

Application of a NoSQL database

by Asgar Ali

FILE	4101_ASGAR_ALI_APPLICATION_OF_A_NOSQL_DATABASE_50131_1038834444.DOCX (31.55M)		
TIME SUBMITTED	03-MAY-2019 11:57 PM (UTC+0100)	WORD COUNT	3323
SUBMISSION ID	106300573	CHARACTER COUNT	19178

PROJECT PROPOSAL: APPLICATION OF A NOSQL DATABASE

Supervisor: Michael Dacey
Course: HND Software Engineering
School of Applied Computing
University of Wales Trinity Saint David

3rd May 2019

Author: Asgar Ali
Email: 1706594@student.uwtsd.ac.uk

3500 WORDS

1 TABLE OF CONTENTS

1.	<i>Introduction</i>	3
1.1	E-commerce product catalogue.....	3
1.2	API opportunities	3
1.3	Inefficiency of SQL for e-commerce	4
1.4	Aims and Objectives.....	5
2.	<i>Background research</i>	6
2.1	The Rising popularity of NoSql.....	6
2.2	Mongodb and Mysql comparison	8
2.3	Related work.....	8
3.	<i>Research and development methods</i>	9
3.1	Philosophies.....	9
3.2	Strategies	10
3.3	Evaluation	11
3.4	Risks	12
3.5	Deliverables	12
4.	<i>Resources and planning</i>	13
4.1	Project Resources	14
4.2	Project planning.....	14
5.	<i>References</i>	17
6.	<i>Bibliography</i>	18
7.	<i>Appendix I</i>	18

FIGURE 1 – RISE OF E-COMMERCE IN 2017 IN THE U.K {FORMATTING CITATION}.....	4
FIGURE 2 – REST (REPRESENTATIONAL STATE TRANSFER) [10].....	4
FIGURE 3 – RELATIONAL E-COMMERCE PRODUCT TABLE [5].	5
FIGURE 4 – GEOSPATIAL PRICING.....	6
FIGURE 5 – INSERT OPERATION RESULTS	9
FIGURE 6 – THE RESEARCH ONION.....	10
FIGURE 7 – THREE COMMON RESEARCH STRATEGIES IN APPLIED COMPUTING.....	11
FIGURE 8- NON-FUNCTIONAL TESTING	12
FIGURE 9 – EXCEL PROJECT TASK LIST	15
FIGURE 10 – THE PROJECT GANTT CHART	16
FIGURE 11 - INFOGRAPHIC STYLE CHART	17

TABLE 1 – COMPARISON OF MONGODB TO MYSQL	8
--	---

1. INTRODUCTION

In this section, the type of project will be considered, as well as arguments for its importance and relevance. The aims and objectives will also be briefly discussed and the project outcomes.

This paper will present a research proposal of a comparative study in developing a RESTful Web API. A simple test product catalogue Cloud-based web application will be used to query a non-relational (MongoDB) and relational (MySQL) database. Evaluation of performance, cost, suitability and scalability will be undertaken between the two databases. The title of the project will be: Application of a NoSQL database.

This purpose of this research is aimed at helping e-commerce retailers to evaluate:

1. Whether migrating an existing relational (SQL) database to a non-relational (NoSQL) database will provide cost and time benefits.
2. If new e-commerce start-ups should adopt NoSQL as opposed to traditional SQL databases.

1.1 E-COMMERCE PRODUCT CATALOGUE

When a buy or sell transaction for goods or services occurs online, and the transfer of money and data to execute these transactions occurs, this is referred to as e-commerce [1]. Retailers like Amazon are an ideal example.

E-commerce retailers require a product catalogue, which is a detailed list of the inventory data related to the products for sale, such as images, descriptions, specifications and price. It will be viewed by customers before making purchasing decision. Therefore, it is important to keeping this data updated, ensure consistency and make it easily accessible regardless of the device it is being viewed on. This entire process is known as product catalogue management [2]. It can also become challenging if there is a large product database.

1.2 API OPPORTUNITIES

Traditional U.K. High Street retailers are increasingly becoming a thing of the past, and many have now shifted to the Web [3]. APIs (Application Programming Interface) are allowing many types of these web-based retailers to interact with customers through iOS, Android, and Web applications [4]. This interaction is now critical for an e-commerce retailer as it will bring many opportunities.

