

MAHESH BALKRISHNA CHAUDHARI

37335 Lantana Common
Fremont, CA 94536
Mobile: (602) 524-0610

mahesh.chaudhari@gmail.com
<http://deparesearch.lab.asu.edu/>
<http://maheshchaudhari.com>

Qualifications

- 3 years of experience in **managing** and **building international team** of engineers to deliver high quality product
- 6 years of hands-on experience in **architecting** and building **complex data integration systems** for processing **terabytes of data**
- 15 years of practical and research experience in designing and developing data stores with **Relational** and **NoSQL** data models
- Expert in building complex **data-processing pipelines for streaming and batch-data** using cloud infrastructure and **cloud technologies**
- Established **mentoring skills** to motivate engineers to undertake responsibilities and learn new technologies

Awards and Honors

- Won the **Graphies Award** for “**The Most Innovative Graph Application in Health Care**” at Graph Connect International Conference 2013
- **Awarded a 3-year National Science Foundation (NSF) research grant** to conduct innovative research in **multiple query optimization** and incrementally maintain **materialized views** over **heterogeneous data sources** in a distributed events and stream processing environment
- Awarded a 3-month NSF supplemental research grant to develop a **hybrid benchmark** with complex queries over relational and XML data sources
- Recognized as the **Preparing Future Faculty (PFF) Emeriti Fellow** for the academic year of 2009-2010 for excellence in research, teaching & mentorship

Professional Experience

- **Principal Software Engineer** 05/2016 - Present
Neustar Inc., San Francisco, CA 94105
 - Principal design engineer for **Data Science Development Kit (DSDK)** to provide fully tagged and unified advertising data to the clients for their personalized data analytics.
 - Improved the performance of the DSDK process by **50% by optimizing** the Spark jobs used in the extract process.
 - Designing the next generation DSDK platform using **Apache Spark, Spark streaming** with **EMR, S3 and Athena**.
 - Designed and delivered new features like **pull-based data model, cross-region, cross-account access, trusted relationships**.
 - Managing a highly motivated international team for **timely delivery of new features for data analytics**.
 - Deep understanding of distributed data processing pipelines over **Spark, and Hadoop infrastructures**.
- **Software Architect** 09/2014 – 04/2016
Zephyr Health, San Francisco, CA 94105
 - Successfully designed and deployed **complex sales analytics** product fetching **multi-million dollar contracts**
 - **Skillfully architected**, designed and delivered the next generation data integration platform using **MongoDB and Neo4j**
 - Designed **Ontology-driven domain Models** for **Disparate and heterogeneous Pharmaceutical data**
 - Designed an **event-based Kafka system** to **stream incremental data** to the domain models
 - Guided the data engineering team on performance improvement using state of the art **non-relational data modeling** and **query design** for enterprise-level multi-tier architecture
 - Migrated the company’s infrastructure from silo multi-client architecture to a **full-stack multi-tenant architecture**
 - Guided the dev-ops team on building automation-based infrastructure for **scalable software deployment** using Bamboo, Stash, bash scripts
- **Senior Software Engineer** 09/2012 – 08/2014
Zephyr Health, San Francisco, CA 94105
 - Design and implementation of **NoSQL** and **Graph-based** database systems for **Big Data Analytics and visualization**
 - Provided **continuous integration** solutions for next generation platform for pharmaceutical and healthcare data
 - Successful deployment of multiple **MongoDB** and **Neo4j** data servers for high-availability on **Amazon EC2 Cloud services**
 - Well adapted to **Scrum/Agile** software development framework using latest tools such as **Jira, Stash, Jenkins**
- **Postdoctoral Research Associate [Supported by NSF Grant CSR #0915325]** 06/2011 – 08/2012
Division of Mathematical and Natural Sciences, ASU, Phoenix, AZ
 - Designed a benchmark for evaluating the performance of a distributed environment containing heterogeneous data sources
 - Transformed the relational TPC-H model into a **hybrid database** containing relational & structured XML data sources
 - Modified the TPC-H data generator program in C to generate valid data for scale factors less than 1 GB
 - Implemented the hybrid database in Oracle 11g, SQL Server and MySQL 5.5 with **varying scale factors**
 - Optimized and translated the **22 complex SQL queries into LINQ queries** to access the heterogeneous data sources
 - Benchmarked the queries over different sizes of the databases with and without caching and published the results in a referred journal
 - Enhanced the database curriculum at Arizona State University to include the benchmarking theory and practical

