Muster mit erfundenen Daten. Alle Bewerbungen sind erfunden. Bewertet wurde geschwärzt, Namen stehen nur im Werkzeug.

Data Engineer (m/w/d)

40 Bewerbungen · 7 Kriterien, davon 2 Muss · Schwärzen an · zwei Prüfstufen

Kurzfassung. Von 40 Bewerbungen erfüllen 23 alle Muss-Kriterien. Die Top 10 liegen zwischen 67 und 93 Punkten, 9 davon bringen Erfahrung mit Sensor-, Zeitreihen- oder Robotikdaten mit. 17 Bewerbungen fehlt mindestens ein Muss-Kriterium, sie stehen am Ende, damit ihr Grenzfälle trotzdem seht.

40	23	10
eingegangen	Muss erfüllt	Top 10

Kriterien und Gewichte

Python und SQL sicher **Muss**	Gewicht 2, mittel	Englisch verhandlungssicher **Muss**	Gewicht 1, wenig
Pipelines, Orchestrierung, Streaming	Gewicht 3, stark	Sensor-, Zeitreihen- oder Robotikdaten	Gewicht 3, stark
Cloud und Infrastruktur	Gewicht 2, mittel	Datenmodellierung und Datenqualität	Gewicht 2, mittel
Start-up und Eigenverantwortung	Gewicht 1, wenig		

So rechnet die Liste: je Kriterium 0 bis 3 Punkte, mal Gewicht, umgerechnet auf 100. Muss-Kriterien brauchen mindestens 2 Punkte.

So ist der Bericht entstanden

Kriterien
Im Telefonat festgelegt: was Muss ist und wie stark jedes Kriterium zählt.

Stufe 1
Claude Haiku hat alle Bewerbungen auf die zwei Muss-Kriterien geprüft und ein Kurzprofil geschrieben.

Belege
Jede Punktzahl beruht auf einem Satz aus dem Lebenslauf. Im Werkzeug steht er bei jeder Bewerbung.

Geschwärzt
Name, Kontakt, Geburtsdaten, Staatsangehörigkeit und Familienstand vor der Bewertung entfernt.

Stufe 2
Claude Sonnet hat jede Bewerbung in allen sieben Kriterien bewertet, mit Beleg, Begründung, Stärken, offenem Punkt und Gesprächsfragen.

Entscheidung
Die Liste ist ein Vorschlag. Wer eingeladen wird, entscheidet ihr.

Top 10

Platz	Bewerbung	Python/ SQL Muss	Englisch Muss	Pipe-lines	Sensor-daten	Cloud	Modellie-rung	Start-up	Punkte
1	**Bewerbung 04** 7 J. Data Engineering für Drohnentelemetrie: Kafka, Flink, Terraform	2	2	3	3	3	3	3	93
2	**Bewerbung 06** 8 J. Data Engineering: Roboter-Telemetrie, Kafka, Iceberg, K8s	2	2	3	3	2	3	3	88
3	**Bewerbung 13** 5 J. Data Engineering für Agrar-Sensor- und Drohnendaten, Azure	2	2	3	3	2	3	2	86
4	**Bewerbung 02** 5 J. Data Engineering für Lidar- und Kameradaten, Databricks, GCP	2	3	2	3	2	3	3	83
5	**Bewerbung 03** 4 J. Data Engineering für Maschinen- und Cobot-Zeitreihen, Dagster	3	2	2	3	2	2	3	81
6	**Bewerbung 01** 6 J. Data Engineering in Robotik: Kafka, ROS-Bags, AWS, Terraform	2	2	3	3	2	1	3	79
7	**Bewerbung 09** 6 J. Data Engineering im Zahlungsverkehr: Kafka, Flink, Terraform	2	3	3	1	3	2	3	76
8	**Bewerbung 05** 3 J. Data Engineering im Ride-Pooling, Airflow, ROS-2-Daten	3	2	2	3	2	1	2	74
9	**Bewerbung 10** 4 J. Data Engineering für Smart-Meter-Zeitreihen, Airflow, AWS	2	3	2	2	2	2	2	69
10	**Bewerbung 11** 9 J. Prüfstands- und Messdaten: Python, Hive/Spark, MDF4	3	2	2	2	1	2	2	67

