Vivek Kaushik

Vivek.kaushik.bit@gmail.com

+91 9831008804

KOLKATA

Technical Skills

|  |  |
| --- | --- |
| Technical Expertise: |  |
| * ETL **Ab-Initio-Talend/Apache-Spark/Sqoop**/**PySpark**/**Scala**/**Hive/Python**. * Oracle, SQL/PL-SQL, MySQL and UNIX. |  |

Professional Summary

* Accomplished IT Analyst with **10** years of experience in providing IT Solutions.
* **7** years **9** months of experience in the **Analysis, Design**and **Development of**Data warehousing solutions and in developing strategies for (**ETL/ELT)**mechanism using **Talend** (Big Data) /**Ab Initio** and working with **Big** **Data/Apache-Spark.**
* Working as an ETL Tech lead for over **5.5** years, leading the development for enhancements in Insurance & Banking domains.
* Having good expertise on **HDFS**, **Map Reduce**, **Hive, Scala** and **Spark** fundamentals including **RDDs** and **Dataframes**.
* **Completed** [**IBM Data Science Professional Certificate**](https://www.coursera.org/specializations/ibm-data-science)**.**
* Machine Learning and Data Scientist enthusiast with a passion for turning **raw** data into products, actionable insights, and meaningful stories.
* ☟ Please have a look at my final report submission on "IBM Data Science Professional Certificate":

[**https://www.linkedin.com/pulse/analyzing-neighborhoods-london-vivek-kaushik/**](https://www.linkedin.com/pulse/analyzing-neighborhoods-london-vivek-kaushik/)

* **2** years **3** months of experience as **Database Administrator** in Application Team.
* Exceptional problem-solving techniques backed up by good written and verbal communication skills.

Experience

**Cognizant Technology Solutions | Pune**

Senior Associate, June**-2019- Till Date** ( **Currently serving Notice Period**)

|  |  |
| --- | --- |
| Project | Data Ingestion & Curation (**Wells Fargo):** (June 2019 – Till Date) |
| Tools | **Python, PySpark, Scala, Hive**, **Talend,** MySQL ,**Oozie** and UNIX. |
| Role | Developer/Lead. |

Working with business analysts to identify the mapping **design** and **workflow** required for implementation. Leading and Coordinating the Offshore team to reach client goals and expectations.

* Working closely with the business users to understand the requirements and converting them into project level technical capabilities.
* In depth understanding/knowledge of **Hadoop Architecture** and its components such as **HDFS**, **Name Node**, **Data Node** and Map Reduce etc.
* Create **PySpark** framework to bring data from Various sources to **Amazon** S3.
* Translate business requirements into maintainable software components and understand impact (Technical and Business).
* Provide guidance to development team working on **PySpark** as ETL platform. Makes sure that quality standards are defined and met.
* Optimize the **PySpark** jobs to run on Kubernetes Cluster for faster data processing.
* Data **ingestion** in target Data lakes from various **file** and DB sources using **Sqoop**/**Spark** jobs.
* Involved in Data **Analysis**, Data **Validation**, Data **Cleansing**, Data Verification and identifying data mismatch using Apache spark **Scala/PySpark** functionalities.
* Experience in validating and analysing Hive underlying files (Reading **Parquet** with **Spark**).
* Involved in creating **Hive** tables, loading it with data and writing hive queries, to create new **views** or **tables**. Also using **Spark** to do the same tasks using **Scala/PySpark** as required.
* In-depth understanding of Spark architecture including the Spark core, **Spark data frame** and **Spark Sql** also including **Scala**.
* Involved in **optimization** and **performance** tuning of **Spark** jobs or using **Hive**.
* Experience in using accumulator variables, broadcast variables and **RDD** concepts.
* Expertise in using Spark **SQL** with various data sources like Json,CSV Parquet,RDBMS and Hive.
* Leading and coordinating **change** **management** processes for consistent and quality deliverables. Monitoring and coordinating development and unit testing along with peer reviews to ensure quality of deliverables.

**Tata Consultancy Services Limited | KOLKATA**

Assistant Consultant, **Sep-2010** – **June-2019**

|  |  |
| --- | --- |
| Project | Customer Data Consolidation**:** (Jan **2015** – June **2019**) |
| Tools | **Python, PySpark, Scala, Hive**, **Talend,** MySQL,**Oozie** and UNIX. |
| Role | Developer/Lead |

* Leading and coordinating **change** **management** processes for consistent and quality deliverables.
* Experience with **ingestion** of data from multiple sources into Hadoop using core Hadoop data loading components (**Sqoop/Spark** jobs).
* Developed **PySpark** and SparkSQL code to process the data in Apache Spark on Amazon EMR to perform the necessary transformations.
* Involved in Data **Analysis**, Data **Validation**, Data **Cleansing**, Data Verification and identifying data mismatch using **Scala/PySpark** concepts.
* Create **PySpark** framework to bring data from various sources to **Amazon** S3.
* Translate business requirements into maintainable software components and understand impact (Technical and Business).
* Provide guidance to development team working on **PySpark** as ETL platform. Makes sure that quality standards are defined and met.
* Optimize the **PySpark** jobs to run on Kubernetes Cluster for faster data processing.
* Experience in **validating**, analysing and **transforming** Big data Hadoop files (Reading and saving **Parquet** with **PySpark/Scala**).
* Involved in creating **Hive** tables, loading it with data and writing hive queries, which will run internally in Map Reduce way using **Scala/PySpark**.
* In-depth understanding of Spark architecture including the **Spark** core, Spark data frame and **Spark** Sql.
* Expertise in using Spark **SQL** with various data sources like **Json**, **Parquet**/**Hive** tables and other **DB** sources.
* Experience in **optimization** and **performance** tuning of **Spark** jobs or using **Hive**.

|  |  |
| --- | --- |
| Project(s) | **BI** Customer Management Repository(Jan **2013** – Jan **2015**) |
| Tools | **Environment:**Ab Initio GDE 3.0 Co-op 3.0, UNIX shell scripting. |
| Role | **Team Lead, Developer** |
|  | * Designed, developed, tested and implemented **Ab Initio** **ETL** graphs for batch processing or extraction of data. * Expertise and well versed with various **Ab Initio Transform, Partition, DE partition**, Dataset and Databasecomponents. * Extensively worked under the **Unix** Environmentusing**Shell** ScriptsandWrapper Scripts. * Responsible for writing the wrapper scripts to invoke the deployed **Ab Initio** Graphs. * Used various component of **Ab Initio graphs** like**‘Partition by Key’, ’Sort’, ’Reformat’, ‘Join’, ‘De-Dup’** etc. to impose business logic on the incoming data. |
|  |  |

|  |  |
| --- | --- |
| Project | **The Hartford Financial Services Group:** (Sep 2010 – Jan 2013) |
| Tools | Oracle 10g, My SQL Server and UNIX shell scripting. |
| Role | Oracle DBA. |
|  | * + Database design and physical **architecture**. Routine daily monitoring and maintenance of customer systems.   + Monitoring and ensuring **SLA** of the various jobs. Resolving and suggesting improvements method for job failures to ensure smooth functionality.   + Create database objects (**tables, indexes**, etc.), assist developers with database activities, and control all migrations of Oracle schema objects.   + Implement and maintain database security. Create and maintain users and roles, assign privileges. |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

Education



Birla Institute of Technology- Mesra, Ranchi

***2010*** B.E. (Electronics and Communication Engineering), CGPA – 7.2/10

Little Flower School-***2005*** Sr. Secondary (XII), 81.20%

Little Flower School-***2003*** High School (X), 86.66%