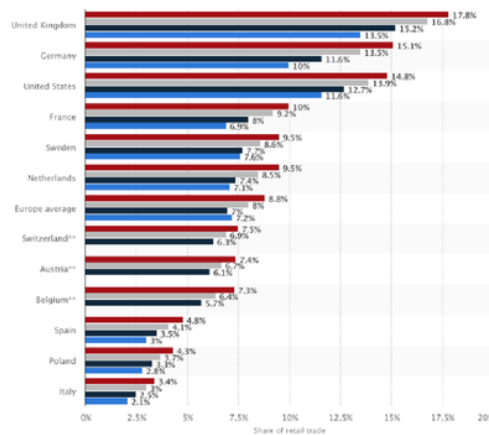


Figure 1 – rise of e-commerce in 2017 in the U.K [5]

An API is based on the REST Architectural standard and is an intermediary application that transfers data between other applications (fig.5). Web API's use the JSON (JavaScript Object Notation) format as it is easy for machines to parse and humans to read and write, for example: {"colour" : "Purple"} [6] [7]. REST compliant systems are often called 'RESTful' [6]. According to rapidAPI.com, the most popular API's are:

- Skyscanner Flight Search – a \$2B company that makes 90% of its money through API's.
- Recipe – Food – Nutrition.
- API Football.
- Open Weather Map.

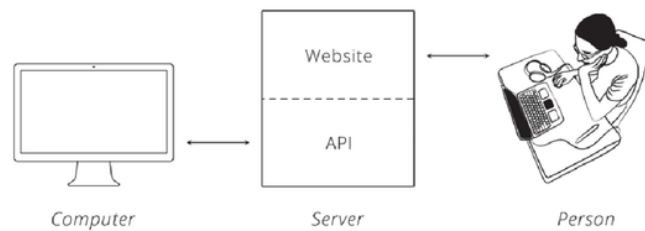


Figure 2 – REST (Representational State transfer) [6].

1.3 INEFFICIENCY OF SQL FOR E-COMMERCE

E-commerce retailers are still overwhelmingly storing their product information in relational databases. MySQL is the most popular open-source relational database management system (RDBMS) as well as being used by

many of the world's most high-volume websites [8]. Relational databases such as MySQL stores data in tables and uses structured query language (SQL) for database access.

The product attribute case

SQL schema is pre-defined with rules to govern relationships between fields in tables, and any changes in schema can be disruptive due to taking the database offline or reduce performance. For example, in Fig.1 each of the five attributes would take up a column in the database table. A row in the table represents the complete set of shirts for sale. Selling shoes and other products would become problematic as more columns would need to be created for other products, which would inefficiently waste memory adding to server hosting costs. This is a common problem for e-commerce retailers [9].

ID	SKU	title	size	color
1	POI0987654321	Awesome Shirt	s	red
2	POI0987654322	Awesome Shirt	m	red
3	POI0987654323	Awesome Shirt	l	red
4	POI0987654324	Awesome Shirt	xl	red
5	POI0987654325	Awesome Shirt	s	blue
6	POI0987654326	Awesome Shirt	m	blue
7	POI0987654327	Awesome Shirt	l	blue
8	POI0987654328	Awesome Shirt	xl	blue

Figure 3 – relational e-commerce product table [9].

1.4 AIMS AND OBJECTIVES

The aim of the project is to compare the effect on cost and performance when building and deploying a NoSQL and SQL database that connects to a RESTful API for a test product catalogue Web application. One Restful API will be built with MongoDB and the other API will be built using a MySQL Database.

1. Build a NoSQL and SQL RESTful API.
2. Evaluate API performance using testing tools.
3. Build a front-end web application using Express.js and Node.js Web technology stack.
4. Evaluate cost and performance using SQL and NoSQL on a PAAS Cloud service.

2. BACKGROUND RESEARCH

2.1 THE RISING POPULARITY OF NOSQL

Over the last 10 years, NoSQL Databases have grown in popularity, with MongoDB being the most popular [10][6][5][3][2]. MongoDB appears to meet modern requirements of high performance, high availability, and automatic scaling [11]. MongoDB uses BSON (Binary JSON) documents to store data and is supported on Windows, Linux and MacOS. The data objects are stored as separate documents inside a collection instead of storing the data into the columns and rows of a traditional relational database. MongoDB is also simple to install and implement. For a product catalogue database, advantages can include:

- Being a JSON document store, MongoDB can help in complex querying of a product catalogue in real time and also keep track of placed orders to accurately calculate the availability of popular products. The data keeps customers aware of available choices at the time of fulfilment and prevents order cancellations later.
- E-commerce retailers need to keep pace with frequent price fluctuations and dynamically adjust product prices for customers. MongoDB has the ability to update data on sites without shutting down or restarting. It has the ability to run geospatial queries using geospatial indexing. For example, this can be used to calculate the correct shipping costs or price based on location (fig.4).

```
db.ATM.find({ location:
  { $geoWithin:
    { $centerSphere: [ [ <longitude> , <latitude> ], 5 / 1123.1 ] } } })

;
{ $centerSphere: [ [ <longitude> , <latitude> ], 5 / 1123.1 ] } } })
```

Figure 4 – geospatial pricing

- MongoDB has an advantage of having geospatial queries to calculate correct shipping estimates based on the location.
- MongoDB is being used by online retailers to consume data from multiple sources and build customised catalogue experiences for individual e-commerce customers on their sites.
- E-commerce retailers can develop customised discounts.

