

I~~P~~RO 313

Spring, 2008

Ultra-High-Speed Market Data Ticker System

Project Plan

Advisors Wai Gen Yee,
Ben Van Vliet

Sponsors ~~Townsend Analytics~~ Illinois Institute of Technology

I~~P~~RO Team Khanh Duong
Jose Acuna-Rohter
Tarun Anupaju
Lance Cooper
Martin Kolodziej
Konstantin Roytman
Oluwaseun Shonubi
Jing Kai Tan
Jong Su Yoon

Illinois Institute of Technology

February 22~~18~~, 2008

1. Objectives

The objective of IPRO 313 is to create a high performance data ticker plant for our sponsor Townsend Analytics which needs to meet or exceed certain performance requirements. Ideally, The the data ticker plant has to have a sustained optimal throughput of three million price quotes per second – the current industry state of the art - and minimize with minimal latency while maintain specific constraints. Toward this end, the team will create a proof-of-concept data ticker plant that processes real data. The initial system will be used as a baseline on which optimizations to specific components will be made. The ticker plant aggregates streaming data for numerous global financial markets and disseminates the data to thousands of users in real time. The data is used in Townsend Analytics' RealTick® Execution Management System (EMS), its flagship institutional product for the financial services industry.

IPRO 313, Spring, 2008, builds on work done in the previous semester. Last Fall, IPRO 313 conducted background research on the market data systems and outlined a basic design and performance metrics for our system. Our goal with IPRO 313 during the Spring 2008 This semester, we will –is to resume system development, implement and work on performance optimize a basic market data ticker system improvements, enhanced functionality, and more detailed benchmarks. Using past semester result, research and development, the group will refine the design, prototype development and benchmark testing. The intent is to also provide detailed performance metrics of the system. This will enable future optimizations and algorithmic enhancements to the system in order to increase system throughput while minimizing latency and variance.

The team has set forth the following objectives:

- Explore competitors' solutions and available technology
 - Know what is currently on the market
 - Know what new technology is available for implementation
 - Understand what works and what does not
- Develop a functioning ticker plant system
 - Analyze ticker protocols used in previous semester
 - Redesign and refine the ticker plant architecture
 - Code a working small system
- Improve system performance
 - Identify performance bottlenecks in the system
 - Design and implement performance optimizations
 - Record performance improvements
- Determine hardware requirements
 - Test off-the-shelf hardware for system
 - Design custom hardware configurations
 - Compare each solution
- Update the technical user manual
- Create a website contain every information regarding this project Web site?

2. Background

With the dawn of every New Year, the speed of business is ever increasing. What used to be performed in months, days or minutes is now being done in milliseconds; and even that isn't fast enough! For businesses to stay competitive within their industries, many will businesses today – particularly in finance – need real time to transition to access to real time data large volumes of data management. This requires vast improvement in information technology infrastructures infrastructures will need to be vastly renovated and millions of dollars will be put into preparing for the huge amounts of data going through the system. The question is: will it be worth the hassle?

Benefits of Real Time Data Integration

Better decisions:

With a more dynamic view of information as it flows in and out of the business, decision makers can better understand business processes than if the information is provided in time-lagged reports

Faster response to material events:

Immediate awareness of changes to acceptable business conditions or to the presence of material events removes latency from business processes for a more agile enterprise.

Unwired operations:

The ability to alert and notify key individuals to changes within the business enables greater flexibility in where business can happen.

Proactively publish data at lower cost and less complexity:

Move data events immediately from multiple databases directly to a messaging infrastructure without the cost of custom coding or changing existing applications.

Preserve operational performance:

When capturing events from database systems, there is zero impact on the source database, preserving operational performance without incurring additional system overhead costs.

Complement existing messaging infrastructures:

Captured events are standardized and distributed to business applications throughout the enterprise via a standards-based messaging infrastructure without coding custom interfaces.

Manage information flows with flexibility:

Sybase Real-Time Data Services facilitates how customers can choose to flow data events through the organization such as enriching events with analytical or historical information before arriving to the decision-maker without slowing down operational system performance.

