Gaining Market Insight & Enhancing Segmentation Through Big Data Analytics

[Abstract 2](#_bookmark0)

[Introduction 2](#_bookmark1)

[Overview 3](#_bookmark2)

[Objectives 3](#_bookmark3)

[Required Resources 4](#_bookmark4)

[Deliverables 5](#_bookmark5)

[Constraints 6](#_bookmark6)

[Legal & Ethical Considerations 7](#_bookmark7)

[Risk Assessment 7](#_bookmark8)

[Project Plan 8](#_bookmark9)

[References 10](#_bookmark10)

[Further Reading 12](#_bookmark11)

[Appendices 13](#_bookmark12)

[Appendix A: The Big Data Stack 13](#_bookmark13)

[Appendix B: Data Science Hierarchy of Needs 14](#_bookmark14)

[Appendix C: Project Plan Gantt Chart 15](#_bookmark15)

[Appendix D: Project Sponsor Agreement Letter 16](#_bookmark16)

2706 words

# Abstract

This document outlines the project management action plan pertaining to a research project in the field of big data science and analytics. A project overview provides a general introduction to the project’s scope, followed by an in depth description of objectives and deliverables. These sections details the research into the identification of behavioural and characteristic traits through an analysis of purchase patterns and history. Specifics are provided on how this research will be applied to the development of a program which facilitates the input and processing of customer data to accurately segment the market based on identified customer characteristics and personalities. Any resources required for successful research or development outcome have been stated, and potential internal and external project constraints have been identified. The legal and ethical considerations associated with undertaking this project have been identified, including issues surrounding informed consent, privacy & security with regards to data storage & processing, as well as reinstating that information generated as a result of such processing is an informed guess as opposed to factual. A project plan is provided, detailing all the stages and sub tasks that need to be addressed to successfully complete all project requirements. This is supported through a visual representation of the project’s timeline in the form of a gantt chart.

# Introduction

This project aims to apply big data analytics techniques to customer data collected from a supermarket involved in the distribution and sale of a variety of consumer goods. In doing so, the client intends to extract customer characteristics and behaviour patterns in meaningful ways capable of accurately segmenting the market and maximising business impact. The global growth of generated data and tracked information is rapidly increasing, allowing companies to utilise its analytical benefits across a wide range of industries ways which provide insight or uncover opportunities (Amarpreet Singh, 2018). Whilst big data refers to the vast and various amounts of continuously generated data being stored and retrieved for standard business operations, the analytics of that data describes the process of examination to discover unknown information such as correlations and patterns, to predict trends and solve specific company problems through informed decision making (Margaret Rouse, 2018).

Overview

This project builds on conclusions drawn from academic works in the field of big data analytics and it’s benefits on business intelligence & performance, with specific importance given to a 2017 literature review on ‘How Big Data Analytics can Contribute to the Marketing Performance of Supermarkets’ by Koen van der Schaaf. Such works identify the significant positive potential impact the analysis of existing big data can have on industries such as supermarkets by providing valuable information capable of supporting critical decision making. Many top company executives have commented on how big data has provided ‘an evolutionary set of capabilities that would have new and sometimes unanticipated uses over time’ (Thomas H. Davenport, 2013). This research project will seek to answer the question: ‘*How can big data science & analysis techniques be most eﬀectively applied to historical customer purchasing data in order to segment and gain insight on the market?*’ by applying relevant market, consumer and behavioural research to the development of a technological application which makes use of optimised algorithms to analyse purchase history, characterise customers and segment the market. Results will then be compared to known data to measure accuracy and eﬀectiveness.

Objectives

Project objectives define the criteria for a successful outcome, and are used as eﬀective goals to monitor progress and motivate the team (Team Clarizen, 2018). It is important for stated objectives to be specific and defined to be completed within a timescale, as vague goals are not easily measurable. An indication to performance and quality objectives should be provided, as the end result should aim to satisfy the stated project purpose and criteria (Brian Miller, 2008). This will also allow for a comparative baseline by which to judge the final outcome.

