also enables configuration of scheduler and certain parameters.

Referring to Fig. 2, a flowchart 200 depicting a data processing and analysis method for report generation is provided in accordance with an embodiment of the present invention. The method comprises the steps of (S210) receiving a set of information from a user through an electronic user interface; (S220) identifying a base data set from a data lake based on the information wherein the data set is an abstraction layer embedded over an underlining data structure such that a plurality of data attributes and the data structure are identified using functional objects thereby enabling the user to define a plurality of components; (S230) processing the information using at least one decision tree data model to extract the plurality of data attributes from the data lake; (S240) capturing by a definition engine, a metadata associated with the plurality of attributes and storing the metadata in a metadata database; (S250) generating at least one query based on the metadata and extracting a plurality of reporting data based on the query from the data lake; and (S260) generating at least one report based on the extracted data.

In an embodiment, Component Analysis Configurator (CAC) is designed to be used by the following types of users in a bank, for the roles or purposes as listed below:

1. Component Definer

- This role can be assigned to any business user in the bank who is trained to use CAC designer to generate plurality of components. The user performing this role should have a good understanding of the type of analysis needed for strategic reporting.

- A user attached to this role would be able to add component types, define the requirements of component analysis and save the definition for execution. The user will also be able to modify, delete and view the component definitions that he/she has created.

2. Reporting user

- This role can be assigned to any business user in the bank who is expected to perform various level of analysis and reports. The user performing this role should have a good understanding of the type of analysis needed for strategic reporting.

- A user attached to this role would be able to generate ad-hoc instances of pre-defined reports for a pre-defined component by-passing the dynamic parameters required for that component.

3. Administrator

- This role can be assigned to any IT (Information Technology) user or administrator whose primary job is to execute specific technical jobs to create data for further business use. The user performing this role should have good exposure to job schedulers and routine administration activities like creating menu creation and user creation.

- A user attached to this role would be able to schedule execution of pre-defined components by passing the dynamic parameters required for that component.

In an embodiment, Component Analysis Configurator (CAC) includes assumptions, namely:

1. Base Data Assumptions

CAC is typically used for generating reports that have multiple levels of drill-down enabled. This necessitates data at a lower granularity level to have maximum levels of roll-up. Account level data availability and references to a lot of key dimensions at the account level are assumed while creating this designer. Components can also be created from a higher aggregated level of data like GL data so on and so forth, but the advanced capabilities of CAC are best demonstrated when account level granularity of data is available.

If there is a need to create a component that has references to dimensions that are not referenced at the account level, then the component definer would need to define complicated joins while creating a base data definition OR enable reference to that dimension in the account level base data.

4. Derived Data Assumptions

For performance reasons, some data may need to be derived/calculated upfront while it is extracted, transformed, and loaded into the warehouse. Some examples of this type of data include Month to Date (MTD), Quarter to Date (QTD), Year to Date (YTD) values for all amount fields available at the account level – both for the fiscal year of the account as well as the calendar year. Similarly, availability of balances in

multiple currencies – accounting currency, reporting currency, base currency etc. enables more varied analyses of measures.

5. Component Re-use Assumptions

Like any resource that is freely available, a risk exists that components may be defined by various users for their specific purpose, without checking the inventory of components that are already available. In some cases, extending a few parameters in an existing component would easily meet the needs of another user. An assumption has been made that some sort of governance would be exercised by the bank while creating new components. It is assumed that a large number of components will be reused by multiple users for their varied reporting needs.

At a later point in time, the system will provide an intelligent prompt to component definers when a newly defined component seems similar to an existing component. This will act as a useful hint to cross-check existing components before defining a new one.

4. Report Re-use Assumptions

Like any resource that is freely available, a risk exists that reports may be defined by various users for their specific purpose, using the same component, without checking the inventory of reports that are already available. In some cases, extending a few parameters in an existing report would easily meet the needs of another user. An assumption has been made that some sort of governance would be exercised by the bank while creating new reports. It is assumed that a large number of reports will be reused by multiple users for their varied reporting needs.

