**SARANSH SRIVASTAVA**

Email: [saranshsk@gmail.com](mailto:saranshsk@gmail.com) Mobile: +91-8130-0318-86

Website: [www.iotnow.co.in](http://www.iotnow.co.in)

I am currently associated with Fidelity International as a Senior Analyst Programmer with an experience of 6.5 years in Data warehousing, I have been part of mainstream projects in financial and healthcare domain. I have practical experience in tools/technologies/languages like Python, Spark, Hadoop /big data stack, Informatica, Oracle and shell script.

Prior to this I was associated with Tata Consultancy Services from Dec 2011 to June 2014. I completed my Bachelor of technology (Computer science) in 2011.

**Core Skills**

|  |  |
| --- | --- |
| Languages | Python, Shell script, SQL, CQL |
| Tools/Packages | Informatica, Apache Hadoop, Apache Spark, DSE Analytics |
| Databases | Oracle 12c/ 11g, HBase, Cassandra, MongoDB, MySQL |
| Job scheduler | Control M |
| Web Server | Apache Tomcat |
| BigData/Hadoop | Multi node Hadoop cluster Setup, PySpark, Pig, Hive, Kafka, Sqoop, SOLR, Spark streaming |
| Deployment | Automated CI and CD processes using Bamboo and TeamCity |

**Current Organization – Fidelity International [June 2014 till present]**

|  |  |
| --- | --- |
| Project | Single customer view 2.0 |
| Objective | Design a flattened database model for NoSQL needs and then migrate data from Oracle 12c to Cassandra database |
| Technology | Cassandra, Python, PySpark, Kafka, Hive, Spark streaming, Sqoop and Informatica |
| Deployment | Automated deployments using CI/CD flows in Bamboo and TeamCity, Bit Bucket |
| Methodology | Agile software development methodology |
| Roles and responsibilities | * Installation of multi node Hadoop cluster with Cassandra, Kafka and Spark packages * Creating flattened data model from the existing relational model * Creating Informatica mappings to load data from Oracle to Cassandra as part of initial load * Developing PySpark code to read data from Kafka stream, apply business transformation and then loading the data into Cassandra as part of delta loading strategy * Creating data files from Cassandra using Python and store them onto HDFS * Creating test cases on the generated data files using Hive query |

|  |  |
| --- | --- |
| Project | Invest sense analytics engine |
| Objective | Design and develop an analytics engine to churn meaningful information such as low balance, dormant cash in account etc out of data stored in Cassandra flattened structure |
| Technology | Cassandra, Python, PySpark, Kafka, Hive, Spark streaming |
| Deployment | Automated deployments using CI/CD flows in Bamboo, Bit Bucket |
| Methodology | Agile software development methodology |
| Roles and responsibilities | * Developing code in PySpark to read data from Cassandra and store in dataframes * Applying business rules upon the data and submit it onto Kafka pipeline for latter consumption using python * Reading Kafka pipes using Spark streaming and dump data into Cassandra tables and text files on HDFS * Validating the text files using Hive query |

|  |  |
| --- | --- |
| Project | Generic file uploader |
| Objective | Design and develop a generic data uploader application which would source text files and dump data into Oracle, HBase, Cassandra tables |
| Technology | Cassandra, Oracle 12c, HBase, Python, PySpark, Hive, HDFS (Hadoop) |
| Deployment | Automated deployments using CI/CD flows in Bamboo, SVN |
| Methodology | Agile software development methodology |
| Roles and responsibilities | * Created PySpark code to read data from source files irrespective of the structure * Created a generic solution to apply user configured business rules on the source data * Created Python code to dump transformed data into Oracle 12c, HBase and Cassandra databases * Generated regression suite using Hive to query on the source files vs. the data in database |

|  |  |
| --- | --- |
| Project | Behavioral Finance |
| Objective | Perform analytics on financial data generated by various in-house systems |
| Technology | HBase, Python, Hive, HDFS |
| Deployment | Automated deployments using CI/CD flows in Bamboo, SVN |
| Methodology | Agile software development methodology |
| Roles and responsibilities | * Created scripts to copy source files onto HDFS * Loaded the data from the files to HBase tables * Churned out information directly from the files in HDFS using Hive queries and simultaneously from HBase stores |

