

# Introduction to Data Used in Workshop

Fethi Rabhi

*School of Computer Science and  
Engineering*

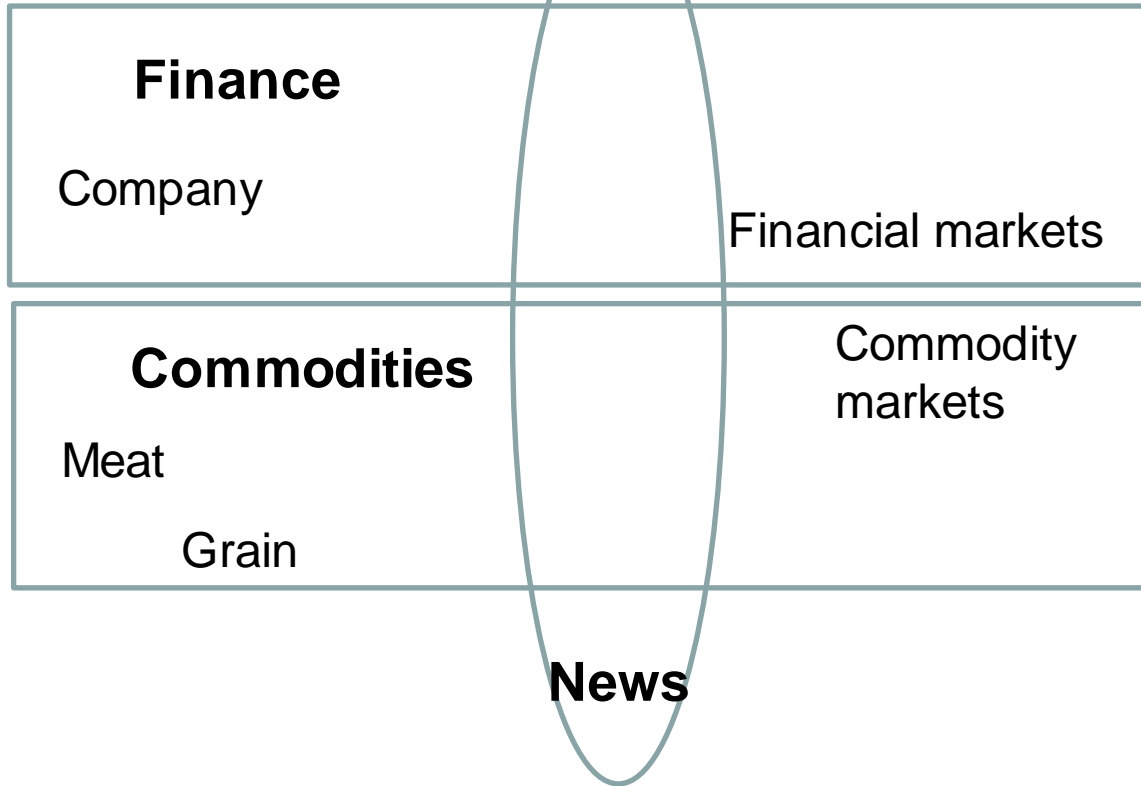
University of New South Wales  
Sydney Australia



# Preliminaries

- Purpose of lecture
  - Look at domains involved in case study
  - Look at different types of datasets that will be part of workshop
- Before that, we need to:
  - Understand the concepts behind the data
  - Understand the processes and context in which data was generated

# Domains involved



# Part 1: Finance

Basic concepts:

