**BI Architecture And Pipeline**

Well, to set up a BI solution from scratch we will have to analyze and understand the goal, needs and situations of the business and business environment.

My technology stack would be:

1. Python modules like “**numpy**” and “**pand**a” for data collection from different sources and do some exploratory analysis and data cleaning and data migration to database.

2. For Database I may use “**PostgreSQL**”. Here, database would be created maintaining necessary schemas for data warehouse.

3.To create data warehouse maintaining **OLAP data cube** I may use **Microsoft Data Studio, Microsoft Analysis Services** for BI and designing data warehouse.

4.From data warehouse I may get necessary insights and pivoted data which will be visualized by **PowerBI** tool and help me to make decisions for my business.

5.Then, whole process would be automated by running **Docker** and Utilize **luigi** to keep me my ETL on track.

**Architecture**

* Data exploration
  + Understanding my data
  + Looking for red flags
  + Utilizing both statistics and data visualization
* Checking my data for issues
  + Identifying things outside of the “normal” range
  + Deciding what to do with NaN or missing values
  + Discovering data with the wrong data type
* How to clean and transform my data
  + Utilize the pandas library
  + Utilize pyjanitor
  + Getting data into tidy format
* Dealing with my database
  + Determining whether or not you actually need a database
  + Choosing the right database
    - Deciding between relational and NoSQL
  + Basic schema design and normalization
  + Using an ORM – SQLAlchemy to insert data
* Building a data pipeline
  + Separate my ETL into parts
  + Utilize luigi to keep me on track
  + Error montitoring