

Towards a GIS-based Decision Support System on the Amazon Cloud for the Modelling of Domestic Wastewater Treatment Solutions in Wexford, Ireland

Nadeem Qazi
National Centre for GeoComputation,
National University of Ireland,
Maynooth, Ireland.
Nadeem.qazi@nuim.ie

David Smyth
National Centre for GeoComputation,
National University of Ireland,
Maynooth, Ireland.
David.n.smyth@nuim.ie

Tim McCarthy
National Centre for GeoComputation,
National University of Ireland,
Maynooth, Ireland.
Tim.McCarthy@nuim.ie

Abstract- The work presented in this paper highlights the development of a web based intelligent Geographical Information System (GIS) spatial decision support tool. Engineered over the Amazon cloud computing environment, the GIS system has been designed to enable local authorities in Ireland to identify solutions for the treatment of domestic waste water in areas of low permeability subsoil. The data used in the geo-spatial modelling includes the XY coordinates of existing treatment systems, topography, soils and subsoil, and road networks. Spatial modelling is performed through ESRI's ArcMap and published as a web service on ArcGIS Server 10.1. This decision support system is made available to the remote user over Amazon's EC2 as a GIS based web service, through Representational State Transfer (REST) architecture. The tool will assist local authorities to better manage domestic wastewater in low permeability areas, thereby reducing the likelihood of pollution of local surface waters. Wexford is the test area.

Keywords: Cloud computing, REST services, Decision support system, Geo spatial modelling, SaaS.

I. INTRODUCTION

Most domestic wastewater in Ireland is managed by public mains sewerage schemes. Out of 1,649,408 households, around two thirds are connected to the mains system (Central Statistics Office, 2011). However, for those not connected to mains sewerage, onsite wastewater treatment is the primary treatment option. The most widespread method in Ireland is the septic tank system, and in 2011 there were 437,652 in operation (Central Statistics Office, 2011). A septic tank system comprises a septic tank and a percolation area. Initial treatment takes place in the septic tank. Solids separate from the wastewater and sink to the bottom to form sludge, while the remaining liquid flows into a second chamber *via* filters. Settling out of the remaining solids occurs in the second chamber. Effluent then flows through pipes into the percolation field. This is the area where the majority of treatment takes place, as the effluent is trickled into the subsoil for purification through biological, chemical and

physical processes. The septic tank system works well in areas of suitable subsoil. However, if there is insufficient permeability and/or subsoil depth to take the effluent load, surface ponding may occur resulting in a risk of effluent discharge and runoff of pollutants to surface waters. This poses a severe threat to surface water quality in such areas. Exacerbating the situation, a large increase in development has been observed in recent years in unsewered areas through the construction of:

- Single dwellings or holiday homes;
- Housing clusters of up to 100 homes served by shared treatment systems;
- Commercial premises such as hotels and guesthouses;
- Light industrial facilities.

The following solutions have been proposed to ameliorate the pollution threat with regard to single dwellings, the subject of this paper:

- Connect houses to a proximate existing main sewer system for wastewater treatment in a centralised treatment plant;
- Construct conventional or small bore sewerage through clusters of single house developments, taking the wastewater to a decentralised plant with consented discharge to water courses;
- Use closed storage tanks with regular tankering of effluent to central wastewater treatment works;
- Use on-site technologies such as willow systems, Low Pressure Pipe (LPP), Drip Distribution (DD) or ecosanitation;
- Discharge effluent through an imported media filter, which percolates down into more permeable subsoil or bedrock.

The selection of an appropriate solution(s) for a given site in a low permeability area depends principally on four major criteria: distance from an existing sewerage network; existing septic tank density (for reasons of economy); the T value (how long it takes effluent to percolate through the subsoil); and the depth to bedrock. These four criteria form the nexus of a modelling

framework that takes as its initial input the proposed site of a new dwelling. Through a series of interconnected GIS geoprocesses coupled with user inputs, the model outputs an appropriate solution(s) for a site and ranks the most suitable solution in terms of sustainability and cost. In this way, the model assists local authorities in better decision making for the treatment of domestic wastewater in low permeability zones. The decision support tool resulting from this current research will be GIS enabled, scalable, capable of online spatial modelling and will be accessible eventually to all local authorities, not just Wexford.