A popular technology stack used alongside MongoDB database to build a RESTful API are Express.js and Node.js [12]. These will be used to build the RESTful API's for the project.

Node.js is an open source JavaScript server environment which runs on Windows, Linux, Unix and Mac OS X. It is designed to be scalable, asynchronous, lightweight and efficient. It has a fast-non-blocking model. It also has the largest open-source library of any programming language, which can be installed using NPM package (Node Package Manager).

Express.js an open-source unopinionated JavaScript web application and API framework that is designed to work on top of Node.js. This will be used to develop the API's and the web application that connects to it.

The intended API will adhere to the following architectural constraints in order to be RESTful:

- **Client-server separation** – By separating the user interface from the data storage, portability is improved across multiple platforms as well as improved scalability. E-commerce retailers may wish to give access to their data in a controlled environment. For example, a Flight booking site such as Skyscanner.com which uses API's to check flight availability and prices from an external Database [13].
- **Stateless** – The server does not remember the API user, for example, if a GET request had already been sent out [13].
- **Uniform interface** – Requests from different clients will look the same e.g. an iOS app or Chrome browser [13].

One disadvantage of an API is there must be an internet connection. Apart from that, there are many advantages to an e-commerce retailer by incorporating API's:

- They can simplify the searching for new customers through different platforms and devices.
- They offer more detailed product information, by connecting the product catalogue with a global database.
- They help customise the website content so it can be tailored to the needs of customers.
- They enable retailers to sell products rapidly, which reduces costs and time-to-market.
- They allow companies to offer their customers a wider range of services in less time.
- They increase innovation by streamline new ideas due to the API's architecture.
- They improve e-commerce retailer's automation processes.
- They provide real-time analytics.

It can be seen that combining an API with a MongoDB database can potentially bring great benefits to small and mid-sized e-commerce retailers, which is the target group for this project, in an ever-competitive business world. The retail API economy is growing especially rapidly, with 1000 available as of 2017, up from 116 in 2013 [14].

2.2 MONGODB AND MYSQL COMPARISONS

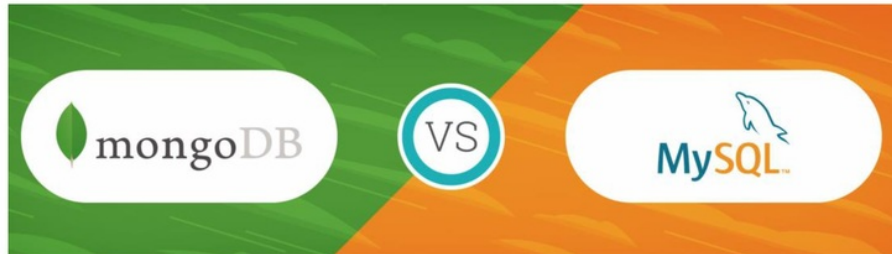


Figure 5 – MongoDB and MySQL official branding [11].

Table 1. illustrates some key differences between the database technologies.

Table 1 – comparison of MongoDB to MySQL

Criteria	MongoDB	MySQL
Storage Model	Document oriented	RDBMS
Agile Practices	Yes	No
Data Schema	Dynamic	Fixed
Scalability	Horizonal	Vertical
XML Support	No	Yes
Concurrency	No	Yes
Map Reduce	Yes	No

2.3 RELATED WORK

There have been many academic studies conducted evaluating the benefits of NoSQL to SQL databases, and a few on MySQL vs. MongoDB. The author believes that this project proposal is unique offering, as there is currently no research available on API benefits and drawbacks between SQL and NoSQL for e-commerce API development.

A paper to a related study of non-relational databases and relational databases was examined by the author. The title of the paper is: A Comparative Study: MongoDB vs. MySQL by GYÖRÖDI et al and was published in June 2015 at Oradea University [15]. It was focused on the implementation of a forum application that used a

MongoDB database and a comparison with MySQL was made. It tried to justify why MongoDB is more efficient than MySQL. For researching performance, four basic operations were used on the databases using a 64-bit Windows 7 computer:

- Insert
- Select (query)
- Update
- Delete.

Fig.5 shows a screenshot of an Insert test result. The MySQL database takes significantly longer to perform this operation at 440 seconds.

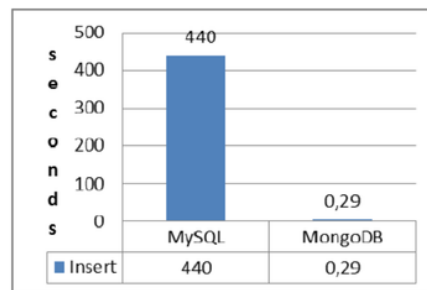


Fig. 4-1 MySQL vs. MongoDB insert

Figure 6 – Insert operation results

They concluded that it was the developers' job to decide which database should be used to meet the application' requirements. For the forum application, they discussed how the test results showed that the most suitable database was found to be MongoDB.

