# K A M A L P A T H A K [kamalpathak10@gmail.com](mailto:kamalpathak10@gmail.com) Pune - India Mb. No. +91 9922377599

## Profile in brief

Total 13+ years of experience in S/W development in Java and other open source tools & technologies like, Apache Spark, Hadoop, Solr, Lucene, Nutch etc. Working experience on various clouds AWS/AZURE/OpenStack - Private Cloud**.** Prior Experience in developing & deploying Linguistic/NLP software and solutions. I have experience of driving a project right from its inception to a successful end.

## EXPERIENCE DETAILS

**COMPANY** : ATOS Global Pvt. Ltd. Pune - India  
**DEPARTMENT** : Business Technology and Innovation Centre

**DURATION** : May 2014 – Till Date

**DESIGNATION** :Technical Architect  
**ROLE** : I’m part of team, that is responsible for Big data and analytics platform, I lead the team to design and develop the platform, orchestrated the deployment across various public and private clouds. My responsibility includes, leading technical team, design & architect solutions on various public & Private clouds. Build solutions using cloud native services. Assisting team’s in responding to customer RFI’s & RFP’s. I have worked on IoT projects for various clients. I’m also responsible for performing various POC’s on IoT & big data.

**COMPANY** : C-DAC Research & Development Center, Pune - India

**DURATION** : October 2005 – May 2014

**DESIGNATION** :Senior Technical Officer  
**ROLE**  : As part of Search Engine team, I lead the team to design and develop distributed search platform. My responsibility includes, design & develop comprehensive, resilient, distributed, highly available and scalable Search engine platform using open source tools & technologies. Provide relevant search on structured and un-structured data. Design and built modules for enabling Indian language search. Provide search as a service on thousands of government websites. I was also involved in writing project proposals and various client demos. Worked with various government departments like ECI, “Patent & IPR division”, NIC, IRCTC, RECs IIIT & IITs. Contributed in winning a project of ₹ 250 million from Election commission of India.

**Technical SKILLS**

|  |  |
| --- | --- |
| **Programming Languages** | Java, J2EE (Servlets, JSP, Spring MVC), Java Script, HTML/CSS |
| **Web Related** | Bootstrap, Highcharts, XML, AJAX, Tomcat Web container, RESTful Web Services |
| **Databases** | HBase, MongoDB, My SQL, PostgreSQL, MongoDB, MS SQL Server |
| **Cloud/Bigdata** | AWS, Azure, OpenStack, Azure HDInsight, Azure ML Studio, Hortonworks Data Platform, Apache Hadoop, Sqoop, Hive, Apache Spark, Solr, Lucene, Nutch, RabbitMQ, Docker, Kubernetes |
| **Source Safe & Others** | Stash, GitHub, Bit Bucket, Jira, Jenkins, Confluence, Crucible |

## EDUCATION

|  |  |  |  |
| --- | --- | --- | --- |
| **Name of Exam.** | **Year of Passing** | **Board/University** | **Grade/Percentage** |
| M.S (Software systems) | 2011 | BITS, Pilani | 8.1(CGPA) |
| PG Diploma (DAC) | 2005 | C-DAC | B+ |
| B.E. (Information Technology) | 2003 | RGPV, Bhopal (M.P) | 70.56 |

## aWARDS/GROWTH PATH

* Best team award for year 2007.
* Designated as “Project Leader” in year 2008.
* Best team award for the project “Patent Analysis & Management System” & project “Localization Project Management System” for year 2010.
* Designated as “Team coordinator” in year 2010.
* Best team award for project “Government of India Search Engine” for year 2012.
* Best performer award for year 2012.
* Designated as Technical Architect in 2016.

## Certification/Workshop/Independent course work

* “AWS Certified Solutions Architect Associate” online course from Udemy.
* “Big Data Analytics with HDInsight” from Microsoft Virtual Academy
* “[Hadoop Essentials” from MAPR Academy - HDE 100](https://kamalpathakblog.files.wordpress.com/2015/10/1443678693.pdf)
* “[HBase Data Model and Architecture” from MAPR Academy - DEV 320](https://kamalpathakblog.files.wordpress.com/2015/10/1443678661.pdf)
* “Hands on workshop on “Hadoop (Admin)” Organized by “Computer Society of India”, Mumbai Chapter in December 2012.
* “[MongoDB for Developers” from MongoDB University - M101P](https://university.mongodb.com/course_completion/d19b157b11574ca9ab8e12d0e716ba75)
* “[MongoDB for DBAs” from MongoDB University - M102](https://university.mongodb.com/course_completion/40f69be119e14e72932402afe5e6b68b)
* Winter school on “Information Retrieval” conducted by “Information Retrieval Society of India (IRSI)”at “Indian Statistical Institute”, Bangalore in December 2011.
* Certification course on “Patent Landscape Analysis” conducted by “Global Institute of Intellectual Property”, Delhi in September 2009.

## National & International Conferences

* Participated as speaker in “16th Annual LRC Internationalisation & Localisation Conference” at Limerick, Ireland in September 2011.
* Participated as speaker in “AGIS 2010” at New Delhi in December 2010
* Organized International conference on “ICT-IPR Trends and Challenges in Globalized World” at New Delhi in April 2011.
* Participated in “Global WordNet Conference” at IIT Bombay in January 2010.

