# Big Data - Case Study

## Subject - Big Data Analytics and Architecture

## PROJECT

## AI Job Market Analysis

**Name**-Adarsh Tripathi
**Roll no.**-1240259002

# AI Job Market Analysis Using Apache Hive

This project analyzes the AI job market dataset using Apache Hive on the Cloudera platform to explore hiring trends, skill demands, and salary patterns across industries. The dataset contains 2,000 job postings with details such as job titles, company information, experience levels, skills required, tools preferred, and salary ranges. Hive queries are used to perform data exploration, aggregation, and trend analysis, including identifying top hiring companies, in-demand AI skills, salary distribution by experience, and popular job locations. The project demonstrates how big data tools like Hive can efficiently process and analyze large-scale structured datasets stored in Hadoop.

Overall, this analysis provides valuable insights into the evolving AI employment landscape, helping professionals and organizations understand current market demands and emerging technology trends.

**Dataset Description**

The dataset ai_job_market.csv contains 2,000 records and 12 attributes, representing real-world job postings from various industries such as Technology, Finance, and Healthcare.

Key attributes include:

- job_id: Unique identifier for each job posting
- company_name: Name of the hiring organization
- industry: Sector of the job (e.g., Tech, Finance)
- job_title: Designation (e.g., Data Scientist, AI Engineer)
- skills_required: Technical skills like Python, TensorFlow, SQL, PyTorch
- exprerience: Level of experience required (Entry, Mid, Senior)
- employment_type: Type of employment (Full-time, Contract, etc.)
- location strings: City and state of the job location
- salary_range_usd: Salary range offered for the position
- company_size: Organization size (Small, Medium, Large)
- tools_preferred: Preferred tools and frameworks used by the company

**Technologies Used**

- Apache Hive
- Hadoop (Cloudera Environment)
- HiveQL (SQL-like Queries)
- CSV File Data Ingestion
- HDFS Storage

**Steps Performed**

1. Created a database and Hive table schema for the job dataset.
2. Loaded CSV data from local storage or HDFS into the Hive table.
3. Executed multiple Hive queries to summarize and visualize insights:
    - SELECT COUNT(*) → total number of job listings.
    - GROUP BY → industry and company analysis.
    - AVG() and MAX() → salary range insights by experience level.
    - ORDER BY and LIMIT → top hiring companies and skill trends.
4. Generated analytical reports summarizing job market trends and data-driven insights.

**Key Insights**

- Identified top industries contributing the most AI job postings.
- Found top companies hiring for AI roles globally.
- Discovered average salary variations across different experience levels.
- Highlighted most in-demand tools and skills for AI professionals.
- Observed growth in AI job postings over recent years.

## Results and Findings

- **Tech and Finance** industries dominated AI job openings.
- **Mid-level and senior professionals** were in highest demand.
- **Python, TensorFlow, and SQL** emerged as the most common skills.
- **Large companies** posted the majority of job listings.
- Salary analysis showed **higher pay for AI Engineers and Data Scientists** in tech-driven firms.

## Conclusion

The *AI Job Market Analysis Using Apache Hive* project demonstrates the practical use of Hive for handling and analyzing structured datasets in a Big Data environment. The findings provide clear visibility into hiring patterns, skill requirements, and salary expectations within the AI industry. By integrating Hive's querying capabilities with Hadoop's storage efficiency, this project showcases how data-driven insights can be extracted effectively to support workforce planning, education strategies, and career decisions in the AI field.

