

Roadmap to Discord Data Engineer Role (3rd Year B.Tech to Placement)

Prepared for: TARUN Soni\ **Target Role:** Data Engineer at Discord\ **Timeframe:** 12 Months (3rd Year to Early 4th Year)\ **Outcome:** Portfolio + Skills to Clear Discord Interviews



Month-by-Month Detailed Roadmap with Weekly Tasks

(Original weekly breakdown maintained above...)



Comprehensive Checklist of Skills, Tools, Technologies, and Tasks

Core Technologies & Languages:

- ☒ Advanced SQL (Joins, Window Functions, CTEs, Indexing, Optimization)
- ☒ Python for Data Engineering (pandas, pathlib, json, requests)
- ☒ Linux (bash scripting, cron jobs, file permissions)
- ☒ Git & GitHub (branches, pull requests, project boards)

Data Engineering Tools:

- ☒ Apache Airflow (DAGs, Operators, Scheduling, Monitoring)
- ☒ DBT (Jinja templates, YAML configs, Incremental models)
- ☒ PostgreSQL (SQL practice, schema design, normalization)
- ☒ BigQuery / Snowflake (SQL, storage, analytics)
- ☒ Looker Studio / Metabase (Dashboards, Reporting)
- ☒ Docker (Containers, Reproducibility)
- ☒ Great Expectations (Data Quality, Audits)
- ☒ Kafka (Intro level, streaming basics)

Foundational Knowledge Areas:

- ☒ Data Lakes vs Warehouses (Architecture differences)
- ☒ ETL Principles (Extract, Transform, Load Design)
- ☒ Partitioning, Sharding, Schema Evolution
- ☒ Data Modeling (Star, Snowflake Schema, OLAP vs OLTP)
- ☒ Data Quality & Anomaly Detection Techniques
- ☒ Logging, Monitoring, Alerts for Pipelines
- ☒ System Design Principles for Data Engineers
- ☒ Designing Scalable, Cost-Efficient Pipelines

Portfolio Projects (Mandatory for Resume & GitHub):

1 YouTube Analytics ETL Pipeline (PostgreSQL) 2 Spotify Listening History Cloud Pipeline (Airflow, DBT, BigQuery) 3 Gaming Analytics Platform (Kafka, Airflow, DBT, BigQuery, Looker) 4 Discord-like Event Tracking System (Data Models, Alerts, Dashboards)

Communication & Documentation:

- ✓ GitHub Clean Structure: Clear READMEs, architecture diagrams, commit history
- ✓ Personal Portfolio Site: Project highlights, visuals, clear explanations
- ✓ Technical Blog Posts: Project walkthroughs, lessons learned, DE concepts

Interview Preparation:

- ✓ SQL / Python Coding Practice (LeetCode, HackerRank)
- ✓ System Design Interview Prep (Data-Intensive Applications, YouTube tutorials)
- ✓ Behavioral Questions (STAR Method, Discord culture fit)
- ✓ Mock Interviews (Peer, Online Platforms)

Networking & Application:

- ✓ Tailored Resume: Quantify impact, list technologies, show portfolio
- ✓ LinkedIn Profile Optimization: Projects, blogs, skills, clear headline
- ✓ Reaching Out to Connections: Discord employees, recruiters, DE community
- ✓ Understanding Discord Culture: Async communication, gaming relevance

Learning Resources:

- **Courses:**
 - DataTalksClub DE Zoomcamp (Airflow, DBT, Cloud)
 - SQL for Data Science (Coursera)
 - Python for Data Engineers (DataCamp)
- **Books:**
 - Designing Data-Intensive Applications (Kleppmann)
- **YouTube Channels:**
 - Seattle Data Guy
 - Data Engineer Things
 - Alex The Analyst
- **Blogs:**
 - Towards Data Science (Medium)
 - dbt Labs Blog
 - Google Cloud Blog (BigQuery use cases)

Final Deliverables Checklist Before Applying

| Deliverable | Status |
|---------------------------|--------|
| LinkedIn optimized | ✓ |
| Resume tailored (Discord) | ✓ |

| Deliverable | Status |
|----------------------------|--------|
| GitHub clean, professional | ✓ |
| Blog post(s) published | ✓ |
| Portfolio website | ✓ |
| 4 Major Projects Complete | ✓ |
| Mock interviews done | ✓ |

End Goal: Complete readiness for Discord Data Engineer role with all skills, tools, deliverables, and proof-of-work projects aligned to industry expectations.

Ready for PDF export with visuals and professional formatting.