## PERSONAL INFORMATION

|  |  |  |  |
| --- | --- | --- | --- |
| **Date of Birth** | April 10, 1981 | **Sex** | Male |
| **Marital Status** | Married | **Languages Known** | English & Hindi |
| **Present Address** | Sr. No 44, Flat No. C-506, Sky Heights, Undri-Pisoli Road, NIBM Annex , Pune | **Permanent Address** | S/O K.C Pathak  Sadbhawana Vihar, Peeli Kothi, Bari Mukhani, Haldwani Nainital, Uttrakhand - 263139 |
| **Email** | [Kamalpathak10@gmail.com](mailto:Kamalpathak10@gmail.com) | **Contact No.** | **+91 9922377599** |

**Refer section “**[***Project Profile***](#_PROJECT_PROFILE)**” for project undertaken**

## PROJECT PROFILE

**Project 1: “Demand Forecasting & Wind Turbine Power Prediction & Predictive Maintenance”**

**Technologies:** Azure Data Factory, Azure HDInsight Spark, Azure Machine Learning Studio, Power BI, Azure Event Hub, Azure Stream Analytics, Hortonworks Data Platform

**Duration:** **Jan** **2018 – Till Date**

**Responsibilities:**

* Designed batch and real-time data ingestion pipelines using Azure cloud native services.
* Designed Spark data transformation job.
* Achieved reduction in data ingestion time by ¼ of the original time by configuring right number of partitions, executors, executor core and re-partitioning of data in Apache Spark.
* Configured optimal partitions and scaling units for Azure Event hub and Azure Cosmos DB to maximize the throughput.
* Developed multi-threaded java utility as event generator simulator for sending events to Azure Event hub.

**Project 2: “Atos Codex – A Data Analytic Platform on Cloud”**

**Technologies:** Hortonworks Data Platform(HDP), Jupyterhub, RStudio, DSW, AWS/AZURE/Openstack

**Duration:** **Jan** **2016 – Dec 2017**

**Responsibilities:**

* Worked as technical lead for this project.
* Blueprinting various components using Java & Shell Script.
* Orchestration of deployment on various clouds using Apache Brooklyn based tool.
* Reduced client’s AWS monthly infra bill from 6,500$ to 3,000$ by hardening security group rules and optimizing ec2 instances types.
* Simplified Java code to remove complexities by code re-factoring and added support for various clouds deployments.
* Containerization of different components of Codex stack using Docker and management using Kubernetes

**Project 3: “Internet of Things (IoT) - Connected Machines”**

**Technologies:** Java, RabbitMQ, HBase, Spring MVC, Spring data template, Spring REST services, Html5, Bootstrap, High charts

**Duration:** **May** **2014 – Dec 2015**

**Responsibilities:**

* Developed multi-thread Java socket server to listen for events on TCP socket.
* Capturing data from sensors & performed ETL operation on data.
* Written Spring MVC based rest services for data storage and data retrieval from database.
* Design and development of a web based responsive user interface using Bootstrap & HTML5
* Designed and developed static and dynamic real-time charts using Highcharts.
* Setup of Distributed HBase cluster as historical datastore.

**Project 4: “Voter Information Search for Election Commission of India”**

**Technologies:** Solr Cloud, Java/J2EE, Keepalived, HAProxy.

**Duration:** **Jan** **2014 – May 2014**

**Responsibilities:**

* Worked as Technical lead & coordinator among different teams for this project.
* Design & Developed a distributed, highly available search service using Solr Cloud.
* Achieved end-to-end response time of 2-3 second for a user query, by configuring appropriate number of Solr shards and memory of systems.
* Enabled and implemented geo-spatial search to get the information of nearest pooling booth.
* 800 million records of voter’s data exported from various state level RDBMS and Imported into Solr index in a short span of time.
* Live system can be accessed by the URL <http://electoralsearch.in>

**Project 5: “Government of India (GoI) Search Engine”**

**Technologies:** Apache Nutch, Lucene, Solr, Hadoop, Stanford NER, JAVA, RESTful Services, VN, Bugzilla, HAProxy.

**Duration:** **March 2011 – December 2013**

**Responsibilities:**

* Writing Project Proposal, Requirement analysis, Design & Development.
* Worked as team-coordinator & single point of contact for client and different teams.
* Designed and developed modules for Indian language search. Support for languages like Hindi, Marathi, Gujarati and Urdu were enabled.
* Implemented NER based faceted search.
* Language detection using N-gram analysis and other Linguistic inputs.
* A distributed cluster of Nutch (which has in-built support of Hadoop) was implemented for crawling of government of India websites.
* A distributed Solr index is used by front-end application to search the web.
* Language identification, lemmatizer, stemmer, spell-checker, cross language search & various NLP modules were designed, developed and integrated in the search platform.
* Contributed in improving the relevance of search by enabling various dictionaries, Ontologies & rule-based approaches.
* Contributed in designing test strategies for recall and relevancy of the search results.