# Topic 2 – Version control

This document is the handbook for Topic 2 – **Version control** – within Module 3 – **Programming and Scripting Essentials**.

The purpose of this document is to guide your learning throughout this topic and help you to maximise the value you get from the materials provided by the BPP School of Technology.

Context

This handbook is for one of 9 topics for this Module.
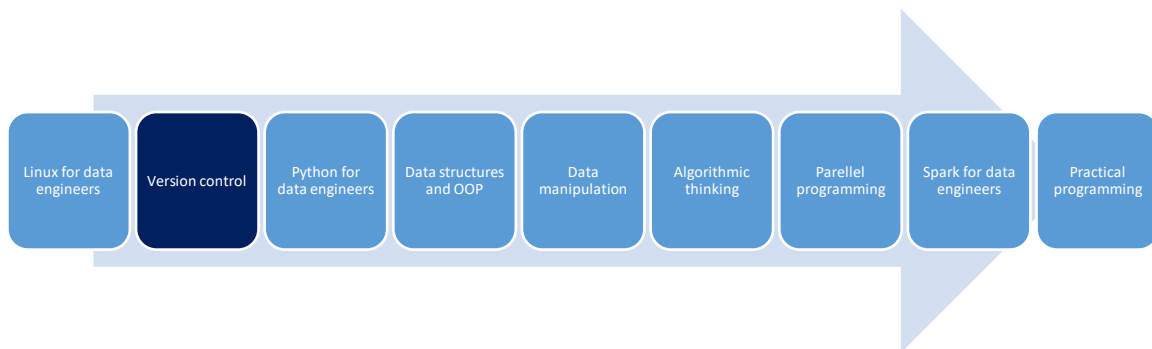
Every topic contributes towards the ultimate learning objectives for the Module, which you will be assessed on at the end of the term.

Module Learning Outcomes

On successfully completing this module, you will be able to:

- **Employ** software development tools and techniques for designing, deploying and maintaining secure data products and pipelines, including debugging, version control and testing.
- **Construct** algorithms that correctly and efficiently handle data at scale whilst mitigating risks.
- **Demonstrate** the knowledge of the steps needed to prepare the code for production.

Module Assessment

The Level 5 Data Engineer EPA has two assessment methods, each with its own mapping of KSBs. The Assessment plan and assessment guidance documents above list the criteria and KSBs that are assessed. The criteria group the KSBs and describe what the apprentice needs to do to achieve a pass or distinction for that assessment method.

Both assessment methods need to be passed by the candidate:

(1)   Project with report

The learner will complete a project and write a report of 3500 words. Project brief submitted at gateway:
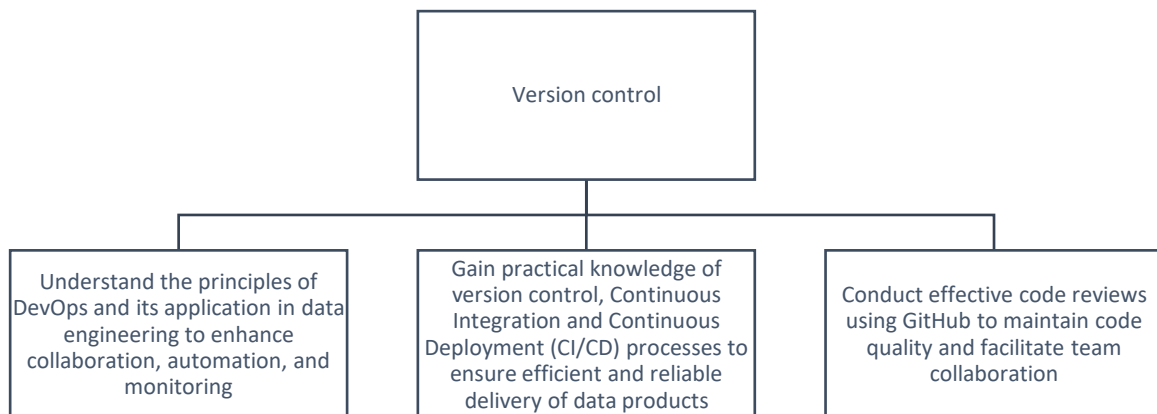
- Learners will have 10 weeks to complete the project and submit the report to the EPAO
- Learners also need to prepare and give a presentation to an independent assessor on their project
- The presentation with questions will last at least 50 minutes. The independent assessor will ask at least 6 questions about the project and presentation
- The project has to have real business application and benefit. Candidates are expected to showcase the use of appropriate standards for sustainability, privacy and security, thoroughly document their data pipeline designs, explain the choice of relevant tooling and demonstrate operational awareness of deployment, access control, risks, and how other stakeholders may be impacted positively and negatively

