Interview with Jonathan Pau - Data Engineer at HackerOne

Background & Career Path

- Jonathan Pau started his career at Nestl in supply chain.

- Moved to data science at Shipt, a tech startup.

- Became a data scientist at a consulting firm, gaining machine learning expertise.

- Later joined HackerOne as a Data Engineer, working on ETL tasks and financial modeling.

- Noted that Data Engineers (DE) dont receive much visibility because companies celebrate dashboards rather than backend work.

- In healthy organizations, product teams should integrate analytics, but in some corporations, analytics teams exist for show without real impact.

- Soft skills are crucial to build trust between the analytics team and product team.

- Data Analysts (DA) must internally market their work to show insights are useful and actionable.

Breaking into Data & Consulting

- Big consulting firms have teams specializing in data science projects.

- Consulting provides diverse experience but can involve projects that dont align with personal goals.

- Important questions to ask before joining consulting firms:

-    - What is the interview process like?

-    - How many rounds are there?

-    - What are the expectations for technical interviews?

-    - What types of teams will I be working with?

-    - How does project assignment work within the firm?

- Recruiters focus on the process, while managers assess cultural fit and team dynamics.

- Some consulting roles require preparation for up to 5 interviews, so preparation guides are helpful.

Career Paths in Data

- Data Engineering (DE): Architect, ETL Engineer, Analytics Engineer.

- Data Science (DS): Decision Scientist, Product Data Scientist.

- Machine Learning Engineering (MLE): Focuses on model deployment and scaling.

- Data Analytics (DA): Business insights, marketing collaboration, dashboard creation.

Data Engineering (DE)

- Responsibilities:
-   - Builds systems to ingest and process real-time data.
-   - Creates systems that ensure database reliability and prevent failures.
-   - Designs models to track user behavior, identity, and analytics.
-   - Ensures security of databases and establishes access control mechanisms.
-   - Sets up monitoring systems to detect failures and optimize processes.
- Tools & Technologies:
-   - Languages: Python, Java, Scala.
-   - Databases: Deep knowledge of SQL, database architecture.
-   - ETL & Data Pipelines: Apache Airflow, Spark, Hadoop.
-   - Streaming Technologies: Kafka.
- Pros:
-   - Satisfaction from building robust data systems.
-   - Field is constantly evolving with new tools and challenges.
-   - Strong job security as companies require data infrastructure.
-   - Engages in solving complex technical challenges.
- Cons:
-   - Work is often overlooked, as the focus is on dashboards rather than backend.
-   - On-call responsibilities can be demanding when data pipelines fail.
-   - Larger companies may assign maintenance of legacy systems instead of building new ones.
- Median Salary: $155K.
- Additional Notes:
-   - Are DEs responsible for transitioning physical databases to the cloud?
-   - How similar are DEs to solutions architects?

Data Science (DS)

- Responsibilities:
-   - Handles ambiguous problems that cannot be solved with simple analysis.
-   - Builds statistical models to determine causality and forecast trends.
-   - Focuses on future predictions, unlike DA, which looks at past performance.
-   - Determines what data is valuable and where to source missing data.
-   - Develops proxies to replace missing or unreliable data.

- - Models are trained for specific applications, such as time estimations in DoorDash.

- Tools & Technologies:

- - Programming: Python, R.

- - Databases: SQL.

- - Machine Learning: Various modeling frameworks.

- Pros:

- - High demand with flexible career paths across industries.

- - Opportunities to work in healthcare, insurance, tech, etc.

- - Engages in ML, hypothesis testing, and statistical experiments.

- - Continuous learning as the field evolves.

- - When models work well, they create a significant impact.

- Cons:

- - High barrier to entry (requires Masters, PhD, or years of experience).

- - A large portion of time is spent on data cleaning.

- - Problems are often ambiguous, requiring strong problem-solving skills.

- - Career transitions between companies can be more difficult than DE or DA.

- Median Salary: $168K.

Machine Learning Engineering (MLE)

- Responsibilities:

- - Uses DS-created models and optimizes them for real-time applications.

- - Ensures models can handle peak usage efficiently.

- - Focuses on productionizing machine learning models at scale.

- Tools & Technologies:

- - Strong software engineering fundamentals.

- - ML frameworks: PyTorch, TensorFlow, Hugging Face.

- - DevOps: CI/CD, systems design, production monitoring.

- Pros:

- - Works with cutting-edge AI/ML technologies.

- - Clear career progression and impact on business outcomes.

- - Role is growing rapidly, especially with LLMs (Large Language Models).

- Cons:

- - Field evolves extremely fast, requiring constant learning.

- - Production failures can be stressful, requiring rapid troubleshooting.

- - The role focuses on making models work at scale rather than inventing new ones.

- Median Salary: Higher than DS, varies by industry.

Data Analytics (DA)

- Responsibilities:

- - Investigates business problems using SQL and data visualization.

- - Works closely with marketing and finance teams.

- - Builds dashboards to track company performance metrics.

- Tools & Technologies:

- - SQL, Python.

- - Visualization: Tableau, Power BI, Excel.

- - Statistics for business insights.

- Pros:

- - Easier entry into the data industry.

- - Skills are applicable across industries.

- - Provides quick insights for business decisions.

- - Can transition into other careers like product management.

- Cons:

- - Involves significant repetitive data cleaning.

- - Typically deals with structured data and predefined problems.

- - Lower salary ceiling compared to DE, DS, and MLE.

- Median Salary: $106K.

How to Choose a Data Career

- Do you prefer building (DE) or discovering insights (DS, DA)?

- Do you want structured work (DE) or flexibility (MLE, DS)?

- Do you prefer working with clearly defined tasks (DE) or ambiguous problem-solving (DS)?

- At smaller companies, DS may include engineering tasks, while larger companies specialize roles.

Tips to Break into Data Careers

- Gain hands-on experience through internships, research projects, and personal projects.

- Not just about technical skillssoft skills and communication are equally important.

- Apply broadly, as hiring processes are often unpredictable.

- Recruiters assess qualifications on a rolling basis.

- Cover letters may help but are often not a deciding factor.