## Use Database:

```
[cloudera@quickstart Desktop]$ hive

Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.p
roperties
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive> show databases;
OK
bca
default
mca
project
Time taken: 0.907 seconds, Fetched: 4 row(s)
hive> use project;
OK
Time taken: 0.082 seconds
hive> show tables;
OK
market
Time taken: 0.224 seconds, Fetched: 1 row(s)
hive> load data local inpath'/home/cloudera/Desktop/ai_job_market.csv into table
market;
```

# Load data:

```
hive> load data local inpath'/home/cloudera/Desktop/ai_job_market.csv' into table market;
Loading data to table project.market
Table project.market stats: [numFiles=1, totalSize=346126]
OK
Time taken: 5.111 seconds
hive> desc market;
OK
job_id                  int
company                 string
industry                string
job_title               string
skill_required          string
exprerience             string
location                string
salary_range_usd        string
post_date               string
company_size            string
tool_preferred          string
Time taken: 0.353 seconds, Fetched: 11 row(s)
```

## Q.1 Total Number of Job Listings

SELECT COUNT(*) AS total_jobs FROM market;

*Insight:* Shows total number of job openings in the dataset.

```
hive> SELECT COUNT(*) AS total_jobs FROM market;
Query ID = cloudera_20251028215656_d0e60fa3-db7c-4654-8062-a4c67f38a128
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1761709441329_0001, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1761709441329_0001/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1761709441329_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2025-10-28 21:57:04,893 Stage-1 map = 0%,  reduce = 0%
2025-10-28 21:57:27,789 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.51 sec
2025-10-28 21:57:41,434 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 5.57 sec
MapReduce Total cumulative CPU time: 5 seconds 570 msec
Ended Job = job_1761709441329_0001
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 5.57 sec   HDFS Read: 354639 HDFS Write: 5 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 570 msec
OK
2001
Time taken: 81.785 seconds, Fetched: 1 row(s)
```

## Q.2 Most Common Job Titles

SELECT job_title, COUNT(*) AS count

FROM market

GROUP BY job_title

ORDER BY count DESC

LIMIT 10;

*Insight:* Identifies the top 10 job roles in demand

```
hive> SELECT job_title, COUNT(*) AS count
    > FROM market
    > GROUP BY job_title
    > ORDER BY count DESC
    > LIMIT 10;
Query ID = cloudera_20251029070808_d6f9de5e-620a-4957-8474-4e40d3891880
Total jobs = 2
```

Output

```
2025-10-28 22:25:23,788 Stage-1 map = 0%,  reduce = 0%
2025-10-28 22:25:37,806 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.61 sec
2025-10-28 22:25:48,057 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 3.97 sec
MapReduce Total cumulative CPU time: 3 seconds 970 msec
Ended Job = job_1761709441329_0009
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1761709441329_0010, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1761709441329_0010/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1761709441329_0010
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2025-10-28 22:26:09,169 Stage-2 map = 0%,  reduce = 0%
2025-10-28 22:26:17,938 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 2.64 sec
2025-10-28 22:26:27,892 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 4.17 sec
MapReduce Total cumulative CPU time: 4 seconds 170 msec
Ended Job = job_1761709441329_0010
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 3.97 sec   HDFS Read: 354115 HDFS Write: 593 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 4.17 sec   HDFS Read: 5612 HDFS Write: 187 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 140 msec
OK
NLP Engineer    178
Data Analyst    178
Quant Researcher        175
AI Product Manager      174
AI Researcher   161
ML Engineer     155
Data Scientist  153
Computer Vision Engineer        147
Finance 105
E-commerce      105
Time taken: 83.079 seconds. Fetched: 10 row(s)
```

## Q.3 Jobs by Experience Level

SELECT experience_level, COUNT(*) AS total

FROM market

GROUP BY experience_level

ORDER BY total DESC;

*Insight:* Helps see which level (Entry, Mid, Senior) has more job opportunities

```
ive> SELECT experience_level, COUNT(*) AS total
   > FROM market
   > GROUP BY experience_level
   > ORDER BY total DESC;
```

**Output**

```
OK
 MLflow 78
 Pandas 77
 GCP    74
 Excel  72
 Power BI       65
 TensorFlow     65
 Hugging Face   64
 SQL    64
 C++    61
 NumPy  60
 PyTorch        60
 Python 59
 Keras  57
 Flask  56
 Azure  55
 CUDA   54
 Reinforcement Learning 53
 LangChain      53
 AWS    53
 Scikit-learn   50
 R      48
 FastAPI        43
"Reinforcement Learning 38
"FastAPI        36
"Azure  36
"Power BI       36
"Keras  35
"Excel  35
"NumPy  33
"Python 33
"LangChain      33
"MLflow 32
"R      31
"CUDA   31
"GCP    30
"PyTorch        29
"Flask  29
"SQL    29
"AWS    28
"Hugging Face   27
"Scikit-learn   27
"TensorFlow     25
"C++    24
"Pandas 22
experience_level        1
Time taken: 59.167 seconds, Fetched: 45 row(s)
hive>
```

