|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Pankaj Kumar** |  | HSR Layout, Sector 2 | |  |
|  |  | Bengaluru/Bangalore, KA 560102 | |  |
|  | **Mu-Sigma,** |  | **(+91) 94808 62869** | |  |
|  |  | [**PK179288@gmail.com**](mailto:PK179288@gmail.com) | |  |
|  | **Data Engineer** |  |  |
|  | (Oct, 2018 – Present) |  |  |  |  |
|  |  |  |  |  |  |



**EXECUTIVE SUMMARY**

I am a results-oriented accomplished professional serving the industry with **6.7** **years of extensive experience in Data Engineering (D3 Stack), Devops/CICD using GITLAB, Docker & Kubernetes, Cloud Computing, EDGE Computing, IoT Analytics, Hadoop Ecosystem, Data Integration & Business Intelligence Tools.**

* I am a fundamental, strong and a promising professional.
* Market/self-driven consistent performer, working towards scaling/expanding the business.
* Can organize constructive meetings with clients to understand the business trends and requirements.
* **Strengths:** quick learner, adaptive, highly motivated, growth &development oriented.
* Demonstrated experience with excellent problem-solving skills, solutioning, high analytical and interpersonal skills.
* I am an IT geek with a never-ending zeal and hunger to learn and grow.

**TECHNICAL SUMMARY**

* End-to-end design, solutioning and deployment experience of **data** **engineering pipelines on AWS and Open Stack or on premise**.
* Research & implementation experience on **IoT Analytics using edge devices** **such as Raspberry PI, Beacons, Cameras, Sensors, etc.**
* Experience in implementing, fully managed high-end **Hadoop, Spark and**

**Kafka clusters on bare metal and cloud using YARN and Kubernetes/Docker for research and development purpose.**

* Aware, engaged & experienced in designing and developing **end to end**

**Data Engineering pipelines on Lambda and Kappa architectures on cloud and on-premise landscape: Data Modelling, Data Ingestion, Data Scrubbing, Data Curation, Data Wrangling/Crunching/Munching, Data Governance & Security, Data Lineage, Data Sciences, Decision Sciences (D3 Stack), Data Consumption, DevOps (CICD), AI/ML and IoT.**

* **2+ years of relevant hands-on experience on Hadoop Technologies -** HDFS,Map Reduce, Pig, Hive, Sqoop, Flume, Oozie/Zookeeper, HBase, Cassandra.
* **5+ years** of relevant hands-on experience **in Teradata – architecture, utilities, performance tuning, stored procs, etc.**
* **4+ years** of relevant hands-on experience **in Informatica PC, architecture, performance tuning, etc.**
* **1+ Year** of relevant hands-on experience **in Tableau & SAP BO.**
* **3+ Years** of relevant experience on **Unix/Linux/Python Scripting.**
* Strong conceptual knowledge on OOP’s & Structured programming languages– Core Java, C++, C, Python, etc.

