

# SIT226 Cloud Automation Technologies

## Pass Task 6.2C

### Database Replication

#### Background

Database Management Systems (DBMS) are critical components in many modern network services and web applications. By using Docker and Kubernetes, we can easily deploy program code, static files, and even share files and folders between replicas of network services and web applications. Databases that support replication are somewhat more complicated. In this task, we explore the support for replication in modern (open source) databases.

#### Get Prepared

*Note: if you are targeting a distinction grade you should also review the requirements of Task 6.3D, which requires you to implement your chosen replicated database.*

There are many DBMS available today, such as Apache Cassandra, Elasticsearch, IBM Db2, MariaDB/MySQL, Microsoft SQL, MongoDB, Neo4J, Oracle, PostgreSQL, Redis, and so on. In this task, you will complete a review of the support for replication provided by one of these databases.

Review the information available regarding replication for several of these databases and then select an appropriate one for this task. Note that the task is addressing databases that manage the distribution and replication of databases automatically, not manually/administratively driven copying of databases.

#### Complete the Task

Page Limit: 1 page of text formatted reasonably, e.g., 2cm margins, 11 or 12 point font, appropriate headings/spacing, etc.

From the understanding you have developed of support for replication your selected database offers, write a brief report according to the following requirements:

1. Provide a brief description of the database, including the type of database (relational, object, key/value, etc.), typical applications (what it's used for), and why the database is preferred/advantageous for those applications (approximately 0.2 page).
2. Explain what support the database provides for replication and any requirements, e.g., any modes/configuration of replication that is possible, hardware and software requirements/dependencies (approximately 0.4 page).
3. Discuss the implications your chosen database when replicated has for applications/software development effort (approximately 0.4 page).

*Hint: to answer this question, consider how an application may need to change being developed for a centralised database vs a distributed/replicated database. Often reviewing/reflecting on how the database handles modifications to data (writes) is revealing here, but also how to improve the overall performance of the application given database replication, e.g., caching, can be relevant.*

## Submit Your Task

Prepare your submission using the word processor of your choice and submit a PDF to OnTrack.

## Taking it Further (Optional)

The world of databases extends vastly beyond the typical introduction to relational databases that you will find in most IT courses. Relational databases have been available and in use for many years, and remain the dominant database type in use today with many significant products you should recognise: IBM Db2, MariaDB, Microsoft SQL, MySQL, Oracle, and PostgreSQL. Although these databases support replication, which allows the database to scale to larger applications, it is limited. Relational databases are also not good at supporting every possible scenario, as the (relational) structure doesn't always match the needs of the application. Consider the following issues, some of which are already touched on in the question:

- Review the different types of databases that are available, e.g., relational, column-oriented, key-value, document, and so on. What are the differences in these types? What are they used for? What are the functional advantages and disadvantages of these database types? (replication is only one function). And so on.
- Consider some of the significant database products you may have seen mentioned, e.g., see the Get Prepared section for this task, but also others you may have heard of such as etcd (used by Kubernetes). What motivated the creation of these products? What are they used for today? Are they only commercial databases or are there open source versions/alternatives?
- Review the database offerings of the major public cloud providers. How do these offerings relate to each other and the databases you may have considered for the previous points? For example, Amazon offers Amazon Aurora, Amazon RDS, and Amazon RedShift alternatives for relational databases, or you could deploy any of the relational databases mentioned previously to a VM. What is the difference between these and why would you choose one over another?
- How are databases used? We know that databases are used by applications, but they are also often used by the operating system, e.g., Windows Registry. What other applications can you find that use databases and why?
- How do databases work? For example, when you do a web search (using any provider) there is clearly a database behind that service. How have those providers used databases to provide near-instantaneous results when searching vast amounts of the Internet? You might actually want to consider how databases are developed (how the code works).

## Citations and Referencing

When completing any work, it is necessary to acknowledge the work of others that you have relied upon. For written assessment, we achieve this through the use of citations and references. Failing to correctly identify the work of others is known as plagiarism and is considered an issue of Academic Integrity.

If your submission to this task has involved the work of others, you must include citations and references where appropriate. Deakin provides a website that explains how to use citations and references, and includes explanations of various referencing styles:

<https://www.deakin.edu.au/students/studying/study-support/referencing>

You may select any style for your citations/references, however, you must be consistent in applying that style in this task (you can use other styles in other tasks if you wish).

Note that any bibliography/list of references is not included in page limits.