

Platform service hosting report

Prepared by NWU/MuST

| | |
|-------------------------------------|-----------|
| Nomenclature | 2 |
| Introduction | 3 |
| Purpose and scope | 3 |
| Hosting | 4 |
| Hosting metrics | 5 |
| User interface | 6 |
| Application server | 7 |
| Speech server | 7 |
| Summary | 8 |
| Commercial hosting options | 9 |
| Technology costs | 10 |
| Platform Setup | 10 |
| Platform monitoring and maintenance | 10 |
| CKEditor | 10 |
| Nanospell | 11 |
| Speech Services | 11 |
| User Interface | 13 |
| Monthly budget | 13 |
| Conclusion | 13 |

Nomenclature

| | |
|-------|--|
| AM | Acoustic Model |
| API | Application Programming Interface |
| APP | Application |
| CPU | Central Processing Unit |
| FTP | File Transfer Protocol |
| FST | Finite State Transducer |
| GB | Gigabyte |
| Gb | Gigabit |
| GHz | GigaHertz |
| GPL | General Public License |
| GPU | Graphical Processing Unit |
| KB | Kilobit |
| HDD | Hard Disk Drive |
| HTTP | Hypertext Transfer Protocol |
| HPC | High Performance Computing |
| IO | Input/Output |
| IP | Internet Protocol Address |
| LM | Language Model |
| MB | Megabyte |
| Mbit | Megabit |
| NCHLT | National Centre for Human Language Technology ¹ |

¹ An initiative funded by the Department of Arts and Culture, which produced a number of text and speech corpora for South African languages. These are used in the development of various technologies, such as language recognition systems.

| | |
|------|--------------------------------------|
| PM | Per Month |
| PP | Per Person |
| RAID | Redundant Array of Independent Disks |
| RAM | Random Access Memory |
| REST | Representational state transfer |
| TB | Terabyte |
| Tb | Terabit |
| UI | User Interface |

Introduction

The service hosting document is drawn up to detail the requirements for hosting the speech transcription platform (STP) beyond the termination of the current development contract. This report will highlight the costs associated with each running component and propose a monthly budget required to effectively host the platform. The document is split into three sections:

- Hosting - a list of platform hosting requirements for the various servers and possible commercial options.
- Technology costs - pricing for technology utilised in the platform.
- Budget - monthly running costs breakdown and additional development and speech services costs.

The information contained herein should be sufficient in aiding a potential end user in making a decision to adopt the services provided by the platform.

Purpose and scope

The purpose of this document is to explore possible hosting solutions for the speech transcription platform. In no way does the development of the service hosting report mean that North-West University (NWU) or IntSyst has to provide hosting, support or maintenance services for the speech transcription platform once the current contract concludes. This falls outside the scope of the current contract. For a clearer understanding, excerpts (Table 1, Table 2 and Table 3) from the signed business contract have been included to show the intent of this service hosting report.

“During year 3, options to ensure the sustainability of a hosted service will be explored (for example, a service hosted by the South African Language Resource Management Agency

(RMA), INTSYST or another partner.) Hosting beyond the current system demonstrator is outside the scope of the current project. (The project will deliver a well-tested prototype system, but not a hosted service.) Requirements for a hosted service and a discussion of possible options will be captured in a service hosting report, included in this work package."

Table 1: Excerpt 1 from the signed business plan

"Future sustainability: the current project addresses the development of a fully functional and well-tested transcription platform; it does not include the long-term hosting of a transcription service nor ongoing user training and support."

Table 2: Excerpt 2 from the signed business plan

"Specific attention will be paid to the future sustainability of the transcription service during year 3, when users' requirements of such a service (as well as hosting capabilities of possible partners) have become clearer."

Table 3: Excerpt 3 from the signed business plan

The only purpose of the report is to provide an external party with information needed to make an informed decision on what related costs and technologies are needed to host the speech transcription platform. The creator of this report is under no obligation to provide hosting services.

