_Pk.png _Pk.bb _Pk.png height

Contribution ID: **52** Type: **Talk**

Processing longitudinal population data using CHPC

Wednesday, 1 December 2021 15:15 (30 minutes)

The South African Population Research Infrastructure Network (SAPRIN) curates longitudinal population data collected by four nodes from a total population of more than 400 000 individuals. Due to the dynamic nature of these study populations data representing episodes of individual surveillance needs to be combined in a way that maintains data integrity and takes into account variations between data collection sites.

We need to deconstruct 4,5 million person years of observation into a day level dataset, requiring the kind of processing and storage capacity provided by a high performance computing environment such as CHPC.

We will describe a data processing pipeline, originally developed in Pentaho and recently converted to the julia programming language which scales well on the CHPC environment.

Student?

No

Supervisor name

Supervisor email

Primary authors: Dr HERBST, Kobus (SAPRIN); MAOYI, Molulaqhooa (SAPRIN); MUTEVEDZI, Tinofa (SAPRIN); COLLINSON, Mark (SAPRIN)

Presenter: Dr HERBST, Kobus (SAPRIN)

Session Classification: NICIS Cloud Projects

Track Classification: NICIS Cloud Projects