Seismic Data Management and Processing with MsPASS

Yinzhi Wang, Gary Pavlis

MsPASS, or Massive Parallel Analysis System for Seismologists, is a framework for seismic data processing. The framework was designed to enable new scientific frontiers in seismology by providing a means to more effectively utilize supercomputers or cloud resources to handle the increasingly large data volume available today. MsPASS leverages several existing technologies: (1) scalable parallel processing frameworks, (2) a NoSQL database management system, and (3) containers. The system leans heavily on the widely used ObsPy toolkit. It automates many database operations and provides a mechanism to automatically save the processing history for reproducibility. The synthesis of these components can provide flexibility to adapt to a wide range of data processing workflows. We demonstrate the system with a basic data processing workflow applied to USArray data. Through extensive documentation and examples, we aim to make this system a sustainable, open-source framework for the community.