**Balaji Devarajan** sendmailbalaji@gmail.com

732-423-1516 <https://www.linkedin.com/in/balajidevarajan>

Balaji comes with 24 years of hands on software/Data Engineering experience with specialization in developing data processing systems for high throughput and high availability applications. His experience includes Life Science & healthcare, Ad-Targeting/Serving, RTB, ML/Behavioral-science, TV/Radio/Media Ad. cost prediction model, Cloud migrations and few successful startups. His expertise includes Big Data solutions, Cloud Solutions, Data Analytics, Solutions Architecture, ETL; Team building, Product developments, Integrations, Automations & Support. He is self-motivated, a Think-Tank, dedicated, strong troubleshooting & analytical, collaborative and a proven data enthusiast.

***Experiences Summary:***

* 18+ years in Data Processing (Data Mgmt./Governance, Migrations, ETL, MySQL-DBA/BI)
* 12+ years in cross-functional data solutions.
* 10+ years in internet advertising data (RTB, Ad-Serv., Ad-client, Desktop behavior)
* **10+ years in Product Management & Development -- Current**
* 10+ years in Project management, Agile, Scrum methodologies (CSM), Team building, mentoring and managing team globally.
* 10 years in AWS Services (Data migration to cloud)
* 8 years in Big Data (Hadoop, Hive, Map/reduce, Pig Latin…more)
* 6 years in Data Science (Data landscape, sampling, regression modeling)
* 5 years in TV/Radio/Cable/OOH data (Rating/Cost analytics)
* 5 years in Life Science & Health Care (Pricing And Contracting Solutions)
* 5 years in Chargeback, Medicaid, Government Pricing, Managed Care Data processing.

***Strengths & Domain Knowledge (by chronological order):***

* At ModelN, I was actively involved in the early stages of the greenfield project “Voyager/Validata/CDC,” where I contributed to defining the scope, designing solutions, creating customized stories, capacity planning, and engaging a global development team. I organized regular sprints and scrums to ensure timely delivery. Developed
* At Deloitte, managed Exalink's (Life science & Health care) global Data Engineering team. Created solutions for multi-platform, multi-client data processing architecture, PaaS/SaaS, client onboarding, data management & data deliveries. Hands on in many data solutions include design & development. Created Dashboards/Alerting system.
* At SQAD, migrated/re-architected their legacy C++ based ETL process into modern Cloud (AWS) platform.
* At Media6, one of the founding technical architect in creating social behavioral Ad-science, Data mining, ML, worked with NYU Stern professors in creating algorithms for targeted audience in RTB platform.
* Been a Principle architect in product development for many **ad-tech** **startups ($2M – $100+M)** on general technology, infrastructure, processes, hiring & team building and building big-data systems using an open-source stack.
* Early **Hadoop/BigData** adapter (2007). Excellent in data layouts for optimal read/writes thru **HDFS** and for better **Map/Reduce** performance & superior in **data Analytics**.
* Strong in **Database** technologies, administrations, performance tuning & optimization.
* Excellent in processing complex, **structured, unstructured** data, created custom ETL solutions with traditional & Big Data systems. Hands on in developing **new & legacy** systems, re-engineering and enhancement.
* Project Management, Estimations, and Hiring, Team building/managing in US & across globe.
* Served as a member of technical due-diligence team and a principle data architect for many Mergers, acquisitions & splits ups.

***Innovative & research oriented with co-own following patents:***

1. [Systems and methods for targeting online advertisements using data derived from social networks](https://www.google.com/patents/US20110041168?dq=balaji+devarajan&hl=en&sa=X&ved=0CCMQ6AEwAWoVChMIssu97NmkyAIVDKSICh1_8wpL)
2. [Methods and systems for providing social networking-based advertisement](mailto:https://patents.google.com/patent/US20110040586A1/en%3Fassignee=balaji+devarajan)

***Education***

* Master of Business Administration @ Madurai Kamaraj University (DLP), Madurai
* Bachelor of Engineering @ Sathyabama Engineering College, Chennai
* P.G.D.C.A at Loyola Institute of Business Administration, Chennai
* Certificate Training on PMP @ Middlesex County College, Edison, NJ
* CSM – Certified Scrum Master @ scrum alliance, USA.
* QA - Mercury Products Training @ Princeton, NJ
* Professional Mgmt Training @ Dale Carnegie, NYC.

