



# Technical Guide: Data Cleansing Tool Development

---

**Date:** 28-08-2023

**Project:** Data Cleansing tool

**Client:** Eastern Region Municipality

**Project Manager:** Eng. Sarah Al-Thabit

## Overview

This document outlines the progress achieved in the development of the Data Cleansing Tool for Eastern Region Municipality. It includes a step-by-step description of the milestones accomplished thus far in the project.

## Milestone 1: Setting up Oracle Linux Development Server

### Description

In this milestone, we successfully set up the Oracle Linux development server on the designated machine with the IP address 10.1.240.99. The server is now fully operational and prepared for further configuration.

### Technical Specifications

Specification	Details
Operating System	Oracle Linux 7.9
Server IP Address	10.1.240.99
RAM	16 GB
CPU	x86_64,64-bit,4 cores
Storage	~250 GB
Network Configuration	inet 10.1.240.99 netmask 255.255.255.0 broadcast 10.1.240.255
Firewall and Security	Managed by the cyber security department
Server Accessibility	Remote access through RDC

**Run the following commands if you want to see some additional specification about the server**

```
#!/bin/bash

# Display Operating System Version
cat /etc/oracle-release
OR
cat /etc/os-release

# Display Network Configuration
ifconfig

# Check IP Address and Subnet Mask
ip addr show ens192

# Check Broadcast Address
ip addr show ens192 | grep broadcast

# Check Total RAM
free -h

# Check CPU Architecture
lscpu

# Check Number of CPU Cores
nproc

# Check Storage Capacity
df -h
```

## Milestone 2: Internet Connectivity Setup

### Description

During this phase, we ensured that the Oracle Linux development server has the required internet connectivity. The following hosts have been enabled for access:

- pypi.org
- files.pythonhosted.org
- github.com
- api.github.com
- yum.oracle.com
- download.fedoraproject.org
- dl.fedoraproject.org

This step guarantees seamless access to external resources that are crucial for the tool's functionality. Additionally, we have identified the potential need for opening additional hosts in the future to accommodate evolving requirements.

## Milestone 3: Python Environment Setup

### Description

In this milestone, we successfully established the Python environment on the Oracle Linux development server. The following components have been set up:

- Python 3
- Pip 3
- Virtualenv
- some additional python libraries

This environment forms the foundation for building and deploying the Data Cleansing Tool.

```
# Update package repository
sudo yum update -y

# Install Python 3 and Pip 3
sudo yum install python3 python3-pip -y

# Verify Python and Pip installation
python3 --version
pip3 --version

# Install Virtualenv
pip3 install virtualenv

# Create a new directory for your project and navigate to it
mkdir my_project
cd my_project

# Create a virtual environment
virtualenv venv

# Activate the virtual environment
source venv/bin/activate

# Install additional Python libraries using Pip
pip install library_name1 library_name2 ...

# Or to install all the dependencies from the `requirements.txt` file
pip install -r requirements.txt
# to Deactivate the virtual environment
deactivate
```

### Conclusion

We are excited about the progress made thus far and look forward to the continued collaboration as we bring the Data Cleansing Tool to its completion. If you have any questions or feedback, please don't

hesitate to reach out.

Sincerely,

---