Hosting

To determine the hosting requirements of the speech transcription platform, the system architecture must be considered. Figure 1 shows a broad overview of the speech transcription platform's system architecture. The platform has three main components:

1. User Interface (UI) - personal computer interface using the Google Chrome web browser.
2. Application Server - an HTTP server utilising a REST API architecture to control information flow between the UI and the databases and speech server.
3. Speech Server - a HTTP server utilising a REST architecture to host diarization, speech recognition and speech-text alignment services.

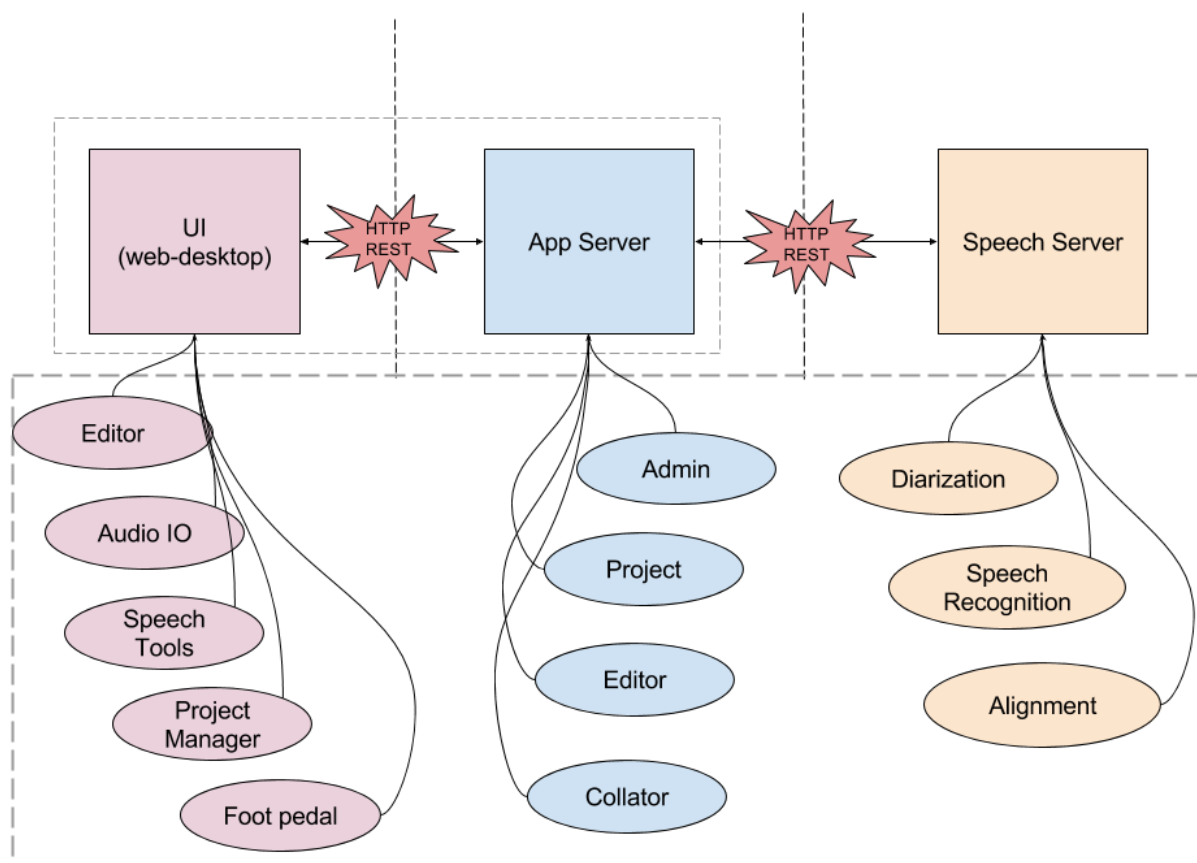


Figure 1: Platform system architecture

All components require hosting support, but the hosting requirements differ depending on the component. The specific requirements will be discussed in the subsequent sections.

