

# Sarthak Nagpal



## About Me

Experienced Data Engineer specializing in cloud-based solutions and big data technologies. Proficient in data warehousing, streaming and batch data processing, data transformation and developing and maintaining robust data pipelines, thus ensuring seamless flow of data across systems.

Collaborative team player and communicator among cross functional teams, experienced in translating business requirements into efficient and scalable data solutions. Proactive in staying abreast of emerging trends in data engineering to drive continuous improvement and delivering high quality data products in fast-paced environments.

## Contact

**Email:** nagpal97.sarthak@gmail.com

**Phone:** +91-7018930292

**Linkedin:** [www.linkedin.com/in/srtk-nagpal/](https://www.linkedin.com/in/srtk-nagpal/)

**Github:** <https://github.com/sarthak2897>

**Address:** Bengaluru, Karnataka

## Skills

### Technical Skills

- **Programming Languages:** SQL, Scala, Python, Java
- **Database Technologies:** PostgreSQL, MySQL, MongoDB, Azure Cosmos DB, Cassandra
- **ETL Tools:** Azure Data Factory
- **Big Data Technologies:** Apache Spark, Azure Databricks, Azure ADLS Gen2, Apache Kafka, Akka,
- **Data Modeling:** ERD, Kimball Dimensional Modeling, Data Vault 2.0
- **Version Control:** Git, SVN
- **Cloud Platforms:** Azure
- **Data Warehousing:** Azure Synapse Analytics, Snowflake
- **Miscellaneous :** Azure Logic Apps, Unix(bash), Docker, Kubernetes, React JS, Jira, Confluence, Grafana, Xmatters

### Business Skills

Communication, Cross-functional collaboration, Leadership, Problem solving, Decision making, Adaptability, Teamwork

## Education

### B Tech in Computer Science Engineering (2014-2018)

- Jaypee University of Information Technology ,Solan
- CGPA - 8.4 / 10

## Certifications

- Azure Fundamentals : AZ-900
- Azure Data Fundamentals : DP-900
- Azure Data Engineer Associate : DP-203
- Snowflake Snowpro Core : COF-C02

## Work Experience

### Software Engineer 2

*Hashedin By Deloitte (May 2021 - Present)*

#### Health and Wellness Data Services (Feb. 2023 - Present)

- Currently working as a Lead Data Engineer in Healthcare domain on data curation and warehousing of Healthcare data consisting of ingesting various file formats and exposing the data as a Data As a Product(DaaP) to customers.
- The work involves designing and creating robust pipelines to ingest data, model, transform and serve data using Microsoft Azure services ensuring the availability and reliability of critical data for analytics and reporting.
- Collaborating with cross-functional teams to design and strategize the processing of different types of feeds related to various drugs ,to be served to Business Partners in efficient and seamless way.

**Technologies used :** Pyspark · SQL · Azure Data Factory · Azure Databricks · Azure ADLS Gen2 · Azure Synapse Analytics · Azure Logic Apps · Azure Log Analytics · Github · Jira

#### Sales Transactions Financial Control System (Jun. 2021 - Jan. 2023)

- Worked as a Data Engineer on real time scalable and fault tolerant distributed systems in retail domain. It was concerned with processing and transformation of approx. 70M sales transactions per day in real time and produce it for reconciliation.
- Worked on scalability of the application by identifying opportunities to improve performance whether through logic or making few performance improvements on the database front (Azure cosmos DB in particular).
- Worked on adding alerting and monitoring mechanisms in the system eg. setting up Kafka consumer lag alerts via Xmatters and creating Grafana dashboards.
- Worked on creating a spring batch job to perform data observability and identify missing points in the system so that the issues can be resolved quickly.

**Technologies used :** Scala · Akka · SQL · Play Framework · Apache Kafka · Azure Cosmos DB · Cassandra · xMatters · Grafana · Spring Batch · Spring Boot · Kubernetes · Github · Jira

### Technology Analyst

*Infosys Limited (Aug. 2018 - Apr. 2021)*

#### Pharmacology Experimentation (Aug. 2018 - Apr. 2021)

- Maintaining applications under Healthcare domain for improved performance by writing efficient and testable code in Java(Core and Spring Boot) and javascript and identifying opportunities for improvement.
- Developed a web application in Spring Boot to validate and store the user data to a remote archive storage , also taking care of the faulty saves which the native application misses out thus, reducing dependency on Business clients and audit reporters for any missed data in the archive storage.
- Created a service script using UNIX shell scripting to monitor certain services of an application as a value add thus reducing an overhead of numerous dollars every month for the Business clients.
- Created an automation to check the health and status of the web application servers of a particular domain and provide alerts to the team on any failure so that the issue can be resolved there and then and the users don't face any downtime.
- Developed a data processing web application in Java where the user can upload and process their biological study data based on the conclusions of their respective experiments with objective being to provide the users a proper validation mechanism to process their sensitive data.

**Technologies used :** Java · Spring Boot · Javascript · CSS · Apache JBoss · Tomcat · Bash

## Projects

- [Healthcare Insights](#) - Creating a data pipeline performing data prep, data modelling, transformation, warehousing and visualization using Snowflake.
- [Business Insights](#) - Created a pipeline which ingests data from Kafka in real time using Akka streams ,data is written to Mongo DB and data is later written to json files to ADLS Gen2 , which is ingested in Snowflake for reporting purposes via Snowpipe.
- [Business Review Insights](#) - Pipeline built in Scala, Akka and Play framework to produce review messages from json file and consume those messages from Kafka, transform and serve it to Cassandra.