**Topic:** What refinements are needed for a "working prototype" to become "production ready"?

In the initial phase of data warehouse development, business requirements are usually unknown, that is why, it is important to apply iterative approach of building prototype in order to succeed in a DW development project.

[Diagram, text

Description automatically generated](http://innovacons.com/wp/data-warehouse/)

(Image Source: <http://innovacons.com/wp/data-warehouse/>)

A working prototype is nothing but a model of a data warehouse or data mart that can be used by a selected group of end users or business analysts. Designing and implementing a working prototype, also known as “**pilot**”**,** is an iterative process as business requirements may change or new requirements may be added as improvements are made.

[Diagram

Description automatically generated](http://dssresources.com/papers/features/demarest08/demarest01252008.html)

(Image source: <http://dssresources.com/papers/features/demarest08/demarest01252008.html>)

There are different methods a prototype can be designed as listed:

* **Non-production Pilot:**

Some organizations may provide the technology (i.e., tools or user interfaces) and opportunity to make the prototype available for selected end-users and decision makers **before** it goes to production. In this case, a **non-production pilot** can be created, revised according to the feedback provided by users, and redeliver it to collect the feedback on latest version until users are satisfied with the results.

* **Hybrid** **Pilot / Rollout:**

Sometimes, data warehouse architects do not have the privilege to get the prototypes tested and collect the feedback from users. In this situation, **Hybrid Pilot/Rollout** is needed. For instance, to get sales information from the data, a new window or a page in UI can be created and made available only to employees in sales department. By monitoring the activity on page or by collecting the feedback, the design can be modified and redelivered until “*product is ready for large-scale deployment*” (Matthew Yates)

**From Prototype to Production:**

When a prototype of data warehouse is built, very small amount of data can be stored in it compared to the actual data warehouse that will be developed in production environment. A data warehouse is bound to provide **a high availability of data and high performance** for complex queries. For instance, thousands of customers may access the information from data warehouse at the same time, and it is expected that the performance is not degraded even in such a case. To achieve this, the prototype must be refined to make it production ready.

**More resources should be allocated** in storage disk or cloud based on the data volume and modelling approach (i.e., multi-dimensional or tabular). Considering the **need of testing and QA environments** along with production environment is also important.

“*Optimization and stabilization are two different things but truly complement one another when identifying the life cycle of a data warehouse*”(Sqlswa5bg). Optimization techniques must be included such as partitioning to serve the reporting requests successfully. Once optimization is completed, the environment must be stabilized. It includes infrastructure resizing and introducing processes such as daily maintenance, backup, and recoverability, etc.

Clear data access and retention policies must be developed in order to prevent the data leak and privacy breaches. Frequency to load the data into data warehouse must be defined so that the latest data is available for analysis and help analysts with an accurate data driven decisions. With the increasing utilization of data warehouse in business, additional activities such as automation of ETL, identification of aggregations to maintain system performance can also be included. To address new business requirements in future, it is reasonable to make provisions for extension of current version to load additional data and introduce new data marts using the existing information.

Data warehouse development is not about just building it. We must know the life cycle of data warehouse and how to maintain it. The key to it is “*to have a good cycle in place for the build then you have a good cycle for the changes and deployments too*”(sqlswa5bg).

**Next steps for “The Wellness Team”:**

The wellness team has been working on designing and developing a prototype of data mart for an online store (eCommerce site) of wellness products and services. All the refinements specified above can be applied to the prototype being developed as “project - part 3” to make it “production ready”.

The data mart can be regarded as the **backend** of our analytics application, and very first step after this would be to integrate the BI tools with the data mart to analyze the data stored in data mart and generate the reports and visualizations as per the requirements. We will also allocate a space to store the clickstream data into this data mart, that will help us identify patterns in customer behavior and identify activities that trigger the order placement.

To increase the customer base in future, we will also include a data mart for “campaign management system” (First we’ll develop a marketing database as another operational source and a workflow for this new system). Campaign management is basically to identify good and loyal customers and provide them with discounts on the products they buy from us. The data extracted from marketing database will be aggregated based on the campaigns and promotions to identify their effectiveness. Employees in marketing department may need be familiar with using BI tools or understand the reports or visualizations, hence we may need to arrange a training for them with business analysts.

**References:**

1. Matthew Yates (Sep 19, 2021). From Prototype to Production for Data Scientists. *Towards Data Science***.** <https://towardsdatascience.com/from-prototype-to-production-for-data-scientists-4d86df6c3c08>
2. Innova Consultant (N.d.) Data Warehouse. Retrieved from [Data Warehouse (innovacons.com)](http://innovacons.com/wp/data-warehouse/)
3. Demarest, M.(Jan 01, 2008), Data Warehouse Prototyping: Reducing Risk, Securing Commitment and Improving Project Governance. *DSSResources.COM*. Retrieved From <http://dssresources.com/papers/features/demarest08/demarest01252008.html>
4. Sqlswa5bg (June 28, 2020). The Life Cycle of a Data Warehouse. *Sqlswagg.com.* Retrieved from <https://sqlswagg.com/2020/06/28/the-life-cycle-of-a-data-warehouse/>
5. tutorialspoint (N.d.). Data Warehousing – Delivery Process. Retrieved from <https://www.tutorialspoint.com/dwh/dwh_delivery_process.htm>