Xplenty

HOW TO BUILD AN END TO END BI SOLUTION

TAKE THE FAST TRACK - BUILD IT IN 1 WEEK!

If you're in charge of planning and/or implementing an analytics solution for your company there is a ton of things you must consider. We've mapped it all out to help you make sure that you haven't missed anything.

FAST TRACK

If you're in a great hurry to start pulling initial insights right away, consider taking the "Fast Track" - here we share with you some useful tips on how to launch your BI project adopting the "Lean approach", evolving it into a more sophisticated one as you go.

BI STRATEGY

You need to start with a plan or a "BI Strategy". Consider your BI strategy as a roadmap; It should chart out the optimal route you need to take to get your data management solution from its current state to the vision you have for it. Here's what you need to consider:

DAY 1

FAST TRACK

1 Report and analytics needs

Data can be consumed in many ways; governed dashboards, ad-hoc analysis, AB testing, real time analytics, etc.

Define 1-2 major reporting needs. If you already have an existing solution simply replicate it in your new BI.

2 Industry KPI's

There are standard KPI's and metrics for most industries today, for example sales, return on investment, profit margin etc. Use industry benchmarks to validate and expand your own KPIs.

Start with the 3-4 most important KPI's

3 Custom KPIs

Unique business derivatives and professional systems require tailor made KPI's to reflect business specific measurements

If the data needed to calculate the KPIs is not complete at the source system, consider implementing these at a later phase

4 Historical data saved as aggregations

Consider the historical depth required for analytics. Define what can be summed up to reflect several dimensions and what needs to be analyzed on a raw data level.

In order to choose the right technological solution it is crucial to evaluate the above, But implementation can be done on a later phase.

5 BI Clients

Consider who will be using your BI solution- partners,customers,high- level executives, analysts, sales representatives, account and marketing managers etc..

Select 1 or 2 clients to begin with , usually starting with the department that is expected to adopt the BI first.

MAP DATA SOURCES

Mapping data sources is the process of selecting the sources of your data and filtering the relevant data from each source. In order to map the data required, consider the following inputs - Existing reports, management requests, industry standard KPI's and of course your vision!

at this point it is advantageous to start from the end-goal and ask yourself: Ultimately, what reports should my BI be able to supply?

FAST TRACK

DAY 2

1 Core data

The data generated by your business, whether it is a mobile app, website, online shop, etc...
This data is typically generated as log files and tends to be too messy to be used directly for analytics.

2 peripheral data

Data generated from purchased products or services. Your CRM software holds additional valuable data which you will likely want to analyze in synergy with your core data. For example: customer geography, purchase history, demographics etc. .
Same goes for your ERP system, Google analytics data, ad campaigns etc.

Initially consider only the critical data that answers essential business questions that need to be answered from day one.
Starting your project with mapping all of the data sources may take a long time. It is often better to start with few specific sources and fields, and add the rest gradually as your BI solution evolves.

3 External data

Gather the data that is "out there"; if you do website crawling, sentiment analysis, gather intelligence on your competitors or create dictionaries, collect this external data to feed your BI solution

Bring only the data that will add significant insight to the BI solution creating actionable decisions not available today.

ETL

Layer that manages the 'Extract Transform Load' Processes.
Can be implemented based either on Scripts (Custom Code) or based on an ETL Tool. Some of the common functions that this layer performs are:

1 Parsing/cleansing

Data generated by applications is saved in various formats such as JSON or CSV. The purpose of the parsing stage is to map the data into a table format with headers, columns and rows while extracting values only from the specified fields.

2 Enrich

To prepare the data for analytics, usually enrichment steps need to be deployed. For example: tweaking, injecting expert knowledge, implementing geo modifications, matching between sources and correcting bugs.

3 History depth

During the initial run, history should be loaded either in bulk or in a loop in order to set the existing data. On the later runs, only incremental ETL is required.

4 Velocity

The frequency of loading data – is it Daily/Hourly. Should existing data be updated or only new data inserted?

Implementing the ETL layer through a SaaS tool is highly recommended, as it facilitates highly effective database management and seamless collaboration within your team.

DWH

There are several crucial criteria of selecting the right database to serve as the DWH. It should load data from the source files in the required velocity, store it in an optimized manner and serve various queries from the BI tool and from any other query you plan to use..

The work should be done in cycles and phases. Load the first batch of the source data as soon as it is ready. Don't wait until all the ETL work has been fully completed.

FAST TRACK

1 Schema design

To make the data accessible for analytics and reporting, a data schema should be deployed to map the types of data to the facts, dimensions, measures and attributes. At this stage, The indexes should be defined and the relationship between the tables needs to be set.

Start with the main entities and expand the schema as you progress.

2 Cloud Vs. On-Premise

Deploying the DWH in the cloud is a highly recommended and popular option as it offers multiple advantages , such as scalability, availability, ease of use and flexible pricing.
Nevertheless, if you have the required hardware and infrastructure already in place, or if most of your source data is stored on-premise it makes sense to maintain an on-premise data warehouse. In some cases your data may be subject to regulatory constraints which may limit your choice to an on-premise solution.

Consider the location of your core data source first

4 DB size

In order to choose the right technological solution one of the most important considerations is the data size

5 Concurrency

The number of users and queries that will run simultaneously on the database should be estimated and carefully planned for.

