

# An Approach to Data Science by ZenLabs

- [Introduction](#)
- [An Approach](#)
  - [Research](#)
    - [Problem Definition](#)
    - [Data Science Solution](#)
  - [Build](#)
  - [Operationalize](#)
- [Conclusion](#)
- [Notes](#)
- [About Zenlabs](#)

## Introduction

At Zenlabs we bring tried and true problem solving methods to bear for solving business problems. We believe in methodological scientific approach. A scientific approach provides a systematic guidance in solving problems. Such systematic guidance provides a razor sharp focus on problem at hand either at the defining stage or at build stage.

We take one such age old method in research which has three stages: define, develop, verify. We modify this method to include operational phase. Thus the new method is "Research (define)", "Build (develop and verify)" and "Operate" stage where we operationalize the verified product.

Our goal is to bring a systematic approach in solving business problem where business gets the highest return possible for the investment. An approach such as this one will enable business to use a data based approach to allocate resources (both people and funds) to their most pressing problems.

The figure below captures the proposed approach. We will describe in detail in the "An Approach" section.

## An Approach

### Research

This phase is to understand and define an actionable business problem. The end state here is a well defined problem along with data sources and elements such that a business valued solution can be achieved during the build phase. The research phase is further sub divided into two stages: Problem Definition and Data Science Solution Exploration.

#### Problem Definition

During the Problem Definition stage, Data Scientists, Business SMEs, and Business collaborate to understand, analyze, and scope the problem. The effort will be focused on defining a problem that brings most value to business.

#### Data Science Solution

During Data Science Solution stage, we do data and model discovery. We find and catalog data sets and machine learning models for the build phase.

This stage goes hand in hand with the Problem Definition phase. As business may challenge the team with a problem that may not be achieved with the available resources (people and funds) at the moment. Thus, Data Scientist and Analyst will provide a feedback on shaping an achievable scope.

In this stage, we will explore all available data and its sources as well as models that are applicable to the problem definition. Here, we do not have to run model against data or clean the data. In this stage, identifying data sets and relevant models is the goal. Therefore, the outcome of this stage is a laundry list of data sets and models.

The goal of this Research phase is to reduce the uncertainty in the build phase and use time available with business SMEs more efficiently.

### Build

This is the phase where we bring concept to the reality. You may think that this is where we clean the kitchen sink as we are getting kitchen sink of stuff from the Research phase. During build phase, the data is sourced, cleaned, and sanitized so that it is ready for applying catalog of models from the Research phase.

There may be many sources of data and may required to be stitched together. In other cases, data sets may be required to be analyzed in steps, adding analytics in each step. Regardless, the goal of this data cleaning stage is to implement reusable modules that connects to databases (i.e. data sources) and implement modules to clean, extract, transform, and normalize data so that we can apply models on the set.

While we are cleaning and preparing the data, we will test our catalog of model to find the best a model or a set of models. As data gets ready, each model shall be tested and verified for its usefulness to achieve the business goals.

