

# SUHAIL REHMAN

---

CONTACT INFORMATION	5514 S Hyde Park Ave, Apt 2 Chicago, IL, 60637 <a href="http://www.suhailrehman.com">http://www.suhailrehman.com</a>	(412)-961-4603 suhail@uchicago.edu
PROFILE	Research-oriented software engineer and developer with instructional experience. Currently pursuing Ph.D. in Computer Science at the University of Chicago. Current Research interests include large-scale data management, lineage inference and distributed systems for the management of data. I've previously worked on projects in cloud computing, virtualization, parallel and distributed computing, general purpose computation on graphics processors (GPGPUs), parallel performance analysis and prediction.	
SKILLSETS	<ul style="list-style-type: none"><li>• <b>Cloud Platforms:</b> Amazon Web Services SDK</li><li>• <b>Parallel &amp; Distributed Computing Frameworks:</b> Apache Spark, Hadoop/MapReduce, Pig, NVIDIA CUDA, MPI, OpenMP</li><li>• <b>Databases &amp; NoSQL:</b> MySQL, HBase, DyanmoDB</li><li>• <b>Web Frameworks:</b> Django, HTML, CSS, JavaScript, D3.js</li><li>• <b>Scripting:</b> Python and Bash</li><li>• <b>Data Science Frameworks:</b> numpy/scipy, pandas, scikit-learn</li><li>• <b>Programming / Scripting:</b> C, C++, Java, Python, bash</li></ul>	
EDUCATION	<b>Ph.D in Computer Science</b> University of Chicago, Chicago, IL	<i>2016 - current</i>
	<ul style="list-style-type: none"><li>• Advisor: <a href="#">Aaron Elmore</a></li><li>• Research Projects: Lineage Inference in large-scale data collections and management of large scale data lakes.</li></ul>	
	<b>Masters in Computational Data Science</b> Carnegie Mellon University, Pittsburgh, PA	<i>May 2016</i>
	<ul style="list-style-type: none"><li>• Capstone: <a href="#">Evaluation of SciDB for Image Processing Tasks</a>, advised by <a href="#">Kayvon Fatahalian</a></li><li>• First-ever MCDS recipient of Departmental Teaching Fellowship, 2014.</li><li>• Sample Coursework: (15640/440) Distributed Systems, (15746) Advanced Storage Systems, Studio in Big Data Systems</li></ul>	
	<b>MS by Research in Computer Science and Engineering</b> International Institute of Information Technology, Hyderabad (IIIT-H)	<i>June 2010</i>
	<ul style="list-style-type: none"><li>• Specialization: High Performance Computing &amp; GPGPU</li><li>• MS Thesis: Exploring Irregular Memory Access Applications on the GPU</li><li>• Worked with Dr. P.J. Narayanan and Dr. Kishore Kothapalli to develop and implement a novel algorithm to solve the list ranking problem on NVIDIA GPUs. This work was also featured in NVIDIA's Annual GPU Technology Conference, 2010 as the State of the Art in GPU Data-Parallel Algorithm Primitives well as on <a href="http://gpgpu.org">gpgpu.org</a>.</li><li>• Worked on a performance prediction model for NVIDIA GPUs.</li><li>• Systems administrator for the research lab, setup and maintained the lab's GPGPU hardware.</li></ul>	
	<b>Bachelor of Technology in Computer Science and Engineering</b> University of Calicut, Kerala, India	<i>June 2007</i>
	<ul style="list-style-type: none"><li>• MES College of Engineering</li><li>• Advanced Courses Taken: Compiler Design, Objected Oriented Programming, Artificial Intelligence</li></ul>	
EXPERIENCE	<b>Research Analyst</b> Carnegie Mellon University in Qatar	<i>Jan 2010 – July 2014</i>
	<ul style="list-style-type: none"><li>• Worked at the Cloud Computing Lab (<a href="#">CCL</a>) under Prof. Majd F. Sakr.</li><li>• Co-Instructor and lead developer for Carnegie Mellon University's online course in Cloud Computing (15319/15619) for the past 5 offerings since 2010. Helped design and impelement several big-data course projects for students to complete on Amazon EC2.</li><li>• Worked on performance analysis and benchmarking of MapReduce applications running on Amazon EC2 as well as vSphere private cloud platforms.</li></ul>	

- Worked on a Human Robot Interaction (HRI) robot receptionist project - **Hala**. Developed a translation system from English-Arabic and vice-versa using the Google Translate API. Also developed a visualization and analytics **dashboard** to view interactions with the roboceptionist using D3.js.

#### **Graduate Research Scholar**

June 2007 – Dec 2009

International Institute of Information Technology, Hyderabad (IIIT-H)

- Master's Thesis: Irregular Algorithms on GPUs

#### **Research Intern**

April 2008 – June 2008

GE Global Research, JFWTC, Bangalore

- Computing and Decision Sciences (CDS) group – exploring GPGPU to accelerate a healthcare application.

## PUBLICATIONS

#### **Safe Double Blind Studies as a Service**

T. J. Skluzacek, **S. Rehman** and I. Foster

In *2017 IEEE 13th International Conference on e-Science (e-Science)*, Auckland, 2017.

#### **A Cloud Computing Course: From Systems To Services**

**M. Suhail Rehman**, Jason Boles, Mohammad Hammoud and Majd F. Sakr

In *Proceedings of the 46th ACM Special Interest Group on Computer Science Education Conference (SIGCSE 2015)*, Kansas City, USA, March 2015.