3 stark belegt · 2 gut belegt · 1 etwas · 0 nicht erkennbar

Knapp dahinter

Platz	Bewerbung	Kurzprofil	Punkte
11	12	3 J. Data Engineering für E-Scooter-Telemetrie: Kafka, dbt, GCP	67
12	16	3 J. Feature-Pipelines (ML), davor 4 J. Robotik-Forschungsdaten	64
13	23	3 J. Data Engineering mit PySpark und BigQuery im Medienhaus	64
14	31	15 J. Daten, zuletzt Leitung einer Datenabteilung mit 18 Personen	64
15	21	4 J. Data Science (Prognosen), Robotik-Datensatz im Forschungsprojekt	62

Die Top 10 im Einzelnen

Platz 1 **Bewerbung 04** 7 J. Data Engineering für Drohnentelemetrie: Kafka, Flink, Terraform **93** / 100

Die Bewerbung zeigt Streaming, Terraform und Rufbereitschaft mit Verantwortung auf einer Drohnen-Datenplattform. Die Tiefe bei SQL-Modellierung und Datenqualität ist im Gespräch zu klären.

Stärken

- Kafka, Flink und Iceberg im Echtzeitbetrieb für Drohnentelemetrie
- Terraform-Module für MSK, EKS und IAM im Betrieb
- Rufbereitschaft mit Leitung, Runbooks und Post-Mortems eingeführt

Offen oder dünn belegt

- SQL-Modellierung und Datenqualität (Tests, Verträge) sind nur knapp beschrieben.

Fragen fürs Gespräch

1. Wie sichern Sie im Pfad von Flink nach Iceberg die Datenqualität ab (Tests, Verträge, Überwachung)?
2. Welche Sensordaten außer GPS und Batterie fließen in die Pipeline, etwa Kamera oder Lidar?

Platz 2 **Bewerbung 06** 8 J. Data Engineering: Roboter-Telemetrie, Kafka, Iceberg, K8s **88** / 100

Die Bewerbung deckt Robotik-Telemetrie, Kafka, Iceberg und Kubernetes mit Aufbauverantwortung ab. Infrastruktur als Code und Cloud-Betrieb sind im Gespräch zu bestätigen.

Stärken

- Telemetrie von Lieferrobotern: Pose, Motorströme, Teleoperation-Vorfälle
- Kafka, Spark und Iceberg auf Kubernetes von Grund auf aufgebaut
- Iceberg-Migration und rund 40 % geringere Kosten bei Spark-Batch

Offen oder dünn belegt

- Terraform und Cloud-Betrieb stehen nur in der Kenntnisliste, nicht in den Aufgaben.

Fragen fürs Gespräch

1. Wie wird die Plattform bereitgestellt (Terraform, Helm, anderes) und in welcher Cloud läuft sie?
2. Welche Datenqualitäts-Prüfungen laufen für Telemetrie und Ereignisse, bevor das Sicherheitsteam sie nutzt?

Platz 3 **Bewerbung 13** 5 J. Data Engineering für Agrar-Sensor- und Drohnendaten, Azure **86** / 100

Die Bewerbung deckt Sensor-Zeitreihen und Drohnen-Flugdaten im Betrieb mit Azure, Databricks und Bicep breit ab. Kubernetes und Kafka-nahe Streaming-Erfahrung sind im Gespräch zu klären.

Stärken

- IoT-Sensordaten von 6.500 Geräten über IoT Hub und Event Hubs
- Metadaten-Katalog für Drohnenflüge mit 1,8 Mio. Kamerabildern
- Great Expectations und Delta Lake für Datenqualität im Betrieb

Offen oder dünn belegt

- Kubernetes, Kafka-nahes Streaming und SQL-Modellierung mit dbt sind nicht belegt.

Fragen fürs Gespräch

1. Wie wird in Ihrer Pipeline mit verspäteten und doppelten Messwerten im Streaming konkret umgegangen?
2. Wie ist die Event-Hubs-Ingestion überwacht, und wer reagiert bei Ausfällen?

Platz 4 **Bewerbung 02** 5 J. Data Engineering für Lidar- und Kameradaten, Databricks, GCP **83** / 100

Die Bewerbung passt fachlich eng zu Sensordaten im großen Maßstab, mit technischer Leitung. Streaming und Container-Orchestrierung sind im Gespräch zu klären.

Stärken

- Lidar- und Kameradaten im Petabyte-Maßstab mit Delta Lake
- Technische Leitung und Architekturentscheidungen zur Metadaten-Plattform
- Datenqualität über Databricks Workflows und Great Expectations

Offen oder dünn belegt

- Streaming und Kubernetes sind nicht belegt, Terraform nur als Grundkenntnis.