5. Dynamic Parameter Assumptions

CAC provides a feature to define dynamic parameters to accept constant values for any filter condition. There is an assumption that this feature will be used extensively during component creation. Thus no filter condition will have hard-coded constant values. Hard-coding values will prevent the component from being re-used for another analysis and results in proliferation of components and resulting execution load on the system.

6. Performance Assumptions

CAC provides a feature to schedule jobs to execute components with some dynamic parameters to pre-create aggregated reporting data. This is primarily aimed at enabling complex analysis on the click of a button. There is an assumption that this feature will be used extensively for canned reports using pre-defined components.

If there are performance issues encountered while executing the component analysis definition, a natural solution would be to create more derived measures during ETL and reference them in the component. This would solve many performance related issues.

7. End-use Assumptions

CAC is primarily meant to be used for strategic reporting at multiple levels of management reporting. The tool should not be used for operational reporting of millions of rows of data. The most optimal use of CAC is when you need to analyse one measure or a set of measures at various levels of aggregation using pre-defined dimensions. The deeper the roll-up hierarchy defined, the more optimal is the use of CAC as a reporting tool.

In an embodiment, the present invention provides a method for data processing and analysis involving Component Analysis Configurator (CAC). The method includes the following steps for configuration of plurality of components and/or analyzing reports:

a) Creating plurality of components and identifying base data set

   - This is the most critical step in the process, as the CAC provides a series of questions to a user (e.g., a component definer) and applies a "decision tree" algorithm to identify a base data set that has to be displayed in a display/screen. The user would be able to create a plurality of components from the data available in the base data set.

b) Selecting measures and specifying the formula for analysis (if any)

   - This step identifies a base output that needs to be reported. For example, if this is an NII (Net Interest Income) Component Analysis, then the base output is Net Interest Income. The relevant measure that has this output needs to be selected. If

some computation needs to be done on a base measure, then a formula can be specified to generate the reporting output.

c) Selecting filters to fine-tune the data requirement

-   This step is necessary to fine-tune the amount of data or range of data that needs to be reported. Most component analysis focuses on specific areas and needs filters to specify the exact area of analysis. Examples of filters include and are not limited to: date filters for specifying a range of dates for analysis, Line of Business (LoB) filters, country filters, scenario filters, and so on. Basically, any dimension, dimensional attribute, or existing measure in at least one base table can be used to filter out certain set of rows and retain only the required rows for analysis.

d) Selecting dimensions for roll up hierarchy

-   This step specifies the exact sequence of roll-up for drill-down that needs to be reported. For example, there could be a base data filter for a country, and additional roll-up hierarchy of Country->LoB->Product Type->Industry->Loan purpose etc. When such a hierarchy is specified, CAC generates aggregated numbers and percentages for the measures specified at each level of drill-down for easy and quick reporting.

e) Executing CAC on-demand

-   This step enables the users (e.g., business users) to execute CAC on demand for a specific component and pass dynamic parameters to it. This is typically done when an ad-hoc analysis needs to be done. Most component executions will be done in batch for periodic reporting needs. Once executed, users can generate an ad-hoc report using the next step or export the generated data into any of the available output formats.

f) Generating ad-hoc report

    - This step enables users (e.g., business users) to generate a canned report based on on-demand execution of CAC in the previous step.

14

g) Scheduling an execution

-   This step enables users (e.g., administrators) to schedule the running of a scheduler for specific components with a specific set of dynamic parameters at a specified frequency in advance. This is particularly useful for generating canned reports on the click of a button (for example, for generating of executive summaries and strategic reports on a daily or on demand basis).

h) Chaining plurality of components

-   This step enables users (e.g., administrators) to schedule the running of the scheduler for specific components with a specific set of dynamic parameters at a specified frequency. This is particularly useful for generating canned reports as with step g).

-   In this step, the invention can perform addition or subtraction of the output of multiple components and generate a report on the new result. This feature is useful in the some scenarios, for example: (i) NII has to be computed for both live and mature accounts, or (ii) group level information have to aggregated from individual country level information.

In an embodiment, the plurality of components can be created using a base data set, which is collection of facts and dimensions visually represented using logical names. The user needs to answer the questionnaire and the CAC applies a decision tree algorithm to identify the base data set.