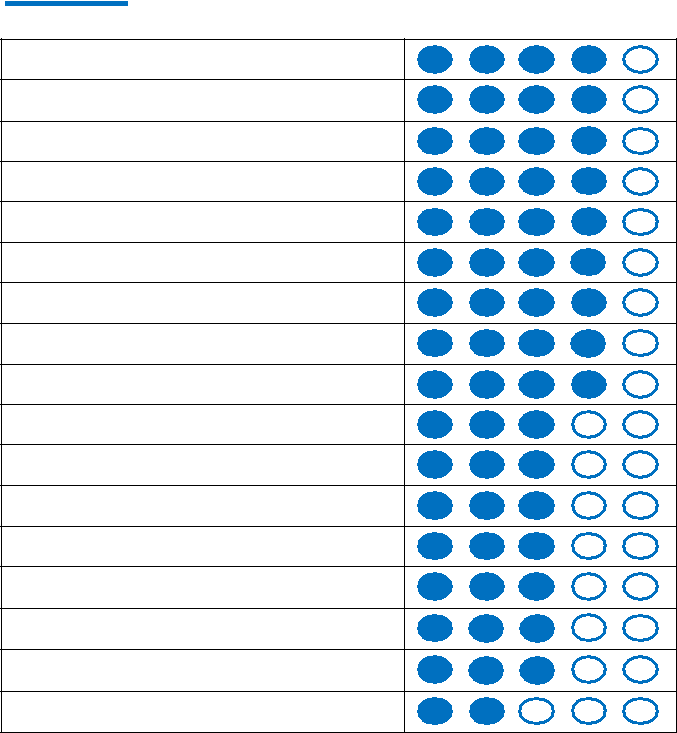
**EDUCATIONAL QUALIFICATIONS**

|  |  |  |  |
| --- | --- | --- | --- |
| **Degree** | **University/College** | **Year of Passing** | **CGPA** |
| **M.Tech** | **Bits Pilani [WILP/WASE] \*** | **2016** | **6.85** |
| **BCA** | **Graphic Era University, Dehradun** | **2012** | **68.80%** |
|  |  |  |  |
| **12th** | **SGRR Public School, Dehradun** | **2008** | **60%** |
| **10th** | **SGRR Public School, Dehradun** | **2006** | **66%** |

*All courses are Full Time*

*\*WILP stands for Work Integrated Learning Program/WASE – Wipro Academy of Software Excellence*

**PROFESSIONAL COMPETENCIES**



**Spark, PySpark**

**Kafka**

**Teradata**

**Informatica**

**Hadoop – Hive, Pig, Sqoop, etc**

**Gitlab/Github**

**Data Engineering**

**Cassandra, HBase, MongoDB**

**Shell scripting**

**Tableau**

**Hyperion Essbase**

**Docker on Kubernetes**

**Jira**

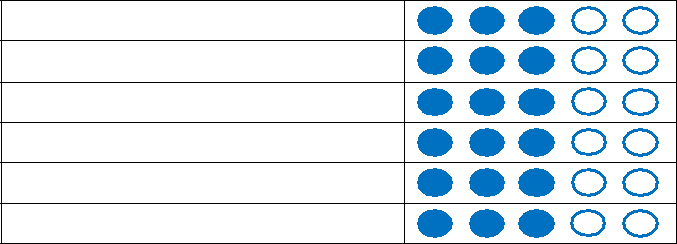
**Jupyter/Zeppelin/Atom**

**Python**

**Applied Stats using R & SAS**

**Scala/Java/C++/C**

**Acquired Skills**



**IoT and Sensor Analytics**

**Machine Learning**

**Decision Sciences**

**Kubernetes and Docker**

**CICD using DevOps**

**Web App Development**

**NOTEWORTHY CREDITS**

* ‘**Early Prodigy & Feather in Cap Award**’.
* Felicitated with ‘**Peoples Champ Award winner**’ for 4 years in a row.
* Got rewarded with **‘Project Execution Excellence** **Award’ and ‘Long Service Award’** in 2017.
* **Major achievements** include Client Satisfaction,multiple appreciations and accolades from clients.
* **Major contributor in Business development & Growth initiatives.**
* **3 things which best describes me** –A Traveler, Areader (Technical and Non-Fiction) & A big foodie.

**PERSONAL DETAILS**

**Date of Birth:** 13-Nov-1990 **Gender:** Male

**Nationality:** Indian

**Languages Known:** English and Hindi

**CAREER PROGRESSION**

**Company - Mu-Sigma, Bengaluru Team - Mu-Sigma LABS (R&D wing of Mu-Sigma) Data Sciences Team (DEFT)**



**Role: Data Engineer**

**October 2018 – Till Date**

**Team Size: 14**



**Spark**  **Kafka**  **NoSQL**  **Kubernetes**  **Docker**  **GCP**  **AWS**



**Alteryx**



**SAS**  **Python & R**  **Teradata**



**OpenStack**



**Data Engineering**



**Informatica BDM**



**Azure**



**EDGE Computing**  **AWS Kinesis**



**IOT Analytics**



**Research and Development wing of Mu-Sigma (DEFT):**

Being part of LABS team in Mu-Sigma, I have helped the organization in growing and building endless capabilities around major and niche-end technologies of the current and future era. I have worked round the clock and helped multiple teams accelerate in Data engineering space. With an aim of business and revenue growth, I have been closely working with presales and sales team to help engage themselves more and more with clients in the space of Data Engineering or D3 Stack.