## Q.4 Top 10 Companies Offering the Most Jobs

SELECT company_name, COUNT(*) AS job_count

FROM market

GROUP BY company_name

ORDER BY job_count DESC

LIMIT 10;

*Insight:* Shows the companies hiring most actively.

```
ost_date, company_size, loot_preferred)
hive> SELECT company_name, COUNT(*) AS job_count
    > FROM market
    > GROUP BY company_name
    > ORDER BY job_count DESC
    > LIMIT 10;
```

Output

```
Total MapReduce CPU Time Spent: 6 seconds 300 msec
OK
"Johnson        16
"Smith  15
"Miller 13
"Williams       13
"Garcia 9
"Brown  8
"Thompson       7
"Gonzalez       6
"Anderson       6
"Walker 6
Time taken: 58.015 seconds, Fetched: 10 row(s)
hive>
```

## Q.5 Job Distribution by Employment Type

SELECT employment_type, COUNT(*) AS total

FROM market

GROUP BY employment_type;

*Insight:* Breaks down jobs by type (Full-time, Contract, Internship, etc.).

```
hive> SELECT employment_type, COUNT(*) AS total
    > FROM market
    > GROUP BY employment_type;
```

**Output**

```
Total MapReduce CPU Time Spent: 3 seconds 50 msec
OK
 AWS    71
 AWS"   12
 Azure  83
 Azure" 8
 C++    68
 C++"   18
 CUDA   83
 CUDA"  17
 Excel  87
 Excel" 18
 FastAPI        70
 FastAPI"       12
 Flask  64
 Flask" 20
 GCP    71
 GCP"   9
 Hugging Face   75
 Hugging Face"  8
 Keras  77
 Keras" 16
 LangChain      76
 LangChain"     12
 MLflow 61
 MLflow"        18
 NumPy  84
 NumPy" 17
 Pandas 88
 Pandas"        14
 Power BI       73
 Power BI"      13
 PyTorch        82
 PyTorch"       17
 Python 70
 Python"        16
 R      77
 R"     20
 Reinforcement Learning 76
 Reinforcement Learning"        16
 SQL    72
 SQL"   16
 Scikit-learn   78
 Scikit-learn"  19
 TensorFlow     80
 TensorFlow"    18
employment_type 1
Time taken: 30.706 seconds, Fetched: 45 row(s)
hive>
```

## Q.6 Average Salary Range by Experience Level

SELECT experience_level,

AVG(CAST(SPLIT(salary_range_usd, '-')[0] AS INT)) AS avg_min_salary,

AVG(CAST(SPLIT(salary_range_usd, '-')[1] AS INT)) AS avg_max_salary

FROM market

GROUP BY experience_level;

*Insight:* Estimates salary differences between junior and senior positions.

```
hive> SELECT experience_level,
    >        AVG(CAST(SPLIT(salary_range_usd, '-')[0] AS INT)) AS avg_min_salary,
    >        AVG(CAST(SPLIT(salary_range_usd, '-')[1] AS INT)) AS avg_max_salary
    > FROM market
    > WHERE salary_range_usd IS NOT NULL
    > GROUP BY experience_level;
```