Technically, GIS systems are data intensive applications, with the capability to share and distribute big GIS data among geographically dispersed users. They require a complex computing environment. Massive spatial data, especially image data, needs high computational power and requires high data storage, hence the need for big infrastructure both in hardware and software for better programme efficiency. The aim of this research is to develop a web-based decision support system coupled with geo-spatial modelling in a cloud computing environment, offering spatial analysis, spatial data storage and a decision support tool as a web service to users. The work presented in this paper proposes such a cloud-based platform, which will provide an online spatial modelling capability with respect to the aforementioned solutions. This paper also investigates the system architecture functions, system integrations and technical problems associated with cloud computing in the context of GIS.

The paper is divided into seven sections: Section I, the introduction; Section II, a brief discussion of the technical details of cloud computing; Section III, spatial modelling in the cloud computing environment; Section IV, the proposed software architecture; Section V, an overview of the design of the geo-spatial services; Section VI, further development considerations; and Section VII, conclusions drawn.

II. THE CLOUD COMPUTING ENVIRONMENT

Cloud computing has been defined by the National Institute of Standards and Technology (NIST) as, "a model for enabling ubiquitous, convenient on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction" (Mell & Grance, 2011: 2). It is a dynamic and scalable utility model of distributed computing, offering shared resources for infrastructure, software application and business processes. One of the principal advantages is that the consumer only has to pay for the time of service used; the hardware and software capital costs are borne by the cloud host. The rising trend in cloud computing has resulted in a shift of a variety of

applications from the use of operating systems as their core to the web as core. This model of distributed computing eliminates the need to install and run applications on user systems, and has been adopted by Google, Amazon, IBM and most recently, Microsoft.

Contemporary authors (e.g. Mell & Grance, 2011) have categorized the cloud environment into three major service models:

- Software as a service (SaaS): This service offers remote access to commercialised software, including operating systems, databases and software applications, eliminating the need for client-side management. Amazon and Google App Engine are examples.
- Infrastructure as a Service (IaaS): In this model, the provider supplies the underlying infrastructural resources, such as network resources, data centres, storage and virtualisation technology. Amazon EC2 is a prime example of this level of service. Other examples are GoGrid, Flexiscale, Layered Technologies, Joyent and Mosso/Rackspace.
- Platform as a service (PaaS): This model provides a code developing, testing and deploying environment. The application development platform includes operating systems, databases, middleware and up-to-date tools, along with the underlying infrastructure resources supplied by the provider. The consumer is responsible for management tasks that relate to the configuration, optimization and updating of platform resources. Google App Engine is a PaaS example.

There are three principal cloud deployment models: public; private; and hybrid. The choice of model depends on the level of accessibility and security required. The public cloud, and all the applications and data on it, is open to all users. Security on the public cloud is managed by both the provider and the consumer. Well known public cloud services are Amazon EC2, Google App Engine, Salesforce and Eucalyptus. Private clouds are owned by organisations for their private use. They are therefore more restricted and generally more secure. The resources and applications are managed at organisational level, rather than by the provider, with the organisation handling its own data storage. The hybrid cloud is a combination of the public cloud and the private cloud. In this model, a private cloud is linked to one or more external cloud services. It is a more secure way to control data and applications than a public cloud, yet allows third party access to information.

III. SPATIAL MODELLING IN THE CLOUD ENVIRONMENT

Cloud computing is making a significant impact in the field of information technology. The area of GIS is no

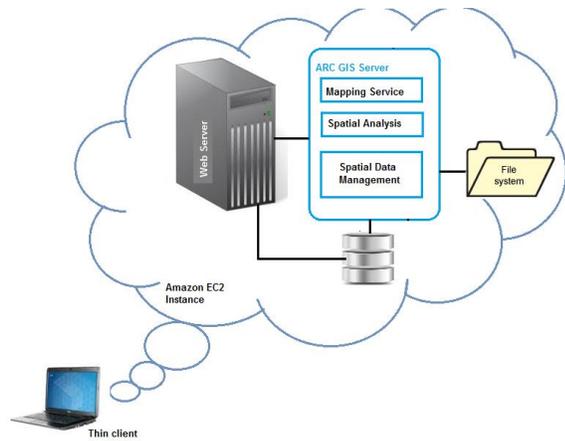


Figure 1. System architecture over Amazon.