- Mentored undergraduate students in database research and guided them for higher education
 - Awarded a 3-month supplemental research proposal by NSF to explore the design and use of benchmarks in assessing the Distributed Event Stream Processing Environment and to introduce benchmarks in database curriculum
 - Explored **research opportunities** in multiple query optimization and materialized views in the **cloud environment** and object-relational mapped (ORM) databases
 - **Enhanced the CS2 course** (ACO 102) with the introduction to Java programming using Android SDK and provided hands-on experience in deploying the programs onto an android-based smartphone
- **Graduate Research Assistant [Supported by NSF Grant CSR #0915325]** 07/2009 – 05/2011
Division of Mathematical and Natural Sciences, ASU, Phoenix, AZ
 - Awarded a 3-year research grant by NSF to support my dissertation and to support an undergraduate student [CSR #0915325, 08/2009-08/2012]. This grant is focused on investigating research issues related to **detecting common subexpressions** for multiple query optimization and defining and maintaining hybrid materialized views over **heterogeneous data sources** in a distributed event stream processing environment
 - Successfully developed a **distributed framework of autonomous agents** to communicate and exchange information using WCF and C#
 - Formalized a **mixed multigraph model** to represent heterogeneous query expressions over relational and XML data sources in a common graph for multiple query optimization using QuickGraph 3.6 and C#
 - Implemented a **heuristics-based algorithm in C#** to detect common subexpressions as the potential candidates for materialized views from the mixed multigraph model
 - Designed an algorithm to define and **incrementally maintain hybrid materialized views** over relational and XML data sources using Magic Sets optimization technique
 - Mentored an undergraduate student in the upcoming areas of events and stream processing, ORM, LINQ, XML and database benchmarks
 - Guided an undergraduate student to develop “Object Manager” application that works with object databases stored either in the main memory or in db4o, an open source object database
 - **Graduate Research Assistant [Partially supported by NSF Grant ITR #0312849]** 08/2004 – 08/2007
PRISM, Department of Computer Science and Engineering, ASU, Tempe, AZ
 - 3D Face Authentication**
 - **Lead database developer** for designing a **large-scale object-relational database** to work in coordination with XML and binary data regarding 3D face scans in Oracle 10g
 - Developed a web application using ASP.NET, C# and Oracle 10g for the demonstration of 3D face authentication and recognition project and the web application was recognized by **National Science Foundation (NSF)** as a “**Nugget**”
 - The system was able to handle **2325 face scans from 1673 participants** of varying age groups and ethnicities for authentication and recognition purposes
 - Analyzed the **performance issues** related to storing XML and binary data in relational databases and chose file pointers as an efficient way to handle the 3D face data files
 - Cumulus Photogrammetric, In-Situ and Doppler Observations (CuPIDO)**
 - Implemented a **graph plotting tool using VC++** for representing sounding data along with other geothermal parameters
 - Developed programs for **volumetric visualization of cloud data** using Interactive Data Language (IDL) software on Mac OS X
 - Successfully converted legacy Fortran programs into VC++ programs for calculating the sounding data and other thunderstorm clouds detection parameters
 - I³DEA and ATIC websites**
 - Designed a database framework to store information related to ATIC such as faculty, staff and students profiles, projects and publications details using MySQL 5.1
 - Developed I³DEA and ATIC websites using ASP.NET and MySQL database
 - **Software Engineer** 09/2003 – 06/2004
Department of Plant and Soil Sciences, Mississippi State University, Starkville, MS
 - Conducted analysis for optimal performance of the **Real-time data acquisition and control system** (SPAR) to process **streaming sensor data** over 180 incoming channels
 - **Redesigned** the data acquisition and control system using Visual Basic 6.0 to communicate over six new HP Agilent 34970A data acquisition/switch units for processing **incoming streams on 180 channels** and **outgoing streams on 300 channels** within a time frame of **10 seconds**
 - Developed **graph-based reporting software** in Visual Basic 6.0 for analyzing the collected data at real-time
 - **Research Assistant** 09/2001 – 08/2003
Mitchell Memorial Library, Mississippi State University, Starkville, MS
 - Genealogy XML Search Engine**
 - Designed and developed software for **automatic conversion and indexing** of Genealogical references from WordPerfect documents into an **XML database**
 - Developed a **web-based search engine** in ASP 3.0 to retrieve records from this XML database efficiently using SAX, XPath and XSLT
 - Presented the genealogy search engine at the Regional Genealogy Fair on June 14, 2003
 - Web-based Time Management System for work-study students**
 - Enhanced Web-based Time and Task management application (TimeClock) that manages time sheets for student workers to allow supervisors to assign and maintain tasks for the student workers and to allow students to enter and maintain their class schedule for each semester

- Created **report generation feature** for supervisors to report money that was spent per department per student, reports on tasks, and web-based project tracking and to provide project status reports to external/internal funding agencies