***Technical Sessions:***

1. Talend Data Integrations @ Talend summit, NYC
2. Spark Development Boot camp @ New Circle, PA
3. Big Data Analytics @NYC
4. Percona Mysql @NYC
5. Many online course works on AWS, Bigdata, Modern ETL, Data solutions...etc.

***Experience Summary:***

***Associate Director, Data & Analytics Platform @ Model N (Jan’21- Current)***

Model N acquired Deloitte’s life sciences pricing and contracting solutions business and underlying technology to better serve a larger segment of the market from pre-commercial to the largest life sciences companies in the world. In my new role at ModelN, I specialize in developing strategic roadmaps and designing scalable platform architectures that align with organizational goals. I lead cross-functional teams in budgeting, project management, and data integration efforts to ensure the effective execution of data initiatives. With a focus on platform design, data governance, and long-term planning, I manage secure and efficient data management solutions that drive business intelligence and operational success. My experience ensures seamless platform development while collaborating with stakeholders to deliver impactful, data-driven results.

At ModelN, I was actively involved in the early stages of the greenfield project “Voyager/Validata/CDC,” where I contributed to defining the scope, designing solutions, creating customized stories, capacity planning, and engaging a global development team. I organized regular sprints and scrums to ensure timely delivery. Out Of all the POCs, the **Nextgen Data & Analytics** product gained significant traction with budget approvals, allowing us to move forward. We began developing the product using various microservices such as **TCM, OS, UAM, AHM, and Datahub**, which opened up the data landscape for Analytics to visualize both raw and curated data.This infrastructure laid the foundation for several product pipelines, including **DaaS, Life Science Analytics, and Formulary.** The technology stack included **Kubernetes, ELK, RDS, S3, EMR, Lambda, DMS, Spark, SQS, Java, HashiCorp Vault, Temporal, Keycloak, Okta IDP, and APIs** to build and scale the platform efficiently.

**Data Engineering Manager @ Deloitte (Feb’18 – Dec’20)**

Principle architect in Exalink’s Data management team focused on creating Data solutions for SaaS & MRS clients. Designed and developed ETL solution to validate, standardize client data and import into Exalink’s system. Developed solutions to process 24/7 data pipelines for multi-tenant, multi-client, Roster, GPO Membership Management data & transactions. Created a cloud based Data lake, with Cloudera/AWS-S3 & HIVE. Created a **dashboard** to monitor the system with **ELK/New Relic** Alerting system.

Identified hotspots, re-engineered some of the monolithic process into efficient ETL solutions involving processing Medicaid, chargebacks, Rebates and Fees, Managed Care, Market Share, Products & Pricing, EDI’s and other data sources. The pipelines are built with Kafka/API, Talend, Python, Java, Postgres, Cloudera, ELK & AWS and the CI/CD in Jenkins/Ansible.

As a leader, created a Data management team, identifying the right candidates (US/USI/Global), training & mentoring. Our team focused on innovating & developing effective & efficient solutions and developing next generation enterprise platform.

Regular conversation with client on various designs discussions to help them build their system more effectively. The collaborative efforts helped mutually to deliver our products on time and meet our goals.

***Data Implementation & Integrations @ Office Practicum, PA (Dec’17 – Feb’18)***

Processing PHR, EMR, EHR data primarily focused on Pediatric health care. Migrating existing PHR data into Office Practicum software. Proposed a new architectural solution to augment the existing implementation process. Worked with many health care providers in gathering data and business needs.

***Sr. Database (BigData) Engineer at SQAD (Tarrytown, New York (Sep’12 – Nov’17)***

SQAD provides dynamic data set for advertisers, agencies, and brands to empower their research, review, planning, management, and reporting across all of their media strategies; including network, local, & cable television, radio, out of home, and internet channels.

The work involved a detailed reverse engineering the legacy C++ based ETL siloes system, research & documentation. Many POC were developed using Hadoop technologies HDFS/HIVE, Python/Spark, Cassandra, Elastic/Kibana, HIVE/EMR. A new cloud based automated Hadoop-Pentaho/Java system was developed and deployed.