#### **Center-of-Gravity Reduce Task Scheduling to Lower MapReduce Network Traffic**

Mohammad Hammoud, **M. Suhail Rehman** and Majd F. Sakr

In *Proceedings of The Fifth International Conference on Cloud Computing (CLOUD 2012)*, Honolulu, Hawaii, USA, June 2012.

#### **VOtus: A Flexible And Scalable Monitoring Framework for Virtualized Clusters**

**M. Suhail Rehman**, Mohammad Hammoud, Majd F. Sakr

In *Proceedings of The 3rd International Conference on Cloud Computing and Science (CloudCom 2011)*, Athens, Greece, December 2011.

#### **Teaching the Cloud - Experiences in Designing and Teaching an Undergraduate-Level Course in Cloud Computing at the Carnegie Mellon University in Qatar**

**M. Suhail Rehman** and Majd F. Sakr

In *Proceedings of the 2011 IEEE Global Engineering Education Conference (EDUCON 2011)*, Amman Jordan, April 2011.

#### **Initial Results for Provisioning Variation in Cloud Computing**

**M. Suhail Rehman** and Majd F. Sakr

In *Proceedings of The 2nd International Conference on Cloud Computing and Science (CloudCom 2010)*, Indianapolis, December 2010.

#### **A Performance Prediction Model for the CUDA GPGPU Platform**

K. Kothapalli, Rishabh Mukerjee, **M. Suhail Rehman**, P.J. Narayanan, Kannan Srinathan, Suryakant Patidar

In *Proceedings of the 16th Annual International Conference on High Performance Computing (HiPC)*. Kochi, India, December 2009.

#### **Fast and Scalable List Ranking on the GPU**

**M. Suhail Rehman**, K. Kothapalli, P.J. Narayanan

In *Proceedings of the 23rd International Conference on Supercomputing (ICS)*. New York, USA, June 2009.

## PROJECTS

#### **Draining the Data Swamp - Crawler**

Build a file-system agnostic crawling interface and pipeline to ingest file system metadata and build a catalog of metadata and inferred structural properties of files.

#### **Co-Instructor and Lead Developer - CMU's Online Course in Cloud Computing**

Worked closely with Prof. Majd Sakr at Carnegie Mellon University in Qatar to develop and deploy CMU's first course in **Cloud Computing**, offered as a conventional course in Spring 2010

and 2012, and then as an Online Course using CMU's Online Learning Initiative (OLI) Platform since Spring 2013. The latest iteration of the course (Spring 2014) has 240 students. Was the lead in development of hands-on projects for the students on AWS. Also assisted in course content development and was the lead Web content developer for the course.

#### **Hala - Bilingual Robot Receptionist**

**Hala** is a robot receptionist designed to explore algorithms for human robot interaction in a mixed cultural setting. Added a new translation feature in the robot that utilized the Google Translation API. Developed tools and analytics to mine data from the Receptionist's Interaction logs, created a Web 2.0 dashboard in D3.js to consolidate and present the interaction logs. Assisted in maintaining the roboceptionist code base.

#### **Multi-Layered Monitoring Framework for Virtualized Clusters (VOtus and Almond)**

Performance analysis in virtualized cluster environments is difficult and cumbersome owing to the number of layers involved (Application, API/Platforms, Guest OS, Hypervisor and/or Host OS, Bare metal Hardware), each having a dedicated tool to collect and analyse metrics. I have built a multi-layered monitoring framework called the All-Monitor Daemon (Almond), which can be used to collect metrics from the different layers of this stack. Almond uses a scalable storage backend (OpenTSDB), which efficiently collects and stores large volumes of metric data.

#### **Center of Gravity Reduce Task Scheduler for Hadoop MapReduce (CoGRS)**

CoGRS is a new task scheduler for Hadoop MapReduce that I helped develop. This scheduler aims to reduce network traffic induced during the shuffle phase of a MapReduce job, by attempting to assign reducers to nodes which either have the data needed, or are closest to it. This approach reduces off-rack network traffic on average of upto 9 - 10 % and yeilds performance improvements of up to 23% on the applications that we have tested.

#### **Irregular Algorithms on the GPU - MS Thesis**

In this project, I had explored the possibility of running irregular algorithms (those that do not have deterministic memory access patterns, such as the list ranking problem) on massively parallel computational settings such as GPGPU. I was able to develop a novel recursive algorithm and reported the fastest performance at the time. This algorithm is a primitive to many others including Euler Tour Technique (ETT)-based parallel graph computations, range queries and bzip2 compression. This work was also featured in NVIDIA's Annual GPU Technology Conference, 2010 as the **State of the Art in GPU Data-Parallel Algorithm Primitives** well as on [gpgpu.org](http://gpgpu.org). I had also contributed to the development of an performance prediction technique to estimate the runtime of GPU kernels through a simple mathematical performance model.

#### TEACHING

**University of Chicago:** Graduate Teaching Assistant:

- CMSC 12100 - Computer Science with Applications (Autumn 2016)
- CMSC 15200 - Introduction to Computer Science II (Winter 2017)
- CAPP 30235 - Databases for Public Policy (Spring 2017/2018)

**Carnegie Mellon University in Qatar / Pittsburgh:** Helped design and develop multiple offerings of CMU's first and only course in *Cloud Computing*, the latest version of which was taught in both Qatar and Pittsburgh campuses.

**IIT-H:** Course tutor for GPGPU/CUDA in *Multicore Architectures* and *Computer Vision* courses. Was a course instructor for a *Parallel Programming Module* in the *Computing Tools* course.