

# 第一章：Databricks简介

# 什么是Databricks？

- Databricks创立于2013年，旨在通过Apache Spark实现高效的数据处理。它在现代数据科学和数据工程中发挥着重要作用。

# Databricks在数据科学和数据工程中的应用

- 数据科学应用包括机器学习工具，数据工程支持数据流、清洗和转换。

# Databricks与Apache Spark的关系

- Databricks由Spark创始人联合创建，提升了Spark在数据处理中的效率。

# Databricks的关键特性和优势

- 自动扩展、数据湖支持，能够处理多种数据格式，提升数据管理能力。