In an embodiment, once the measures, filters, and/or hierarchies are defined, the user will be able to define the required format/template for the report. The format/template can be in the form of sunburst or tabular.

In embodiment, steps a), e)-g) can be performed by a processor, steps b)-d) and h) can be performed in a screen or display.

In an embodiment, the steps above can run on any Unix platform. Also, the steps are tested and compatible with Red Hat Linux, Sun Solaris or Windows operating system; Oracle 12c database; Tomcat or Weblogic Application Server; and/or Google Chrome or Internet Explorer.

In an embodiment, the invention includes a plurality of queries is executed and the plurality of reporting data is aggregated to generate a visualization report for time ranges.

In an embodiment, the invention includes connecting an output of one report to another report by using mathematical formula. This is also known as chaining where data from multiple components are combined using mathematically formulas. The mathematical formula can be a financial formula.

In an embodiment the method includes scheduling and pre-processing execution of frequently used reports.

In an embodiment, the method includes generating reports based on a report type selected by the user.

In an embodiment, the method includes generating actionable reports wherein the user selects data points in a first report, enters comments to record actions that are assigned to the user such that a second report when generated is compared with the first report based on the recorded actions.

In an embodiment, the components include measures, filters, hierarchies and extraction formats.

In an embodiment, the measures identify a set of financial parameters on which a drill down analysis is performed. A measure is generated by applying a mathematical operation on one or more attributes that are available for selection from a base data set.

In an embodiment, the filters enable filtering for the base data set used for analysis by identifying predicate conditions. The filters are configured to restrict the data set used for analysis. There are two types of filters:

1. Static filters

   - The value of these filters is defined at a component setup. For example, selecting a business vertical such as "Conventional Banking".

2. Dynamic Filters

   - These filters require input filter value(s) when a component is executed.

16

The invention is configured to be capable changing a static filter into a dynamic filter, and vice versa.

In an embodiment, the hierarchies identify specific aggregation layers at which measures are analyzed.

5   In an embodiment, the extraction format identifies the format in which reporting data is extracted for analysis.

In an embodiment, any attribute available in base data set are selected as measure, filter or hierarchy. Executed components are visualized using reports. CAC tool includes built in reporting templates that are highly configurable.

10   In some implementations, the machine learning data model is configured and trained to map attributes data and entity information with data processing rule to generate reports. For example, output generated over the model may provide an indication of whether a particular object or class of objects is present, and optionally user instructions. In some implementations, the machine learning model is configured and trained to map minimum data

15   for processing and report generation. Accordingly, in those implementations a single pass over a single machine learning model may be utilized to detect whether each of multiple objects is present. For example, output generated over the model may provide an indication of whether the attribute is available in the base data set.

In an advantageous aspect, the present invention enables a user without any prior

20   programming knowledge to configure reports without any assistance from developers. Further, chaining Components helps users to combine the output of one or components using mathematical operators (+ / -). Also, pre-scheduling executions enables execution of time-consuming queries and made available for analysis.

The present invention provides ability to define Banded Dimensions as Hierarchies.

25   The invention enables externalization of configuration (sub-totals, provide filters in reports, determine underlying customers / accounts) in the reporting template to allow the user to select-deselect required features.

In an example embodiment, the present invention provides the system and method that is used by any team (Risk / Finance) that has requirement to perform drill down analysis. The

30   system application is used by balance sheet management application to generate NII

Component Analysis report. The system is also used by a pricing application to compute the RWA for Customer and efficiency frontier to compute multiple risk ratio.

It must be apparent that different aspects of the description provided above may be implemented in many different forms of software, firmware, and hardware in the implementations illustrated in the figures. The actual software code or specialized control hardware used to implement these aspects is not limiting of the invention. Thus, the operation and behavior of these aspects were described without reference to the specific software code—it being understood that software and control hardware can be designed to implement these aspects based on the description herein.

Fig. 2a shows a high-level process flow diagram 200a of the data processing and analysis of the present invention in accordance with an embodiment of the present invention. The process flow is performed in sequence to generate drill down reports. The sequence includes identifying the base data set by answering a questionnaire. Then, defining component using the attributes available in the base data set takes place. If required, the sequence includes creating a new component by chaining one or more components. Also, the sequence includes selecting a reporting template and executing the component to view report. If required, the sequence includes schedule of the component so that it can be pre-executed and made ready for analysis.