Fragen fürs Gespräch

1. Wie würden Sie Ihre Batch-Ingestion auf kontinuierliche Streams (Kafka) übertragen, falls Echtzeit nötig wird?
2. Wie viel Infrastruktur betreiben Sie selbst (Terraform, Kubernetes) und wo liegen dabei Ihre Grenzen?

Platz 5 **Bewerbung 03** 4 J. Data Engineering für Maschinen- und Cobot-Zeitreihen, Dagster **81** / 100

Zeitreihen- und Cobot-Daten mit Dagster und dbt in Produktion passen direkt zur Stelle. Streaming und Infrastruktur als Code sind im Gespräch zu klären.

Stärken

- Roboter- und Maschinendaten (OPC UA, MQTT) in Produktion
- Dagster, dbt und TimescaleDB für Zeitreihen im Betrieb
- Aufbau der Plattform für 14 Kundenwerke auf Azure mit AKS

Offen oder dünn belegt

- Streaming mit Kafka oder Flink und Infrastruktur als Code sind nicht belegt.

Fragen fürs Gespräch

1. Wie gehen Sie mit Streaming-Last um, wenn Signale nicht gebündelt, sondern kontinuierlich eintreffen?
2. Wie wird die AKS-Umgebung bereitgestellt: per Infrastruktur als Code oder manuell?

Platz 6 **Bewerbung 01** 6 J. Data Engineering in Robotik: Kafka, ROS-Bags, AWS, Terraform **79** / 100

Die Bewerbung deckt Streaming, Orchestrierung und Robotik-Sensordaten in Produktion breit ab, mit fachlicher Verantwortung. Offen sind Kubernetes und Datenqualität im Sinne von Tests oder Lakehouse-Formaten.

Stärken

- Kafka-Telemetrie für rund 400 Fahrzeuge in Produktion
- ROS-Bag-Aufbereitung nach Parquet mit Glue-Katalog
- Airflow-Migration von 140 DAGs auf MWAA

Offen oder dünn belegt

- Kubernetes sowie Datenmodellierung mit Tests oder Lakehouse-Formaten sind nicht belegt.

Fragen fürs Gespräch

1. Wie sichern Sie im S3-Data-Lake Datenqualität und Schemata ab (Tests, Verträge, Überwachung)?
2. Wo laufen Ihre Dienste: Container-Orchestrierung oder Managed Services, und wer betreibt sie?

Platz 7 **Bewerbung 09** 6 J. Data Engineering im Zahlungsverkehr: Kafka, Flink, Terraform **76** / 100

Die Bewerbung zeigt sehr tiefe Streaming-, Terraform- und Rufbereitschafts-Erfahrung mit Verantwortung. Der Domänenbezug zu Sensor- und Robotikdaten fehlt.

Stärken

- Kafka und Flink im Betrieb bei 15k Transaktionen pro Sekunde
- Exactly-once und idempotente Senken nach realem Vorfall eingeführt
- Gemeinsame Terraform-Module für drei Produktteams

Offen oder dünn belegt

- Sensor-, Zeitreihen- und Robotikdaten sind nicht belegt, die Erfahrung liegt bei Zahlungsereignissen.

Fragen fürs Gespräch

1. Wie würden Sie Ihre Streaming-Erfahrung auf Sensordaten mit Zeitstempel-Drift und Ausfällen übertragen?
2. Wie viel Kubernetes (EKS) betreiben Sie tatsächlich im Alltag Ihrer Pipelines?

Platz 8 **Bewerbung 05** 3 J. Data Engineering im Ride-Pooling, Airflow, ROS-2-Daten **74** / 100

Beide Muss-Kriterien sind erfüllt; belegt sind Airflow im Betrieb mit großem Volumen sowie ROS-2- und Fahrzeug-GPS-Daten. Offen bleiben Streaming, Kubernetes und Infrastruktur als Code.

Stärken

- Airflow-Pipelines mit großem Datenvolumen im laufenden Betrieb
- ROS-2-Aufzeichnungen (rosbag2/MCAP) in der Masterarbeit zu Parquet verarbeitet
- Fahrzeug-GPS-Spuren bereinigt und per Map Matching aufbereitet

Offen oder dünn belegt

- Streaming (Kafka, Flink) sowie Kubernetes und Infrastruktur als Code sind nicht belegt, AWS wird als Grundlagen bezeichnet.

Fragen fürs Gespräch

1. Wie lief die Verarbeitung der Fahrzeug-GPS-Daten ab, Batch oder Streaming, und wie wurden Ausfälle abgefangen?
2. Welchen Anteil gab es an der Umstellung auf S3 und Athena, und wo laufen die Airflow-Container?