## Output

```
Total MapReduce CPU Time Spent: 5 seconds 450 msec
OK
 AWS    NULL    NULL
 Azure  NULL    NULL
 C++    NULL    NULL
 CUDA   NULL    NULL
 Excel  NULL    NULL
 FastAPI        NULL    NULL
 Flask  NULL    NULL
 GCP    NULL    NULL
 Hugging Face   NULL    NULL
 Keras  NULL    NULL
 LangChain      NULL    NULL
 MLflow NULL    NULL
 NumPy  NULL    NULL
 Pandas NULL    NULL
 Power BI       NULL    NULL
 PyTorch        NULL    NULL
 Python NULL    NULL
 R      NULL    NULL
 Reinforcement Learning NULL    NULL
 SQL    NULL    NULL
 Scikit-learn   NULL    NULL
 TensorFlow     NULL    NULL
"AWS    NULL    NULL
"Azure  NULL    NULL
"C++    NULL    NULL
"CUDA   NULL    NULL
"Excel  NULL    NULL
"FastAPI        NULL    NULL
"Flask  NULL    NULL
"GCP    NULL    NULL
"Hugging Face   NULL    NULL
"Keras  NULL    NULL
"LangChain      NULL    NULL
"MLflow NULL    NULL
"NumPy  NULL    NULL
"Pandas NULL    NULL
"Power BI       NULL    NULL
"PyTorch        NULL    NULL
"Python NULL    NULL
"R      NULL    NULL
"Reinforcement Learning NULL    NULL
"SQL    NULL    NULL
"Scikit-learn   NULL    NULL
"TensorFlow     NULL    NULL
experience_level        NULL    NULL
Time taken: 322.627 seconds, Fetched: 45 row(s)
hive>
```

Q.7 Most Popular Tools Preferred by Companies

SELECT tools_preferred, COUNT(*) AS count
FROM market
GROUP BY tools_preferred
ORDER BY count DESC
LIMIT 10;

**Insight:** Most frequently mentioned AI/ML tools.

```
hive> SELECT tools_preferred, COUNT(*) AS count
    > FROM market
    > GROUP BY tools_preferred
    > ORDER BY count DESC
    > LIMIT 10;
```

**Output**

```
Total MapReduce CPU Time Spent: 5 seconds 740 msec
OK
Internship      130
Full-time       127
Contract        118
Remote  116
Entry   62
Senior  58
Mid     48
 IQ"    9
 PL"    8
 GQ"    7
Time taken: 82.832 seconds, Fetched: 10 row(s)
hive>
```

## Q.8 Most Required Skills

SELECT skills_required, COUNT(*) AS count

FROM market

GROUP BY skills_required

ORDER BY count DESC

LIMIT 10;

*Insight:* Shows which skills appear most frequently in job descriptions

```
hive> SELECT skills_required, COUNT(*) AS count
    > FROM market
    > GROUP BY skills_required
    > ORDER BY count DESC
    > LIMIT 10;
```

**Output**

```
Total MapReduce CPU Time Spent: 4 seconds 820 msec
OK
ML Engineer     95
Data Analyst    93
NLP Engineer    87
Data Scientist  85
AI Product Manager      84
Computer Vision Engineer        83
Quant Researcher        76
AI Researcher   76
"FastAPI        75
"NumPy  71
Time taken: 52.568 seconds, Fetched: 10 row(s)
hive>
```

## Q.9 **Number of Job Postings by Year**

SELECT SUBSTR(posted_date, -4) AS year, COUNT(*) AS job_count

FROM market

GROUP BY SUBSTR(posted_date, -4)

ORDER BY year;

- *Insight:* Reveals trends in job postings over the years.

```
hive> SELECT SUBSTR(posted_date, -4) AS year, COUNT(*) AS job_count
    > FROM market
    > GROUP BY SUBSTR(posted_date, -4)
    > ORDER BY year;
```

**Output**

```
Total MapReduce CPU Time Spent: 4 seconds 510 msec
OK
NULL    2001
Time taken: 49.33 seconds, Fetched: 1 row(s)
hive>
```

## Q.10 Which industry has the highest number of AI-related job postings?

**Hive command:**

SELECT industry, COUNT(*) AS job_count

FROM market

GROUP BY industry

ORDER BY job_count

DESC LIMIT 1;

**Insight:**
This tells you which industry (e.g., Tech, Finance, Healthcare) is leading in AI job opportunities.

```
hive> SELECT industry, COUNT(*) AS job_count
    > FROM market
    > GROUP BY industry
    > ORDER BY job_count DESC
    > LIMIT 1;
Query ID = cloudera_20251029071717_119700e0-ac98-4c66-815d-81e2bb7a1415
Total jobs = 2
Launching Job 1 out of 2
```

