# Cloud computing architecture for scaling biometrics research for big data

Sudeep Sarkar and Ravi Panchumarthy
Department of Computer Science and Engineering,
University of South Florida, Tampa

**The Need:** The identification or verification of persons from physiological and behavioral attributes, such as fingerprint, face, iris, and gait, is the central theme of the field of biometrics. Applications abound from securing personal mobile devices to securing the national borders. Research in this field is essentially empirical and data driven. Algorithms and theory are vetted and designed based on performance on datasets, whose diversity and size become a critical factors. The sizes of commonly available datasets have been growing very fast from MB to GB to TB to PB [1]. In order to perform large scale exhaustive testing of a biometric algorithm, one needs to build a robust system that requires expertise in several computer technologies. This creates a barrier for groups without adequate resources. There is also the important issue of uniformity of experimental protocols. There is great variance in the literature about how experiments are designed and reported for the same dataset.

**Current Approach:** Several groups/evaluators built systems using grid-computing technologies that handle everything from data acquisition till result generation, but these systems require the evaluator to acquire and maintain lot of computational resources [2, 3, 4]. Though these systems solve the issues in standardization, reliability and fault tolerance, building a new system similar to them for a new biometric area of research is expensive, time consuming and a technical challenge. Also regular system maintenance, ensuring highest possible utilization of the hardware and scaling recourses to handle unexpected traffic peaks is hard.

**Solution:** The cloud-computing paradigm offers the best solution to these problems. The key advantage of using cloud-computing architecture is that it can host a common dataset and allow the submission of algorithms either as code or executable, to enforce a standard experimental protocol, and to provide results in a standard format. This facilitates comparing algorithms with each other and benchmarking progress. In order to efficiently service these algorithms, we need powerful systems. Having such systems eliminates the need for procurement of datasets and resources for experimentation, thus lowering the barrier for engaging in biometrics research. The user can lease only the required amount of software and/or hardware recourses for the desired duration without incurring capital and maintenance expenses and also dynamically scale resources if needed.

**Prior Work:** We have conducted a pilot feasibility study [5], with an Amazon Web Services (AWS) education grant, and implemented a prototype of cloud-based evaluation system for gait algorithms. The Human-ID Gait Challenge dataset [6] is one of the principal standards in evaluating gait recognition algorithms. The dataset consist of persons walking in elliptical paths in front of the camera(s) and is about 1.2 TB in size, with 1870 sequences from 122 subjects spanning 5 covariates. The structure of the challenge experiments is in terms of gallery and probe sets, patterned after the Face Recognition Technology (FERET) evaluations [7]. There are 12 different experiments with increasing difficulty. The gallery contains 122 subjects and each experiment contains varying number of probes. This is a method to enforce neutrality between the algorithms being developed. The algorithm to be evaluated must be executed 116,876 times to complete 12 experiments. We used the Infrastructure as a Service (IaaS) offering from Amazon Web Services to build the HumanID Gait Challenge problem on the cloud that allows the user to use this web-app as Software as a Service (SaaS). The current implementation of the cloud-based evaluation system hosts the HumanID dataset (silhouettes), allows submission of algorithms either as code or executable, evaluates on a standard experimental protocol and provides results in a standard format. On a dual core 3GHz machine, running all 12 experiments using baseline gait algorithm took 18 hours whereas using a basic distribution algorithm on AWS, we could complete in 4 hours on a 26-core machine.

**Proposed Use of NSFCloud:** We will use the NSFCloud to develop a general-purpose biometric evaluation platform that can be customized to handle different biometric modalities and diverse dataset types. This will enable future researchers to easily port their biometric evaluation methods onto local or commercial clouds with minimal technical cloud knowledge. The biometric research enthusiasts can focus on the development of algorithms and use the cloud evaluation system to evaluate their algorithm.

