Digital Research Alliance of Canada: Project Manuals, Tutorials, and User Guides

Tutorial 1: Getting Started with Alliance HPC Systems

Tutorial Objective:

Set up and run a Python-based fake news detection model on Alliance's Niagara supercomputer using SLURM.

Section 1: Accessing the Alliance Supercomputers:

1. **Request an Account**:
   - Apply via CCDB.
   - Sponsor: Your PI or lab manager.

```
# After approval, log in via SSH:
ssh username@niagara.computecanada.ca
```

2. **Multi-Factor Authentication (MFA)**:
   - Follow Alliance's MFA guide.

Once you have utilized the MFA, go to the next section.

Digital Research Alliance of Canada: Project Manuals, Tutorials, and User Guides

Section 2. Transferring Data

Example: Upload a fake news dataset (CSV) to Niagara:

```
# From your local machine:
scp ~/data/fake_news_dataset.csv username@niagara.computecanada.ca:~/project/data/
```

Note: The most used datasets collected for this purpose can be found in our practical GitHub repository.

Section 2. Submitting a SLURM Job

Python Script (`train_model.py`):

```
import pandas as pd
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.linear_model import LogisticRegression

# Load data
data = pd.read_csv("fake_news_dataset.csv")
X = data["text"]
y = data["label"]

# Train model
vectorizer = TfidfVectorizer()
X_vec = vectorizer.fit_transform(X)
model = LogisticRegression().fit(X_vec, y)
model.save("fake_news_model.pkl")
```

Digital Research Alliance of Canada: Project Manuals, Tutorials, and User Guides

SLURM Job Script (`submit_job.sh`):

```
#!/bin/bash
#SBATCH --nodes=1
#SBATCH --cpus-per-task=4
#SBATCH --time=1:00:00
#SBATCH --mem=8G

module load python/3.10 scipy-stack
python train_model.py
```

Submit Job:

```
sbatch submit_job.sh
```

Output:

(Use `squeue -u username` to monitor)

The preprocessed datasets and the train/test scripts can be found in our repository.

For a step-by-step tutorial on creating an Account on Digital Research Alliance of Canada (Compute Canada), see this video.