



MICRO-SCHOOL

NITheCS

National Institute for
Theoretical and Computational Sciences

Getting Started with Data Engineering with Apache Spark

Small Tshithavhana (University of the Witwatersrand)

Fri, 13 September 2024 | 12h00-12h30 SAST

Attend online

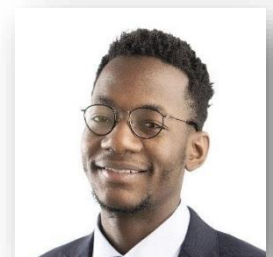
ABSTRACT

In this micro-school, we explore the fundamentals of Data Engineering with a focus on parallel and distributed computing. Our goal is to build a foundational understanding of how large-scale data processing works, utilizing one of the most prominent frameworks in this field: Apache Spark. We will delve into Spark's capabilities for handling vast amounts of data efficiently and introduce its practical applications through a hands-on session using Databricks—a powerful environment for Spark-based data processing.

This micro-school is designed for those who are new to Data Engineering or those who work with large amounts of data for their research and are seeking to enhance their knowledge in parallel and distributed computing. It is ideal for individuals who are interested in learning how to manage and analyse large datasets effectively.

BIOGRAPHY

Small is a MSc candidate at the University of the Witwatersrand. He initially did his undergraduate degree majoring in Mathematics and Physics at the University of Venda. He forwarded his studies in Honours, majoring in Mathematics and Computer Sciences, conducting his research focused on Inner Product on the Space of Complex Square matrices. Currently he is doing his Masters, which is focused on the application of Bayesian Networks and data sciences in banking or fraud detection.



REGISTER: <https://bit.ly/4dOt7QI>