Further, certain portions of the invention may be implemented as a "component" or "system" that performs one or more functions. These components/systems may include hardware, such as a processor, an ASIC (Application Specific Integrated Circuit), or a FPGA (Field Programmable Gate Array), or a combination of hardware and software.

The word "exemplary" is used herein to mean "serving as an example." Any embodiment or implementation described as "exemplary" is not necessarily to be construed as preferred or advantageous over other embodiments or implementations.

No element, act, or instruction used in the present application should be construed as critical or essential to the invention unless explicitly described as such. Also, as used herein, the article "a" and "one of" is intended to include one or more items. Further, the phrase "based on" is intended to mean "based, at least in part, on" unless explicitly stated otherwise.

Each of the above identified processes corresponds to a set of instructions for performing a function as described above. The above identified programs or sets of instructions need not

be implemented as separate software programs, procedures or modules, and thus various subsets of these modules may be combined or otherwise re-arranged in various embodiments. For example, embodiments may be constructed in which steps are performed in an order different than illustrated, steps are combined, or steps are performed simultaneously, even

5   though shown as sequential steps in illustrative embodiments. Also, the terminology used herein is for the purpose of description and should not be regarded as limiting. The use of "including," "comprising," "having," "containing" and variations thereof herein, is meant to encompass the items listed thereafter and equivalents thereof as well as additional items.

The above-described embodiments of the present invention may be implemented in any of

10   numerous ways. For example, the embodiments may be implemented using various combinations of hardware and software and communication protocol(s). Any standard communication or network protocol may be used and more than one protocol may be utilized. For the portion implemented in software, the software code may be executed on any suitable processor or collection of processors, whether provided in a single computer or distributed

15   among multiple computers. Such processors may be implemented as integrated circuits, with one or more processors in an integrated circuit component, or any other suitable circuitry. Further, it should be appreciated that a computer may be embodied in any of a number of forms, such as a rack-mounted computer, a desktop computer, a laptop computer, single board computer, micro-computer, or a tablet computer. Additionally, a computer may be

20   embedded in a device not generally regarded as a computer but with suitable processing capabilities, including a Personal Digital Assistant (PDA), a smart phone or any other suitable portable or fixed electronic device.

Also, the various methods or processes outlined herein may be coded as software that is executable on one or more processors that employ any one of a variety of operating systems

25   or platforms. Additionally, such software may be written using any of a number of suitable programming languages and/or programming or scripting tools or a combination of programming languages, and also may be compiled as executable machine language code or intermediate code that is executed on a framework or a virtual machine. In this respect, the invention may be embodied as a computer readable storage medium (or multiple computer

30   readable media) (e.g., a computer memory, one or more floppy discs, compact discs (CD), optical discs, digital video disks (DVD), magnetic tapes, flash memories, circuit configurations in Field Programmable Gate Arrays or other semiconductor devices, or other

tangible computer storage medium) encoded with one or more programs that, when executed on one or more computers or other processors, perform methods that implement the various embodiments of the invention discussed above. As is apparent from the foregoing examples, a computer readable storage medium may retain information for a sufficient time to provide

5   computer-executable instructions in a non-transitory form.

The terms "program" or "software" are used herein in a generic sense to refer to any type of computer code or set of computer-executable instructions that may be employed to program a computer or other processor to implement various aspects of the present invention as discussed above. Additionally, it should be appreciated that according to one aspect of this

10   embodiment, one or more computer programs that when executed perform methods of the present invention need not reside on a single computer or processor but may be distributed in a modular fashion amongst a number of different computers or processors to implement various aspects of the present invention. Computer-executable instructions may be in many forms, such as program modules, executed by one or more computers or other devices. Also,

15   data structures may be stored in computer-readable media in any suitable form. Any suitable mechanism may be used to establish a relationship between information in fields of a data structure, including the use of pointers, tags, or other mechanisms that establish relationship between data elements.