exception. For example, Li *et al* (2010) developed a project on Microsoft's Azure platform for downloading and re-projecting satellite imagery; Dominkovics *et al* (2011) designed a decision support system based on a geoprocessing web service and spatial density map to analyse cases of tuberculosis in Barcelona; Nymon Piarsa *et al* (2012) created a spatial decision support system for searching the commercial market place; and Martin *et al* (2011) developed a GIS-based Spill Management Information System (SMIS) to effectively manage the risks associated with the accidental or intentional release of hazardous materials into an inland waterway. In the commercial sphere, ESRI users have been utilising cloud mapping services through ArcGIS Online, and ESRI has recently released ArcGIS Server 10.1 for the Amazon platform. ESRI has also provided enterprise geodatabase machine images and a preconfigured ArcGIS server for Amazon's cloud infrastructure, making a cloud-based GIS system more accessible for consumers.

IV. SYSTEM SOFTWARE ARCHITECTURE

A basic GIS web application over a cloud computing environment follows a server-browser architectural system. In such systems, consumers (termed 'thin clients'), who may be widely geographically distributed, can access server resources from individual locations through their browsers. ESRI's ArcGIS Server, the system used in this project, gives the ability to create, manage and distribute GIS services over the web to support desktop, mobile and web mapping applications. It is a public model, providing Infrastructure as a Service (IaaS) and Software as a Service (SaaS). Utilising this model and services, the architecture of the proposed cloud-based geo-spatial decision support system was constructed. It uses a Windows 2008 server, ArcGIS10.1 server and an internet information server over Amazon's EC2, as shown in the Figure 1. An instance of elastic Amazon cloud (EC2) was created for this purpose through Amazon's AWS

management console. The ArcGIS Server10.1, along with the Windows server 2008 operating system, was then installed on EC2 through ESRI-provided Amazon Machine Images (AMIs). The application was developed using Representational State Transfer (REST) architecture, offering a scalable distributed web application on a cloud platform. Such an application architecture offers consumers direct interaction with the available resources through a combination of Hypertext Transfer Protocol (HTTP) and Uniform Resource Identifier (URI), and standard formats such as Hypertext Markup Language (HTML) and Extensible Markup Language (XML). Client requests to the application are made through a URL, and the requested resource is assigned to the task. By using REST architecture, the client connects not to the server but to the requested resource.

V. SYSTEM DESIGN OF DECISION SUPPORT TOOL

The objective was to design a simple, scalable and independent rich internet application. This was made possible through REST architecture provided by ESRI's ArcGIS Server 10.1. The ArcGIS server exposes its resources to clients through REST application protocol using an HTTP URL, which means that it is accessible to clients through common HTTP methods such as GET and POST. By exploiting common web standards such as HTTP and URI, this approach provides several benefits for application development simplicity, including; scalability, since the addition of new resources does not break with existing client applications due to the uniformity of the access interface; and independency, since servers and clients are loosely coupled and can evolve independently without limiting interoperability. Following this approach, the service architecture of the proposed system was made of a presentation layer, service layer and data layers as shown in Figure 2. The presentation layer at the client end interacts with the set of RESTFUL services in the service layer, which provides.

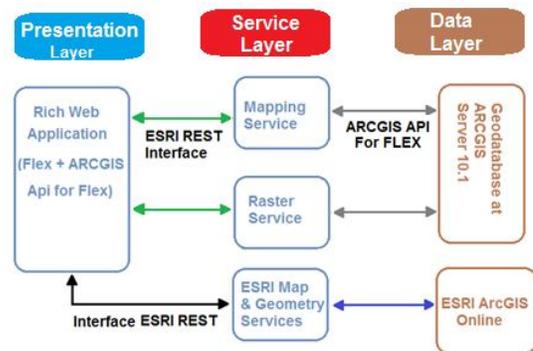


Figure 2. Software architecture.

the image mapping service and geo-processing services to the web application at the client