3. RESEARCH AND DEVELOPMENT METHODS

3.1 PHILOSOPHIES

There are two main research philosophies from Epistemology – Positivism and Phenomenology. As this project is about the development of software, positivism is the appropriate philosophy, where research is based on observations which produce data that is quantifiable. It also encompasses the paradigms of objectivism and experimentalism (see fig.3):

- Qualitative data results.

- Uses a deductive approach by testing the theory.
- Specific and precise data.
- Reliability is if often high.

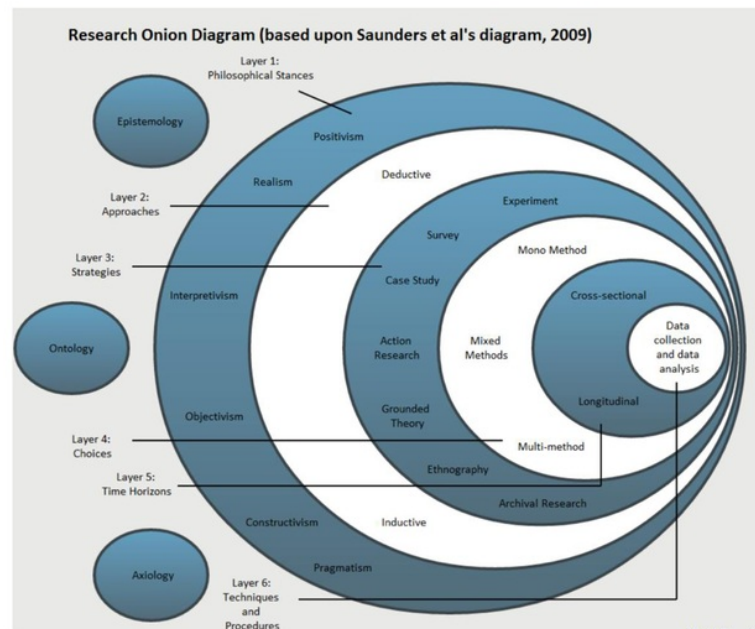


Figure 7 – the Research Onion

An Inductive approach will be used. This approach allows evaluating theories based upon specific examples, e.g. a description of how businesses use Facebook for advertising based upon a survey of local businesses.

3.2 STRATEGIES

The three research strategies from fig.6 are required for this project so that results can be triangulated.

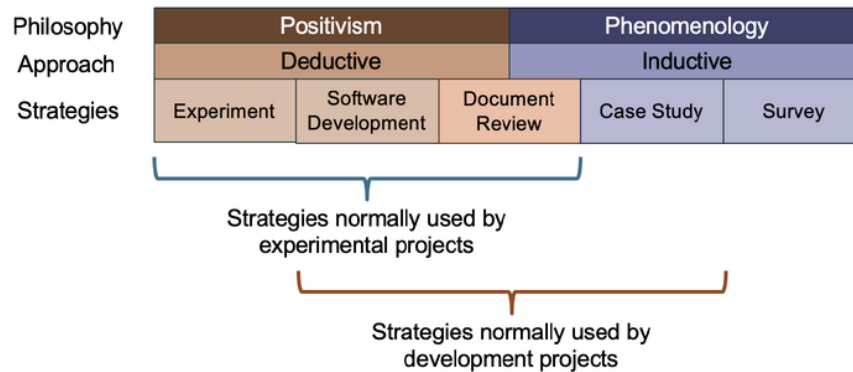


Figure 8 – three common research strategies in applied computing.

1 Literature review - The aims of this strategy are to:

- Identify the main theories and research in the field.
- Risks that may affect the project.
- Confirm that the project is realistic and practical.
- Provide something that you can compare your results to
- Identify if the work will support or contradict the findings of previous studies.

2 Software development - This strategy involves developing software and testing that it works effectively.

3 Experimentation

Sets out to examine the relationship between two factors (or variables) by manipulating one whilst measuring changes in the other. For instance, we may test out two different software to measure the effect on usability.

3.3 EVALUATION

The main data that will be evaluated and tested are performance, cost, and scalability. Suitability of each database technology will be evaluated based on the data.

Performance testing – is a non-functional testing technique used to determine how an application will behave under various conditions. The goal is to test its responsiveness and stability in real user situations. Automated API testing tools will be implemented. A focus on the speed of CRUD operations with comparative bar charts provided will be generated. These will show the time taken for different operations.