Financial Industry – Quick Intro and Classical Approach

Trading is the voluntary exchange of goods or service and it exists due to regional comparative advantages and specialization/divisions of labor. Trades are based on an auction market paradigm where a potential buyer bids a specific price and a potential seller asks a specific price:

Classic Trade Flow

Step 1: Decision

~~Portfolio manager uses research, fundamentals and historical market data to make investment decision~~

~~Step 2: Place Order~~

~~Buy-side trader calls the sell-side trader and communicates trade over the phone~~

~~Step 3: Interact with market~~

~~Sell-side interacts with the marketplace or takes on position~~

~~Step 4: Relay Trade~~

~~Sell-side calls buy-side trader with execution information.~~

~~Step 5: Enter in Ledger~~

~~Buy-side books the trade on the position ledger.~~

Introduction of Electronic Trading

~~In~~ The 1975 Security Acts ~~1975~~ Amendment, enacted by Congress in 1975, ~~there was request to~~ gave the SEC the authority to create an efficient national market infrastructure (read electronic trading technologies). ~~encourage maximum reliance on computer and communications technologies.~~ The reason for this was to eliminate as much of the redundant paper work that slowed the trading process down as possible. The implementation of this however also resulted in considerably faster access to market data, by the millions! As a result, people had access to substantially larger volumes of data with significantly less latency.

The new trading infrastructure changed trading both quantitatively and qualitatively. Quantitatively, much more market data are available with lower latencies. A March, 2007 report by the Tabb group indicated that the number of messages transferred per day in global options and equities markets has been growing by 139% annually from 2005 and is projected to continue until at least 2010. Furthermore, volatility in the market has decreased as any sudden "spikes" in the market can be handled very quickly with the new market technologies.

According to TowerGroup, a research firm, \$480m is likely to be spent in America this year on developing technology for algorithmic trading. In 2005, providing fastest data to customers had such an impact that the financial industry increased spending on computers and software to \$26.4 billion and in the past years, the compound annual growth rate for algorithm use from 2004 through 2007 was projected at 34%. Such is the focus on speed that even location counts. Servers positioned nearest to a trading venue can shave milliseconds of the timing of a trade and get a better price.

What can anyone possibly do with information arriving in such quantities and speeds on these "fire hose" data feeds? The obvious answer is that people alone cannot do anything with that much information. The receiving end always involves a computer. The kinds of tasks we can do with computers have grown so incredibly that, to use them effectively, we have to change the way we think about them.

By severely reducing the amount of human intervention, electronic trading has radically changed the nature of the classical trade flow.

Trading Is Automated So Where Is the Hassle?

Automated trading is, automated. Within one second between 5,000 and 50,000 changes can occur. That means that within one year, terabytes of data will need to be handled by specially prepared warehouses full of zero-latency databases. A sudden transition of an ordinary warehouse of database to the different style of processing required, data volume, velocity, and rate of change required for real time processing would certainly cause a database overload. Gradual shifts need to be made for a healthy implementation.

Townsend Analytics

Townsend Analytics (TAL), a direct-access trading system vendor, provides connectivity to a multitude of electronic communications networks and stock exchanges:

Servicing the global capital markets for over 20 years, Townsend Analytics Trading Services offers world class trading and trade order management solutions to the institutional-portfolio management and broker-dealer marketplace. In today's competitive institutional trading environment, portfolio managers and traders are under increasing pressure to access a wider variety of liquidity sources, employ more aggressive yet cost-efficient trading strategies and achieve best execution faster than ever. Achieving this demands the right service bureau partner, powerful order management/reporting tools and direct access to a variety of electronic markets across multiple asset classes.

TAL has already developed a means for managing real time data which has captured both the United States market as well as the European market:

Powered by RealTick®, Townsend Analytics' flagship institutional product, RealTick EMS (Execution Management System) is the institutional financial service industry's leading multi-asset, multi-broker, multi-routing and multi-regional market data, analytics and direct market access (DMA) trading platform.

Electronic Trading Market

Current Market Competitors

There exist several vendors of the technologies necessary to participate effectively in this industry, including Townsend Analytics, Exegy, Reuters, Bloomberg, RMDI and Wombat are examples of such. These companies provide various services, such as real-time price data as well as market analyses. Their current generations of systems, however, are incapable of handling the impending mountains of data.

Market Value

According to TowerGroup, a research firm, \$480m is likely to be spent in America this year on developing technology for algorithmic trading. In 2005, providing fastest data to customers had such an impact that the financial industry increased spending on computers and software to \$26.4 billion and in the past years, the compound annual growth rate for algorithm use from 2004 through 2007 was projected at 34%. Such is the focus on speed that even location counts. Servers positioned nearest to a trading venue can shave milliseconds of the timing of a trade and get a better price.