Through big data science and analytics techniques, the supermarket will gain a greater insight on its evolving customer base, with the aim of intelligently segmenting the market, allowing for more informed business decisions and targeted marketing approaches. This insight is to provide valuable business and marketing strategy information compiled as a result of algorithmic data interpretation and analysis.

The project timeline will run for approximately eleven weeks, starting at the end of February and running through till the beginning of May. This will incorporate time for necessary research, development, testing and documentation and is detailed in the gantt chart under the project plan. Appendix A illustrates ‘the big data stack’ - a conceptual model which identifies the various layers comprising a big data application. Whilst technological development will focus on the outcome of the big data model as a whole, research will be conducted to determine the most accurate and eﬃcient logic structuring algorithms, code, functions and services; the processes which are at the core of interpreting, segmenting and characterising groups of customers.

Research into consumer characteristics & purchasing behaviour will identify the appropriate segmentations in which to group customers, as well as the criteria required for the most accurate classification to be made. These classifications will be based on each individual customer’s shopping history, where each item will contribute a weighting towards a characteristic, and a combination of characteristics will determine the segment for the customer to be assigned.

A basis must be defined for which items or purchase traits define a characteristic, such as gender, income bracket, family size etc. These behaviours will be determined through second hand research, and data accuracy will be ensured through trial and error with tailored A/B data alongside surveyed participants within the local market. These items will then be compared to a shoppers purchase history during data analysis in order to categorise the customers into market segments, allowing for informed targeting.

The client has specified an average of 4000 registered customers will require processing and analysis for segmentation, each containing between two and forty purchase transactions. Future transactions will be uploaded into the system over time as they become available, and these records should adjust how aﬀected customers are segmented appropriately.

Required Resources

The company currently collects structured customer data through both online sales and in store transactions, and maintains these records for three years. This includes an itemised list of products, order total, location (online/in-store), and a timestamp among other miscellaneous pieces of data. This information is to be processed alongside a product catalogue, where each item contains ranked values for each personality trait or characteristic. The ranked product catalogue does not currently exist, and is to be compiled through analysis of purchase history and psychological consumer research to determine appropriate ratings for various product groups.

Whilst A/B test data may be generated and used to verify that assigned characteristics accurately align with the algorithmically programmed definitions, it would be beneficial to source identifiable respondents who can verify such results through means of a survey or interview. The participants results are to be compared to the analytically generated data to assess for accuracy and identify potentially mis-weighted items or behavioural traits.

This testing requirement shall be used to ensure the validity and accuracy of the behavioural research data gathered on the local market, and to assign categorical item weightings to characteristics or purchase traits.

The application should be hosted form a cloud or local dedicated server, running custom code on top of open source software to deliver the required capabilities to multiple simultaneous clients over a network connection. Depending on its existing availability within the company, this may incur purchasing and setup costs. Although development will be carried out on existing shared servers currently hosting a variety of applications, it is recommended that a dedicated server be set up for deployment due to access to higher level controls, as well as to mitigate payment and general data privacy concerns.

Deliverables

A project’s deliverables are defined by any output obtained through work conducted over the project’s duration. Outputs must meet some criteria to be classified as a deliverable, such as being within the project scope, having a definite role in accomplishing a stated objective, be the result of deliberate work and be agreed to be all stakeholders (Sylvia Moses, 2018). Whilst a project’s objective’s are conceptual outcomes such as statements which describe the desired outcomes and accomplishments, deliverables include all the in/tangible products or services produced as a result of the project’s undertaking. This Amy include research, software, creations/work, documentation and management/ administrative resources among other resources (Ashley Marron, 2017).