Automated Costing & Predicting modeling process with HDFS/HIVE/EMR/(R)-Studio into scalable production process using AWS’s PDI, EC2, VPC, RDS, Aurora, S3, Lambda, SNS...etc. Automated the reporting system and created a new data delivery mechanism. Involved in creating POC for hosting Nielsen & Rentrack’s high volume data in EMR, Hive...etc.,

Extensively involved in AWS infrastructure planning, cost estimating & implementations. Deep dived into Pentaho DI 5.5, 6.0, Python 2.7 & 3.6, Bash scripting, building util’s in java and explored most of the AWS services and used EC2, Lambda, RDS, Aurora, Athena, Redshift, Dynamo DB, EMR-Hive, EMR-HDFS, Data Pipeline, Route53, S3, SNS, Security groups, IAM Roles, Cloud Trail, Cloud watch, explored Elastic Search/Kibana for hosting Rentrack data. Managed the task & deliverables using JIRA, Dataplane Reports, Confluence...etc.

***Hadoop/BigData consultant at* Dstillery *(NYC, New York) (Feb’07- Aug’12)***

Being one of technical founding member of Dstillery formerly media6degrees, I was responsible for many things like research, proto-type, POC, design, development, testing, implementation, infrastructure, demo’s, monitor, reports, presentations…etc. Later, started focusing on data management, involved in designed the overall ETL process, primarily using Bash, Java, Hadoop, Pig Latin, HIVE & MySQL and NFS for archives, this includes log rotations, downloads, aggregations, prod-sync, BI, Prediction analysis, statistical modeling, scoring...etc. I also extended my role in technology operations, monitoring, troubleshooting and biz analysis.

Designed, developed and maintained all the Data-scientist’s predictive analysis (algorithms) into production using Bash, Python, R, Java, Hadoop, HIVE & MySQL. Helped migrating almost all the process from MySQL to Hadoop/SQOOP & HIVE (HDFS). Wrote many utilities to support, monitor and auto-fixing things for both MySQL & HIVE.

Designed, developed and implemented a multi-level (2-tier) replication, MySQL-HA with DRBD & Linux Heart beat with Auto-failover. Introduced Innodb, federated engine, Blackhole. Designed, developed and managed many monitoring, alerts, beeper module…etc., I like all my documents to be with a drawing, so extensively used Omnigraffle for all my drawing.

MySQL 5.5, Hadoop 0.20.2, HIVE-0.7.0, Piglatin, R, WorkBench, Nagios, Cacti, PHP Myadmin, Atlassian JIRA, Confluence, crowd, Sales force, Omnigraffle, Agile, Scrum, SysBench, PHP.

***Sr. DBA/Analyst at Direct Revenue (NYC, New York) (Sep’02 – Jan’07)***

Direct Revenue was providing useful content and free software in exchange for the opportunity to deliver highly targeted behavioral messages. The Direct Revenue family includes Best Offers Network and Soho Digital International, Providing full-service planning and brokerage services to advertising partners around the world.Involved in extensive Cab/Exe for ActiveX installer and eventually moved to wise installers. Created a process to monitor Ad client behavior, loss, reconnaissance, data mining, Life Time Value Projections. MySQL DBA activities like **Master/Slave** relationships for **High Availability**, DB/structure/user management, backup/recovery/performance/monitoring, replication and master/slave infrastructure. Also created many alert modules, Stored procedure, functions for optimizing the performance of the application.

***Sr. System Analyst at TrueData.org (NYC, New York) (Feb’01 – Aug’02)***

This Company was a desktop advertising company. This was a very small team of 6-8 people. The business model was quite simple but the technology lying under was very interesting, here every day I felt I was learning something new. There were always new challenges for us to solve. This venture was not very profitable but gave a lot of experience for both the Stakeholder and technologist like us.

***System Analyst at amexb2b.disk11.com (NYC, New York) (Sep’00 – Jan’01)***

Disk11 does various projects for American Express. One of the projects is a web-based directory to allow Card Holders to find merchants that accept the American Express Card. Another Project Supplier Match System, both of these Projects were using **JAVA/JSP/Oracle 8**. Developed in HTML/JSP/Servlets/JavaScript in Model View Controller (MVC) Architecture.

***Lecturer/IT Support Engineer, India (Feb’95 – Jun’00)***

Lecturer, IT/CAD/CAM Tech, Placement officer, MIS portal support.