Output

```
Total MapReduce CPU Time Spent: 4 seconds 360 msec
OK
Automotive      202
Time taken: 47.91 seconds, Fetched: 1 row(s)
hive>
```

**Q.11 Jobs by Company Size**

SELECT company_size, COUNT(*) AS total
FROM market
GROUP BY company_size
ORDER BY total DESC;

**Insight:** Job share by organization size (Small/Medium/Large)

```
hive> SELECT company_size, COUNT(*) AS total
    > FROM market
    > GROUP BY company_size
    > ORDER BY total DESC;
Query ID = cloudera_20251029071919_fc5c1fa9-03a5-4380-a6ab-09a97507ef51
Total jobs = 2
Launching Job 1 out of 2
```

Output

```
"East Carolport 1
"East Bruce     1
"East Anthony   1
"East Ana       1
"Duartebury     1
"Donnaland      1
"Daymouth       1
"Dawnmouth      1
"Davishaven     1
"Danieltown     1
"Cooperstad     1
"Collinsland    1
"Cliffordview   1
"Christianland  1
"Charlottechester       1
"Cervantesmouth 1
"Catherineshire 1
"Carrollview    1
"Campbellmouth  1
"Caitlynmouth   1
"Bryceport      1
"Bryantton      1
"Brownburgh     1
"Brianshire     1
"Brandonville   1
"Bonnieview     1
"Benjaminview   1
"Bendershire    1
"Beckerberg     1
"Bairdmouth     1
"Arnoldmouth    1
"Anthonyshire   1
"Angelafurt     1
"Andreside      1
"Andreashire    1
"Amyside        1
"Amandabury     1
"Allenchester   1
"Alexchester    1
"Aimeestad      1
"Adamtown       1
"Adamfort       1
"Maxwellchester 1
Time taken: 94.346 seconds, Fetched: 381 row(s)
```

## Q.12 Average Salary by Industry

```
SELECT industry,
    AVG(CAST(SPLIT(salary_range_usd, '-')[0] AS INT)) AS avg_min_salary,
    AVG(CAST(SPLIT(salary_range_usd, '-')[1] AS INT)) AS avg_max_salary
FROM market
WHERE salary_range_usd IS NOT NULL
GROUP BY industry
ORDER BY avg_max_salary DESC;
```

**Insight:** Highest paying industries for AI jobs.

```
Time taken: 94.346 seconds, Fetched: 381 row(s)
hive> SELECT industry,
    >        AVG(CAST(SPLIT(salary_range_usd, '-')[0] AS INT)) AS avg_min_salary,
    >        AVG(CAST(SPLIT(salary_range_usd, '-')[1] AS INT)) AS avg_max_salary
    > FROM market
    > WHERE salary_range_usd IS NOT NULL
    > GROUP BY industry
    > ORDER BY avg_max_salary DESC;
Query ID = cloudera_20251029072121_f3261c8b-ba43-43c7-92df-019ab5376c0b
Total jobs = 2
Launching Job 1 out of 2
```

