A Database Administrator (DBA) for the cause

Which one ? Yours !

db-analyzer presentation

(META , DATA , SCHEMA , SAMPLER)

- The Genesis
- What is this for ?
- What this is not ?
- What are the roles the it tries to fullfill ?
- For whom this is ?
- The pilars of analysis

- Metadata pilar
- Data analysis pilar
- Schema pilar
- Data sampler pilar
- Database support
- Installation notes
- Few key takeaways

The genesis

○ The origin of the project goes back more than 3 years ago. I worked on it in my free time (mostly at night). I designed this mainly for new employee in mind, developer, database administrator etc.. starting in a company in almost any position really.

○ I feel that quite often the time to get "confortable" in the beginning of any new job experiences is long and most of the time non efficient. Above all, the time to know the databases echosystems, their infrastructure etc…

○ I always thought that it will be good to have a material, a basic piece of information. That could be produced and refreshed **automatically** and be used to ease the all integration process. The key being to provide the **right amount of the right information !**

○ Once the foundation was built, the project has grown naturally, and more pilars of analsysis added on different aspects of the databases.

What is this for ?

In a nutshell this is a tool to assist you in knowing better YOUR data and have and a kind of an inventory in the following main aspects of your database server.

- Metadata
- Data analysis
- Schema
- Data sampler

All the result information is available in different files formats addressing themselves to potentially different people. Json and csv (excel) for detailed analysis . Png and Svg for the visualization modules.

In addition to the analysis files generated a special attention has been ported to gather results within a pure HTML dashboard called **Reporter** so we can access all the analysis results clearly, simply, and see evolution over time and more.

What this is not ?

○ **Is this a magical tool that will provide some unicorn powers ?**

Nope, this is just a program to provide you insights on your database that you may or may not be aware of. And once gathered, those insights could be shared easily and will remain as baseline snapshot in a point in time.

○ **Is this unique and one of a kind, a master-piece ?**

Nope, the information that will be gathered and reports that will be generated could be retrieved in various other ways and with many programs that already exist but for which mainly you'll will have to do it manually and/or pay. This is just one alternative presented in an original, simple way and ,above all, driven for a good cause, that's all.

What are the roles it tries to fulfill ?

○ Inventory

○ Documentation

○ Modelisation helper

○ Database dictionary
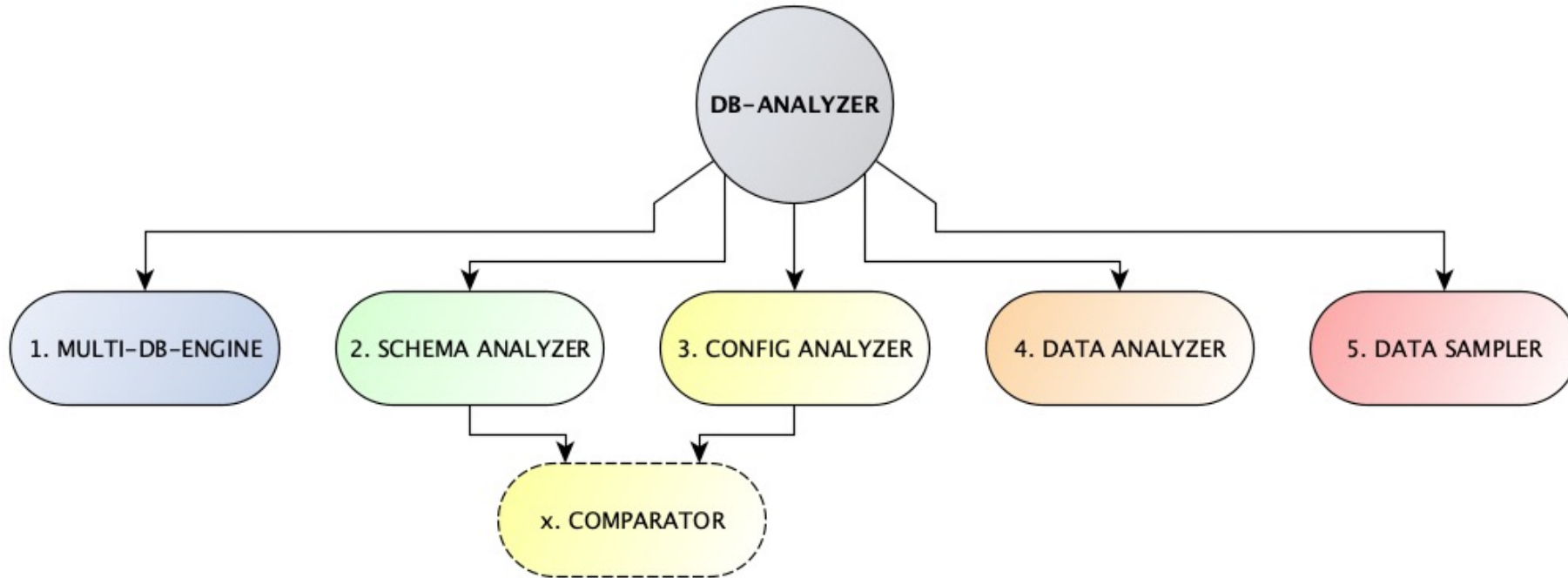
○ Offline data repository

