

Big Data Analysis as a Service at the UW

What should we be teaching today's students:

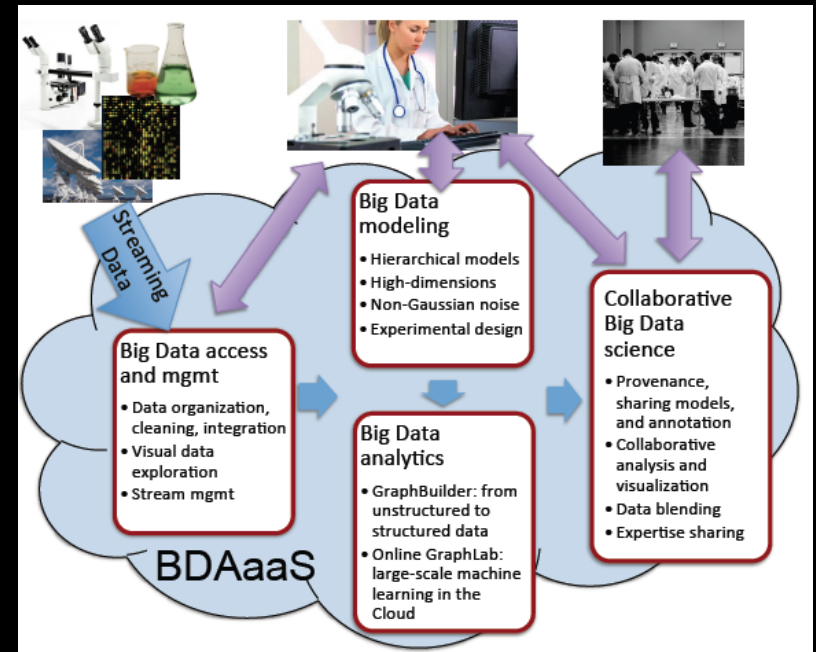
Data Wrangling: data processing, visualization, cloud based approaches (most don't work out of the box)

Data Analytics: scalable machine learning techniques, graph-based analyses, accounting for uncertainty and missing data.

Data Modeling: accounting for heavy tailed distributions, structured non-parametric approaches, multimodal likelihoods, iterative or active learning

Collaborative Science: provenance, annotations, collaborative visualization and workflows

The first steps: eScience Institute and shared expertise, software carpentry, big data certification, IGERT program, internships, emphasis on quantitative undergrad programs



A common development environment