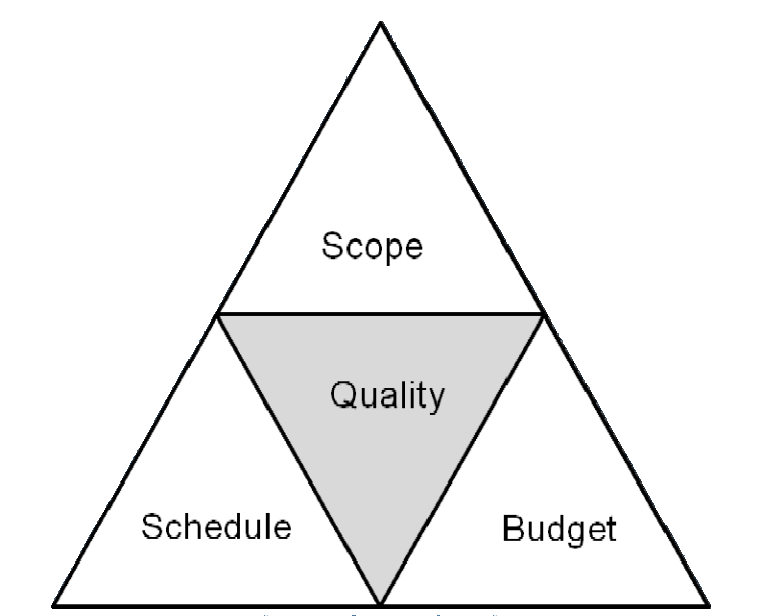
The project research and development will result in a program who’s interface will facilitate the periodical upload of identifiable customer purchase history. Based on an analysis of this data, customers will be assigned biological and/or personality characteristics, then categorised into pre-defined marketing segments. This data will be presented in a user friendly format, as well as be available for export to facilitate further analysis and documentation.

Any research and development outcomes shall be documented and made available to the project client. Through the research, a technological development shall be created which produces accurate data worthy of basing business and marketing decisions upon. A detailed listing of test results will be provided against both tailored and real data, allowing a fair judgment of the software’s accuracy.

The client shall also be provided with the source code and any open source software required to achieve full functionality, as well as technical and user documentation detailing how to set up, maintain, and use the delivered system. Whilst hardware is not provided as part of the project, a demonstrative set up may be requested on provided hardware which meet the specifications outlined in the project’s documentation. Any completed or updated reports, meeting minutes, and project plans or documents of a similar nature shall be shared with the client over the duration of the project.

Constraints

A project management constraint is any hindrance or restriction which may pose limitations to one or more of the project’s stated outcomes. These can consist of both internal and external factors, such as in the forms of cost or physical environment respectively. The ‘triple constraint’ concept depicted by the project management triangle below identifies that all projects are executed under three constraints: schedule, budget and scope (PM4DEV, 2011). It is the aim of the project’s manager to ensure that the project is delivered within the timeframe, on budget and in alignment with the outlined objectives & deliverables (Suresh Krishnan, 2017).



Project Management Triangle

With the project deadline set for just under three months, time will play a major constraint in the planning, research, development and testing required to achieve a successful project outcome. For this reason it is important to ensure a comprehensive project plan is devised, whilst allowing for inevitable minor setbacks. This will be a major consideration in the choice of application development methodology to be adopted, and the client has been made aware that once initial criteria have been defined, it would be challenging to alter them during development without extending the project deadlines.

Data science deals in six main categories, as identified by the hierarchy of data science needs in Appendix B. These are collect, move/store, explore/transform, aggregate/label, learn optimise and apply AI & deep learning algorithms to the data. Due to technical limitations and time constrains in learning such a vast new topic, AI and machine learning techniques will not be implemented into the characterisation and segmentation software.