Ultimately, because the need for speed of information will only continue to grow, if TAL can develop a unique and reliable solution, it will provide a vital product for a demanding future market.

3. Methodology/Brainstorming/Work Breakdown Structure

A. Defining the problems

Development team

Developing software based on the architecture laid out by the previous team to create an efficient ticker system. Necessary tasks include:

- Redesign the data generate to also read real world data and pass it along for testing
- Developing head end component
- Implement different hashing algorithm
- Improve the last value cache implementation
- Develop a data distribute capable of handling multiple clients

Research/Optimization team

Look at the base system developed, and then have it optimized for maximum reliability and performance.

- Compare the different hash functions present the performance of alternative data access structures
- Analyze the behavior of real market data, to find one that best suits our needs
- Find the right hardware solutions
- Handling system failures

IPRO/Web design team

- Research effective website design
- Collect information from other sub-team
- Development an information website about this project

[SHOULD WE SAY SOMETHING ABOUT THE WEB TEAM?]

B. Defining how the team will go about solving those problems

1. *Understanding the requirements of the IPRO*

During the first three weeks of the semester, the team conducted background research and conducted receives presentations on on the project goals, project background, past results and IPRO requirements. Based on these presentations, the team and discussed defined the project objectives and execution of our system approaches to achieving these objectives and allocated resources to various tasks.

2. *Initial Organization*

The team initially looked at the majors of everyone on the team and assigned roles to them according to the relevancy to their respective degrees, experiences and interests. The team then formed 3 sub teams: Development, Research/Optimization and IPRO/Web Design. Each of the sub groups had a leader to ensure that everything progressed in a timely manner. The team assigned officer responsibilities such as appointing Project Leader, Minute taker, and Master scheduler on February 4th, 2008. During the initial meetings the team defined the problem and how to approach it using various technologies (different hashing algorithms, different hardware, and different software development methodology). Along with these needs, we identified the information necessary such as real world data.

3. *Developing the project plan*

The team prepared a project plan/methodology to implement, improve and optimize a basic market data ticker system and to create a website for IPRO 313. Each team contributed requested material and the IPRO/Web Design team compiled the information into the Plan. We used MS Project, iGroups, and MS Word to organize the information

4. *Assigning individual team member to sub-groups and tasks*

Please refer to section 7 and section 8 for details

5. *Implementing a ticker system*

Based on the architecture laid out by the previous semester team, the Development team will develop different components and implement a working ticker system.

6. *Performing research*

While the Development ~~team~~ Team works the system with current available solutions, the Research/Optimization team will continuously perform research using industry journals, and different sources to find better solutions as well as to improve the current implementation. They will work closely with the Development Team to ensure that the research is relevant and compatible with the system being developed.

7. Review the progress

Each sub-team will give an initial design plan and subsequently, present a bi-weekly progress reports to the rest of the class.

8. Test the product and perform analysis

The team will perform tests on the system being developed using real world data as well as randomly generated synthetic data. To do this in a scientifically rigorous way, the team will develop a standard set of benchmark tests. Results will be compiled and included in the final report. Using the test result, the team will identify bottlenecks and specific problematic areas that can be worked on to improve performance. Performance bottlenecks in the basic system will be identified and drive further efforts.

C. Analysis/Testing

Development team

All system components will undergo thorough unit testing by each engineer through every stage of the development process. This combined with peer code review and different variations of white box and black box testing will comprise the testing strategies used during the development process. Instrumentation will also be a necessary part of the system in order to provide insight into runtime performance of the system.

Research/Optimization team

Testing will be done by physical implementation of the code against a controlled set of variables, e.g.: to find the best hash function, real data will be used, to be hashed and will then be checked for collisions and also the time taken to create the hash table. Furthermore, analytical models of real data will be created.

D. Documentation

Each process teams will maintain a record of their respective goals, bench mark data, and specific questions that they should ask to test their output. The actual testing and result report could be performed by other peer process groups or a third party to minimize the biases in the analysis. To keep the team abreast of the progress, each sub-team will give an initial design plan and subsequently, a bi-weekly progress report. The goal of these progress reports is to timestamp progress and to solicit input from other team members. Results will be compiled in the final report.

Further documentation included a technical document and a user document. The technical describes design decisions and is appropriate for an engineer who wishes to technically evaluate or further modify the system. The user document is intended for the end user who wishes to use the system.