* Worked on creating capabilities around real-time and batch-based data pipelines using Lambda and Kappa architectures on cloud and on-premise infrastructure. Heavily engaged in handling multiple clients on data engineering portfolios from different domains such as Telecom, Retail, Manufacturing, Banking, E-Commerce, Healthcare and from different geographies across the globe as well.
* Worked with teams on operationalizing, curating and creating meaningful insights from data of up to 40+ TB’s by suggesting architectures, landscapes and deploying pipelines in client environments;
* Implemented and created reusable artefacts from the research and experiments around multiple cloud vendors (AWS, GCP, Azure);
* Benchmarked different open-source and proprietary tools on different API’s, platforms, versions using different resource managers;
* Benchmarked Apache products like Spark, Flink, Kafka, etc. using YARN/Kubernetes on cloud and on-premise;
* Created and shared a dozen of artefacts within the org and with the clients for showcasing capabilities in space of Data Engineering and cloud;
* Created real-time and batch-based data pipelines on Lambda and Kappa architecture using Kafka, Spark, AWS Kinesis, Cassandra, Redshift and Python, keeping compliance, InfoSec and security protocols as top concern;



**Client: World’s Largest Networking Company [B2B] Project: Retail/Finance ASBI Domain: Manufacturing [NEPC] Metrics – Finance and Retail**



**Role: Senior Data Analyst**

**June 2017 – Till Date**

**Team Size: 12**

**Wipro Limited, Bengaluru**



 **Hive**



 **Sqoop**



 **Teradata**



 **Cassandra**



 **Impala**



 **Oracle**



 **Informatica**



 **Spark**



 **Scala**



 **Python**



**Project Description: Setting up batch and real time data pipeline for data analysis**

This is a Teradata to Hadoop and SAP HANA migration project. The objective of this project is to induct the data into Hadoop environments from existing legacy systems [Teradata, Oracle, etc.]. The project deals with Finance and Revenue metrics. With source of truth as Teradata for business users, providing solutions to the customers where in optimal data ingestions techniques have to be laid down. Major challenges involved in this project are: first is to integrate various metrics from Teradata in a single Hadoop platform without hampering the data; secondly to keep up the current reporting system with incremental and parallel releases in Hadoop environment; implementing an automated synchronous data load mechanism to reduce the manual intervention while loading data to multiple platforms including Hadoop.



|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Client: World’s leading electronics manufacturer [B2C/B2B]** | | | | |  | **Role: Data Analyst** | |  |  |  |
| **Project: GBI TURBO & GBI SPA** | |  |  |  |  | **Oct 2012 – May 2017** | | |  |  |
| **Domain: Manufacturing** | |  |  |  |  | **Team Size: 16** | |  |  |  |
| **Metrics - Finance/Retail/Operations/Sales** | | |  |  |  | **Wipro Limited, Bengaluru** | | |  |  |
| **Teradata** | **Informatica** | **SAP BO** | **Essbase** | **Tableau** | **Shell Scripting** | **Python** | **Hive** | **Oracle** | **Cassandra** | **Sqoop** |



**Project Description: Global Business Intelligence [GBI] is an ecosystem of multiple Analytical tools and technologies which helps Business users to make decisions, forecasting, perform analysis, etc.**

The objective of this project was to fill gaps in the reporting and analysis tools available to Finance analysts. The project will integrate units, revenue, and margin data into Essbase and automate reporting and analysis. This engagement started with providing world-class Essbase finance solutions using Informatica PC and Teradata as ETL Tools to the finance and retail business users. The project continues to provide seamless reporting experience.

**Projects/Application Implementations (Oct,2012 – May,2017):**

|  |  |  |
| --- | --- | --- |
| SPA (Ship Plan Attainment) | Developed Corp Units | STF Actuals and Forecast |
| Essbase DR Setup | EQM [Essbase Quality Management] | FAST [Finance Analytics Spreadsheet Tool] |
| Essbase Alias Optimization | RAST [Retail Analytics Spreadsheet Tool] | FINOPEX [Financial Operating Expenses] |
| Developed URM & GM/PL | Hyperion Upgrades and enhancements | Supported EDW migrations/up-gradations |