There is a possibility that the data set provided will not be complete enough to determine real world accuracy, increasing the reliance on A/B test data to prove accuracy; therefore also increasing the time required in the testing phase. Newly inputted data over time provides the opportunity to improve the analytical model and categorisation weighting values using backwards propagation techniques. However, this touches upon concepts of

artificial intelligence and neural networks, which may be beyond the scope of this project

- therefore is it likely that manually updatable category weighting will be used (Margaret Rouse, 2015). This therefore requires the deployed program to be periodically improved for accuracy based on real world comparative analysis and future research studies into behavioural and predictive shopping.

Legal & Ethical Considerations

Due to the nature of data being analysed, it is important to inform the involved customers that their data will be collected for processing, and how it shall be used to the extent allowed by law. The customers should be made aware of this prior to data collection, and should be given the option to opt out. Customer data involved in this project’s research and development was collected as part of a membership agreement program between the company and its consumers. All data is processed in accordance with the company’s customer data policy, and is anonymised to the extent possible whilst maintaining research legitimacy and development functionality.

Privacy is of paramount importance as informed guesses on personal characteristics are being assigned to persons as a result of the analysis required. Some of these characteristics may be personal to the customer, such as political views, allergy requirements and relationship status. For such reasons, the information acquired as a result of this project must be stored securely and not only used within the limits allowed by law, but also in an ethical manner. It is also to be remembered that any information generated by the application is not to be treated as fact, but an educated guess based on a potentially limited data source.

Risk Assessment

There exists the risk that the software will not produce an accurate characteristic & personality assessment of the customers in the operating target market, ergo eﬀecting segmentation results. This may either be due to a fault in programming logic, or an inaccuracy in the research data which the software shall be based on as a result of market discrepancies. This can be minimised through a beta launch eﬀecting a small portion of customers, in which results will be closely monitored and any issues may be addressed prior to the final launch. Once a period of time has passed without any significant noticeable discrepancies between generated customer assumptions and observed/surveyed results, the risk of inaccuracy in the given target market will be minimised.

There also exists an unlikely situation in which the client may not be able to dedicate the necessary level of input or feedback to develop or evaluate the project to its full potential, although this risk is mitigated due to ensured mutual benefits gained as a result of the

LP30483A1 Research Project Action Plan

utilising realistic mock data generation as a replacement in the technological development phase. Where the client is unable to provide evaluation on the final project outcome or to further assist development, objectives & deliverables will be independently evaluated for quality and successful completion. These shall be determined using pre established metrics and best circumstantial evidence to justify the determined level of project success.

Project Plan

The data science research and development project will take place over 11 weeks, and shall incorporate planning, research, development and testing. Although development will be time consuming, the amount of time dedicated to research is almost equally as large, as gaining an in depth understanding will shape the development requirements and reduce the required time to develop the segmentation software.



A corresponding gantt chart can be found under Appendix C.

The project will incorporate at least one meeting per phase to approve completed work and gain insight into upcoming project objectives. Additional meetings may be included if and when deemed appropriate. An initial meeting will determine the objectives and deliverables, risk assessment, among other details. These will then be incorporated into a formal project action plan to be approved at the meeting scheduled for the end of the planning phase, after which research will begin.

Although this phase begins with identifying and categorising items existing client, these stages can be performed at any time prior to development, as they do not impact personality weighting, behavioural analysis or sorting models to be used. Once all research has been carried out, the development plan can be refined to incorporate any required alterations as a result of research discoveries which have led to new or improved knowledge.

# References

Koen van der Schaaf (2017). How Big Data Analytics can Contribute to the Marketing Performance of Supermarkets. Available at: <http://edepot.wur.nl/422370> Accessed on February 25th, 2019

Monica Rogati (2017). The AI Hierarchy of Needs. Available at: https://hackernoon.com/ the-ai-hierarchy-of-needs-18f111fcc007 Accessed on February 25th, 2019

Team Clarizen (2018). What Are Project Objectives in Project Management?. Available at: [https://www.clarizen.com/what-ar](http://www.clarizen.com/what-are-project-objectives-in-project-management/)e-pr[oject-objectives-in-project-management/](http://www.clarizen.com/what-are-project-objectives-in-project-management/) Accessed on February 25th, 2019