## **Company**

Entity that requires funding, usually listed company

## **Equity**

A share of that company, can be bought or sold

## **Company code**

A unique identifier that represents the company when processing data

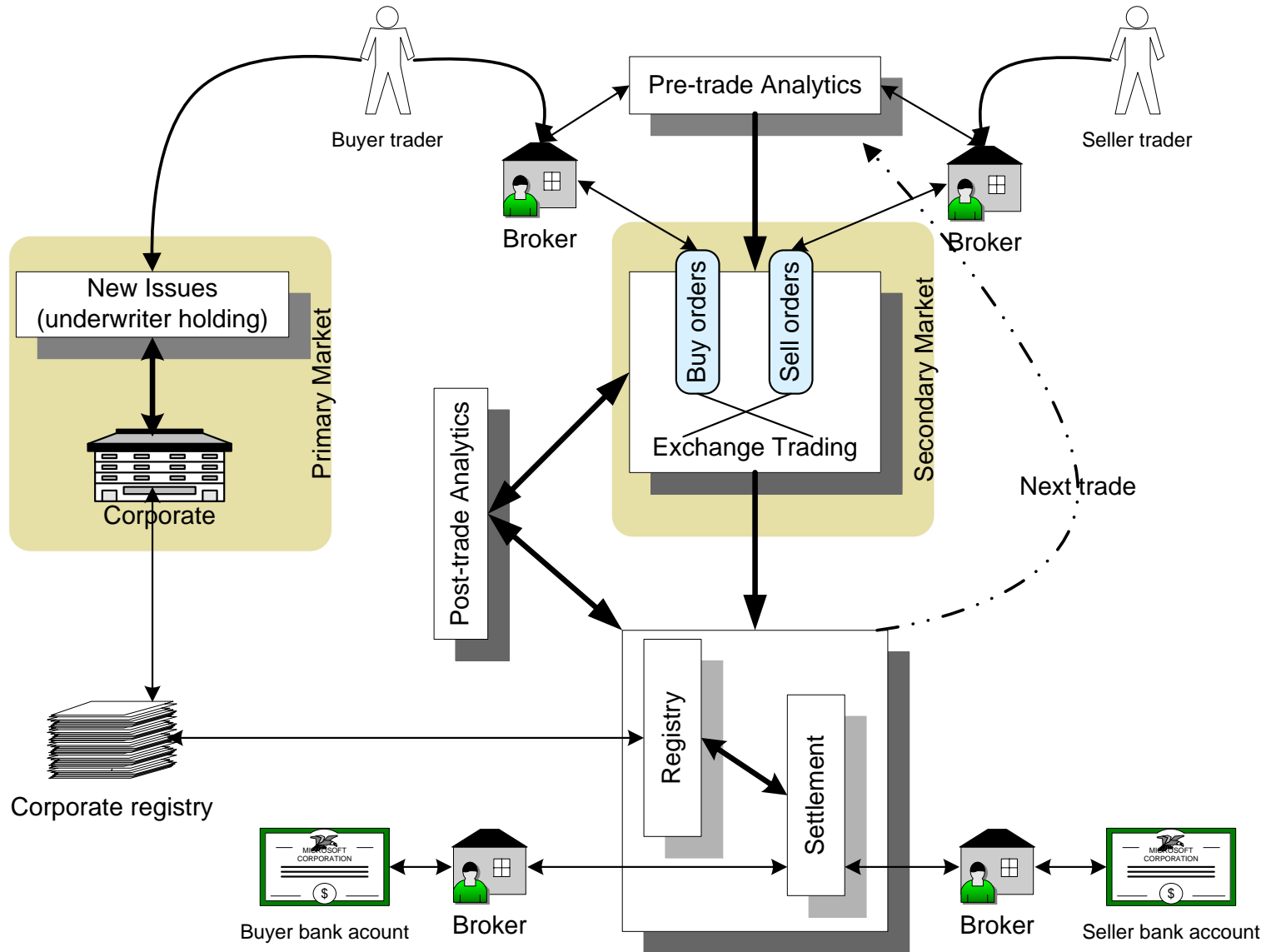
## **Derivative**

Contracts on future price of company (options, futures)

## Part 2: Financial markets

- Characteristics
  - Allow the buying and selling of financial instruments
  - Financial instrument = equity or derivative
  - Most financial markets trading occurs through electronic exchanges
- Financial market trading cycle
  - Pre-trade analytics
  - Trading
  - Post-trade analytics
  - Settlement and Registry

