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 with Required Technologies

✂️ YouTube Analytics ETL Pipeline (PostgreSQL)

Goal: Build an end-to-end ETL pipeline to collect, store, and analyze YouTube video data.

Technologies / Skills Needed:

- PostgreSQL (Database setup, SQL)
- Python (API calls, ETL scripting)
- YouTube API (Data extraction)
- GitHub (Project management)
- Linux (Scheduling with cron)

Key Deliverables: - ERD Diagrams - Data pipeline flow - SQL reports (views, analytics queries)

🖋️ Spotify Listening History Cloud Pipeline (Airflow, DBT, BigQuery)

Goal: Automate a cloud data pipeline for Spotify listening analytics using modern orchestration and transformation tools.

Technologies / Skills Needed:

- Spotify API (Data extraction)
- Python (API, ETL scripting)
- Google Cloud (BigQuery, Storage)
- Airflow (Orchestration)
- DBT (Transformations)
- Looker Studio (Dashboards)

Key Deliverables: - Cloud architecture diagrams - Airflow DAGs & DBT models - Analytical dashboards (listening patterns)

🛏️ Gaming Analytics Platform (Kafka, Airflow, DBT, BigQuery, Looker)

Goal: Build a simulated event-based analytics system mimicking Discord or Steam analytics for games.

Technologies / Skills Needed:

- Kafka (Stream simulation, optional for realism)
- Python (Data simulation, API handling)
- Airflow (Batch orchestration)
- DBT (Warehouse transformations)

• Google Cloud (BigQuery)
  • Looker Studio / Metabase (Dashboards)
  • Great Expectations (Data quality checks)

Key Deliverables: - Event schema design (DAU, MAU, retention) - Analytical dashboards for KPIs - System diagrams

---

🌂 Discord-like Event Tracking System (Data Models, Alerts, Dashboards)

Goal: Build a robust event tracking pipeline similar to what Discord might use for user analytics, focusing on data quality and operational excellence.

Technologies / Skills Needed:

  • PostgreSQL (Schema design, storage)
  • Python (ETL, anomaly detection scripts)
  • Airflow (Orchestration)
  • DBT (Transformations)
  • Great Expectations (Data audits)
  • Looker Studio / Metabase (Business reporting dashboards)

Key Deliverables: - Anomaly detection reports - Quality audit documentation - Alerts for data issues (email/slack)

🔗 Final Deliverables Checklist Before Applying

Deliverable	Status
LinkedIn optimized	✅
Resume tailored (Discord)	✅
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.