Other significant contributions to the University Library

- Analyzed and enhanced the efficiency of large databases at the library using **normalization techniques** and **query optimization**
- Developed an **integration framework** to collaborate library databases with university level databases especially for authentication purposes using LDAP
- **Transformed new ideas** into successful web-based database-driven projects such as “MSU Authors”, and “Books Order Online”

• **Software Engineer**

07/1999 – 02/2001

Indian Institute of Technology (IIT), Mumbai, India

- Developed the **multimedia-based object-relational database** in Oracle 8.0 to maintain information about different parts of the aircraft in various formats such as text, jpeg, audio, VRML and animation
- **Integrated** the Illustrated Parts Catalogue (IPC) project designed for automating aircraft spare parts maintenance with Aircraft Systems Maintenance Simulator (ASMS) project so that the ASMS project can retrieve information from the common object-relational database for simulation purpose

Technical Skills

- **Database Systems** : MongoDB, Neo4j, MySQL 5.6, PostgreSQL 4.2, Redshift, Athena, HDFS, MS SQL, Oracle 11g
- **XML Technologies** : XML, XPath, XSLT, SAX, DOM
- **Query Languages** : SQL, Cypher, LINQ, XQuery, NoSQL
- **Programming Languages** : C#, Java, C/C++
- **Technologies** : Kafka Event Framework, Map Reduce, Spark, Spark Streaming, AWS Console, AWS Command Line, Sybase Complex Event Stream Processor, Windows Communication Foundation (WCF), Oracle BPEL Process Manager, Android SDK, QuickGraph 3.6
- **Web Development** : ASP.NET 3.5, PHP5, HTML, CSS, Groovy and Grails
- **Development Tools** : IntelliJ, Eclipse, NetBeans, Microsoft Visual Studio

Teaching Experience

- **Faculty Associate & Teaching Assistant** 08/2011 – 05/2012
Division of Mathematical and Natural Sciences, ASU, Phoenix, AZ 08/2007 – 05/2009
- Responsibilities:**
 - **Designed the course curriculum**, prepared instructional material and assignments
 - Enhanced the course curriculum by introducing new topics:
 - Java programming using **Android SDK** for smartphones
 - C# programming using Visual Studio 2008/Visual Studio 2010
 - Used the **online assessment tool WileyPlus** for assigning the student with additional practice and assignments on Java programming
- Courses:**
 - ACO 100: Overview of Applied Computing (Fall 2007)
 - ACO 101: Introduction to Computer Science (Spring 2009, Fall 2008, Spring 2008, Fall 2007)
 - ACO 102: Principles of Computer Science (Spring 2012, Fall 2011, Spring 2009, Fall 2008, Spring 2008, Fall 2007)
 - ACO 320: Database Systems and Transaction Processing (Fall 2007)
 - ACO 420: Object Databases (Spring 2008)

Education

- **Ph.D. in Computer Science and Engineering** May 2011
Arizona State University, Tempe, AZ
Dissertation: “Materialized Views over Heterogeneous Structured Data Sources in a Distributed Event Stream Processing Environment”
Committee Members: Dr. Suzanne W. Dietrich (**Chair**), Dr. Susan D. Urban, Dr. Hasan Davulcu, and Dr. Yi Chen
- **M.S. in Computer Science and Engineering** August 2003
Mississippi State University, Starkville, MS
Project: “Conversion of Legacy Genealogy Index into XML Database”
Committee Members: Dr. Edward Allen (**Chair**), Dr. Thomas Philip, and Dr. Lynne Mueller
- **Post Graduate Diploma in Advanced Computing** July 2001
Centre for Development of Advanced Computing (C-DAC) Mumbai, India
Project: “Online Shopping Cart” using ASP and Oracle 8i
- **Bachelor of Engineering in Computer Engineering** May 1999
University of Mumbai, Mumbai, India
Project: “Inventory Control Systems” using Visual Basic 6.0 and Oracle 8i
Project Supervisor: Assistant Professor Merly Thomas

Publications

Book Chapters

- Anshuman Razdan, Gerald Farin, Myung Soo-Bae and **Mahesh Chaudhari**, State of 3DFace Biometrics for Homeland Security Applications, Book chapter in *National Security (Part of the Elsevier Publishing Handbooks on Information Systems)*, Eds Hsinchun Chen, T. S. Raghu, Ram Ramesh, Ajay Vinze and Daniel Zeng, 26 April 2007, pp. 73-99.

