Applied Data Science for Research

Technical Skills

A non-computational non-data driven researcher. Practices and methods used often prevent scaling

Desktop-based computational researcher able to apply local computer methods and techniques.

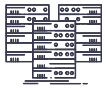
HPC user or beginning Cloud user. Has a mental model for computing at-a-distance. Is able to

A skilled computational practitioner who is able to connect discrete systems for business and research

Compute

Spreadsheet skills, mostly unstructured use of spreadsheets and other data sets. Beginning to

Able to implement scripts and simple programs which aid the researcher in autommating repetitive



Able to think about discrete servers and systems and how they could contribute to a research or business

Can leverage multiple cloud systems and or geographic locations to research or economic benefit.

Storage

Considers data and storage as something local and contained within a device caried around. Limited

Able to consider throughput and redundancy as important considerations to achieve research

Able to make decisions about storage based on performance and latency and design systems to

Able to build/use tiered storage systems which trade off economics, timeliness and accessibility/

Software

Mostly point and click interface usage. Limited models for reasoning about

Beginer use of version control and tracking changes in scripts and

Able to collaborate on version control systems and begin using continuous

Able to build integrated systems which are continuously integrated and invite external contribution

Workflows

Makes checklists with little to no integration into how items interelate or are scripted.

Able to connect multiple scripts or programs together in simple

Able to consider asynchronous workflow systems and the use of

Able to build systems of systems, interconnected and interoperable to build

Methods

Applies statistical models over small and individual data sets. Limited ability to apply similar models over

Growing understanding of Machine Learning (ML) and ANNs. Able to apply stat methods across splits

Advanced applications of ML and considerations of custom hardware, GPUs, FPGAs etc. Able to

Able to connect and interconnect flows of machine learning and statistical insight into

People Skills

Actively listens. Able to effectively communicate with technical and non-technical people. Seeks continuous

Build empathy with end-user. UI/UX design skill such as prototyping, wire frames, & personas. Makes usable and

Competent in behavioral science skills. Able to build tools and systems aware of human behavior, psychology

Competently navigates ethical landscape around appropriate data science usage and the pitfalls such