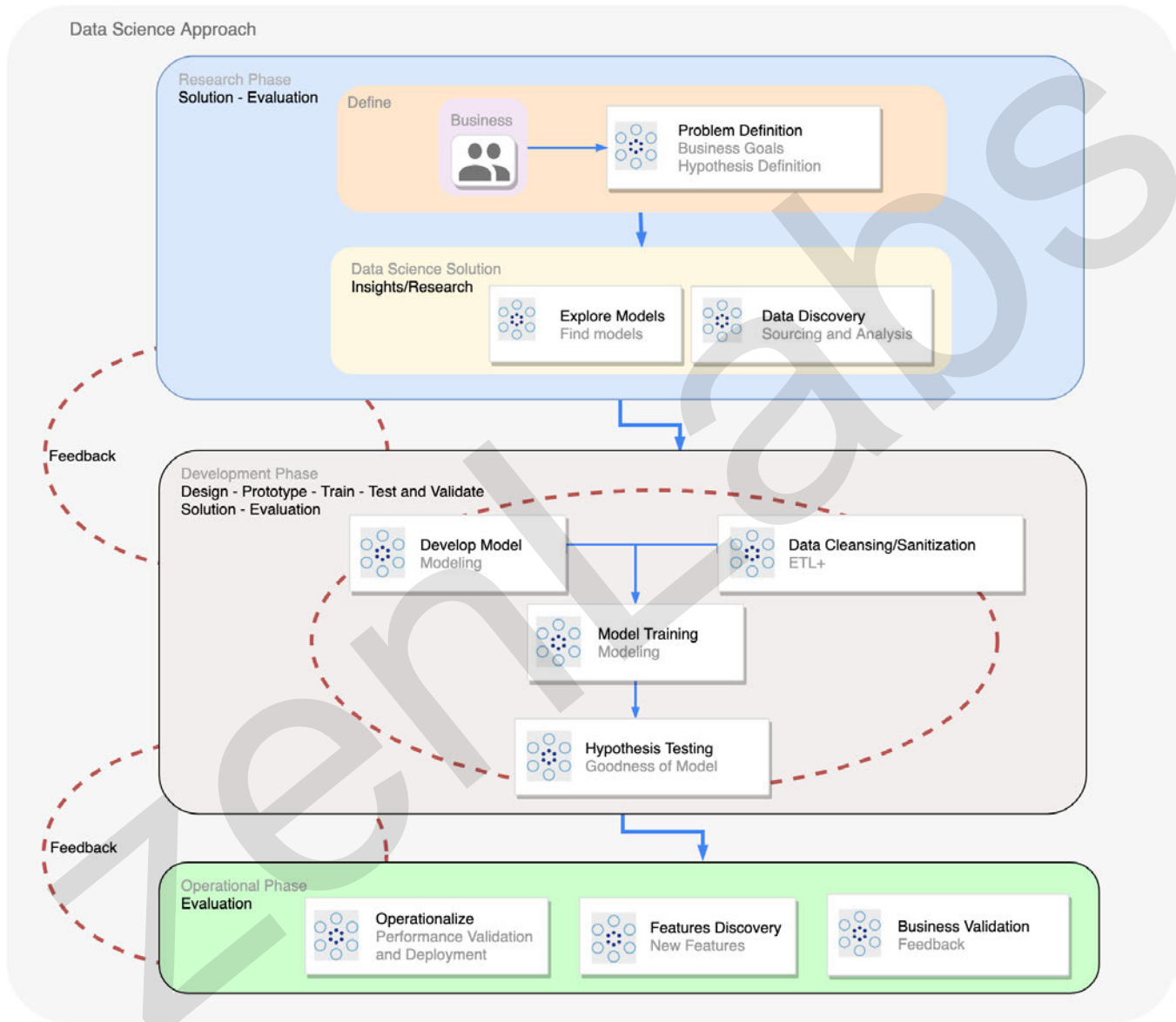
Data Scientists bring their magic wand to find and choose a model or a set of models solve the problem. They iterate through modeling, training, and hypothesis testing to validate and verify that a model is genuine. Through various testing statistics, the model will be blessed for operationalize that is a solid verification is a key contributor to success in this stage.

The outcome of this stage is a proven and trained model with all the data sources integrated into production

## Operationalize

In the operationalize phase we test and do necessary hardening of implemented code for deploying into production. A key success criteria here is gathering performance metrics of the models against volume of data that are 1x, 2x, and 3x of the production volume. Here we size and measure of hardware and software environment necessary for production deployment with success. Such a stress testing allows us to discover any issues with our environment before deploying to production so that we can take corrective action before it becomes an issue in the production.

Zenlabs' Data Science platform provides a seamless mechanism to perform this phase cost effectively.



## Conclusion

This approach a team can use on their Data Science journey. It is very imperative that a well defined achievable problem is taken into consideration to be successful in this journey. It is very important for business to understand, acknowledge the investment and possible return. The business shall be able to recognize the return on investment on each project. In the "Build Out" phase of Data Science Platform, the business will incur a higher investment. However, the investment in the Data Science Platform will be realized in near term from having insights into your business through data and realize "Data is the Next Oil."

## Notes

- During build phase we integrate the data sources. An example of platform for such data integration is the industry leading platform of choice Composable (<https://composable.ai>)

## About Zenlabs

Zenlabs Data Science platform enables businesses to realize their Data Science Journey from concept to operation. The platform provides a plug and play enterprise environment that covers new or existing micro services. The platform provides metrics for you to refine your investment and journey. Zenlabs goal is to make you (the Business, Data Scientists, and technologist) successfully arrive at your destination with as less pain as possible.

ZenLabs