Referred Journals

- **Mahesh B. Chaudhari**, Suzanne W. Dietrich, Jennifer Ortiz and Spencer Pearson, Towards A Hybrid Relational and XML Benchmark for Loosely-Coupled Distributed Data Sources, *Journal of Systems & Software*, November 2015, pp. 78-87, DOI: <http://dx.doi.org/10.1016/j.jss.2015.07.029>
- **Mahesh B. Chaudhari** and Suzanne W. Dietrich, Detecting Common Subexpressions for Multiple Query Optimization over Loosely-Coupled Heterogeneous Data Sources, *Journal of Distributed and Parallel Databases*, Springer Publications, December 2014, pp. 1-27, DOI: <http://dx.doi.org/10.1007/s10619-014-7166-6>.

Referred Conference Proceedings

- Jennifer Ortiz, Suzanne W. Dietrich and **Mahesh B. Chaudhari**, Learning from Database Performance Benchmarks, *Journal of Computing Sciences in Colleges*, 27, 4 (Apr. 2012), 151-158 (In Proceedings of Consortium for Computing Sciences in Colleges, Southwestern Region, Stockton, CA, March 2012).
- Suzanne W. Dietrich and **Mahesh Chaudhari**, LINQ ROX! Integrating LINQ into the Database Curriculum, In *Proceedings of ACM SIGCSE International Conference on Computer Science Education*, Dallas, Texas, March 2011, pp. 293-298.
- **Mahesh B. Chaudhari**, A Distributed Event Stream Processing Framework for Materialized Views over Heterogeneous Data Sources, *VLDB 2010 Ph.D. Workshop*, Singapore, September 13-17, 2010.
- **Mahesh B. Chaudhari** and Suzanne W. Dietrich, Metadata Services for Distributed Event Stream Processing Agents, In *the 19th International Conference on Software Engineering and Data Engineering (SEDE2010)*, San Francisco, June 16-18, 2010, pp. 307-312.
- Suzanne W. Dietrich and **Mahesh Chaudhari**, The LINQ between XML and databases: a gentle introduction, *Journal of Computing Sciences in Colleges*, 25, 4 (Apr. 2010), 158-164 (In Proceedings of Consortium for Computing Sciences in Colleges, Southwestern Region, Thousand Oaks, CA, March 2010).
- Suzanne W. Dietrich and **Mahesh Chaudhari**, The Missing LINQ between Databases and Object-Oriented Programming Languages: LINQ as an Object Query Language for a Database Course, *Journal of Computing Sciences in Colleges*, 24, 4 (Apr. 2009), 282-288 (In Proceedings of Consortium for Computing Sciences in Colleges, Southwestern Region, San Diego, CA, April 2009).

Conference Presentations

- **Mahesh B. Chaudhari**, Integrating Diverse Healthcare Data using MongoDB and Neo4j, at *GraphConnect*, San Francisco, CA, October 21, 2015.
- **Mahesh B. Chaudhari**, Ontology-driven modeling of healthcare data using a graph database, at *the Stanford Medicine X Conference*, September 26, 2015.
- **Mahesh B. Chaudhari**, Modeling Ontology-Driven Healthcare Data over NoSQL Solutions, at *the Fifth Annual NoSQL Now! Conference*, San Jose, California, USA, August 18, 2015.
- **Mahesh B. Chaudhari**, Designing an adaptive ontology-driven model for Healthcare data in MongoDB, at *the 2nd Annual Global Big Data Conference*, Santa Clara, California, USA, September 10, 2014.
- **Mahesh B. Chaudhari**, Avoiding Deadlocks in Neo4j on Zephyr Analytics Platform, at *the Fourth Annual NoSQL Now! Conference*, San Jose, California, USA, August 12, 2014.
- **Mahesh B. Chaudhari**, Case Study: Avoiding Deadlocks and Other Lessons Learned in Graph Database Implementation at Zephyr Health, *TDWI Solution Summit: Delivering Business Value from Big Data Analytics*, Savannah, GA, USA, March 21, 2014.
- **Mahesh B. Chaudhari**, Avoiding Deadlocks in Neo4j on Z-Platform, at *GraphConnect*, San Francisco, California, USA, October 12, 2013.

Poster Presentation

- **Mahesh B. Chaudhari**, Combining Teaching, Research, Learning & Service, at *the Preparing Future Faculty Capstone Fair*, April 24, 2009.

Reviewer

- Consortium for Computing Sciences in Colleges
- SIGCSE Technical Symposium
- ITiCSE - Innovation and Technology in Computer Science Education
- GPSA: Jumpstart Research Grant 2010, Graduate Research Support Program 2009

Professional Memberships

- Member of The Upsilon Pi Epsilon Honorary Society
- Member of Association for Computing Machinery (ACM)
- Member of Institute of Electrical and Electronics Engineers (IEEE)