## Collaborative Analytics Environment (CAE)

The Collaborative Analytics Environment (CAE) provides cloud services for data ingestion, transformation and preparation, as well as data exploration and computation. It includes tools for collaborative analytics, machine learning environments, and data visualization capabilities. Notebook environments and virtual machines provide analytical capabilties using a variety of statistical software such as R, Python, SAS, etc. The CAE leverages Microsoft Azure Platform as a Service (PaaS) and Software as a Service (SaaS) offerings.

## Environment Overview

We are currently testing different use cases against the platform. Each use case can be onboarded into the **main** or a new **private** environment can be created.

### Main (Shared) Environment

Shared with users from several use cases. When granted access to this environment, users can view / share data across use cases.

### Private Environment

A private environment configured upon request so only named users can access workspace files.

## Data Ingestion

Data enters the platform via an external storage account. Once inside the platform, the data is stored an internal storage account (Data Lake). Publicly available data sources may be ingested directly via one of the platform tools.

### External Storage Account

Users will be able to access the external storage account from the Internet, and use it to upload / download data in and out the environment. In some private environments, restrictions or additonal vetting processes may be implemented for data upload / download.

### Internal Storage Account (Data Lake)

Files that are uploaded into the external storage account are automatically moved to an internal Data Lake. This Data Lake is located in a secure virtual network, and is only accessible from platform services and virtual machines.