|  |  |
| --- | --- |
| Project | Investment Platform/Single customer view – Integration |
| Objective | Migrate existing EDW data warehouse flow to consume real time data feeds from upstream sources |
| Technology | Oracle 12c, HBase, Python, HDFS (Hadoop), Hive, Kafka |
| Deployment | Automated deployments using CI/CD flows in Bamboo, SVN |
| Methodology | Agile software development methodology |
| Roles and responsibilities | * Installed multi node Hadoop cluster with HBase, Kafka and Hive packages * Created python code to simulate the work of legacy Informatica mappings * Created code to consume real time data feeds through Kafka pipes and dump data onto Oracle and HBase * Created regression test cases using Hive to query on the source feed files against the data loaded into databases |

|  |  |
| --- | --- |
| Project | EDW Upgrade |
| Objective | Bringing down the mammoth batch execution timing of ~14 hours to fewer than 4 hours |
| Technology | Informatica, Oracle 12c, Unix Shell script and Control-M |
| Deployment | Automated deployments using CI/CD flows in Bamboo, SVN |
| Methodology | Agile software development methodology |
| Roles and responsibilities | Implemented various tuning methodologies such as:   * Informatica – Session partitioning, optimal memory provisioning, pushdown optimization and persistent caches * Oracle – Table partitioning, DOP, indexing and gather statistics |
| Value adds | * Developed a user notification framework for data rejected during ETL loads * Created a Control-m xml pre-validation utility to point out issues before CI deployments * Developed a generic database statistics gathering framework for Oracle 12c |

|  |  |
| --- | --- |
| Project | FATCA – Regulatory Reporting |
| Objective | Design and develop a data warehouse to report tax defaulters as per FATCA norms |
| Technology | Informatica, Oracle 12c, Unix Shell script and Control-M |
| Deployment | Manual deployments using custom built shell scripts |
| Methodology | Waterfall methodology |
| Roles and responsibilities | * Developed a data warehouse for storing client and transaction information * Generated a data mart on top of the warehouse tables to enable report generation |

**Previous Organization – Tata Consultancy Services [Dec 2011 – June 2014]**

|  |  |
| --- | --- |
| Project | Data warehouse for large US based Pharmaceutical firm |
| Objective | Design and develop data warehouses for sales and clinical trial data |
| Technology | Informatica, Oracle 11g, Unix Shell script, HDFS (Hadoop), Python |
| Deployment | Manual deployments using custom built shell scripts |
| Methodology | Waterfall methodology |
| Roles and responsibilities | * Developed Informatica mappings to load data into Oracle tables after applying business rules * Developed a data management framework for keeping track of the erroneous records during load * Generated data files using Python code and stored them into HDFS for downstream systems |

|  |
| --- |
| **Certifications** |

|  |  |
| --- | --- |
| 1. Hadoop Platform and Application Framework |  |
| 1. Introduction to Big Data |
| 1. Big Data Integration and Processing |
| 1. Machine Learning With Big Data |
| 1. Red Hat Certified Engineer (RHCE) Certificate Number (805010725050216) Score - 98.45% |
| 1. 1Z0-007 Introduction to Oracle 9i: SQL certification with 98% score |

|  |
| --- |
| **Achievements** |

|  |
| --- |
| 1. Effective technologist of the year 2015 – Fidelity International |
| 1. Business Unit (BU) Champion – Fidelity International |
| 1. 5 times Director award – Fidelity International |
| 1. 5 times On-the-spot certificate – TCS |
| 1. Best project award for final year project during B-Tech curriculum |

|  |
| --- |
| **Trainings** |

|  |  |
| --- | --- |
| 1. Oracle 12c Performance tuning | 1. RHS333 Redhat module |
| 1. J2SE and J2EE from E-soft technologies | 1. Python & PySpark |

**Self-Initiative**

|  |
| --- |
| [www.iotnow.co.in](http://www.iotnow.co.in) (Up on self-built home server): *Iotnow* is a platform aimed at easing people's internet experience. It provides free of cost web hosting with 21 TB of space and technology support in Java, .NET, Oracle, Cassandra, MySQL and MongoDB. The website also provides Hadoop/Big Data video tutorials |

|  |  |
| --- | --- |
| **Personal Details** | |
| Languages | English and Hindi |
| Mobility | Willing to work anywhere in India or Overseas (Indian Passport no. J7652393) |
| Address | Flat - 902, Tower – B5, Orris Carnations Residency, Sector - 85, Gurgaon, India |