The principal modelling datasets were: topography; soil and subsoil; rivers; urban and rural sewerage area; wastewater treatment plants; depth to bedrock; likelihood of inadequate percolation; and XY locations of assumed existing domestic wastewater treatment systems, namely, septic tanks. These datasets were packed as file geodatabases on the ArcGIS Server in the form of a series of tables holding feature classes, raster datasets and attributes. The first step after data assimilation was to identify areas of low permeability. The likelihood of inadequate percolation layer was displayed in ArcMAP, onto which existing septic tank locations, roads, rivers, sewerage areas and centralized and de-centralised wastewater treatment plants were added. This was then published as a map service to the ArcGIS Server. The current version of the system offers modelling to identify locations suitable for the first two onsite solutions: 1) connection to a centralised wastewater treatment plant; 2) the construction of conventional or small bore sewerage through clusters of single house developments, taking the wastewater to a decentralised plant with consented discharge to water courses. Two services were designed for this purpose:

- For solution 1): an Online Distance Measurement Service to measure distance between the sewerage area and septic tanks was provided as a combination of geometry and a mapping service.
- For solution 2): an Online Evaluation of the density of septic tanks per square kilometre was performed. This was accomplished by calculating the kernel density of septic tank point data using the Spatial Analyst in ArcMAP. It calculated the density in the neighbourhood around each output raster cell, with the aim of assessing whether it would be economically feasible to recommend a clustered solution. Conceptually, kernel density was determined by fitting a smooth curve over each feature point, such that the surface value was highest at the point and diminished when moving farther from the point. The volume under the surface equals the population field value for the point. The density at each output raster cell is calculated by adding the values of all the kernel surfaces where they overlay the raster cell centre. The output was generated as a raster layer, which was published as a web mapping service on the ArcGIS Server.

VI. CLIENT APPLICATION USING FLEX WIDGETS

All the above-mentioned services were modelled as geo-processing tasks and then hosted on the ARCGIS

Server. A rich internet application was developed using an ARCGIS API for Flex under the Flash Builder environment, to enable access to these services from client browsers. Flex widget technology was used for this purpose. Flex widgets are the software components that run on the client, and communicate and share data with each other using Event Bus architecture as shown in the Figure 3. The Event Bus is a viewer component of the FLEX application that centrally manages the complete communication between viewer components. Another Flex application component (shown as Config Manager in Figure 3) makes the configuration data available for widgets and other modules of the Flex application. The widgets share data with each other through the events dispatched by the base widgets and Data Manager, following the sequence illustrated in Figure 3.

Three widgets have been created. The first, the Buffer widget, measures the distance between the sewerage areas and septic tanks using the geometry service. This widget passes the measured distance to the second widget, the Decision Support Tool widget, using the bus architecture and mechanism defined in Figure 3.

A third widget, called the Identify widget, is used to determine the point kernel density of the septic tanks from the raster layer provided by the map service published on the ArcGIS server. This widget also passes its data to the Decision Support Tool widget.

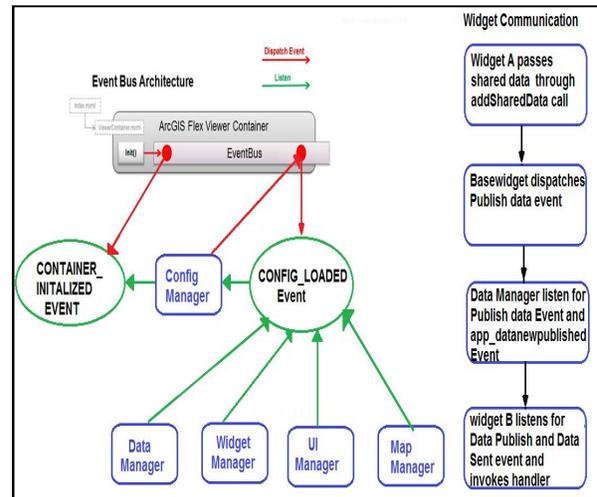


Figure 3. Client application using Flex widgets.

The main decision making is done in the Decision Support Tool, and is based on user-input variables such as the minimum required distance to an existing sewerage network. If the conditions are not met that satisfy this condition, the model automatically calculates for the next condition; septic tank density. The user interface and output of this DSS tool is shown in Figure 4.

VII. CONCLUSIONS

A web-based decision support system has been developed and deployed on the Amazon cloud. This tool will assist local authorities in the process of decision making, for the identification of appropriate domestic wastewater treatment solutions in zones of low permeability subsoil. More software development is planned for evaluation of the remaining strategies, which will also incorporate a sustainability and cost index to indicate preferred solutions for a given situation. In addition, the model could be adapted to assist local authorities in deciding the optimum locations for new decentralised wastewater treatment plants and identifying which facilities should be upgraded should the need arise.

