Introduction

• Mass spectrometry workflow is sequential (Figure 1)

• Protein identification [2] in real time, see Figure 2

Cloud Backend Fast Data Architecture

• Fast Data architecture [3]
• Stream producer (Figure 5-1)
• Horizontally scalable scorer (Figure 5-2)
• Distributed structured protein knowledge base (Figure 5-3)
• Smart validator of peptide-spectrum-matches (Figure 5-4).

MSDataStream

• Collaboration with Bruker Daltonik GmbH.
• MSDataStream reads single spectrum data during measurement
• Sending the data to the cloud
• Enables real time diagnostics on the cloud

References


Figure 1: Sequential mass spectrometry pipeline.

Figure 2: Principle of protein database search. (source: https://www.nature.com/articles/nbt0710-659)

Figure 4: MSDataStream user interface

Figure 5: Fast Data architecture