# Financial Market Systems



# Part 3: Commodity markets

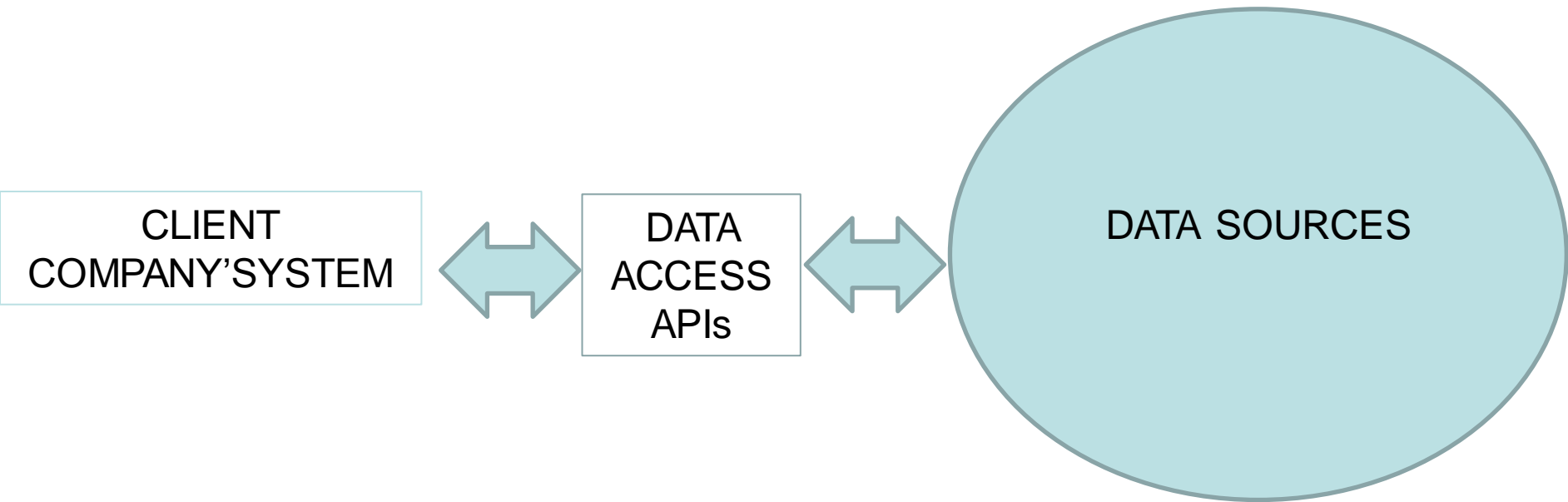
- Characteristics
  - Allow the buying and selling of commodities
  - Divided into different categories
  - Commodities are traded via exchanges or OTC
- Commodities trading cycle
  - Similar to financial markets cycle
  - Involves physical goods delivery
  - Price depends on logistics, weather etc.