Platz 9 **Bewerbung 10** 4 J. Data Engineering für Smart-Meter-Zeitreihen, Airflow, AWS **69** / 100

Die Bewerbung bringt Zeitreihen-Pipelines mit klarer Qualitätsprüfung im Produktivbetrieb mit. Streaming und Roboter-Sensordaten sind im Gespräch zu klären.

Stärken

- Zeitreihen im großen Maßstab: 36 Mio. Messwerte pro Tag
- Airflow-Validierung und Lückenfüllung mit rund 60 % weniger Handkorrekturen
- AWS-Betrieb mit Lambda und RDS, als Runbooks dokumentiert

Offen oder dünn belegt

- Streaming, Kubernetes und Infrastruktur als Code sind nicht belegt, Robotikdaten fehlen.

Fragen fürs Gespräch

1. Wie gehen Sie mit lückenhaften oder verspäteten Sensordaten um, wenn Pipelines kontinuierlich laufen müssen?
2. Wie wird die AWS-Umgebung bereitgestellt: manuell oder als Code?

Platz 10 **Bewerbung 11** 9 J. Prüfstands- und Messdaten: Python, Hive/Spark, MDF4 **67** / 100

Die Bewerbung bringt tiefe Messdaten- sowie Python- und SQL-Erfahrung aus der Fahrzeugtechnik mit. Cloud, Streaming und Infrastruktur als Code müssten neu aufgebaut werden.

Stärken

- Python-Importer für MDF4 und ASAM ODS mit Parquet in HDFS
- Messdaten von 38 Prüfständen, rund 1,2 TB pro Woche
- SQL-Data-Mart ersetzt rund 200 Excel-Auswertungen

Offen oder dünn belegt

- Cloud, Container und Streaming sind kaum belegt, die Plattform läuft im eigenen Rechenzentrum auf Hadoop.

Fragen fürs Gespräch

1. Welche Teile Ihres Hadoop-Wissens würden Sie auf Cloud und Container übertragen, und wie?
2. Wie viel Erfahrung gibt es mit kontinuierlicher statt nächtlicher Verarbeitung von Messdaten?

Muss-Kriterium fehlt

Diese Bewerbungen stehen nicht in der Rangliste. Die Punkte zeigen, ob ein Grenzfall dabei ist, den ihr trotzdem ansehen wollt.

Nr.	Kurzprofil	Was fehlt	Punkte
39	4 J. Robotik-Daten (ROS 2, MCAP) auf AWS mit Terraform	Englisch	81
35	5 J. Data Engineering mit Kafka, AWS und Sensordaten (Windenergie)	Englisch	74
36	6 J. Data Engineering auf der JVM: Spark, Kafka Streams, AWS	Python/SQL	62
24	5 J. Big-Data-Batch mit Spark/Scala und Hive, On-Premise	Python/SQL	48
37	8 J. Messdatenauswertung (CAN, MDF) im Elektroantrieb mit MATLAB	Python/SQL	40
20	5 J. ETL und Data Warehousing im Telekom-Umfeld: Informatica, Oracle	Python/SQL, Englisch	38
22	4 J. Plattformbetrieb auf AWS mit Kubernetes und Terraform	Python/SQL	38
17	12 J. BI und Data Warehousing: SQL Server, SSIS, Power BI	Python/SQL	36
33	Quereinstieg, Bootcamp 2026: Airflow, Kafka, dbt, Python, SQL	Python/SQL	36
25	7 J. Embedded-Entwicklung (C++, ROS 2) in der Robotik	Python/SQL	29
27	10+ J. Webentwicklung mit PHP, Laravel und MySQL	Python/SQL, Englisch	29
34	16 J. SAP BW/BW4HANA, Datenmodellierung und Teilprojektleitung	Python/SQL, Englisch	29
28	6 J. Produktmanagement für Datenprodukte, SQL, kein Pipeline-Bau	Python/SQL	26
29	10+ J. Controlling, SQL auf ERP-Daten, Excel/VBA-Automatisierung	Python/SQL, Englisch	21
38	10+ J. Reporting, Tableau-Dashboards und Excel im Handel	Python/SQL, Englisch	21
32	10+ J. Linux-Betrieb mit 140 Servern, Monitoring und Ansible	Python/SQL, Englisch	19
40	Quereinstieg: Online-Kurse in SQL/Python, Berufserfahrung in Marketing	Python/SQL, Englisch	14