Most probably concurrency will not be a big issue on day one, as It takes time until all departments and applications adopt a new DWH.

6 Scaling

The rate at which the volume of data is expected to grow. If the growth is linear scaling out is usually the prefered option (adding hardware to the existing db), if the growth will be exponential scaling up might be a better solution (replacing the db)..

Scaling is very simple to execute on cloud platforms. With BigQuery for example, scaling requires no action at all.

Choosing a BI Tool

The BI reporting tool will be the face of the entire BI solution, Serving dashboards, data visualizations and client facing reports. Here are the issues to consider when choosing a reporting tool.

DAY 6

1 Clients (internal & external)

Who are the departments/business functions in the organization and what type of use will they require from the BI tool? Are there also partners/clients that will require access to reports?

2 Self-service

An environment for users to do their own analytics and create custom reports

3 Type of analytics

There are different methods of creating reports and there are BI tools excelling at each one. Some of the types can be: Lists and tables, Visualizations, Governed dashboards and ad hoc analysis?

4 Integration with DWH

is the BI tool optimized to query the selected DB?

The internet is full of benchmarks, POC's and comparisons between different tools, narrow down to 2-3 tools as finalists and run a short POC on 1 or more of them.

Designing Dashboards and Reports

Creating dashboards, reports and analytics might quickly become more complex than initially anticipated. Below are the main points to consider

DAY 5

1 Dashboards design

When designing a dashboard, ask the following questions: What are the elements in the dashboard? How do they interact with each other?
Is all the data available from the same datasource?
Which filters should be used?
What kind of visualization is most suited for each element in the dashboard?
Utilize existing dashboards and simplify them as needed. Do not try to stretch the limits of your solution at the initial stage of development.. Keep it simple so your users can easily understand what has been done.

2 self-service environment

Self service means that a user of the BI is able to login to the BI tool and perform an ad-hoc analysis or generate custom reports.
 Having the ablility to share the reports and add comments is considered as a nice to have.
At the first stage choose an initial model out of the DWH schema to be public for self service

3 performance tuning

In many cases long loading times of the carefully designed dashboards will occur.. Take into account that performance tuning and optimization for the

DWH, schema, query and dashboard development will be required and will take several days to implement.
Use in memory capabilities or aggregate the data prior to loading it into the BI tool.
Don't use too many elements in your initial dashboard design.

4 adoption plan

We believe that this is THE most crucial point to consider. Define who are the end users for each report. Prepare a training program to familiarize the users with the reports and with the interface and functionality of your BI tool. Measure ROI figures for specific reports whenever possible. Your main task is to make the users realize the benefits of the data - mostly its reliability. Teach them the main functions of the DWH schema and how they can access it through the BI tool. Your final goal is to educate and train your users to become "self served". It is recommended to involve the end users in the development of the BI tool already at the planning stage They should stay involved also during the development phase as active QA and acceptance testers.
Choose your first users carefully. As your BI solution evolves, the early adopters will become the advocates and active promoters of your analytics solution within your company.

TEAM

BI teams differ between organizations in both size and internal composition

Plan your recruiting in stages and consider hiring an external contractor to temporarily fill in the skill gap.

FAST TRACK

1 Head of BI/ Director/ VP of analytics/ Team leader

The person responsible for making a firm truly data-driven. Equipped with both business and technological skills, they establish and execute strategies that generate insights, creating self-service analytic platforms and other business intelligence solutions.

Often the right candidate is someone who already works in your company.Current employees could have the edge because they know the business inside out and frequently have the right mix of technical and managerial skills..

2 BI developer

Data professionals who design and build data pipelines to integrate data from various sources, assuring the data is extracted, transformed, and loaded into databases or data warehouses.

Consider using an ETL tool to be able to generalise and scale this position easily.

3 Data/business analyst

Data analyst acquires, processes, and summarizes data. By using specific queries and existing analytical tools, data analysts construct insights from data. They supply their organizations with reports, summaries, visualizations, thereby transforming the data into digestible insights to be used by non-technical personnel. Their work helps management make data-driven decisions and set goals based on evidence. Furthermore, data analysts run unique queries for stakeholders..

Make sure the candidate you consider for this position has the right mix of business and technical skills and is not proficient in only one of them.

4 DBA

Orchestrates the capacity planning, design, installation, configuration, performance monitoring, migration, and troubleshooting of all things database related. DBAs maintain database systems, create new database applications, support existing database applications, and manage an organization's data and metadata..

Knowing the inner workings of the specific database your company utilises is crucial for this role Therefore it is advised to select the DB first and then to start the recruitment process.

5 Data scientist

data scientists have the ability to invent new
algorithms to solve complex analytical problems.
They utilize computer programming, statistics,
analytical tools, and sophisticated machine learning
to pull out actionable insights from big data.
The range of qualifications required to fill this role is
often very wide. Prioritise the skills most important
for your company.

Creating a full stack business analytics solution
for your company is a process that requires
careful planning and meticulous execution.
This infographic walks you through the tasks
you will need to tackle on the way, from
formulating a strategy to hiring the right team.
Keep this cheat-sheet handy to refer to as
needed.

**To learn how easy it is to integrate your data
with Xplenty [click here](#)**