# Part 4: News

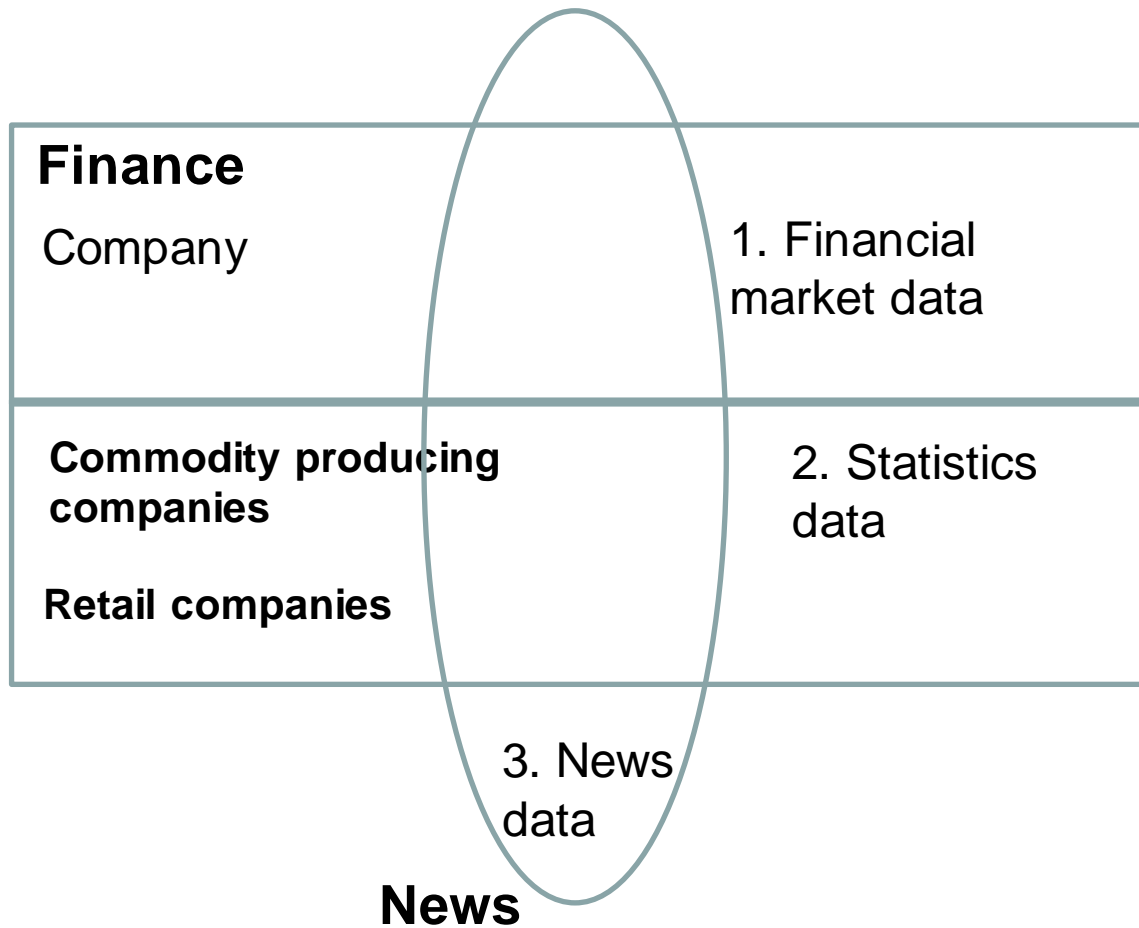
- Characteristics
  - News about companies, trading, commodities etc.
  - News about economic conditions, interest rates, regulations, indicators etc.
  - Political news
- News cycle
  - News sources
  - News aggregators
  - News analysts
  - News distributors



# Context



# Datasets involved



# 1. Company trading data

- Trading data is a source of knowledge about
  - Company's performance
  - Trends and patterns
  - Impact of news
- Examples of commercial data providers
  - Real-time
    - Thomson Reuters
    - Bloomberg
  - Each provider can maintain multiple *datasets*
- Recent years
  - Rise of social networks
  - Open data initiative
  - More data available free

# Yahoo Finance

- Instrument naming scheme
  - Yahoo Instrument Code
- Asset classes
  - Extensive coverage
  - All worldwide exchanges
- Coverage
  - End of day data

## 2. News data

- Thomson Reuters News
  - Access will be provided via a CSE server
  - High frequency data, each message containing a piece of news
  - Different types of items: alerts, headlines, body
  - Has a timestamp
  - Has Company Codes (RICs)
  - Has Topic Codes
  - Text of the news
- Other
  - Students are free to use other news sources
  - Must be good quality and contain company-related information

### 3. Statistics data

- From Australian Bureau of Statistics (ABS)
- Retail data
  - Turnover by industry groups
  - Region-specific information
- Commodities data
  - Merchandise exports
  - Region-specific information

# Summary of Learning Outcomes

- Learning a new domain
  - Finance and commodities trading
  - News processing
  - Economic data
- Application development
  - Building a data analysis dashboard for traders
  - Incremental building (from API to application)
  - Exploiting publicly available data (teams free to use additional data sources)
- Software Development Skills
  - Business Analysis
  - API design and Implementation
  - Coding and Testing
  - Software Reuse