***Technical Skills & Work Summary (actually used at some point in last 20+ years):***

|  |  |
| --- | --- |
| Big Data | Apache Hadoop, HDFS, Map/Reduce, Dremio, Hive, Sqoop, PigLatin, MongoDb, Cassandra, Spark |
| AWS | Ec2, Lambda,SNS, VPC, IAM, AMI, EBS, Security Groups, Key pairs, S3, RDS, Glue, (MySQL, Aurora, MSSQL), Dynamo DB, Athena, Redshift, EMR/Hive/Spark, Route53, Cloud Trail, SNS...more |
| Databases/Tools | Maria, Postgres, MySql, Dremio, Hive, MSSQL, SQLite, WorkBench, Sybase Power Designer, SQL Power architect, TOAD, Dbeaver, SysBench, PhpMyAdmin. RazorSQL, SQL\*Plus, SQL\*Loader |
| ETL Tools | Nifi, Talend, Pentaho, Tableau, SSIS, built many in-house custom ETL tools |
| Tools | Ambari, Go-anywhere, SSO, Solr, AWS services, VPN/VPC, Okta, IAM, okta, Omnigraffle, Penguinet, Visio 2000, Eclipse, Drawio, SSIS, SSR |
| Pub/Sub | Kafka, AWS-SQS, SNS |
| Workflow | Talend, Pentaho, Airflow, Temporal, Nifi, |
| API Tools | Postman, Karate, Curl, Talend, |
| Data Analytics | Tableau, Anthena, quicksight, Drill, Solr, R, Python, Pentaho, Qlik, Looker, Dremio |
| Operation System | Linux, Red hat, Centos, Mac OS, Windows |
| Source control | Github, Subversion(svn), CVS. |
| Network/Monitoring | SSH, SCP, FTP, Telnet, Port forwarding, Rsync, Nagios, Ganglia, cacti |
| Scripting | Bash, Python, JShell, JavaScript |
| Language | PLSQL, R, Java, PHP |
| Web | Apache, Tomcat, Resin, HTML, CSS, JavaScript, XML |
| Q/A | Postman, Test Automation, Jrun, VMWare, JIRA, confluence, WinRunner, LoadRunner, TestDirector, Bugzilla, Selenium, Karate, TestNG, Zypher. |
| Project Management | Certified Scrum Master, Agile, Waterfall, Confluence, JIRA, Kanban, Teams, , Project, TaskLand |
| Automation | Jenkin, Zypher, TestNG, Build, Deploy, Test, Scheduler, Orchestrator, |
| Logging/monitoring/Alerts | Graylog, grafana, Kibana, Elastic, logstash, sidecar, open search, New Relic, Prometheus, nagios, cacti. |
| Ci/CD | Jenkins, Ansible, Groovy, Helm, |
| Data platform | SaaS platform, DaaS, targeted audience prediction and scientific model, cost and trend analysis. CDC |
| Data governance | Talend Data Fabric, Rentrack, Nielsen and Comscore, Dun & bradstreet for lookups. Meta-data, Data policies, tie-outs, reconciliations. |
| Security & Tools | Checkmarks, SAST, DAST, SCA, Pen test. ISO27001, ISMS, GDPR,SOC1, SOC2, SAPA, Kevin Mitnick SAT. |
| Microservice | Kubernetes, Docker, EKS, Argo-CD, |
| Project Management | JIRA, Confluence, MS project, Excel, word, Pivot tables, |

***Product Portfolio (In my tenure):***

|  |  |
| --- | --- |
| Life Science | Exalink, CDL, Validata, Data Platform, LSDNA, TCM, OS, DNA, DH, UAM, AHM (GP, Medicaid, Chargebacks, MDM, PbN, PbF), PEDIATRIC EHR |
| Ad-Tech | Traffic master, Ad-server, campaign manager, Crawlers, Adc-lient, RTB, DSP, lookalikes, Adsafe, |
| Network/Media | MediaCost, Mediatools, MediaLogic, |

***Global Team:***

|  |  |
| --- | --- |
| PM, Architects, Leads, Software engineers, Quality Assurance engineers | US (EST, CST, PST), Brazil, UK, Ukraine, Poland, Russia, Thailand, Philippines, Singapore, India |