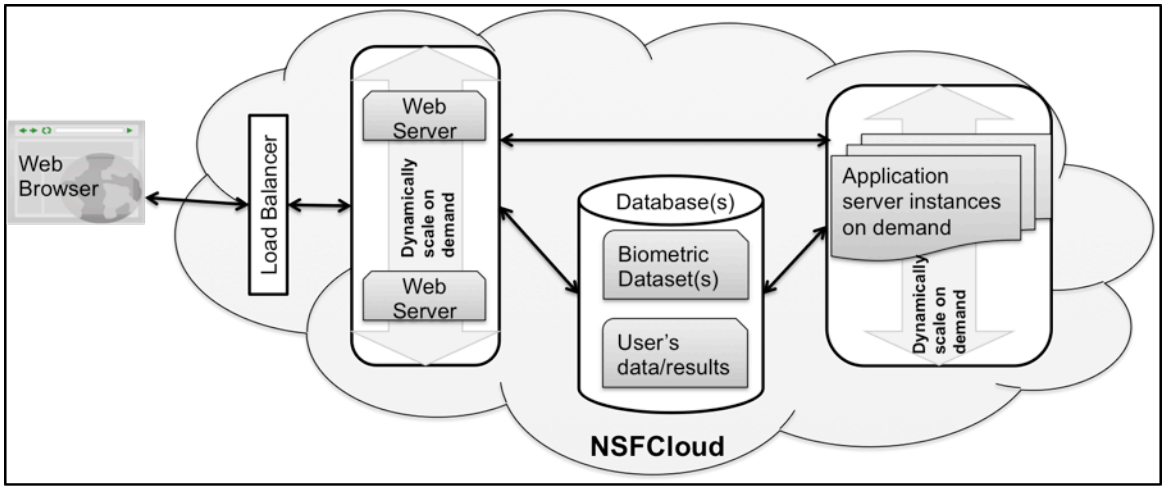


**Figure 1: Proposed cloud architecture for biometric evaluation system**

Evaluation of multimodal biometric systems - which use multiple biometric traits to complete the recognition task – can be leveraged easily as this cloud platform will be a central repository of datasets with different biometric modalities. With NSFCloud projects such as Chameleon and CloudLab, continuing development to production quality architectures and experimenting new cloud architectures towards enhanced performance will be possible. For instance, running all 12 experiments of Human-ID Gait Challenge dataset that require 116,876 executions of algorithm could be distributed to 116,876 cores and complete the evaluation in few minutes as compared to 18 hours on a dual core 3GHz machine.

Simultaneous evaluation of all the user's algorithms being submitted on the entire dataset (1.2TB for Human-ID gait dataset) is computationally intensive which require high processing power. The proposed cloud infrastructure is shown in Figure 1 and the key components are described below:

**Load Balancer:** Manage incoming web-traffic and dynamically scale the web-servers.

**Web Server(s):** Handle incoming user web-requests and deploy the execution of experiments on required number of application servers/nodes in parallel. These web servers must be dynamically scaled up or down based on the demand to ensure high availability and optimize costs.

**Application Servers/nodes:** Evaluate the biometric algorithm based on the preconfigured benchmark experiment protocols. Application servers/nodes are dynamically provisioned to evaluate user's biometric algorithm and decommissioned immediately after completing the evaluation. NSFCloud is a perfect test bed to experiment different distributed architectures such as All-Pairs, MapReduce, MPI-MR, Phoenix, Spark or Apache Flink to get optimum throughput of evaluating user's biometric algorithm.

**Database(s):** Repositories that store biometric dataset(s) and user's data/results.

**Potential Impact:** This system will act as a one-stop look up for all the biometric algorithms and would enable development of progressive increase in algorithm performance. The exhaustively tested algorithms on this system will help the biometric companies focus on developing systems that use these algorithms.

**References:**

[1] Romero, J. J. (2012). India's big bet on identity. Spectrum, IEEE, 49(3), 48-56.

[2] Bui, H., Kelly, M., Lyon, C., Pasquier, M., Thomas, D., Flynn, P., & Thain, D. (2009). Experience with BXGrid: a data repository and computing grid for biometrics research. Cluster Computing, 12(4), 373-386.

[3] Maio, D., Maltoni, D., Cappelli, R., Wayman, J. L., & Jain, A. K. (2004). FVC2004: Third fingerprint verification competition. In Biometric Authentication (pp. 1-7). Springer Berlin Heidelberg.

[4] Yeung, D. Y., Chang, H., Xiong, Y., George, S., Kashi, R., Matsumoto, T., & Rigoll, G. (2004). SVC2004: First international signature verification competition. In Biometric Authentication (pp. 16-22). Springer Berlin Heidelberg.

[5] Panchumarthy, R., Subramanian, R., & Sarkar, S. (2012, November). Biometric evaluation on the cloud: A case study with humanid gait challenge. In Proceedings of the 2012 IEEE/ACM Fifth International Conference on Utility and Cloud Computing (pp. 219-222).

[6] Sarkar, S., Phillips, P. J., Liu, Z., Vega, I. R., Grother, P., & Bowyer, K. W. (2005). The humanid gait challenge problem: Data sets, performance, and analysis. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 27(2), 162-177.

[7] Phillips, P. J., Moon, H., Rizvi, S. A., & Rauss, P. J. (2000). The FERET evaluation methodology for face-recognition algorithms. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 22(10), 1090-1104.