Cloud-Based Server Design, Specification, and Costing

Integrated Information System for Flood Control and Water Management

By Dr. Dilum Bandara
Dept. of Computer Science and Engineering, University of Moratuwa

1. Introduction

The proposed Integrated Information System for Flood Control and Water Management is a multi-stakeholder project, which aims to forecast and control flooding and water resource management in Sri Lanka. This system is expected to integrate diverse forms of data, such as satellite, weather station, sensor (e.g., river and dam levels), and historical weather and climate data from several data sources. Some of these data feeds to be connected in real time and while others may be accessed from achieving sources. The key functions of the system include monitoring of rainfall, river levels, and street inundations; integration of data from different sources (e.g., point and gridded data); simulation and forecasting; real-time dissemination and control; and short and long term risk assessment. This requires a heterogeneous computing and storage infrastructure due to the diversity of data and programs used to process those data. It is envisioned that the required computing and storage infrastructure to be built using cloud computing, providing benefits such as rapid deployment, low initial cost, scalability, reliability, and agility.

2. System Design

Figure 1 illustrates a high-level design of the proposed computational and storage infrastructure. Servers required to provide desired computational functions and their specifications are listed in Table 1. While the storage to be allocated from the block storage provided by the cloud service provider, each virtual machine instance needs its own disk space to install the operating system and applications. Table 2 list a summary of required servers based on the required operating system, CPU cores, and memory.

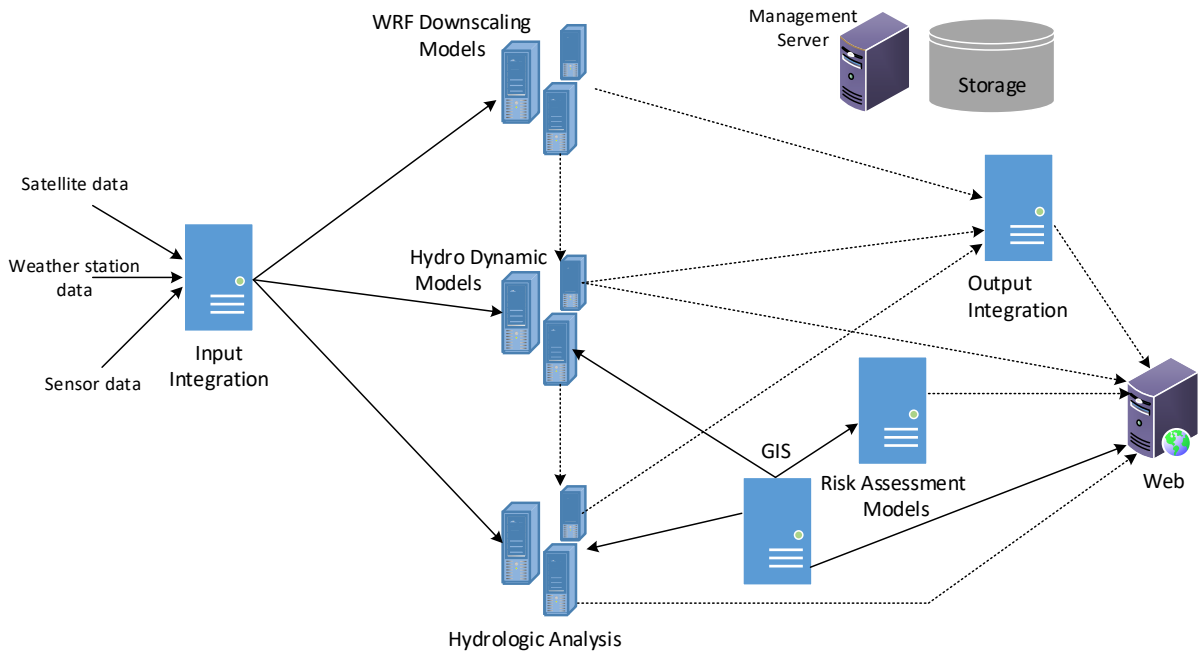

Figure 1 – High-level server architecture.

Table 1 – Server details and specifications.

