

# Generating Synthetic Ansible Programs with Probabilistic Methods

HORIZON EUROPE -  
Cloudstars Project

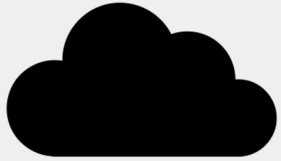
Nathanael Nussbaumer<sup>1</sup>, Louis Mandel<sup>2</sup>,  
Guillaume Baudart<sup>3</sup>, Martin Hirzel<sup>2</sup>,  
Jürgen Cito<sup>1</sup>

<sup>1</sup>TU Wien, Austria,

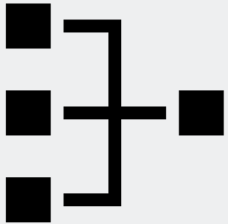
<sup>2</sup>IBM Research, USA,

<sup>3</sup>ENS - PSL University - CNRS - Inria,  
France

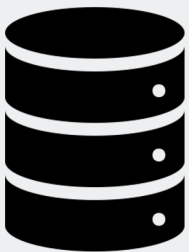
# Project Motivation



- **Ansible is often used for Cloud orchestration and managing infrastructure**



- **LLMs have shown success in helping developers write code in multiple programming languages**



- **Compared to more popular programming languages, publicly available training data is sparse for Ansible**

## Project Motivation

**To help in the creation of LLMs finetuned for Ansible we need a way to circumvent this data sparseness**

# Project Context - Ansible

```
- name: Update web servers
  hosts: webservers
  remote_user: root

  tasks:
  - name: Ensure apache is at the latest version
    ansible.builtin.yum:
      name: httpd
      state: latest

  - name: Write the apache config file
    ansible.builtin.template:
      src: /srv/httpd.j2
      dest: /etc/httpd.conf
```

Example Ansible Playbook from [https://docs.ansible.com/ansible/latest/playbook\\_guide/](https://docs.ansible.com/ansible/latest/playbook_guide/)

- **Ansible scripts are called “Playbooks”**
- **Playbooks have one or more tasks**
- **A task runs an Ansible Module**
- **Modules have multiple Parameters**
- **We have more data for some Modules than for others**

# Project Context – Ansible Templates

```
- name: Update web servers
  hosts: webservers
  remote_user: root

  tasks:
  - name: Ensure apache is at the latest version
    ansible.builtin.yum:
      name: httpd
      state: latest

  - name: Write the apache config file
    ansible.builtin.template:
      src: /srv/httpd.j2
      dest: /etc/httpd.conf
```

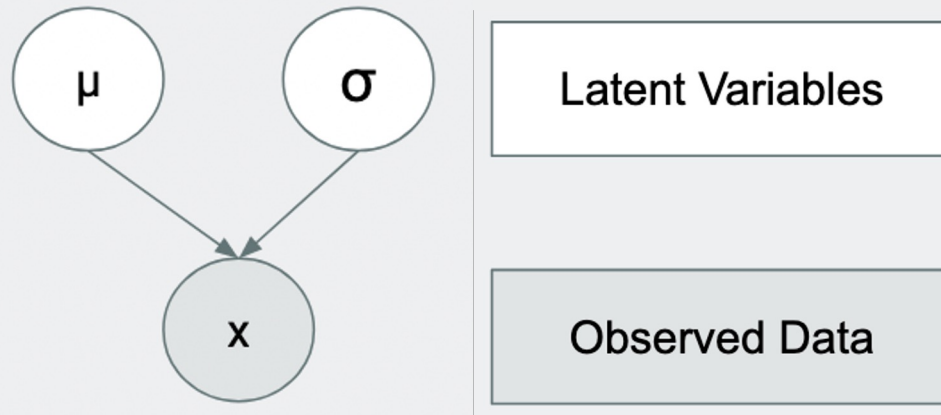
```
- name: Update web servers
  hosts: webservers
  remote_user: root

  tasks:
  - name: Ensure apache is at the latest version
    ansible.builtin.yum:
      name: httpd
      state: {$1}

  - name: Write the apache config file
    ansible.builtin.template:
      src: /srv/httpd.j2
      dest: /etc/httpd.conf
```

# Project Context – Probabilistic Models

$\mu \sim \text{Normal}(0, 10)$   
 $\sigma \sim \text{LogNormal}(0, 5)$   
 $x \sim \text{Normal}(\mu, \sigma)$

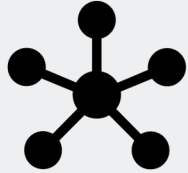


- **Are composed of random Variables**
- **Are best suited for problems with inherent uncertainties**
- **Allow for uncertainty quantification**

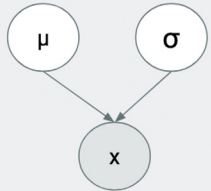
# Project Goals

- **Generate meaningful synthetic values for Ansible Parameters that can be used in templates**
- **Build an End-to-End Pipeline for synthetic value generation**
- **Evaluate and compare different models**

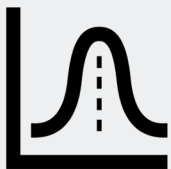
# Approach



- **Cluster parameters based on their description**
  - Using different LDA variations



- **Use a probabilistic model to transfer knowledge from better known parameters to less known parameters**



- **Sample new values**



# A Quick Example

**mysql\_db**



login\_port: **integer**

$\{x : x \in \mathbb{Z}\}$

$\{1, 2, \dots, 65535\}$

**bigip\_monitor\_http**



port: **string**

$\{?\}$

# A Quick Example

**mysql\_db**



**login\_port: integer**

Port of the MySQL server. Requires *login\_host* be defined as other than localhost if *login\_port* is used.

**{ 1, 2, ... , 65535 }**

**bigip\_monitor\_http**



**port: string**

Port address part of the IP/port definition. If this parameter is not provided when creating a new monitor, then the default value will be '\*'. Note that if specifying an IP address, a value between 1 and 65535 must be specified.

# A Quick Example

## **login\_port: integer**

Port of the MySQL server. Requires *login\_host* be defined as other than localhost if *login\_port* is used.

## **port: string**

Port address part of the IP/port definition. If this parameter is not provided when creating a new monitor, then the default value will be '\*'. Note that if specifying an IP address, a value between 1 and 65535 must be specified.

Clustering  
Algorithm

```
[  
  0.1,  
  0.04,  
  0.8,  
  0.06  
]
```

```
[  
  0.2,  
  0.01,  
  0.77,  
  0.12  
]
```

# A Quick Example



## Current State

- **End-to-End Pipeline built**
- **Multiple Clustering Algorithms integrated**
- **Developed our own modification of LDA**
- **Multiple probabilistic models developed**

## In Progress

- **Evaluating different combinations of models and clustering algorithms**
- **Full evaluation of results**