(2)   Professional discussion underpinned by a portfolio of evidence

- Learners will have a professional discussion with an independent assessor. It will last 80 minutes
- They will be asked at least 10 questions about Data Engineering
- The portfolio of evidence will be used to help answer the questions
- We expect the candidates to demonstrate examples of working with data teams on data projects and data products, showcase ideas for future-proofing data, be clear on applying problem-solving skills, show regulatory awareness, and sensitivity towards data quality, data governance and areas for continuous improvement, both personal and organisational

Topic Learning Outcomes

As a step towards build your skills towards the final module assessment, the learning objectives for this topic are:

```
                          ┌─────────────────────────┐
                          │     Version control     │
                          └─────────────────────────┘
```

Understand the principles of DevOps and its application in data engineering to enhance collaboration, automation, and monitoring	Gain practical knowledge of version control, Continuous Integration and Continuous Deployment (CI/CD) processes to ensure efficient and reliable delivery of data products	Conduct effective code reviews using GitHub to maintain code quality and facilitate team collaboration

Introduction

In the rapidly evolving world of technology, the ability to manage and track changes in code is a critical skill. This is where the concept of 'Version Control' comes into play. Version control systems are a category of software tools that help a software team manage changes to source code over time. They allow you to keep track of what was changed, when it was changed, and who changed it.

For Level 5 data engineering students, understanding and implementing version control is particularly relevant. As data engineers, you will be working with large volumes of data and complex algorithms. Changes and improvements to these algorithms are a constant part of the job. Without a robust version control system in place, managing these changes can become a daunting task.

BPP

Consider the real-world case of a multinational corporation with a global team of data engineers working on the same project. Without version control, coordinating tasks, tracking changes, and maintaining the integrity of the project would be nearly impossible. However, with a version control system like Git, changes from multiple sources can be seamlessly integrated, tracked, and even rolled back if necessary.

By the end of this course, you will have gained a deep understanding of version control systems, particularly Git. You will learn how to track changes, resolve conflicts, and manage branches in a collaborative coding environment. These skills are not just limited to your academic projects; they are highly sought after in the professional world. Whether you're working in a small startup or a large corporation, version control is an essential part of the software development lifecycle. Mastering it will give you a significant edge in your future career as a data engineer.

Structure

Topics for this programme follow a Prepare-Collaborate-Apply structure:

Prepare

This is the stage where you build the knowledge to underpin your learning. This might involve completing interactive e-learning packages, watching videos, or working through reading materials.

It is essential that you make the most of the learning materials provided before attending webinars, as this will allow you to test your knowledge and stretch you understanding further.

Collaborate

This is where you will receive guidance from our expert tutors and coaches to shape and refine your understanding through in-depth explanation, discussion, testing and carrying out more advanced practical and realistic tasks. This also helps to develop valuable team-working skills.

Apply

You now apply the knowledge you have developed to real-world tasks.

Off-the-job learning tasks

This stage is all about ensuring you truly grasp and retain what you've learned. Through completion of off-the-job (OTJ) revision tasks and tests, you'll get plenty of practice applying your knowledge. Plan to dedicate 6-8 hours each week to guided study and portfolio work, with sessions typically on the same day each week.

Task 1 brief:

Further guidance can be found here:

Link

This handbook is for one of 9 topics for this Module.
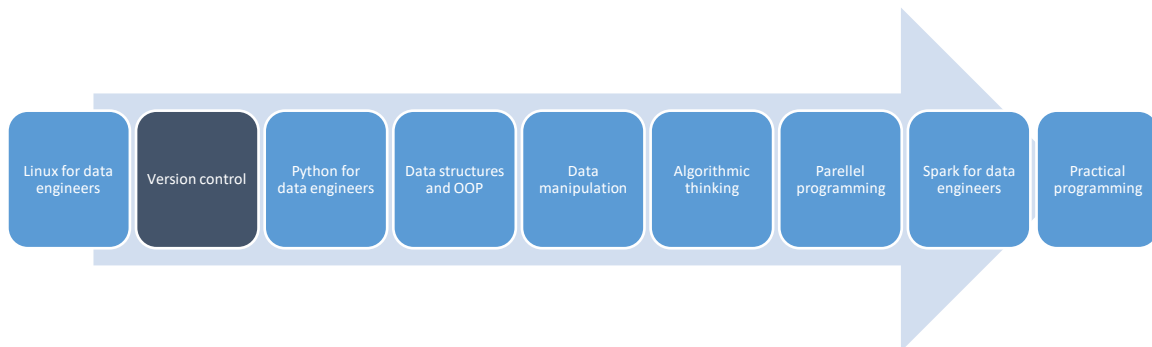
The sequence of topics in this module is carefully designed so that your knowledge and skills will develop as you progress.

The next topic is **Python for data engineers**.