It is to be understood that the above-described embodiments are only illustrative of the

20   application of the principles of the present invention. The illustrative discussions above are not intended to be exhaustive or to limit the invention to the precise forms disclosed. Various modifications and alternative applications may be devised by those skilled in the art in view of the above teachings and without departing from the spirit and scope of the present invention and the following claims are intended to cover such modifications, applications,

25   and embodiments.

CLAIMS

1. A method for data processing and analysis by component analysis configurator for report generation comprises the steps of:

    receiving a set of information from a user through an electronic user interface;

    identifying a base data set from a data lake based on the information wherein the data set is an abstraction layer embedded over an underlining   data structure such that a plurality of data attributes and the data structure are   identified using functional objects thereby enabling the user to define a   plurality of components;

    processing the information using at least one decision tree data   model   to extract the plurality of data attributes from the data lake;

    capturing by a definition engine, a metadata associated with the   plurality of attributes and storing the metadata in a metadata database;

    generating at least one query based on the metadata and extracting a   plurality of reporting data based on the query from the data lake; and

    generating at least on report based on the extracted data.

2. The method of claim 1 wherein a plurality of queries is executed and the plurality of reporting data is aggregated to generate a visualization report for time ranges.

3. The method of claim 1 further comprises the step of connecting an output of one report to another report by using mathematical formula.

4. The method of claim 3 further comprises the step of scheduling and pre-processing execution of frequently used reports.

5. The method of claim 4 further comprises the step of generating reports based on a report type selected by the user.

6. The method of claim 5 further comprises the step of generating actionable reports wherein the user selects data points in a first report, enters comments to record actions that are assigned to the user such that a second report when generated is compared with the first report based on the recorded actions.

7.   The method of claim 1 wherein the components include measures, filters, hierarchies and extraction formats.

8.   The method of claim 7 wherein the measures identify a set of financial parameters on which a drill down analysis is performed.

9.   The method of claim 8 wherein the filters enable filtering for the base data set used for analysis by identifying predicate conditions.

10.   The method of claim 9 wherein the hierarchies identify specific aggregation layers at which measures are analyzed.

11.   The method of claim 10 wherein the extraction formats identifies the format in which reporting data is extracted for analysis.

12.   A system for data processing and analysis by component analysis configurator for report generation, the system comprises the steps of:

        an electronic user interface configured for receiving a set of information from a user;

        a data lake storing a plurality of data attributes;

        an abstraction base data set layer embedded over an underlining data structure such that the plurality of data attributes and the data structure are identified using functional objects thereby enabling the user to define a plurality of components wherein the base data set is identified from the data lake based on the information;

        a processor configured for processing the information using at least one decision tree data model to extract the plurality of data attributes from the data lake;

        a controller coupled to the processor and encoded with instructions enabling the controller to function as a bot for controlling multiple components of the system for data processing and analysis;

        a definition engine configured for capturing a metadata associated with the plurality of attributes and storing the metadata in a metadata database; and

an executing engine configured for generating at least one query based on the metadata and extracting a plurality of reporting data based on the query from the data lake,

wherein at least on report is generated based on the extracted data.

13. The system of claim 12 wherein the definition engine is configured to identify a base data set, built components and trigger dynamic executions using the user interface.

14. The system of claim 12 wherein the execution engine is configured for processing data from the definition engine and extract reporting data into format that is utilized for report.

15. The system of claim 14 further comprises a reporting layer consisting of pre-defined reporting templates that combine with the extracted data to generate different types of visualization.

16. The system of claim 15 further comprises a scheduler configured to pre-execute components when required thereby enabling the report to be readily available.

17. A computer-readable non-transitory storage medium storing executable program instructions for data processing and analysis by component analysis configurator which when executed by a computer cause the computer to perform operations comprising:

receiving a set of information from a user through an electronic user interface;

identifying a base data set from a data lake based on the information wherein the data set is an abstraction layer embedded over an underlining data structure such that a plurality of data attributes and the data structure are identified using functional objects thereby enabling the user to define a plurality of components;

processing the information using at least one decision tree data model to extract the plurality of data attributes from the data lake;

capturing by a definition engine, a metadata associated with the plurality of attributes and storing the metadata in a metadata database;

generating at least one query based on the metadata and extracting a plurality of reporting data based on the query from the data lake; and

generating at least on report based on the extracted data.