Margaret Rouse (2015). Backpropagation Algorithm. Available at: https:// searchenterpriseai.techtarget.com/definition/backpropagation-algorithm Accessed on February 26th, 2019

Sylvia Moses (2018). The Complete Project Manager’s Guide to Project Deliverables. Available at: [https://www.workamajig.com/blog/project-deliverables-guide](http://www.workamajig.com/blog/project-deliverables-guide) Accessed on February 26th, 2019

Ashley Marron (2017). Project Management 101 – Objectives and Deliverables. Available at: [https://www.arraspeople.co.uk/camel-blog/project-management/pr](http://www.arraspeople.co.uk/camel-blog/project-management/project-)oject- management-101-objectives-and-deliverables/ Accessed on February 26th, 2019

Brian Miller (2008). The Purpose of Project Management and Setting Objectives. Available at: [https://www](http://www.projectsmart.co.uk/purpose-of-project-management-and-setting-).pr[ojectsmart.co.uk/purpose-of-project-management-and-setting-](http://www.projectsmart.co.uk/purpose-of-project-management-and-setting-) objectives.php Accessed on February 26th, 2019

Suresh Krishnan (2017). Project Management Constraints and Examples. Available at: https://pmalphabet.com/project-constraints-examples/ Accessed on February 27th, 2019

PM4DEV (2011). Project Management for Development Organizations . Available at: [https://www](http://www.pm4dev.com/resources/docman/pm4dev-articles/14-the-project-constraints/).pm4dev[.com/resources/docman/pm4dev-articles/14-the-project-constraints/](http://www.pm4dev.com/resources/docman/pm4dev-articles/14-the-project-constraints/) file.html Accessed on February 27th, 2019

SAS Institute Inc & Thomas H. Davenport (2013). Big Data in Big Companies. Available at: [https://www.sas.com/content/dam/SAS/en\_us/doc/whitepaper2/bigdata-](http://www.sas.com/content/dam/SAS/en_us/doc/whitepaper2/bigdata-)

bigcompanies-106461.pdf Accessed on February 27th, 2019

Amarpreet Singh (2018). What is the Importance of Data Analytics for Business. Available at: https://brandlitic.com/importance-of-data-analytics-for-business/ Accessed on February 28th, 2019

Margaret Rouse (2018). Big Data Analytics. Available at: https:// searchbusinessanalytics.techtarget.com/definition/big-data-analytics Accessed on February 28th, 2019

# Further Reading

Mikolaos Katsaras, Paul Wolfson, Jean Kinsey, Ben Senaeur (2001). Data Mining: A Segmentation Analysis of U.S. Grocery Shoppers. Available at: http:// citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.500.7787&rep=rep1&type=pdf Accessed on February 26th, 2019