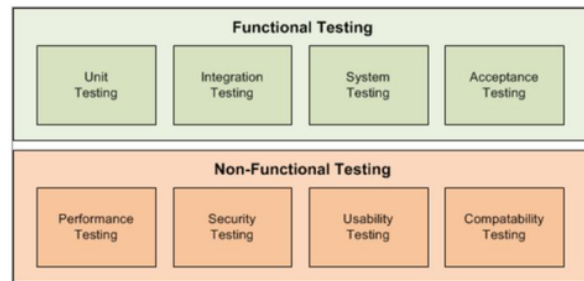


Figure 9- non-functional testing

Cost – This will be calculated based on how much Cloud server resources the SQL based API and NoSQL based API will consume, taking into consideration the pricing policy of the host platform.

Scalability – How difficult it would be to scale the RESTful API databases with greater data and traffic being generated. Again, popular API test automation tools such as Postman or SoapUI are available for this purpose [16] [17].

Suitability – how each database technology is suited to the product catalogue web application and API development in general.

3.4 RISKS

There are several risks to the project that have been identified and will need to be resolved before and during the project:

- Poor time management leading to project deadlines not being met.
- Workload constraints leading to project deadlines not being met.
- Requirements of project not being met due to unrealistic goals.
- Inconsistency in performance data/results between different automated testing software/services.

Ethical issues can also be considered a risk, although only test product catalogue data will be used for the research. When considering databases, these may hold confidential or sensitive information, therefore step must be put in place to keep this data from going to third parties and industry standard security measures should also be implemented and considered when designing the software and the front end website that will be used to query information from the MongoDB and MySQL databases. For example, using prepared statements when designing the authorisation and authentication of the front-end website that will connect to the API's on the cloud service.

3.5 DELIVERABLES

An SQL and NoSQL API will be created on the authors computer using Visual Studio Code. A Bash Terminal will be used for installing MongoDB. The API will be deployed to a Cloud service. A popular Node.js Platform-as-a-Service (PAAS) provider will be considered such:

- Heroku
- Google Cloud
- Microsoft Azure

API testing will be performed to determine if they meet expectations for functionality, reliability, performance, and security. This will be the minimum success criteria for the project.

At the end of the project, it is expected that a JavaScript Web Application that will consume the RESTful API will be designed and built. Options will be created for registering, updating, deleting and searching product information. Twitter Bootstrap will be used for the Website styling. Login will be created for authorisation.

API documentation will be provided with the project report, which will be submitted to the University, for an overview of how everything works and will provide instruction on common tasks:

- Overview and use case
- Getting started
- Key Concepts
- Workflow
- Architecture
- Authentication
- Tutorial
- Sample Code
- References

The report will also include screenshots, feedback from testing, recommendations for improvement and final evaluation of SQL vs NoSQL for API development of a product catalogue.

4. RESOURCES AND PLANNING

4.1 PROJECT RESOURCES

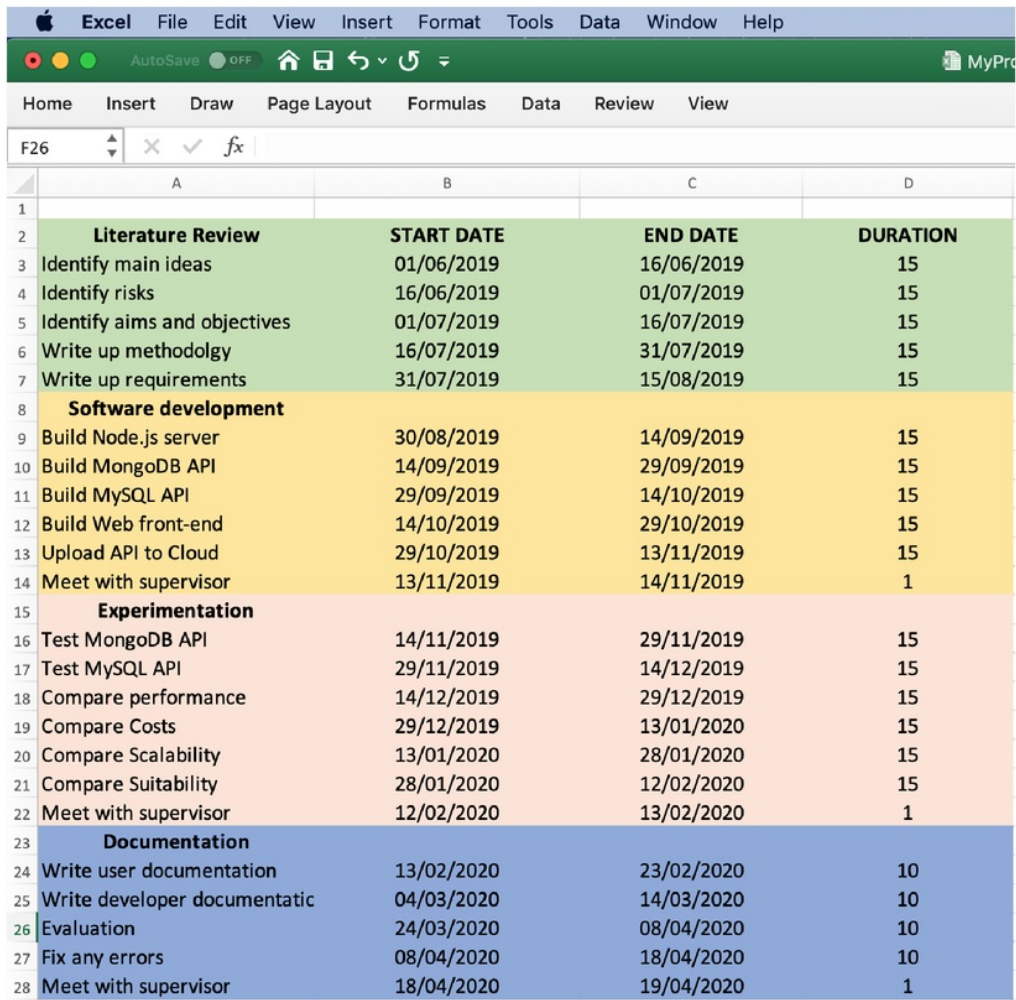
In order to successfully complete the project, the following resources will be required