Output

```
Bradford and Sharp"     NULL    NULL
Boyer and Thompson"     NULL    NULL
Boone and Stein"        NULL    NULL
Bonilla and Mclaughlin"         NULL    NULL
Bonilla and Cohen"      NULL    NULL
Bonilla and Caldwell"   NULL    NULL
Bond and Myers"         NULL    NULL
Blake and Gardner"      NULL    NULL
Black and Ferrell"      NULL    NULL
Bishop and Wade"        NULL    NULL
Bishop and Silva"       NULL    NULL
Bernard and Flores"     NULL    NULL
Bennett and Stokes"     NULL    NULL
Beltran and Lucero"     NULL    NULL
Barrett and Colon"      NULL    NULL
Barrera and Winters"    NULL    NULL
Barnett and Robles"     NULL    NULL
Barnes and Mercado"     NULL    NULL
Barnes and Howard"      NULL    NULL
Barber and Young"       NULL    NULL
Barajas and Hughes"     NULL    NULL
Ball and Tyler"         NULL    NULL
Ball and Owens"         NULL    NULL
Baker and Sanchez"      NULL    NULL
Baker and Parks"        NULL    NULL
Baker and Ortega"       NULL    NULL
Bailey and Hebert"      NULL    NULL
Bailey and Banks"       NULL    NULL
Austin and Smith"       NULL    NULL
Austin and Robinson"    NULL    NULL
Atkinson and Durham"    NULL    NULL
Arellano and Porter"    NULL    NULL
Archer and Lynch"       NULL    NULL
Anthony and Woods"      NULL    NULL
Anderson and Robinson"  NULL    NULL
Anderson and Brown"     NULL    NULL
Anderson and Barber"    NULL    NULL
Allison and Ryan"       NULL    NULL
Allen and Watts"        NULL    NULL
Allen and Washington"   NULL    NULL
Allen and Horton"       NULL    NULL
Aguilar and Jackson"    NULL    NULL
Adkins and Peterson"    NULL    NULL
Adams and Robertson"    NULL    NULL
Adams and Dominguez"    NULL    NULL
Adams and Carroll"      NULL    NULL
Time taken: 119.345 seconds, Fetched: 686 row(s)
hive>
```

**Q.13 Locations with Most Job Opportunities**

SELECT location, COUNT(*) AS job_count
FROM market
GROUP BY location
ORDER BY job_count DESC
LIMIT 10**;**
Insight: Top cities or regions for AI jobs.

```
hive> SELECT location, COUNT(*) AS job_count
    > FROM market
    > GROUP BY location
    > ORDER BY job_count DESC
    > LIMIT 10;
Query ID = cloudera_20251029072525_2dcbdedc-9b8e-4db1-a5ae-8be2a4c1e5a9
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
```

Output

```
Total MapReduce CPU Time Spent: 9 seconds 800 msec
OK
 Pandas 88
 Excel  87
 NumPy  84
 Azure  83
 CUDA   83
 PyTorch        82
 TensorFlow     80
 Scikit-learn   78
 R      77
 Keras  77
Time taken: 107.268 seconds, Fetched: 10 row(s)
```

Q.14 Count total jobs available
SELECT COUNT(*) AS total_jobs FROM market;

```
hive> SELECT COUNT(*) AS total_jobs FROM market;
Query ID = cloudera_20251029081919_16d79c66-5ac9-4bce-b810-125ea8e38cd4
Total jobs = 1
Launching Job 1 out of 1
```

Output

```
Total MapReduce CPU Time Spent: 6 seconds 960 msec
OK
2001
Time taken: 84.278 seconds, Fetched: 1 row(s)
```

## Q.15 Find the Top 10 Highest Paying Job Role

```
SELECT job_title, salary_range_usd
FROM market
WHERE salary_range_usd IS NOT NULL
ORDER BY CAST(SPLIT(salary_range_usd, '-')[1] AS INT) DESC
LIMIT 10;
```

**Insight:** Displays the 10 job titles offering the highest salary ranges.

```
ost_date, company_size, tool_preferred)
hive> SELECT job_title, salary_range_usd
    > FROM market
    > WHERE salary_range_usd IS NOT NULL
    > ORDER BY CAST(SPLIT(salary_range_usd, '-')[1] AS INT) DESC
    > LIMIT 10;
Query ID = cloudera_20251029082626_e7a92d86-16a7-4fbe-baf4-e62cfcca6bde
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
```

## Output

```
Total MapReduce CPU Time Spent: 10 seconds 410 msec
OK
Quant Researcher        Senior
Quant Researcher        Senior
Education       Scikit-learn
AI Product Manager       C++
Finance  Power BI"
AI Researcher    C++"
NLP Engineer    Senior
Computer Vision Engineer         PyTorch"
Quant Researcher         MLflow
Computer Vision Engineer         Scikit-learn"
Time taken: 79.023 seconds, Fetched: 10 row(s)
hive>
```