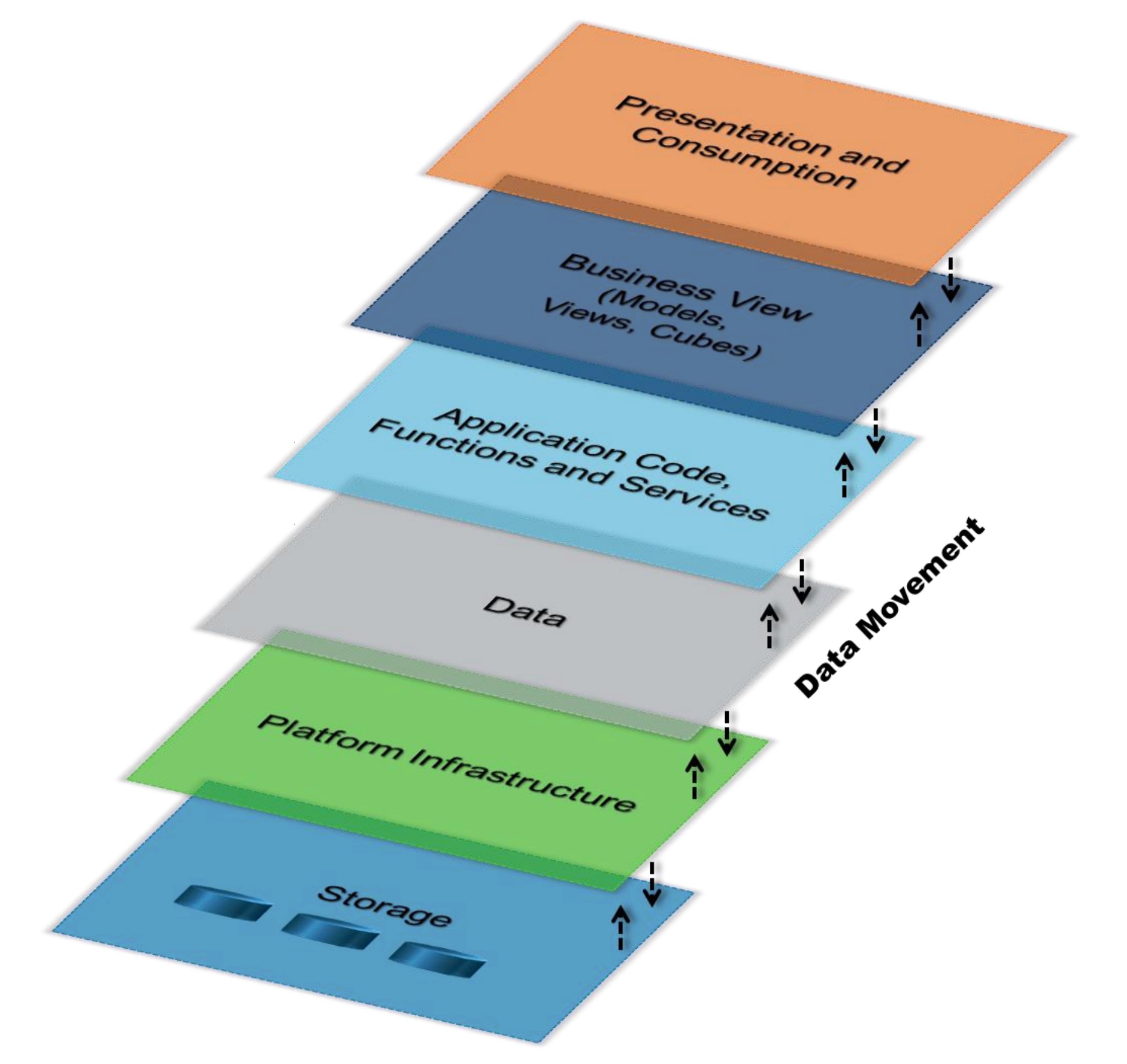
Tim Bowen (2017). How Supermarkets Are Using Big Data & Predictive Analytics To Win. Available at: https://expert360.com/resources/articles/supermarket-retail-big-data Accessed on February 26th, 2019

Bernard Marr (2016). Big Data in Practice: How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results (John Wiley & Sons, 2 May 2016).