18.   The computer-readable storage medium of claim 17 further comprises executable program instructions in a memory to be executed for generating decision tree data model.

19.   The computer-readable storage medium of claim 17 further storing instructions that cause the processor to automatically add storage for storing the received data.
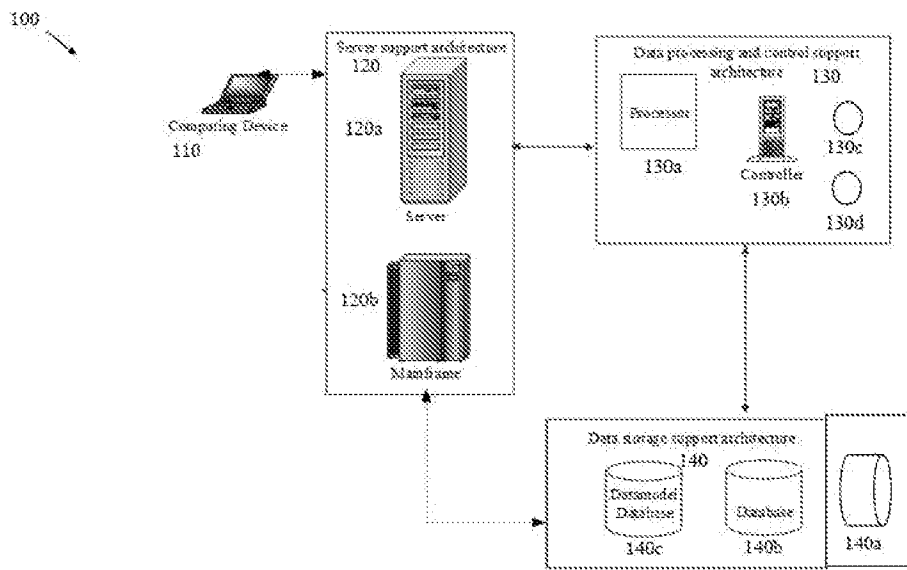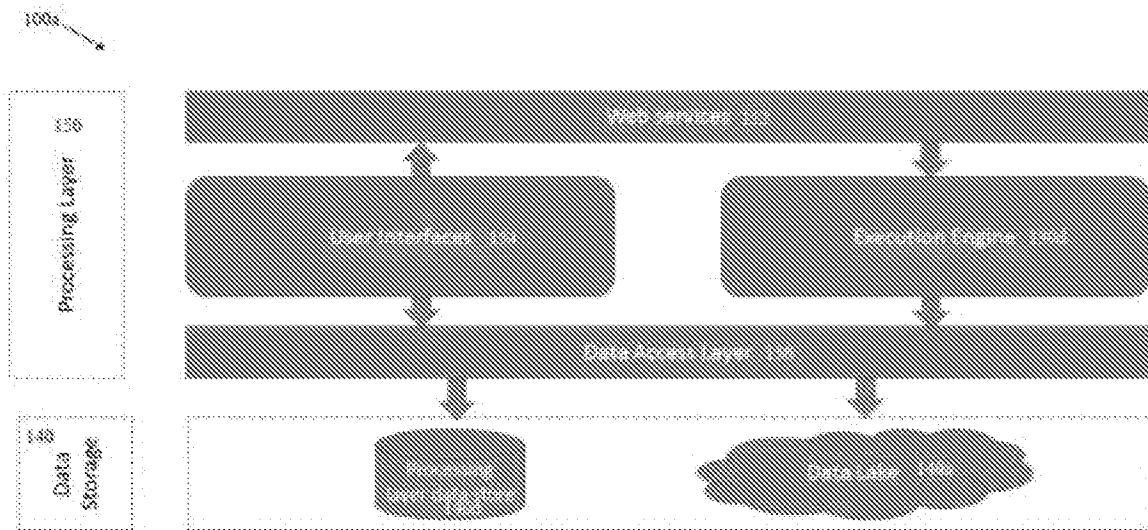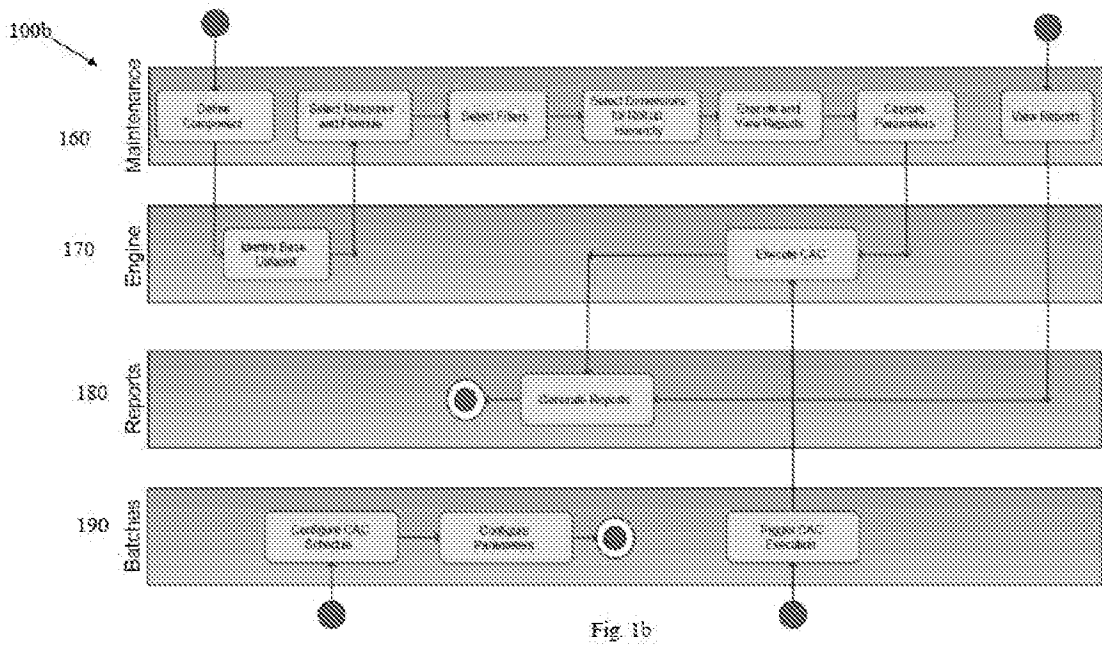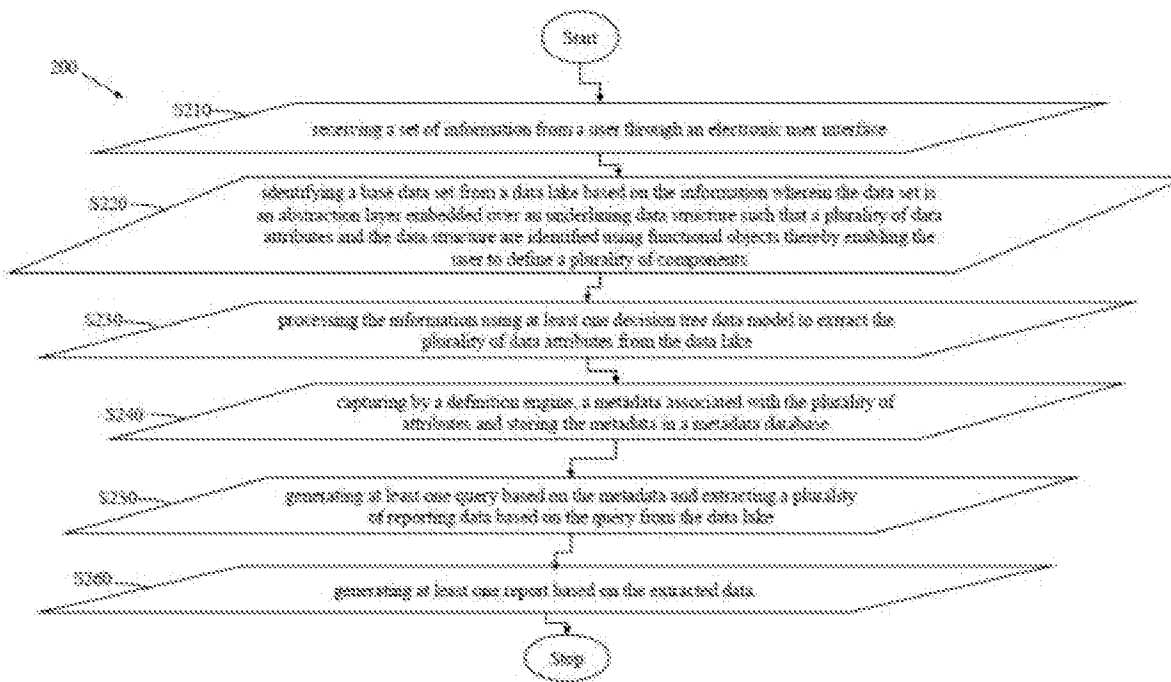
100