For whom this is?

My IT background make me often oversee things with technical eyeglasses. Sometimes using the basic but rather efficient approach : identify the problem & find a solution.

This project tries to reflect on the database information not only in a technical point of view but with a business touch as well . Therefore this project can be of use for :

- Managers
  - Understand / get confortable with data overall
  - Provide **support** for presentation to new employee

- Analyst & developer
  - Database development and implications
  - Schema information

- Database Administrator
  - Day to day jobs
  - Sizing capacity
  - Trends and baseline

The pilars of analysis

Metadata pilar

This pilar focus essentially on the metadata attributes of the database as well as its configurations parameters

- Server & database(s) information

- Server & database(s) configurations

- Tables & their columns

- Views & their columns

- Materialized views & their columns

- Jobs, jobs details

Data analysis pilar

This pilar focus essentially in the data analysis aspect of the database. Data distribution as well as null and not null statistics gathering.

- Highly parameterized filtering engine
  (tables and columns inclusions and exclusions etc.)

- Level of fields analysis (up to 3 fields combined)

- **Automatic** Data distribution for each combination

- **Automatic** Null and not null statistics gathered for each combination

Schema pilar

This pilar focus essentially in the modelisation aspect of the database. We tried to implement different type of presentation; one may fit better a need than other.

- Entity Relationship Diagram (ERD)
- Use case
- Neo4j cypher (Graph database)

Data sampler pilar

This pilar focus essentially in the data sampling aspect of the database.

- Highly parameterized filtering engine
  (tables and columns inclusions and exclusions etc.)
- **Automatic extract random x records from tables and columns specified**

Database support

- RDBMs database type
  - Mssql
  - Mysql (MariaDB)
  - Oracle

- Nosql database type / document oriented
  - ElasticSearch (OpenSearch)

Installation notes

- Easy to setup , docker machine with all librairies already pre-installed and tested
- Installation time is minimal. 0 to 10mins
- Configuration has parameters to allow you
  - To be really detailed if you want, if you are after something specific
  - Or you can let the default attributes for standard analysis
- Recommendations is to start analyzing the metadata
  - Then you could setup up to 5 schema analysis fully parameterized
  - Then you could setup data analysis for your required tables and fields
  - Finally, you could setup sampler analysis for your required tables and fields

Few key takeaways

○ With the fully parameterized filtering engine, you can for data analysis and data sampler remove any sensible tables and fields that you may have and still generate the information all the non-sensible information that could be shared.

○ Usage of query verse is implemented to allow flexibility with compatible engine that one may be using as well as providing clear view of what will be executing by the application .

○ Some additional information have been added to the json and visualization output for some analysis to provide some insight and be able to compare over time (i.e execution time duration for data sampler)

○ Special attention has been ported so that the application can react accordingly, if there is some issue here and there with part of the system database queries.

○ Want to know more ? proceed with the reporter result output package include in the communication….

Ubuntu :
I am, because you are