National College of Ireland

Configuration Manual

MSc Research Project
MSc In Cloud Computing

Sunil Suresh Gadhe
Student ID: x22179607

School of Computing
National College of Ireland

Supervisor: Shaguna Gupta

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Sunil Suresh Gadhe
Student ID:	x22179607
Programme:	MSc In Cloud Computing
Year:	2024
Module:	MSc Research Project
Supervisor:	Shaguna Gupta
Submission Due Date:	12/08/20
Project Title:	Configuration Manual
Word Count:	854
Page Count:	5

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Sunil Gadhe
Date:	12th September 2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Configuration Manual

Sunil Suresh Gadhe
x22179607

1 Introduction 1

Dynamic Scaling of Apache Solr Clusters on Azure Cloud Using Custom Metrics like Query Rate and a Tailored Scaling Algorithm.

2 Hardware Requirements

- Total VM: 3
- Size: 1 CPU, 2GB RAM
- Storage: 30GB
- OS: Ubuntu

3 Steps to Deploy Scheduler and Zookeeper Service

- Download the Zookeeper packages from the official website and extract the contents from the zip file.

  ```
  https://zookeeper.apache.org/releases.html#download
  ```

- Install OpenJDK packages using the following commands:

  ```
  sudo apt update -y
  sudo apt install openjdk-21-jre-headless -y
  ```

- Verify the existence of the configuration file by running:

  ```
  ls -l /mnt/solr-data/zookeeper/apache-zookeeper-3.9.2-bin/conf/zoo.cfg
  ```

  (If the file does not exist, create it.)

- Create the configuration file using a text editor (nano or vim) with the following content:

```
tickTime=2000
dataDir=/mnt/solr-data/zookeeper/data
clientPort=2181
initLimit=5
syncLimit=2
4lw.commands.whitelist=mntr,conf,ruok
```

- Check directory permissions to ensure the necessary permissions:

```
sudo chmod -R 755 /mnt/solr-data/zookeeper/apache-zookeeper-3.9.2-bin
```

- Verify the data directory and ensure it exists and is writable:

```
mkdir -p /mnt/solr-data/zookeeper/data
sudo chmod -R 755 /mnt/solr-data/zookeeper/data
```

- Retry starting ZooKeeper:

```
sudo bash <path>/zkServer.sh start
```

If issues persist, check the ZooKeeper logs for detailed error messages:

```
ls -l /mnt/solr-data/zookeeper/apache-zookeeper-3.9.2-bin/logs/
```

4 Setup Standalone Solr Machine

To deploy Solr on the standalone Solr virtual machine, follow these steps:

- Install Java OpenJDK:

```
sudo apt update -y
sudo apt install openjdk-21-jre-headless -y
```

- Download and extract the Solr package:

```
sudo wget https://dlcdn.apache.org/solr/solr/9.6.1/solr-9.6.1.tgz
sudo tar zxf solr-9.6.1.tgz
```

- Make the Solr Admin Dashboard accessible from any server. Edit the ‘solr.in.sh‘ file and add:

```
SOLR_JETTY_HOST="0.0.0.0"
```

- Start the Solr service and point it to ZooKeeper:

```
sudo /mnt/solr-data/solr-9.6.1/bin/solr start -cloud \
-s /data/solr-9.6.1/solr-home -p 8987 \
-z 10.0.0.4:2181 -force
```

5 Create Solr Base Image on Azure

To create a Solr base image for VMSS, first deploy a virtual machine and perform all steps from the standalone Solr machine setup. A separate Solr machine is needed for image creation to avoid rendering the original machine useless. Follow these steps:

- Deploy a new Solr virtual machine (VM) for the image. This ensures the original machine remains functional.
- Follow all steps outlined in the "Setup Standalone Solr Machine" section.
- Configure the 'rc.local' file to automatically start Solr upon machine deployment in VMSS:

```
sudo touch /etc/rc.local

Add the following entries to /etc/rc.local:

#!/bin/bash
sudo apt update
sudo /mnt/solr-data/solr-9.6.1/bin/solr start \
-cloud -s /mnt/solr-data/solr-9.6.1/solr-home \
-p 8987 -z 10.0.0.4:2181 -force
exit 0
```

  Change permissions:

```
sudo chmod +x /etc/rc.local
```

  Start the 'rc-local' service:

```
sudo systemctl start rc-local
```

- Create an image from the Solr VM in the Azure portal:

```
Navigate to the Azure Virtual Machines dashboard and
select the VM to create the Solr base image.
Go to Capture -> Image.
Fill in the required details and select
"Yes, share it to a gallery as a VM image version."
Choose the target Azure Compute Gallery and
provide the necessary details, ensuring the
operating system state is set to "Specialized."
```

6 Deploy Virtual Machine Scale Set (VMSS) Using Solr Base Image

To create an Azure VMSS using the Solr base image, follow these steps:

- Step 1: Navigate to the Azure compute galleries dashboard.
- Step 2: Select the Azure compute gallery storing the Solr base image then select image definition then select image version.
- Step 3: Navigate to the "Create VMSS" option and select it.
- Step 4: Create VMSS with the following settings: Orchestration mode: Uniform. Scaling mode: Manually update the capacity. VNet: Should be the same as the Standalone Solr VM for communication between all VMs.

7 Configure Azure CLI on Scheduler VM

To install Azure CLI on your system, follow these steps:

- 1. Update the Package List Ensure your package lists are up-to-date. sudo apt-get update
- 2. Install Required Dependencies

  sudo apt-get install -y ca-certificates curl apt-transport-https lsb-release gnupg
- 3. Add the Microsoft Repository

```
curl -sL https://packages.microsoft.com/keys/microsoft.asc | sudo apt-key add -
echo "deb [arch=amd64] https://packages.microsoft.com/repos/azure-cli/
$(lsb_release -cs) main" | sudo tee /etc/apt/sources.list.d/azure-cli.list
```

- 4. Update the Package List Again sudo apt-get update
- 5. Install Azure CLI sudo apt-get install azure-cli
- 6. Verify the Installation az –version

8 App Registration to Access and Control Solr VMSS

To create and control Azure VMSS, follow these steps:

- **Step 1:** Navigate to Azure App registration.
- **Step 2:** Register a new application:
  Go to "New registration" Click "Register" Navigate to "All applications" Select the app registration created for Solr VMSS.
  Navigate to "Certificates and secrets" Create secrets.
  Note: Store secrets securely as they cannot be recovered once lost.
- **Step 3:** Install Azure CLI packages on the scheduler machine.
- **Step 4:** Configure App registration/Service Principal on the scheduler machine:

```
az login --service-principal -u <app-ID> \
-p <password-or-value> --tenant <tenantID>
```

9 Schedule cronjob to run monitoring script every 10 minutes

```
*/10 * * * * /usr/bin/python3 /mnt/solr- data/autoscale_scripts/
solr_autoscaling.py
```

10 Additional Instructions

- use "find" command to get path of any file
- Once standalone Solr node is ready, create collection and add any sample ".csv" datasets for experiment.
- Scaling algorithm script name "solr_autoscaling.py"

```
Conf script to set threshold for scaling "autoscale_conf.py"
```

- auto scale logs path : /home/azureuser/logs/auto_scale_logs
- credentials to access infra has been provided in the Artefact zip file(readme.txt)