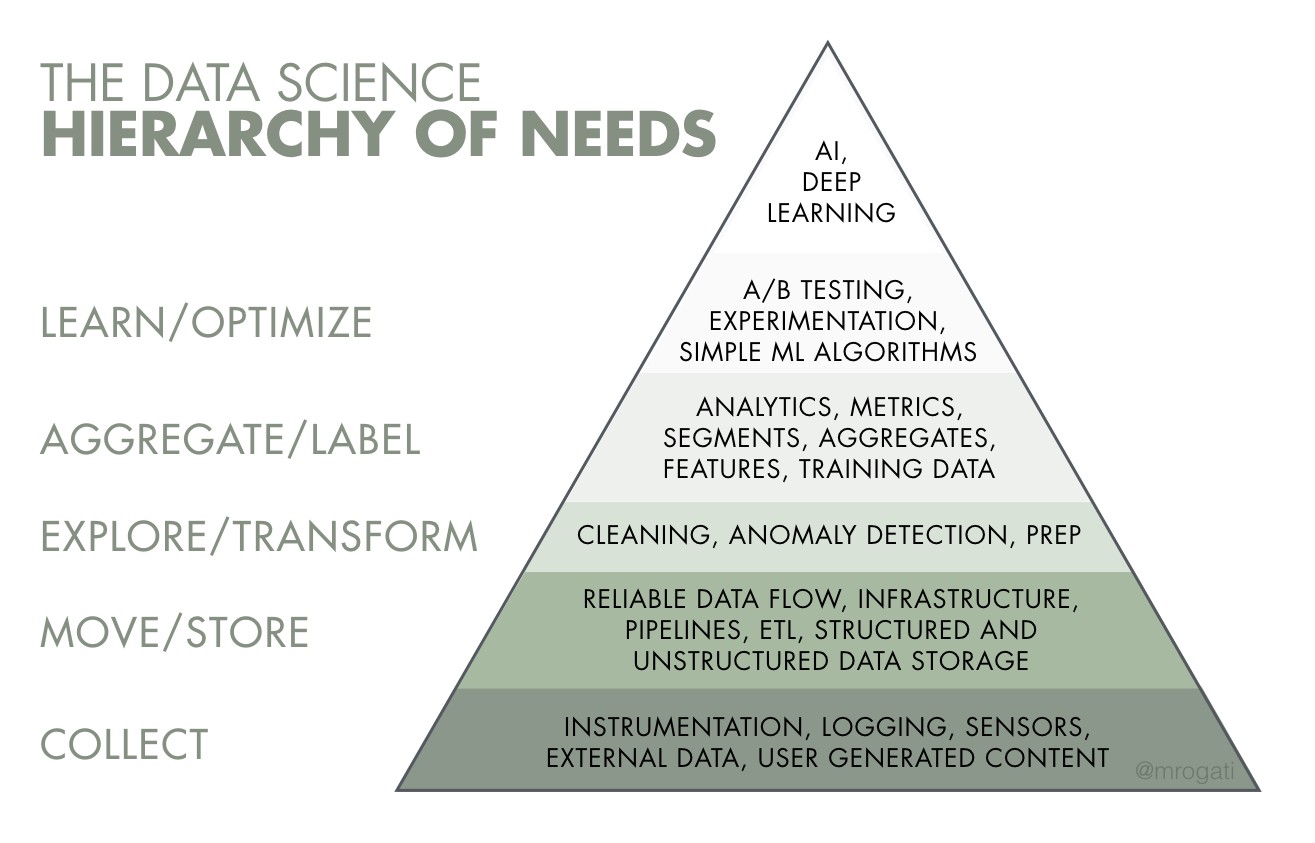
John Appleby (2014). Do supermarkets really need big data to engage customers?. Available at: [https://www.bluefinsolutions.com/insights/john-appleby/february-2014/do-](http://www.bluefinsolutions.com/insights/john-appleby/february-2014/do-) supermarkets-really-need-big-data-to-engage-cus Accessed on February 27th, 2019

Usama Fayyad, Gregory Piatetsky-Shapiro, Padhraic Smyth (1996). From Data Mining to Knowledge Discovery in Databases. Available at: [https://www](http://www.aaai.org/ojs/index.php/).aaai.or[g/ojs/index.php/](http://www.aaai.org/ojs/index.php/) aimagazine/article/viewFile/1230/1131 Accessed on February 27th, 2019