Hosting metrics

To evaluate the available hosting options, a set of hosting and system metrics are required. This will make the task of assessing the different options more tractable. The following standard list of metrics will be used during the evaluation:

- Uptime - a measure of the time a service or system has been working and is available. Measured in percentage e.g. 99.99%
- RAM size - the amount of system memory. Measured in GB e.g. 16 GB.
- Hard disk size (HDD) - the amount of long term data storage. Measured in TB e.g. 2 TB.
- Storage redundancy - the RAID (a hard drive configuration design for data redundancy and/or performance improvement) level of the storage is important as this will determine the recovery possibilities if a single or multiple hard disk drive(s) fail. The exact properties are determined by the RAID level e.g. RAID 0, 1, 2, 3, 4, 5, 6.

- Data backup - data storage independent of the system. Measured in GB or TB e.g. 50 GB or 1 TB.
- Network speed - the network uplink and downlink speed to the system. Measured in Mb or Gb e.g. 100 Mb or 1 Gb.
- Network bandwidth - the amount of network traffic (data) that can be transferred to and from system. Measured in TB e.g. 4 TB.
- CPUs - the number of physical CPUs, the number of cores and the speed of the processing units. CPU speed is measured in GHz e.g. 2.0 GHz.
- Cost - the monthly cost of renting the provided server/system, plus the out-of-contract costs for items like bandwidth or additional backup.

The metrics can roughly be split into three categories: Uptime, Network (speed and bandwidth) and System components (CPU, RAM, HDD, Backup). The last metric, Cost, is directly related to these categories and increases as quality of each or all the dimensions are improved.

The Uptime is a really important service measure and is generally specified in percentage by the system service provider. For instance, a 99.9% uptime per month would be related to a service going down for about 0.7 hrs in that month. Factors that contribute Uptime which are under control of the provider include:

- Network connectivity
- Hardware failure recovery
- Uninterruptible power supply

When the system service provider estimates the uptime percentage, these factors are taken into consideration.

The network metrics include speed and bandwidth. Speed relates to how quickly the data can be moved to and from the server, and, bandwidth determines how much data may be transferred to and from the server. Once the bandwidth has been exceeded for a month, then out-of-contract data rates apply.

Lastly, the system components determine the performance of the server. An increase in the amount of a specific component generally relates to a performance boost.

User interface

The UI codebase is statically served by an HTTP server which is interpreted by a client's web browser. The web-based browser is installed on the client's personal computer and the UI is specifically written for the Google Chrome browser. This browser is free to install and does not require any additional support costs. Updates and fixes are periodically pushed by the Google Chrome developers.

The UI codebase is installed on the server but does not require much storage space. The current size of the code is 40 MB. Other important hosting metrics are Uptime and network

speed -- the UI should be accessible by a client whenever requested and the speed of the access should be fast.

Application server

The application server has three main functions:

1. Manage requests from the UI
2. Handle data flow between the UI and databases (storage)
3. Manage requests between the UI and speech server

The important metrics that relate to the main functions are Uptime, network speed, network bandwidth and the number of CPUs/cores.

Uptime is most important as the client, via the UI, needs to access the application server ideally at any time. A client's frustration would increase if the UI is inaccessible and this deteriorates the UI experience.

Network speed is also crucial; requests should not be delayed due to network congestion, as this will slow down the UI experience and make it appear sluggish and impact on the client's perception of the system. The time it takes for a user to upload audio data is also affected by the network speed. The quicker this operation can conclude, the better the UI experience will be. It is hoped that using compressed audio will help mitigate this.

Network bandwidth plays a smaller role, but influences the amount of audio data that users can transfer from the UI to the platform and from the platform to the speech server.

The application server makes use of UWSGI to provide the HTTP server. Each request that is made to the server is handled by a separate thread, which UWSGI spawns on startup. These threads are run on different CPUs or cores if available to the system. Therefore, having access to many cores would increase the number of requests that could be handled and speed at which the requests are concluded. The CPU/core speed, measured in GHz, also plays a part: generally the quicker the CPU/core (the larger the GHz value), the faster the request can be processed. However, storage speed would influence any storage-based request as storage access speeds are much slower compared to CPU or RAM access.

