

DuckDB Loader

Contents

- Specifying Which Columns are Content vs Metadata
- Adding Source to Metadata

Load a DuckDB query with one document per row.

```
from langchain.document_loaders import DuckDBLoader
```

```
%%file example.csv  
Team,Payroll  
Nationals,81.34  
Reds,82.20
```

Writing example.csv

```
loader = DuckDBLoader("SELECT * FROM read_csv_auto('example.csv')")  
data = loader.load()
```

```
print(data)
```

```
[Document(page_content='Team: Nationals\nPayroll: 81.34', metadata={}),  
Document(page_content='Team: Reds\nPayroll: 82.2', metadata={})]
```

Specifying Which Columns are Content vs Metadata

[Skip to main content](#)

```
loader = DuckDBLoader(  
    "SELECT * FROM read_csv_auto('example.csv')",  
    page_content_columns=["Team"],  
    metadata_columns=["Payroll"]  
)  
  
data = loader.load()
```

```
print(data)
```

```
[Document(page_content='Team: Nationals', metadata={'Payroll': 81.34}),  
Document(page_content='Team: Reds', metadata={'Payroll': 82.2})]
```

Adding Source to Metadata

```
loader = DuckDBLoader(  
    "SELECT Team, Payroll, Team As source FROM read_csv_auto('example.csv')",  
    metadata_columns=["source"]  
)  
  
data = loader.load()
```

```
print(data)
```

```
[Document(page_content='Team: Nationals\nPayroll: 81.34\nsource: Nationals',  
metadata={'source': 'Nationals'}), Document(page_content='Team: Reds\nPayroll:  
82.2\nsource: Reds', metadata={'source': 'Reds'})]
```