Prakash Chauhan

+91-9718707585

prakashchauhan2010@gmail.com

**EXPERIENCE:**

* Over 13 years of experience in the field of Information Technology, principally in development and designing in Java, Apache Hadoop, Apache Spark & Apache Cassandra.
* An analytical senior-level software developer with deep expertise in Java/J2EE and Big Data technologies.
* Thorough understanding of software development processes and worked in Waterfall as well as AGILE methodologies.
* Team leading and delivery experience as being SCRUM MASTER for few projects.
* Have a good exposure to various Design Patterns and scalable architecture.
* Have good knowledge and experience in working with multi-tier architectures and enterprise applications for batch as well as streaming applications.
* Handled varied data sources, performed orchestrations, and loaded into targets including RDBMS, NoSql Databases, etc.
* Actively interacted with business partner and product owner to get low level details on business requirements and also to discuss the design and architecture.

**TECHNICAL SKILLS:**

* **Big Data Technologies**: Palantir Foundry, Apache Spark 2.x, Apache Hadoop 2.x, Apache Kafka, Apache Cassandra 2.1.x, Apache NIFI, Apache Sqoop
* **Languages / Frameworks**: Scala, Java
* **IDE**: IntelliJ, Eclipse
* **Build Tool**: Maven
* **Repository**: GitHub
* **Testing frameworks:** JUnit 4

**PROFESSIONAL EXPERIENCE:**

Nov’17 - Till date Impetus Technologies (India) Pvt. Ltd.

Mar’14 - Oct’17 Ericsson India Global Services Pvt. Ltd.

Jun’12 - Feb’14 MPS Ltd.

Jul’10 - May’12 Tenon & Groove Software Pvt. Ltd

Apr’09 - Jul’10 Valid Page

**PROJECT DETAILS:**

#### Company: Impetus Technologies (India) Pvt. Ltd. Aug’20 to till date

**Designation/Role:** Lead Software Engineer

**Platform & Skills:** Apache Spark 2.4, Scala, MySQL, NoSQL databases, Apache Kafka.

**Brief:** CS Streams is a Platform for real-time, near real-time and batch ingestion and compute operations. It has the capability to read the real-time streaming data from multiple sources like solace/Kafka, apply compute on top of the input data using multiple modules and eventually publishing the data to multiple

downstream systems like HDFS, Kafka, MapR DB, Elastic Search. It supports plugin-based architecture and various modules can be plugged-in with a simple metadata configuration. It also provides the inner source capability so that the platform users can write code that can be plugged-in at run time in streaming job. The platform also has the capability to switch to DR cluster in disastrous situations.

My key responsibilities are:

* Worked on the development of Spark structured streaming platform using Scala for real-time ingestions and compute operations. The platform is fully metadata driven and the metadata is

specified using JSON document.

* Added MapR DB / HBase publisher capability in Structured streaming platform that reads data from Kafka, applies modules like message parser, payload parser and eventually publishes the data to MapR DB.
* Added Kafka publisher capability in Structured streaming platform. CS Streams Kafka Publisher receives message stream from Kafka/Solace, apply modules like message parser, payload parser and publish to Kafka.
* Added ES publisher capability to the platform. ES Publisher Module takes the input dataset and writes it to the Elastic Search. It can write to both CSRT ES and External ES.

#### Company/Client: Impetus Technologies (India) Pvt. Ltd. Apr’19 to Aug’20

**Designation/Role:** Module Lead Software Engineer

**Platform & Skills:** Palantir Foundry, Apache Spark 2.x, Python 3, Java

**Brief:**

Sirax is a Passenger Revenue Accounting (PRA) system which process United Airline’s daily passenger ticket and ancillary sales and usage data feeds for accounting and revenue reporting. There are 3 daily output files produced by Sirax:

**MIS Sales:** Daily sales output of United (016) tickets that have been successfully processed in Sirax.

**MIS TCN:** Daily other airline sales for future UA travel and 016 ticket itineraries.

**MIS Revenue:** Daily flown revenue from United and other airline (OA) tickets flown by United.

My key responsibilities are:

* I have developed Sirax core module to parse big raw sequential files using Apache Spark and converting the raw data into JSON datasets to create the clean layer datasets. Clean layer contains multiple python transformations which reads the JSON input from raw layer and populate datasets with more granular information like ticket, coupon, tax etc.
* UPSERT module is SCD type 4 (Slow Changing Dimensions) implementation.
* Python transformations in UPSERT module ensures that only the latest and greatest data goes to the final dataset and all stale data goes into the history datasets for tracking purpose.
* Developed the ontology module for accounting and revenue reporting.

#### Company: Impetus Technologies (India) Pvt. Ltd. Dec’17 to Mar’19

**Designation/Role:** Module Lead Software Engineer

**Platform & Skills:** Apache Spark, Apache Hive, Apache Kafka, HBase

**Brief:** Project is about change data capture (CDC). Application receives events from distributed messaging system and inserts/updates the same in Hive / HBase for reporting purpose.

My key responsibilities are:

* Worked on importing data from relational database to Hadoop HDFS using Apache SQOOP.
* Worked on CDC module to capture the change data and then inserting the same into Hive / HBase database.
* Collected logs from multiple sources via Apache Nifi and the same is pushed into the Kafka topic which is then processed by Apache Spark.

**TRAININGS DELIVERED:**

|  |  |  |
| --- | --- | --- |
| **Technology** | **Client Name** | **Location** |
| Big Data Hadoop Analytics | Samsung | Noida |
| Big Data: Apache Spark | Nagarro | Gurgaon |
| Big Data Analytics with Hadoop | Boston Consulting Group | Gurgaon |
| Apache Cassandra | Netcracker | Online (4 yrs till now) |
| Apache Cassandra | Simplilearn(AMEX) | Online |
| Apache Cassandra | Intellipaat | Online |
| Apache Cassandra | Edureka | Online |

**EDUCATION:**

2008 Bachelors in Computer Science Dr. B.R. Ambedkar University, Agra