

**FBDQA**课程编号-01510243



# 金融数据访问

智能系统实验室

清华大学基础工业训练中心

# 数据访问

- 本地数据访问
  - 文件读取
- 云端数据访问
  - 网络访问

# 本地访问

Python高性能接口

# 本地数据 I/O

- Numpy 的 I/O

Numpy 库提供了读取和写入文件的接口

- Pandas 的 I/O

Pandas 库提供了丰富的遍历函数及方法，读取以不同格式存储的数据（例如 CSV、JSON），或将数据写入不同格式的文件中。

- PyTables 的 I/O

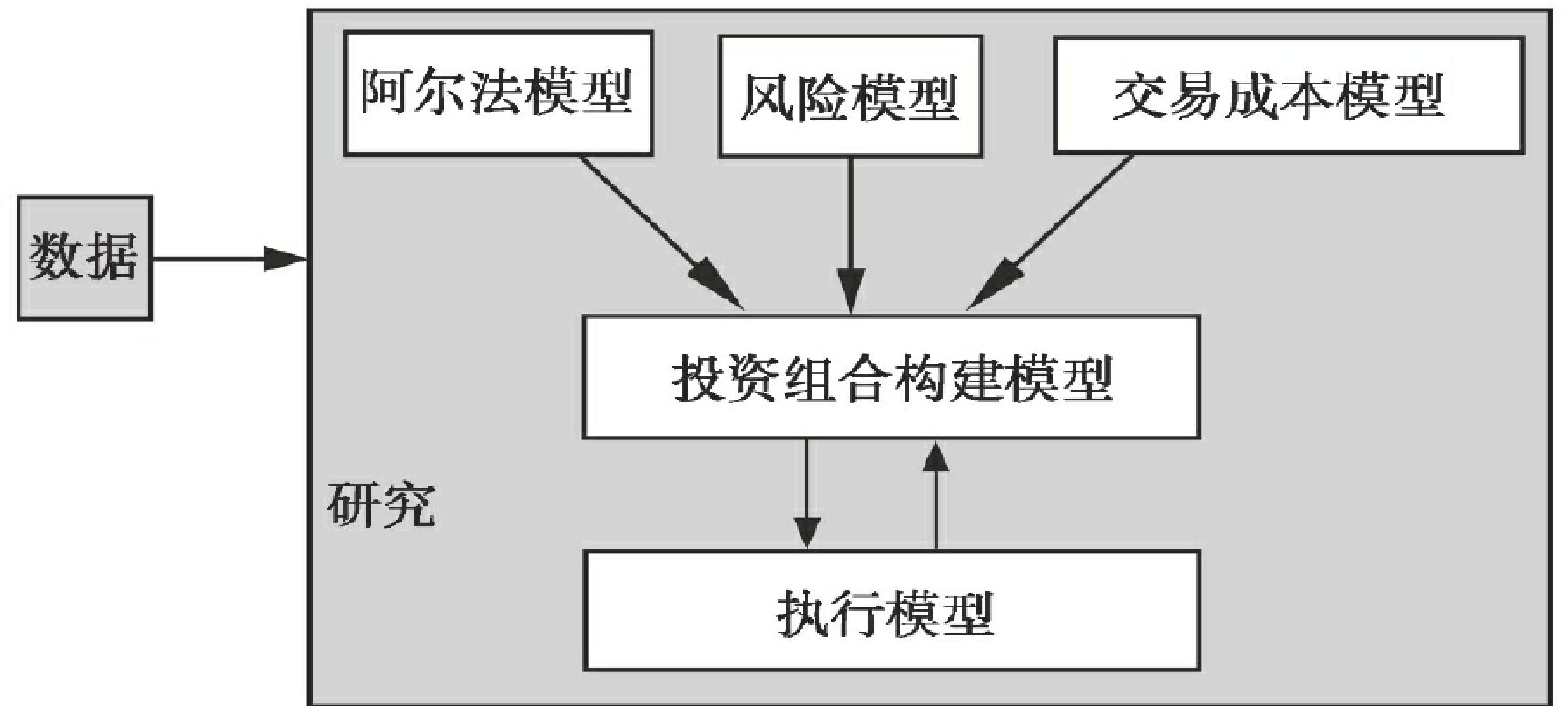
PyTables 使用 HDF5 标准实现大数据集的快速 I/O 操作；其速度往往只受到所使用硬件的限制。

# 云端访问

服务器端

# 数据系统

- 数据系统是交易系统的基础
  - 数据来源多样，格式不一
  - 数据质量差异，数据清洗
  - 数据修正（前视偏差 lookahead bias）



# 大数据系统

- 数据管理语言
  - SQL数据管理语言
- 数据系统软件：
  - PostgreSQL数据库，简称pg
  - 基于pg集群的Greenplum分布式数据库，简称gp
- 数据分析
  - 数据分析工具介绍
  - Greenplum MAD-lib

# 数据管理

- 过去：
- Flat File 带索引的文件（专用）
- 数据系统的形式
- 数据库形式（云端）
  - 关系型数据库（RDBMS）（通用）
  - 数据立方（Data Cube）