Lastly, the codebase footprint for the application is quite small: 388 KB, which is negligible in respect to current storage sizes.

Speech server

The speech server hosts three speech services:

- **Diarization** - breaks up a long audio file into smaller portions, usually on silence boundaries.
- **Speech recognition** - generates a transcription of what was said in an audio recording.

- Speech-Text alignment - aligns a text transcription and the associated spoken audio i.e. assigns the start and end times of the words.

Nowadays, speech servers require high performance computing (HPC) setups with additional GPUs. Service providers that offer these systems are limited and, if provided, the costs for renting such a system are quite high. One possible workaround is to make use of dedicated servers. These systems are usually equipped with large system memory, medium number of CPUs and large storage facilities. However, not many dedicated server commercial options have, as standard, a GPU installed. For the current platform this is not a hinderance, but for future releases this missing hardware component could become crucial.

The important metrics associated with the speech server are: Uptime, RAM size, CPU/core count and speed, HDD size, and network speed and bandwidth. Again, Uptime is vital as clients would prefer to access the speech server at any time of day.

The amount of system RAM could potentially play a delaying role in how long it takes to complete a speech task (particularly speech recognition). The speech recognition system utilises Kaldi, a toolkit for speech recognition, to perform the recognition task. During a recognition operation, Kaldi loads a finite state transducer (FST) which is used to model a language -- models the grammar and word order in a probabilistic manner. These FST models can be rather large and consume large amounts of memory. For instance, a single recognition job that uses the current SA English FST expands to close to 12GB in RAM. This size limits the amount of concurrent recognition jobs the speech server can launch, which means these types of tasks must wait in a queue and thus cause delays.

Similarly, the CPU/core count determines how many speech tasks may be run concurrently. The CPU speed also plays a role in how long it takes to run a task. But the most important factor is overall CPU/core count.

The HDD size will determine how much data can be stored. The currently installed speech server size is roughly 40 Gb. Additional space is needed to store speech task audio data that is uploaded by the user. Provision should be made for scaling.

Lastly, sufficient network speed and bandwidth is needed to allow: (1) the quick uploading of audio data to the speech server, and, (2) large amounts of data to be uploaded during a month. Network speed should not affect the overall latency much, as it takes about 1.0x - 3.0xRT to complete a speech task. Enforcing the use of compressed audio should increase the effective amount of data that can be uploaded to the speech server.

Summary

Given the different hosting requirements for the platform components, the speech server requires the most system resources and will heavily influence which system will be chosen. All servers require a high Uptime, and network speed and bandwidth and must be kept in mind when choosing the server option. The speech server requires a large amount of RAM, many CPUs/cores and large storage which will be sufficient for the interface and application server. The choice of the system

parameters is constrained by the cost. Given the speech server requirements, hosting options will focus on dedicated servers which will closely match the requirements.

Commercial hosting options

There are several dedicated server service providers that advertise their services on the web. Table 4 shows a list of possible rental options -- included are the system specifications and pricing.

| Provider | System Specifications | Pricing |
|---|--|-------------------------------------|
| Afrihost https://www.afrihost.com/site/product/dedicated_hosting | Web traffic (bandwidth): 4TB CPU: 3.2GHz Quad-Core Xeon RAM: 16 GB DDR3 1333MHz Hard Drives: 2x 1TB RAID Level: RAID 1 Backup: 50 GB FTP Data Transfer Overage: 59c per GB 99,999% network uptime | R1,450.00 PM |
| WebAfrica https://www.webafrica.co.za/dedicated-server-hosting/ | Bandwidth: 4TB CPU: Quad-Core E5 2603 1.8 Ghz RAM: 64 GB Hard Drives: 2x1TB RAID Level: RAID 1 Network: 100 Mbit connection Data Transfer Overage: 50c per GB 99.9% Uptime Guarantee 24/7 Expert Support | R1,499.00 PM R599.00 setup fee |
| Hetzner https://hetzner.co.za/dedicated-servers/ | Bandwidth: 4TB CPU: Intel Xeon E5-1620 3.5GHz RAM: 32 GB ECC RAM Hard Drives: 2 x 1TB RAID Level: Software RAID 1 Backup: 50 GB FTP Network: 100 Mbit connection Data Transfer Overage: 60c per GB | R1,595.00 PM R695.00 setup fee |
| GoDaddy https://za.godaddy.com/hosting/dedicated-server | Bandwidth: Unmetered bandwidth CPU: 4 CPU Cores @ 3.1 GHz RAM: 16 GB memory Hard Drives: 2 TB storage RAID Level: RAID 1 Access: 3 dedicated IPs 99.9% Uptime Guarantee | R 1 830,99 PM for 12 Month contract |