Fig. 1

Fig 1a

Fig 1b

Fig. 3

A. CLASSIFICATION OF SUBJECT MATTER

G06F 16/20 (2019.01)   G06Q 40/00 (2012.01)   G06N 20/00 (2019.01)

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

PATENW: IPC: G06F16/20; Keywords – report, generate, create, template, format, decision tree, abstraction, layer, level and like terms.

Google, Google Patents and Google Scholar: similar keywords as above.

Applicant/Inventor name searches were performed in Google & Espacenet websites and internal databases provided by IP Australia.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
	Documents are listed in the continuation of Box C	

| X | Further documents are listed in the continuation of Box C | X | See patent family annex |

*	Special categories of cited documents:		
"A"	document defining the general state of the art which is not considered to be of particular relevance	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"D"	document cited by the applicant in the international application		
"E"	earlier application or patent but published on or after the international filing date	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O"	document referring to an oral disclosure, use, exhibition or other means	"&"	document member of the same patent family
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search	Date of mailing of the international search report
19 January 2021	19 January 2021

Name and mailing address of the ISA/AU	Authorised officer
AUSTRALIAN PATENT OFFICE PO BOX 200, WODEN ACT 2606, AUSTRALIA Email address: pct@ipaustralia.gov.au	Wendy Hou AUSTRALIAN PATENT OFFICE (ISO 9001 Quality Certified Service) Telephone No. +61262832579

	INTERNATIONAL SEARCH REPORT	International application No.
C (Continuation).	DOCUMENTS CONSIDERED TO BE RELEVANT	**PCT/MY2020/050106**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 110083645 A (HITHINK ROYALFLUSH INFORMATION NETWORK CO., LTD.) 02 August 2019 Figs.6-7; [0009], [0011], [0019], [0049], [0052], [0068], [0074], [0081], [0083], [0086], [0091], [0093]-[0095], [0098], [0101], [0103], [0111]	1-19
X	WO 2015/134518 A1 (SYSTEMS IMAGINATION, INC.) 11 September 2015 Figs.1&3; [0022], [0031], [0033]-[0034], [0041], [0043], [0048], [0050], [0056]	1-19
X	US 2019/0172564 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 06 June 2019 Fig.3; [0034], [0078]	1-19
X	US 2016/0055594 A1 (BUILDFAX, INC.) 25 February 2016 [0126], [0219]	1-19

This Annex lists known patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent Document/s Cited in Search Report		Patent Family Member/s	
Publication Number	**Publication Date**	**Publication Number**	**Publication Date**
CN 110083645 A	02 August 2019	CN 110083645 A	02 Aug 2019
		US 2020356588 A1	12 Nov 2020
		US 2020356915 A1	12 Nov 2020
WO 2015/134518 A1	11 September 2015	WO 2015134518 A1	11 Sep 2015
US 2019/0172564 A1	06 June 2019	US 2019172564 A1	06 Jun 2019
US 2016/0055594 A1	25 February 2016	US 2016055594 A1	25 Feb 2016
		US 2015235322 A1	20 Aug 2015
		US 2019156427 A1	23 May 2019
		US 2019295180 A1	26 Sep 2019

End of Annex

Due to data integration issues this family listing may not include 10 digit Australian applications filed since May 2001.