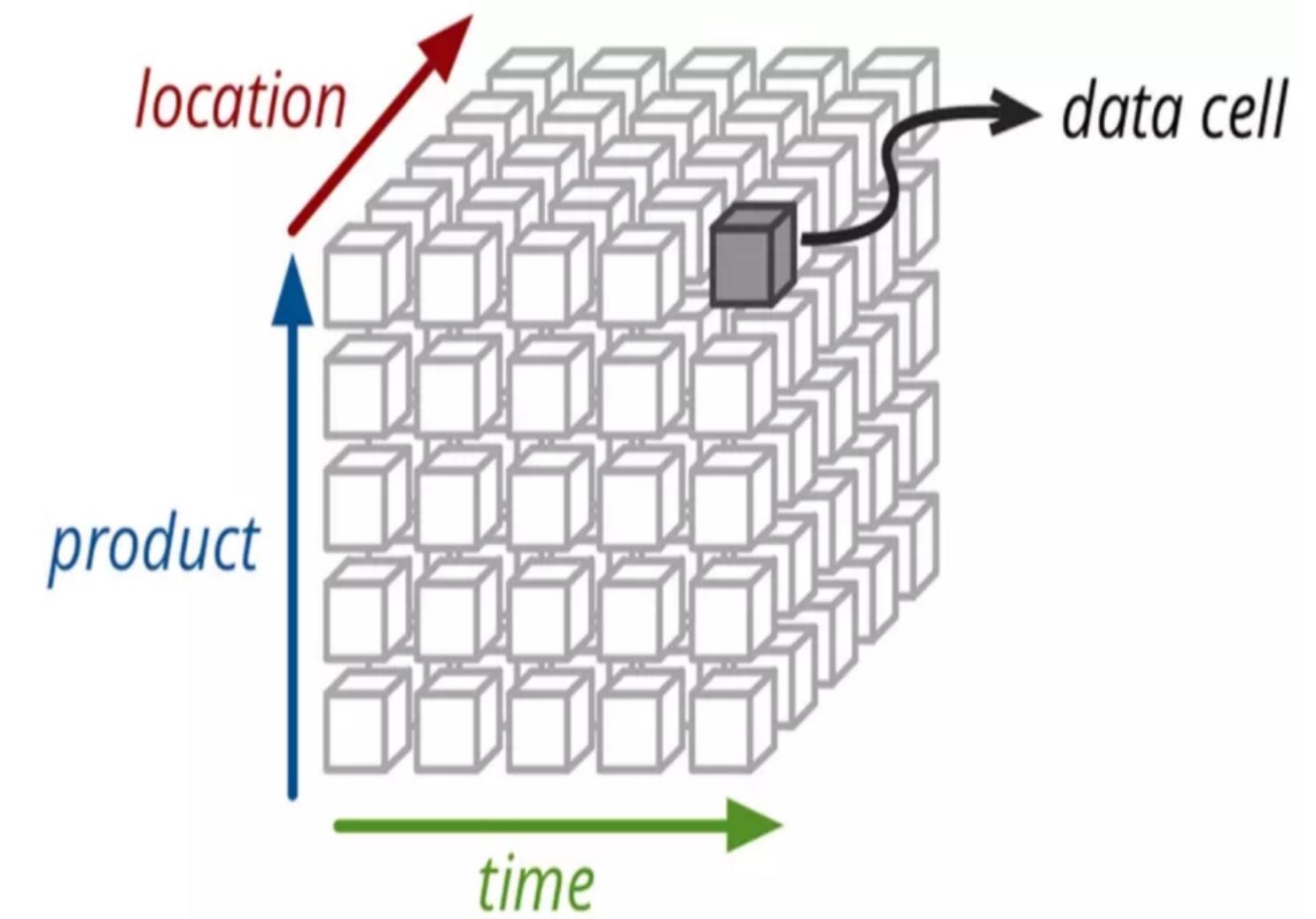
# 基于SQL的关系型数据库

# 关系型数据库 (RDBMS) -最常用

- 统一的数据库管理语言 : SQL(Structured Query Language)
- 关系型数据库软件
  - MySQL: <https://dev.mysql.com/doc/refman/8.0/en/sql-statements.html>
  - PostgreSQL: <https://www.postgresql.org/docs/12/sql.html>
  - SQLite: <https://www.sqlite.org/lang.html>
- 数据库排名 : [https://db-engines.com/.](https://db-engines.com/)

# 数据立方-一种数据管理系统

- 数据立方体(Data Cube)，把所有金融产品所有属性的取值放在一个三维（更多维）数据表
  - Data Cube : 时间、地点、产品
- 比如对某个日期，所有的金融产品，产品的给定取值属性（收盘价），其他属性（当日每股收益）
- Apache Kylin
  - <http://kylin.apache.org/cn/>



# SQL实 践路 线

SQL统一的数据管理语言

# 本地SQLite访问使用

- SQLite使用量最大的本地SQL数据库
- Python 操作SQLite学习
- 熟悉SQL语言的语法和操作

# 云端数据访问

- Python—SQL—RDBMS
- Greenplum分布式数据库集群
- 内部搭建的实验环境

# 参考书目

- [德] 伊夫·希尔皮斯科 (Yves Hilpisch) 著, 姚军 译, Python金融数据分析, 人民邮电出版社, 2015.
- 第7章-金融数据高性能IO-Python.

# 参考书目

- 《打开量化投资的黑箱》 Inside the Black Box: A Simple Guide to Quantitative and High-Frequency Trading, 作者：里什 纳兰 (Rishi K. Narang), 出版社: 机械工业出版社; 第1版 (2016年5月1日)
- Chapter 8. Data

谢谢指正！