| | | |
|--|---|-------------------------------------|
| ParadigmSolutions https://www.paradigmsolutions.co.za/linux-dedicated-servers/ | Bandwidth: 30 TB CPU: Intel® Xeon® E3-1270 v3 Quadcore RAM: 32 GB ECC Hard Drive: 2 x 2 Tb SATA 6 GB/s 7200 rpm Network: 1 Gbps Software-RAID 1 class Enterprise Backup: 100 GB | R1,749.00 PM R2,699.00 setup fee |
|--|---|-------------------------------------|

Table 4: Local dedicated server options (Accessed: 24 July 2017)

Technology costs

Most components used to develop the speech transcription platform are open source. The licences are predominantly General Public License (GPL) and Apache 2.0. However, there are certain software packages that require a commercial licence for use, which adds to the cost of the platform.

Platform Setup

It will take approximately three working days to set up and test the platform. The average rate of a skilled technician is around R500/h. Therefore it will cost R12,000.00 to perform the setup.

Platform monitoring and maintenance

To monitor the ongoing use of the platform will require about minimum of an hour a day to possibly 5 hours depending on issues. This ranges from R10,000.00 to R50,000.00 per month.

Software maintenance defined by the IEEE is “*the process of modifying a software system or component after delivery to correct faults, improve performances or other attributes, or adapt to a changed environment*”. At this stage it is difficult to assign a monetary value to the maintenance task, as it will depend on the nature of the task. A separate contract will have to be drawn up for the system maintenance between the client and external service provider.

CKEditor

The transcription editor makes use of CKEditor to facilitate the editor of text. CKEditor has predefined commercial licences which are shown in Figure 2. The pricing in Figure 2 is per year. Given the size of Parliament, the \$499 fee per year should be sufficient; this allows 50 users and 10 support queries a month. Given a ZAR/\$ exchange of R14.00 the fee would amount to R6,989.00 per year and R583.00 per month.














|  STANDARD |  PLUS |  BUSINESS PLUS |  ENTERPRISE <small>OEM, WEBSITES AND SAAS</small> |
|--|--|--|---|
| <input type="checkbox"/> Add  CKFinder with a discount | <input type="checkbox"/> Add  CKFinder with a discount | <input type="checkbox"/> Add  CKFinder with a discount | <input type="checkbox"/> Ask for  CKFinder special offer |
| <div>Buy now </div> 99/USD | <div>Buy now </div> 499/USD | <div>Buy now </div> 999/USD | <div>  Call us now or  Email us now </div> <div> Toll-free (US/CAN): +1 800 643 7519 Internationally: +1 650 353 3268 </div> |
| Company size: 10 (employees + contractors) | Company size: 50 (employees + contractors) | Company size: 200 (employees + contractors) | Company size: Up to unlimited (employees + contractors) |
| Domains + subdomains: 1 | Domains + subdomains: 10 | Domains + subdomains: 50 | Domains + subdomains: Up to unlimited |
| Support: 2 requests / month | Support: 10 requests / month | Support: 20 requests / month | Support: Up to unlimited |
| SAAS/OEM: No | SAAS/OEM: No | SAAS/OEM: No | SAAS/OEM: Yes |
| License: Open Source | License: Commercial | License: Commercial | License: Commercial |

Figure 2: <https://cksource.com/ckeditor/buy> (accessed: 20 April 2017)

Nanospell

The spell checking feature is provided by the nanospell software package which integrates with CKEditor. Table 5 shows the annual fee for a licence compatible with Parliament's needs. This amounts to R3,486.00 per year and R219.00 per month (assuming 14ZAR/\$).