Server	Description	Operating System	No of Instances	CPU Cores	Memory	Disk
Input Integration	To integrate data from multiple sources using Data Integration and Analysis System (DIAS)	Linux	1	2	8 GB	15 GB
WRF Downscaling Models	To run downscaling models on Weather Research and Forecasting (WRF) Model	Linux	3	8	32 GB	15 GB
Hydro Dynamic Models	To run hydro dynamic models	Linux and Windows	4	2	8 GB	15 GB
Hydrologic Analysis	To run hydrologic analysis models	Windows	2	2	8 GB	15 GB
GIS	To run GIS data sources	Windows	2	2	8 GB	15 GB
Risk Assessment Models	To run risk assessment modules	Windows	1	2	8 GB	15 GB
Output integration	To integrate outputs from multiple modules	Linux	1	2	8 GB	15 GB
Web	To run the web server to visualized the generated data from multiple modules. Apache or NGINX	Linux	1	2	8 GB	15 GB
Management Server	To manage the cloud environment, deployments, software versioning, and server availability	Linux	1	2	4 GB	15 GB
Storage	To store production and archive data	-	1	-	-	3 TB

Table 2 – Summary of servers and respective specifications.

Resource Type	No of Instances	Operating System	CPU Cores	Memory	Disk
Linux – HPC	3	Linux	8	32 GB	15 GB
Linux – Medium	6	Linux	2	8 GB	15 GB
Linux – Small	1	Linux	2	4 GB	15 GB
Windows – Medium	6	Windows	2	8 GB	15 GB
Storage	1	-	-	-	3 TB

Note – Actual server capacities (CPU cores, memory, and disk space) may need to be further refined based on the data volumes and workload once the system is in operation.

3. Cloud Hosting Options

Cloud hosting options from three popular cloud service providers are considered next. Google Cloud, Microsoft Azure, and Amazon Web Services (AWS) are considered for the evaluation, as they can be accessed from Sri Lanka, as well as due to proven to technology and system administrator and developer familiarity. Table 3 summarizes the cost to deploy and operate servers and storage listed in Table 2 for a period of one month and one year respectively assuming a 24x7 operation. Therefore, options such as spot instances, pre-emptible, and per minute billing are not considered, as they tend to be more expensive for long term user. Pricing indicated for the Asian region are considered in the comparison. 100 GB of bandwidth utilization for a month is assumed.

Based on the pricing, hosting the computational and storage infrastructure in the Google Cloud seems to be the most economical solution. While the second best pricing option from Amazon Web Services have virtual machines with the exact memory requirement, memory on Google Cloud virtual machines are slightly lower. However, this should be considered as a minor deviation as all the aspects of the specification is substantially met. This minor deviation contributes to a saving of approximately USD 10,000/- per year.

When estimating the budget, it is advised to keep a 15 – 20% margin during the first three months of operations, as unforeseen requirements, complications, and expenses may be required during the configuration stage. Moreover, there will be other minor expenses such as the cost of obtaining a domain name, digital certificate, and database as a service (for web server).

Depending on the platform chosen about 200 hours will be required to setup and configure the cloud-based system. Hence, it is advised to allocate another USD 2,000 to 3,000 for such expenses.

While 3 TB is estimated as the storage it does not contribute significantly to the overall cost of operation. Hence, it is advisable not look for a hosted/in-house solution only on the basis of need for large storage.

Table 3 – Deployment cost based on Google Cloud, Microsoft Azure, and Amazon Web Services (AWS) pricing as of October 3, 2016.

Resource Type	**Price in USD (instance type)**		
	Google Cloud	**Microsoft Azure**	**Amazon Web Services**
Linux – HPC (per hour)	0.312 (n1-standard-8)[1]	0.692 (D4 v2)[5]	0.498 (m4.2xlarge)
Linux – Medium (per hour)	0.078 (n1-standard-2)[2]	0.173 (D2 v2)[6]	0.125 (m4.large)
Linux – Small (per hour)	0.039 (n1-standard-1)[3]	0.116 (A2)[7]	0.064 (t2.medium)
Windows – Medium (per hour)	0.078 + 0.08 (n1-standard-2) [4]	0.322 (D2 v2)[6]	0.216 (m4.large)
Storage (per GB per month)	0.04 Only OS space is free	0.016 Application space included in VM	First 1TB 0.0300 then 0.0295
Bandwidth (per GB)	0.12	0.087[8]	0.12
Cost per month (USD)	**1,860.56**	**3,774.44**	**2,698.02**
Cost per year (USD)	**22,326.72**	**45,293.29**	**32,376.19**

1 – Only 30 GB RAM
2 – Only 7.5 GB RAM
3 – Only 1 CPU core and 3.75GB RAM
4 – Core-based Windows License for 2 cores
5 – Only 28 GB RAM and include 400 GB disk space
6 – Only 7 GB RAM and include 100 GB disk space
7 – Only 3.5 GB RAM and include 60 GB disk space
8 – First 5GB free