

第三章：数据集成与管理

数据源的连接与数据导入

- 支持多种数据源连接并进行数据导入管理。
 -

数据清洗与预处理

- 利用Spark DataFrames执行数据清洗和预处理。

数据存储与格式

- 使用Delta Lake、Parquet等格式保存数据，保证数据一致性和效率。

数据管理最佳实践

- 通过版本控制和数据整理，保证数据管理的可追溯性。