The modelling remains complex, however, and is designed to be a decision support tool; the final decisions on treatment solutions are dependent upon on-site measurement and assessment.

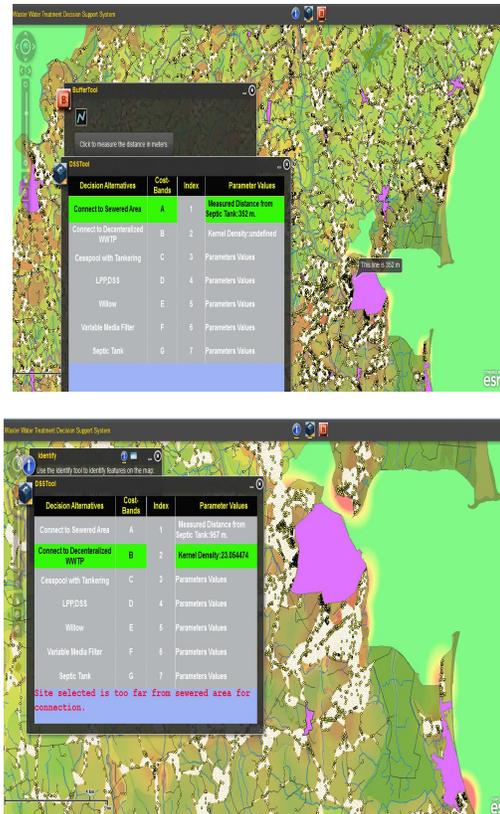


Figure 4. User interface of the system

REFERENCES

- [1] ArcGIS Server on Amazon Web Services <http://resources.arcgis.com/en/help/main/10.1/index.html#//00rq00000020000000>
- [2] Central Statistics Office (2011) Private Dwellings in Permanent Housing Units by Type of Sewerage Facility, Regional Authority, Period in which Built, Aggregate Town or Rural Area and CensusYear. [Online.] Available at: <http://www.cso.ie/px/pxeirestat/Statire/SelectVarVal/saveselectons.asp> (accessed 1 December 2012).
- [3] Dominkovic, P., Granell, C., Perez-Navarro, A., Casals, M., Orcau, A. and Cayla, J. A. (2011) Development of spatial density maps based on geoprocessing web services: application to tuberculosis incidence in Barcelona, Spain. International Journal of Health Geographics, doi: 10.1186/1476-072X-10-62.
- [4] Google App Engine <http://code.google.com/appengine>
- [5] Li, J., Humphrey, M., Agarwal, D., Jackson, K., van Ingen, C. and Ryu, Y. (2010) eScience in the cloud: A MODIS satellite data re-projection and reduction pipeline in the Windows Azure platform, in Parallel Distributed Processing (IPDPS), IEEE International Symposium, April 2010, pp. 1 –10.
- [6] Martin, H. P., LeBoeuf, E. J., Daniel, E. B., Dobbins, J. P. and Abkowitz, M. D. (2004) Development of a GIS based Spill Management Information System. In: Journal of Hazardous Matter, 112(3):239-52.
- [7] Mell, P. and Grance, T. (2011) The NIST Definition of Cloud Computing Recommendations of the National Institute of Standards and Technology. In: National Institute of Standards and Technology Special Publication 800-145. [Online.]
- [8] Nyoman Piarsa, I, Kompiang Oka Sudana, A. A., Gde Wahyu, M.,Gunadi (2012) Web-based GIS by using Spatial Decision Support System (SDSS) Concept for Searching Commercial Marketplace using Google MAP API, International Journal of Computer Applications (0975 – 8887), Volume 50 – No.7, July 2012.
- [9] Paul H. Martin, Eugene J. LeBoeuf, Edsel B. Daniel, James P. Dobbins, Mark D. Abkowitz „Development of a GIS-based Spill Management information System ,Elsevier,2004.

ACKNOWLEDGEMENTS

This project gratefully acknowledges the support of the Irish Environmental Protection Agency and Science Foundation Ireland's StratAG programme.