- **Apple MacBook Pro 15" Laptop** – to be used for the building of RESTful API and Web application.
- **Visual Studio code** – A popular source-code editor that will be used to create the API and Web Application. Node.js, NPM and Express.js will be installed on it via the command line tool.
- **Bash Shell Terminal** – This is pre-installed on MacOS and will be used to install MongoDB.
- **MongoDB database** will be downloaded from MongoDB.com.
- **Parallels 14** – This is a commercial Virtual Machine that will be used to interact with Windows 10.
- **MySQL** – for the build of the actual system accessed via a web front end
- **Kindle and Physical Books** – These will be purchased from Amazon.co.uk: *Restful API design Rulebook*; *Restful Web API design with Node.js*; *Web development with Node and Express*.
- **Postman** – A free API Tool
- **Swagger** – A free API Tool
- **Curl** – A Command Line Tool.
- **Benchmarking tool**
- **Microsoft Office** – For project management.

It is possible that during the course of the project a resource may be required that has not been previously considered.

4.2 PROJECT PLANNING

Microsoft Excel was used to break up parts of the project into manageable 15-day timeslots. A task list was initially created, then this was used to construct a Gantt chart that showed the project timeline.



	A	B	C	D
1				
2	Literature Review	START DATE	END DATE	DURATION
3	Identify main ideas	01/06/2019	16/06/2019	15
4	Identify risks	16/06/2019	01/07/2019	15
5	Identify aims and objectives	01/07/2019	16/07/2019	15
6	Write up methodology	16/07/2019	31/07/2019	15
7	Write up requirements	31/07/2019	15/08/2019	15
8	Software development			
9	Build Node.js server	30/08/2019	14/09/2019	15
10	Build MongoDB API	14/09/2019	29/09/2019	15
11	Build MySQL API	29/09/2019	14/10/2019	15
12	Build Web front-end	14/10/2019	29/10/2019	15
13	Upload API to Cloud	29/10/2019	13/11/2019	15
14	Meet with supervisor	13/11/2019	14/11/2019	1
15	Experimentation			
16	Test MongoDB API	14/11/2019	29/11/2019	15
17	Test MySQL API	29/11/2019	14/12/2019	15
18	Compare performance	14/12/2019	29/12/2019	15
19	Compare Costs	29/12/2019	13/01/2020	15
20	Compare Scalability	13/01/2020	28/01/2020	15
21	Compare Suitability	28/01/2020	12/02/2020	15
22	Meet with supervisor	12/02/2020	13/02/2020	1
23	Documentation			
24	Write user documentation	13/02/2020	23/02/2020	10
25	Write developer documentation	04/03/2020	14/03/2020	10
26	Evaluation	24/03/2020	08/04/2020	10
27	Fix any errors	08/04/2020	18/04/2020	10
28	Meet with supervisor	18/04/2020	19/04/2020	1

