User-Oriented Modelling of Scientific Workflows for Big Data Analysis

Aarthi Natarajan
PhD Student, UNSW
Big Data: Meet the New Boss

- Massive information streams from different sources
- Information is critical to make wise decisions
- Volatile business context, change is constant
- Ability to respond and anticipate on market-trends

Financial Markets:
- Measure volatility, correlation - pricing and risk management strategies

Patient Health Monitoring Systems:
- Prompt analysis, Detect patterns, Make inferences

SMART SERVICES CRC
Big Data is a blessing and a curse ...
Research Interest…

**What** is our research focus?
- Temporal Events - A subset of Big Data dealing with high frequency event streams

**Why** temporal events?
- A time of change and opportunity

**How** do we analyse this data today?
Many technologies..... Great promises, but many drawbacks
Proposed User Perspective

- User-defined conceptual analysis models
  - Collaborative analysis
  - Focus on science
- Incremental Adoption
  - Can use old tools
  - Use pre-packaged functions
  - Seamlessly integrate multiple tools
- Template based approach
  - Re-run pre-existing models
  - Consistent reproducibility of analysis results
  - Adapt & Customise process templates to build new models
How do we realise this vision?

- Propose a hybrid solution
  - Scientific Event Processing (SEP)
  - Uses Workflow-like structures to deal with events

- Define a new methodology and architecture
3 step process to conceptual modelling

1. Define Abstract Service Components
2. Compose abstract Event Processing Network
3. Map EPN to concrete workflow

A SCA Based Architecture

Conceptual Model

Abstract Process Template

Concrete Model

ABSTRACT SERVICE COMPONENTS

EXPORT SERVICES

PROCESSING SERVICES

IMPORT SERVICES

CONCRETE SERVICES

CEP
BPEL
MATLAB
WEB SERVICE
JAVA

Executable Process
Applying SEP to Financial Market Domains Market
Various Sources of Financial Market Data and News

• Market data providers
  – Thomson Reuters, Bloomberg
• Regulators
  – RNS, EDGAR
• Printed media
  – Financial Times, Wall Street Journal
• Newswire services
  – AP, AFP, Dow Jones newswires, PR Newswire, Business Wire
• Aggregators
  – Factiva, Google Finance, Yahoo! Finance
Case-Study:

Detecting Abnormal Returns on Stock Data
User Defines Conceptual Model

User defines conceptual model

Abstract Event Processing Network

Mapping to a complex, concrete workflow

Select Event Source → Detect Abnormal Price Jumps → Visualise Price Jumps

TRTH Import (stock data) → Time Series Building → TRTH Import (index data) → Time Series Building → Merge (stock & index time series) → Price Jump Detection → Visualization (as chart)

SMART SERVICES CRC
User Visualizes Analysis Results
Future Work

• Short term research
  – Formalise data model for SEP framework
  – Build an initial prototype
  – Expand range of services
  – Investigate user-driven composition
  – Make tools available to financial analysts / academics

• Long term research
  – Apply framework to other eResearch domains / business analytics
  – Integrate cloud infrastructures
  – Apply semantic capability to assist service discovery
Acknowledgments

• Supervisor: Prof. Fethi Rabhi
  – f.rabhi@cse.unsw.edu.au
  – http://soc.cse.unsw.edu.au/

• SIRCA