

POSITION	Senior Data Engineer / Data Engineer
LOCATION	Bangalore/Mumbai/Kolkata/Gurugram/Hyd/Pune/Chennai
EXPERIENCE	1.3 to 7.5 Years

ABOUT HASHEDIN

We are software engineers who solve business problems with a Product Mindset for leading global organizations.

By combining engineering talent with business insight, we build software and products that can create new enterprise value.

The secret to our success is a fast-paced learning environment, an extreme ownership spirit, and a fun culture.

WHY SHOULD YOU JOIN US?

With the **agility of a start-up** and the **opportunities of an enterprise**, every day at Hashedin, your work will make an impact that matters.

So, if you are a **problem solver** looking to thrive in a dynamic **fun** culture of inclusion, collaboration, and high performance – Hashedin is the place to be!

From learning to leadership, this is your chance to take your software engineering career to the next level.

So, what impact will you make?

Visit us @ <https://hashedin.com>

JOB TITLE: Senior Data Engineer / Data Engineer

OVERVIEW OF THE ROLE:

As a Data Engineer or Senior Data Engineer, you will be hands-on in architecting, building, and optimizing robust, efficient, and secure data pipelines and platforms that power business-critical analytics and applications. You will play a central role in the implementation and automation of scalable batch and streaming data workflows using modern big data and cloud technologies. Working within cross-functional teams, you will deliver well-engineered, high-quality code and data models, and drive best practices for data reliability, lineage, quality, and security.

Mandatory Skills:

- Hands-on software coding or scripting for minimum 3 years
- Experience in product management for at-least 2 years
- Stakeholder management experience for at-least 3 years
- Experience in one amongst GCP, AWS or Azure cloud platform

Key Responsibilities:

- **Design, build, and optimize scalable data pipelines and ETL/ELT workflows** using Spark (Scala/Python), SQL, and orchestration tools (e.g., Apache Airflow, Prefect, Luigi).
- Implement efficient solutions for **high-volume, batch, real-time streaming**, and event-driven data processing, leveraging best-in-class patterns and frameworks.
- Build and maintain **data warehouse and lakehouse architectures** (e.g., Snowflake, Databricks, Delta Lake, BigQuery, Redshift) to support analytics, data science, and BI workloads.
- Develop, automate, and monitor **Airflow DAGs/jobs** on cloud or Kubernetes, following robust deployment and operational practices (CI/CD, containerization, infra-as-code).
- **Write performant, production-grade SQL** for complex data aggregation, transformation, and analytics tasks.
- Ensure **data quality, consistency, and governance** across the stack, implementing processes for validation, cleansing, anomaly detection, and reconciliation.
- Collaborate with Data Scientists, Analysts, and DevOps engineers to **ingest, structure, and expose structured, semi-structured, and unstructured data** for diverse use-cases.
- Contribute to **data modeling, schema design, data partitioning** strategies, and ensure adherence to best practices for performance and cost optimization.
- Implement, document, and extend **data lineage, cataloging, and observability** through tools such as AWS Glue, Azure Purview, Amundsen, or open-source technologies.
- Apply and enforce **data security, privacy, and compliance requirements** (e.g., access control, data masking, retention policies, GDPR/CCPA).
- Take ownership of end-to-end **data pipeline lifecycle**: design, development, code reviews, testing, deployment, operational monitoring, and maintenance/troubleshooting.
- Contribute to **frameworks, reusable modules, and automation** to improve development efficiency and maintainability of the codebase.
- Stay abreast of industry trends and emerging technologies, participating in code reviews, technical discussions, and peer mentoring as needed.

Skills & Experience:

- **Proficiency with Spark (Python or Scala), SQL, and data pipeline orchestration** (Airflow, Prefect, Luigi, or similar).
- Experience with **cloud data ecosystems** (AWS, GCP, Azure) and cloud-native services for data processing (Glue, Dataflow, Dataproc, EMR, HDInsight, Synapse, etc.).

- Hands-on development skills in at least one programming language (Python, Scala, or Java preferred); solid knowledge of software engineering best practices (version control, testing, modularity).
- **Deep understanding of batch and streaming architectures** (Kafka, Kinesis, Pub/Sub, Flink, Structured Streaming, Spark Streaming).
- Expertise in **data warehouse/lakehouse solutions** (Snowflake, Databricks, Delta Lake, BigQuery, Redshift, Synapse) and storage formats (Parquet, ORC, Delta, Iceberg, Avro).
- **Strong SQL development skills** for ETL, analytics, and performance optimization.
- Familiarity with **Kubernetes (K8s), containerization (Docker)**, and deploying data pipelines in distributed/cloud-native environments.
- Experience with **data quality frameworks** (Great Expectations, Deequ, or custom validation), monitoring/observability tools, and automated testing.
- Working knowledge of **data modeling (star/snowflake, normalized, denormalized)** and metadata/catalog management.
- Understanding of **data security, privacy, and regulatory compliance** (access management, PII masking, auditing, GDPR/CCPA/HIPAA).
- Familiarity with BI or visualization tools (PowerBI, Tableau, Looker, etc.) is an advantage but not core.
- Previous experience with **data migrations, modernization, or refactoring legacy ETL processes to modern cloud architectures** is a strong plus.
- Bonus: Exposure to open-source data tools (dbt, Delta Lake, Apache Iceberg, Amundsen, Great Expectations, etc.) and knowledge of DevOps/MLOps processes.

Professional Attributes:

- Strong analytical and problem-solving skills; attention to detail and commitment to code quality and documentation.
- Ability to communicate technical designs and issues effectively with team members and stakeholders.
- Proven self-starter, fast learner, and collaborative team player who thrives in dynamic, fast-paced environments.
- Passion for mentoring, sharing knowledge, and raising the technical bar for data engineering practices.

Desirable Experience:

- Contributions to open source data engineering/tools communities.
- Implementing data cataloging, stewardship, and data democratization initiatives.
- Hands-on work with DataOps/DevOps pipelines for code and data.
- Knowledge of ML pipeline integration (feature stores, model serving, lineage/monitoring integration) is beneficial.

EDUCATIONAL QUALIFICATIONS:

- Bachelor's or Master's degree in Computer Science, Data Engineering, Information Systems, or related field (or equivalent experience).
- Certifications in cloud platforms (AWS, GCP, Azure) and/or data engineering (AWS Data Analytics, GCP Data Engineer, Databricks).
- Experience working in an Agile environment with exposure to CI/CD, Git, Jira, Confluence, and code review processes.
- Prior work in highly regulated or large-scale enterprise data environments (finance, healthcare, or similar) is a plus.