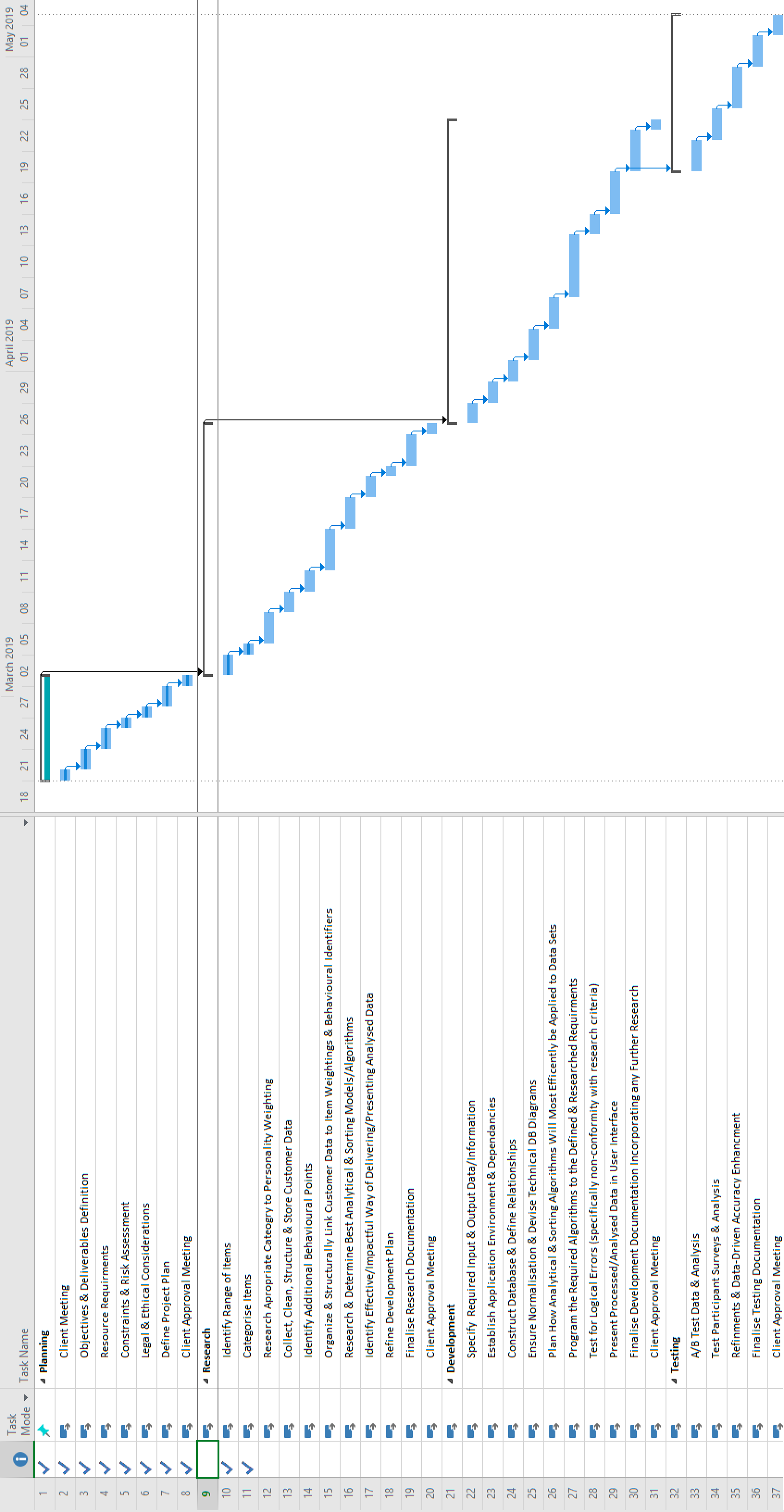
Appendix A: The Big Data Stack



(SAS Institute Inc, 2013)



(Monica Rogati, 2017)



*The project sponsor will shortly be forwarding a document indicating agreement to participation, as well as specifying the amount of time and resources they are able to provide. This document will be submitted separately when it becomes available.*