Data Engineering Interview prep

**1. Python Data Stack**

* Pandas and Numpy for data manipulation
* PySpark or Dask for handling large datasets
* Scikit-learn or similar for featurizing and processing data
* Best practices for automating data tasks (e.g., using Python scripts, scheduling jobs)

**2. Data Onboarding and Reconciliation**

* Data exploration techniques
* Automating data quality checks and validations
* Processes for onboarding datasets and classifying data
* Methods for reconciling data discrepancies (diff checks, error tracing)

**3. ETL Pipelines**

* ETL tools (Airflow, Prefect, Luigi, etc.)
* Building and maintaining automated ETL pipelines
* Handling and transforming financial datasets during ETL processes
* Monitoring and debugging ETL pipelines in a production environment

**4. Data Debugging and Anomaly Detection**

* Techniques for finding anomalies in datasets (e.g., outlier detection, validation logic)
* Tracing issues to the root cause in a data pipeline
* Deductive reasoning to debug complex data flows

**5. Data Analysis and Featurization**

* Cleaning and preprocessing financial data
* Featurization techniques (e.g., time-series analysis, aggregation)
* Data enrichment processes
* Using SQL for data analysis and querying large datasets

**6. SQL and Databases**

* Proficiency in at least one SQL dialect (PostgreSQL, MySQL, MSSQL)
* Writing complex SQL queries (joins, subqueries, window functions)
* Query optimization techniques
* Handling database operations on financial data

**7. Linux Command Line**

* Navigating the Linux file system
* Shell scripting and basic system administration tasks
* Managing processes, logs, and environment variables on Linux

**8. Production Support and Monitoring**

* Techniques for proactive monitoring of data pipelines
* Handling incidents and resolving issues in live trading environments
* Communication with internal customers regarding production issues

**9. Financial Datasets**

* Familiarity with Refinitiv, S&P, and Bloomberg data structures
* Understanding financial data conventions (tickers, pricing, metadata)
* Working with market data APIs

**10. Soft Skills**

* Communicating technical issues and resolutions with stakeholders
* Collaboration across teams in a trading environment

These topics align with the job description and should prepare you for a range of questions during the interview. Let me know if you'd like to dive deeper into any of these areas!