Figure 10 – Excel project task list

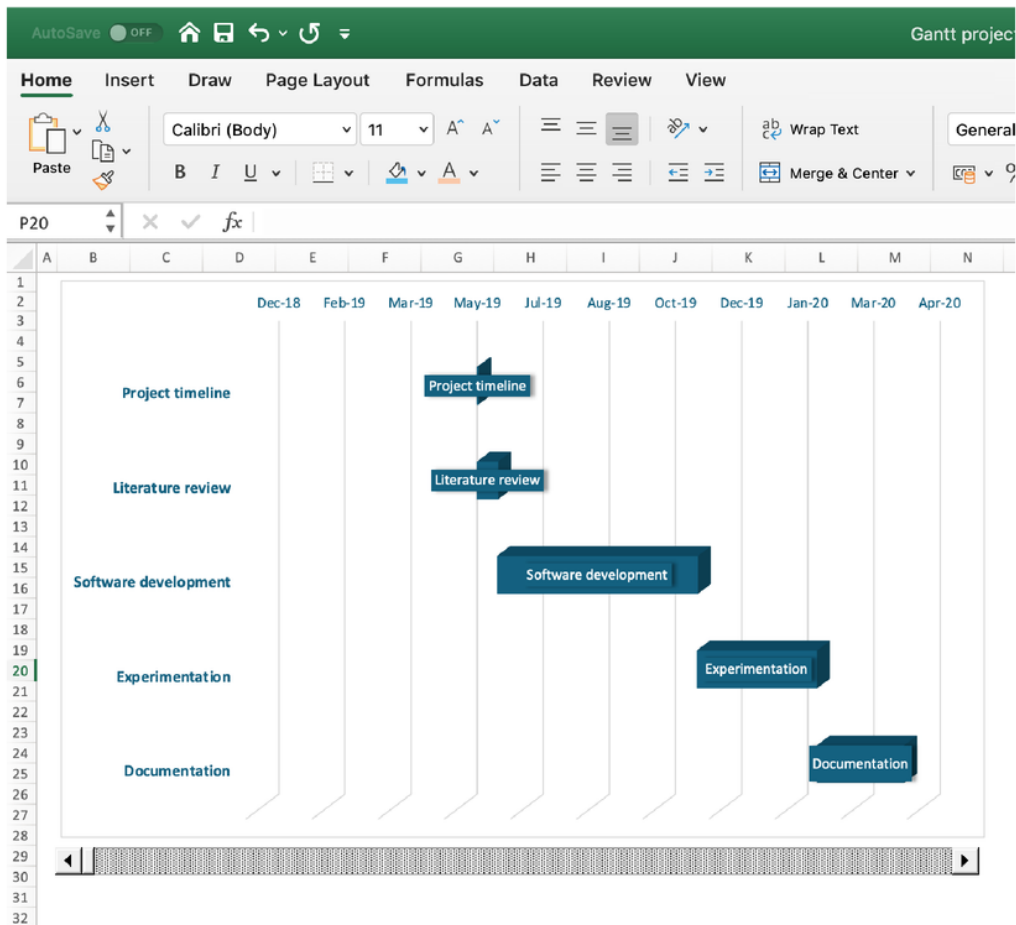


Figure 11 – the project Gantt chart

Project management will be needed to successfully complete the project. In order to plan and manage each step of the project a Gantt chart has been constructed (see figures 4-8).

The use of the Microsoft Project's Gantt chart with PERT analysis tools should help limit the risks surrounding time management and unrealistic goals. By using a Gantt chart, it should be possible to identify the critical path through the project and the use of the PERT analysis with the calculation weighted more heavily towards the pessimistic should allow for a more realistic timeline to be proposed. Unfortunately, there isn't much that can be done about workload constraints other than being mindful of what is due in the coming months.

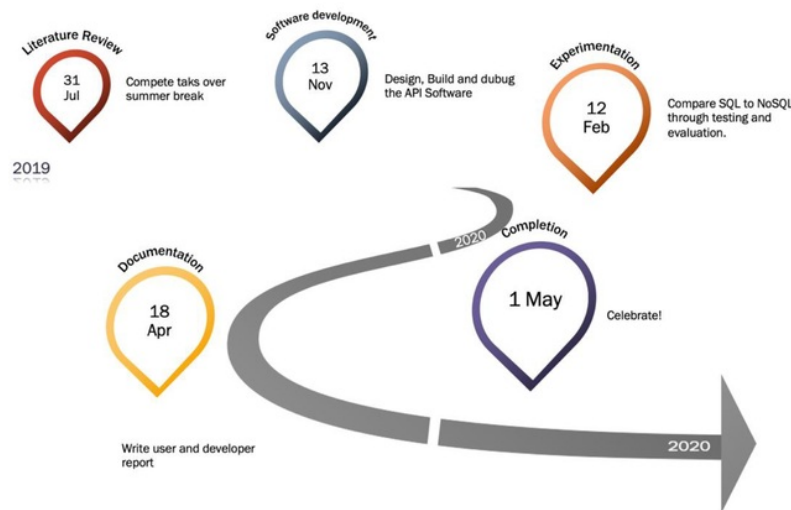


Figure 12 - PERT style chart designed on Excel.