\$249 Enterprise License

This license allows the use of NanoSpell within ONE public website domain and up to TWO private intranet or extranet domains. It covers only the usage for ONE end user organization. All machines on which the Software is used for intranet or extranet usage must physically located in the same country. This license is designed to be ideal for organizations with intranets and extranets.

Table 5: <http://ckeditor-spellcheck.nanospell.com/license> (accessed: 20 April 2017)

Speech Services

The speech services offered are diarization, speech recognition and text-audio alignment. Currently, diarization is language independent but may require some customisation over the product lifecycle. Optimisations to the recognition and alignment may be in the form of adding

an additional language or improving the service accuracy. Table 6 shows an estimated cost of performing the recognition and alignment optimisations.

| Task | Duration | Cost |
|--|-----------|-------------------------------------|
| Adding a baseline NCHLT South African language recognition service | 1 week | R40,000.00 PP |
| Customising an existing language model for a recognition service <i>(NOTE: this is only possible if the data is available and duration depends on the quality of the data and the amount of manual intervention required)</i> | 1-2 weeks | R40,000.00 - R80,000.00 PP |
| Customising an existing acoustic model for a recognition service <i>(NOTE: this can only happen if the data is available and duration depends on the quality of the data and the amount of manual intervention required)</i> | 1-2 weeks | R40,000.00 - R80,000.00 PP |
| Adding a new South African language to the alignment service | 1-2 weeks | R80,000.00 - R160,000.00 (2 people) |

Table 6: Duration and cost estimates for adding or customising the speech services

Table 7 shows a breakdown of the subtasks involved adding or customising the speech services.

| Task | Subtasks |
|--|--|
| Adding a baseline NCHLT South African language recognition service | <ul style="list-style-type: none"> • Build Kaldi LM and AM using NCHLT data • Test Kaldi system • Integrate into platform • Test integrated service |
| Customising an existing language model for a recognition service | <ul style="list-style-type: none"> • Process text data -- normalisation • Build LM, FST and test • Integrate into platform • Test integrated service |
| Customising an existing acoustic model for a recognition service | <ul style="list-style-type: none"> • Process audio and text data • Adapt existing Kaldi AM • Integrate into platform • Test integrated service |

| | |
|--|--|
| Adding a new South African language to the alignment service | <ul style="list-style-type: none"> • Process text data -- normalisation • Build G2P model for language • Build custom expansion FST • Build Kaldi AM using NCHLT data • Test alignment system • Integrate into platform • Test integrated service |
|--|--|

Table 7: subtasks required to complete the main speech service additions and customisations

User Interface

Over the course of the platform lifecycle certain fixes or additions may be needed to improve the user interface. These changes may require a day to a week to implement. Therefore the cost of altering the UI will be in the range of R4,000.00 to R20,000.00 per addition.

Monthly budget

Based on the server hosting fee and technology costs, it will cost a minimum of R32,347.00 per month to keep the platform running. Table 8 shows a breakdown of the cost. The first month's cost will be R44,374.00 due to the installation fee.

| Item | Cost |
|-----------------------------|--------------------|
| Dedicated server rental | R 1,500.00 |
| Average platform monitoring | R30,000.00 |
| CKEditor fee | R 583.00 |
| Nanospell fee | R 291.00 |
| TOTAL MONTHLY COST | R32, 374.00 |

Table 8: Monthly running platform cost

Conclusion

This report details the requirements to host the speech transcription platform developed by IntSyst with funding from the Department of Arts and Culture. Based on the hardware rental costs and various technology subcomponents' costs, an estimated minimum monthly cost of R32,347.00 will be needed to maintain the platform in its current state.