



ESP Technical Specifications

server sizing, software dependencies, and network considerations

Prepared by

Bob Zambarano





Revision History

Revision	Date	Author	Notes
1	April 7, 2017	Bob Zambarano	Initial version.
1.1	9/18/2020	J Miller	Minor updates, including ref to git lab repo

Table of Contents

Revision History.....	2
ESP Overview	3
Sizing Hardware or Virtual System	3
Storage.....	3
Memory	4
CPU	4
OS and software stack	4
Network Requirements and Data Security	5



ESP Overview

ESP is an open-source, free-license software originally developed under a CDC Centers of Excellence grant to the Harvard Medical School's Department of Population Medicine. Information is available at <http://www.esphealth.org/>. Software source code is available for download at <https://gitlab.com/esp-project/esp.git>

ESP performs notifiable and chronic disease case detection against a DataMart of patient clinical data. ESP deals only with clinical data including: patient demographics, patient visit and visit diagnoses, medications, lab test results, immunizations and social behaviors (smoking and alcohol use).

ESP is typically installed and maintained with the health care data partner's data center.

The standard model of operation is to extract all new or updated patient clinical data from the EHR on a nightly basis, and load this into ESP for processing. ESP then uses a set of plug-in disease detection modules to identify disease cases, identify all relevant available data, generate an HL7 message, then encrypt and transfer this message to MDPH. ESP also tracks what case data has been sent so when additional information for a given disease case is collected for a patient, it can trigger the creation and transfer of a case update message to MDPH.

Once disease detection plugins are configured and validated, ESP requires very little maintenance beyond mapping new lab test types when these are added to the EHR system that feeds ESP.

Sizing Hardware or Virtual System

Storage

At its core, ESP is a DataMart of patient EHR data. Storage requirements can be significant. Storage requirements primarily depend on number of patients to be included in the ESP system, and the number of years of data the ESP system will need to support (going back historically, and going forward for the planned life of the server).

Having worked with several large, primarily outpatient health care organizations, a good sizing rule is:

50GB base + (1GB * Number of active patients in your system/5000) *Number of years of patient data kept in the ESP system

For example: An organization with approximately 200,000 active patients intends to maintain up to 12 years of patient data in the ESP DataMart. This DataMart would require:



$50\text{Gb} + (1\text{GB} * 200,000/5000) * 12 = 530\text{Gb}$ of storage.

A third factor to consider when sizing storage is the density of patient data, meaning the number of distinct observations, orders and results collected per patient. If your organization deals with long-term care of very sick patients, patient data density will be much higher and will require more storage capacity. Alternatively, a health care organization that primarily deals with transient short-duration patients with limited care provision will have much lower data density and will require less storage.

ESP does not support any significant transactional processing in the database, so disk I/O performance need only be moderate.

Memory

A good rule of thumb for adequate performance is 2Gb of memory for each 50Gb of storage. A system with 500Gb of storage, for example, would require 20Gb of memory.

CPU

ESP can run disease detection algorithms in parallel threads, so more CPUs will increase performance up to the point where disk I/O or memory reach capacity. Two CPU cores are a minimal requirement. For each 100Gb of disk storage, an additional CPU core will be useful.

OS and software stack

ESP runs on Linux OS. It is developed on Ubuntu Server LTS systems, and has been run on many other Linux distributions including RedHat, SuSe and CentOS.

A basic Linux server would need the following additional software:

- Administrative logins from anywhere but the system console require Open-SSH service.
- Iptables, a Linux computer firewall, should be installed and configured to manage and restrict system access according to policy.
- Git is used for ESP distribution.
- The ESP DataMart requires PostgreSQL as the RDBMS.
- ESP software is developed using Python and the Django ORM. The ESP installation uses the Python virtual environment infrastructure.
- The ESP System has an administrative web interface that uses the Apache web server.
- The MDPH provided encryption software requires Java.



Network Requirements and Data Security

ESP will store and provide access to HIPAA regulated PHI data, so network design for data security issues are of critical importance. The ESP server must exist within a firewalled and access controlled network environment.

ESP software is hosted on gitlab.org and uses Python modules. A firewall rule enabling outgoing requests to the following sites are necessary for installation on periodic updates:

<https://gitlab.org> , <https://pypi.org> , and <https://pypi.python.org>

Data provisioning for ESP occurs via a set of extracted and delimited text files, generated nightly or weekly, so there must be a network connection between the ESP server and the machine that will generate the EHR extract files. Data flows to the ESP system should use some encrypted protocol (Open-SSH supported transfer encryption such as SFTP or SCP, or other custom built encryption based on asymmetric/public-key cryptography).

Data reporting from ESP to MDPH utilized an MDPH-built encryption and transfer software. This will require an outgoing rule in the data center firewall to allow traffic from ESP to the MDPH reporting servers.