5. REFERENCES

- [1] "What is ecommerce? Ecommerce explained with examples." [Online]. Available: <https://ecommerceguide.com/guides/what-is-ecommerce/>. [Accessed: 02-May-2019].
- [2] "(1) What is meant by Product Catalogue and Product Catalogue Management? - Quora." [Online]. Available: <https://www.quora.com/What-is-meant-by-Product-Catalogue-and-Product-Catalogue-Management>. [Accessed: 01-May-2019].
- [3] "How APIs are Driving Innovation in Retail | API Management Blog - Akana." [Online]. Available: <https://blog.akana.com/apis-in-retail/>. [Accessed: 02-May-2019].
- [4] "What is REST – Learn to create timeless RESTful APIs." [Online]. Available: <https://restfulapi.net/>. [Accessed: 30-Apr-2019].
- [5] "• Online share of retail trade in countries 2017 | Statistic." [Online]. Available: <https://www.statista.com/statistics/281241/online-share-of-retail-trade-in-european-countries/>. [Accessed: 03-May-2019].
- [6] "An Introduction to API's – The RESTful Web." [Online]. Available: <https://restful.io/an-introduction-to-api-s-cee90581ca1b>. [Accessed: 02-May-2019].
- [7] "What is a REST API? – Charlie Brooke O'Hara – Medium." [Online]. Available: <https://medium.com/@charlie.b.ohara/what-is-a-rest-api-79c78de3a6fc>. [Accessed: 30-Apr-2019].
- [8] "DB-Engines Ranking - popularity ranking of database management systems." [Online]. Available: <https://db-engines.com/en/ranking>. [Accessed: 30-Apr-2019].
- [9] "A Better Way to Store Ecommerce Product Information | Practical Ecommerce." [Online]. Available: <https://www.practicalecommerce.com/A-Better-Way-to-Store-Ecommerce-Product-Information>.

[Accessed: 02-May-2019].

- [10] “Why MongoDB is the most popular NoSQL database today | Packt Hub.” [Online]. Available: <https://hub.packtpub.com/mongodb-popular-nosql-database-today/>. [Accessed: 30-Apr-2019].
- [11] “The SQL vs NoSQL Difference: MySQL vs MongoDB – Xplenty Blog – Medium.” [Online]. Available: <https://medium.com/xplenty-blog/the-sql-vs-nosql-difference-mysql-vs-mongodb-32c9980e67b2>. [Accessed: 02-May-2019].
- [12] “How To Build Simple RESTful API With NodeJs, ExpressJs And MongoDB.” [Online]. Available: <https://medium.com/@dinyangetoh/how-to-build-simple-restful-api-with-nodejs-expressjs-and-mongodb-99348012925d>. [Accessed: 02-May-2019].
- [13] “What is REST — A Simple Explanation for Beginners, Part 2: REST Constraints.” [Online]. Available: <https://medium.com/extend/what-is-rest-a-simple-explanation-for-beginners-part-2-rest-constraints-129a4b69a582>. [Accessed: 30-Apr-2019].
- [14] “The Value of Retail APIs | MuleSoft.” [Online]. Available: <https://www.mulesoft.com/resources/api/retail-apis>. [Accessed: 02-May-2019].
- [15] Györödi, “A Comparative Study: MongoDB vs. MySQL,” *13th International conference on Engineering of Modern Electrical Systems, Oradea, 2015*. [Online]. Available: https://www.researchgate.net/publication/278302676_A_Comparative_Study_MongoDB_vs_MySQL. [Accessed: 02-May-2019].
- [16] “Postman | API Development Environment.” [Online]. Available: <https://www.getpostman.com/>. [Accessed: 02-May-2019].
- [17] “The World’s Most Popular API Testing Tool | SoapUI.” [Online]. Available: <https://www.soapui.org/>. [Accessed: 02-May-2019].

6. BIBLIOGRAPHY

7. APPENDIX I

Application of a NoSQL database

GRADEMARK REPORT

FINAL GRADE

GENERAL COMMENTS

Instructor

55/100

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9

PAGE 10

PAGE 11

PAGE 12

PAGE 13

PAGE 14

PAGE 15

PAGE 16

PAGE 17

PAGE 18

GRADING FORM: NEW ASSESSMENT

ASGAR ALI**55****ENGAGEMENT WITH LITERATURE**

An understanding of some of the importance of relevant technologies such as MongoDB and REST has been demonstrated. I think the project title may need to be amended to reflect the importance of REST in this project.

13**KNOWLEDGE AND UNDERSTANDING**

The aims and objectives could have been written a bit more formally, but the project has merit. The introduction could have concentrated more on the potential benefits of NoSql instead of digressing into eCommerce.

5**COGNITIVE, INTELLECTUAL SKILLS**

There is no discussion of the software design approach that will be used. Some tools are listed in the resources section.

6

PRACTICAL SKILLS

 The research methods are briefly discussed but I would like to have seen a clearer description of the overall research process and how the results will be obtained/evaluated. 13 A project plan is provided and if achieved then the project will be successful. Some of the planning does seem a bit ambitious. 10

23

TRANSFERABLE SKILLS

 The presentation is a little scruffy and the flow of the written work could have been better..

8