Santosh Kumar Pandit

(+91) - 9439501171 | [santosh.pandit006@gmail.com](mailto:santosh.pandit006@gmail.com)

Developer with 6+ years of IT experience. Create detailed design artefacts, work on development and perform code reviews, implement validation and support activities in line with architecture requirements and participate in knowledge management activities with the objective of ensuring the highest level of service offering to clients in technology domain.

## Professional Experience

Infosys Limited – India. May’14 – Present

Technology Analyst

* I have around 2 years of experience in Big data and cloud technologies. Have designed and developed volume based scalable spark application. Experience on data ingestion for incremental and full load.
* I have more than 3 years of experience in Informatica and Microstrategy. Requirement gathering, Design analysis and impact analysis to the system, development of critical mappings and workflows, code review were among the main responsibilities. Used Microstrategy to create enterprise reports.
* Have strong experience in Oracle SQL. Have worked on optimization, partitioning at etc. Have experience in Unix scripting and data modelling.

Georgia Tech University – Atlanta, USA.

Graduate Student | Masters in Computer Science | Machine Learning Specialization

* Have created models using different ML Algorithms. Used multiple data preparation techniques and used several methods for model comparison. I have used Python as well as R for programming and analysis.

Work Experience

Current Employer: Infosys Ltd. (Oct’ 2014 – Till Date)

Client: Apple Inc, Carefirst, HCSC and Amerigroup.

Current Role: Technology Analyst

Technical Skills

|  |  |
| --- | --- |
| Tools | Spark, Kafka, Hive, AWS S3, Informatica, Oracle, MSTR. |
| Language/Script | Scala, SQL and Unix. |
| Analytics Tools | Python, R, ARENA, |
| IDE | Scala IDE, IntelliJ, PyCharm, RStudio ad Visual Studio. |

Projects

Big Data:

* Have developed scala applications and spark jobs with scalability based on the volume of the data. Developed scheduling application using kafka and APIs. Have utilized amazon S3 for storage. Have applied optimization techniques.
* I have worked on real time data consumption and ingestion projects. Have implemented full load and incremental load using sqoop, kafka, spark, hive, hbase and unix. Have worked with kafka offset management. Have developed Scala Unit test case with code coverage of 85-90%.

ETL & Reporting:

* Have worked in different informatica projects to implement data mart. Have worked with complex transformations and tasks in designer and workflow. Have also implemented partitioning at informatica and database level. Also, implemented different type of SCDs.
* Have developed several Enterprise level reports using Microstrategy tool. Have done data modelling as part of reporting requirement analysis.

Machine Learning:

* I have developed AI agent to solve RPM problem using python. Have developed the ML algorithms from scratch like DT, Random Forest, etc. and used for stock trading portfolio using python.
* I have used several factor-based models and probability based model for data analysis. I have applied analytic techniques like normalizing, scaling, cross-validation, Outlier Analysis, change detection, AIC, BIC, PCA, exponential smoothing, ROC, variable selection method, design of experiments, etc. for model and data analysis using R Studio.

## Certification:

* Introduction to Data Science in Python – University of Michigan
* Applied Machine Learning in Python – University of Michigan
* The Data Scientist’s Toolbox – John Hopkins University

## Education:

B-Tech in Electrical Engineering. Apr’10 – Apr’14

C.V. Raman College of Engineering, Odisha.

Masters in Computer Science. Jan’19 - Present

Georgia Tech University, Atlanta, USA.