E. Analysis of the expected results

The expected results will be analyzed in two major ways:

1. How the system performs. Assume maximum data flow, how much data can the system process in a set period of time. By testing using real world data from OPRA, we should will have a measure of see how system effectiveness the system is. Graphs and charts will be constructed to give an easy overview of the effectiveness of the system. Our basic metrics are average message throughput and latency. The base case system, pre-

optimization, will provide baseline performance. The baseline system is derived from the system designed and implemented in last semester's IPRO. We expect performance with individually optimized components as well as the integrated system to improve in terms of both metrics. Metrics will be recorded on a standard hardware platform.

2. How the system perform over extended period. The system need to remain bug free (at least in major components) and capable of processing data over an extended period of time.
3. Consistencies of data model. For the analyses of the real data, we expect that our models are consistent with time and accurately represent real data.

Regarding research results: research articles, with the help of class presentations will be used as evidence to back up opinions voiced.

F. Production of IPRO Deliverables

The IPRO deliverables will be constructed in the following matter:

1. Software artifacts: One important A major deliverable will be the actual is the code for the developed program system that was fine-tuned this semester. Code will be documented. The Development Team will lead in this.
2. Documentation: Another important deliverable will be the Technical and user documents will be prepared by the respective teams. Technical-These documents detailing how to use the system as well as the inner working mechanism so that a financial company could easily utilize the system to their advantage as well as improving it.
3. Benchmarks: Developed benchmarks should be applicable to future improvements of the system as well as general market data systems.
- 3.4. Web Site. One extra deliverable compare to last semester would be the IPRO 313 Website, which will details all the information regarding the project.
4. Although the content of each deliverable will be created by the respective teams, The the production of the final IPRO Deliverables will be handled by the Web/IPRO Team who are responsible for collecting data and making IPRO deliverables

G. There are no documents that need to be attached

4. Expected Results

The goal of IPRO 313 is to produce a data ticker plant which needs to meet or exceed certain performance requirements. The ultimate goal is to have it exceed the performance of current state-of-art ticker plant which can handle 2 millions updates per second.

Development team

The results that this IPRO will generate include a working data ticker plant with reasonable performance on real world data. Applied research and data gathered, such as technology, industrial standards, criteria and financial values will also benefit to the enhancement of the current architecture. Testing on the other hand, will allow us to work with industries and quickly identify issues and required modification. All gathered comments and suggestions will benefit to the improvement of the system and can potentially increase its performance.

The Development team will be implementing the system based on the architecture from the previous semester IPRO. The Research/Optimization ~~team~~ Team will also collect data from

the existing model to assist the Development ~~team~~Team. . The expected deliverables that we are planning to produce are:

- working market data system
- user manual containing instructions on how to use the system
- technical manual detailing system design
- suite of metrics with which we test each component or the entire system
- compilation of experimental results and analyses

Research/Optimization team

The goal is to provide the software team with an optimized solution for the system, which then they could implement. The ~~research~~Research/e~~Optimization~~ ~~team~~Team will look at various possible ways to make the system reliable and also see to it that the system stays up to date (i.e. keep with the increase in load data that can be seen as time progress, and also advances in hardware technologies.)

All information gathered in the research phase will be also available for potential customers. With the successful completion of the mentioned above activities and results from them IPRO 313 will be proud to benefit the financial industries.

Although our major goal is to create a stable system, we also expect it to have reasonable performance. Quantitatively, based on the literature, we see room for significant improvement. A report by Google Labs, for example, indicates that some recent data access structures improve on the current state of the industry by up to 30%. We expect to see performance improvements of this magnitude for certain components. Performance for the entire system, however, may be limited by system bottlenecks.

IPRO/Web Design Team

The result should be a functional website that contain all the useful information related to the project. The website should be user friendly and light. It should also be made easy to update any portion of the website.

[Web site? You might say something to the effect of – we expect it to be easy to use and have tons of useful information. We will check with our sponsors to see if the information is okay.]

5. Budget

IPRO Day	\$300.00
Miscellaneous	\$200.00
TOTAL	\$500.00

6. Schedule of Tasks and Milestone Events

Development team

	Task	Start Date	Target end date	Lead
	1. Data Generator	2/06/08	4/12/08	Kenny
1.1	Obtain OPRA data.	2/06/08	2/08/08	
1.2	Implement code to stream OPRA data into RAM	2/06/08	4/15/05	
1.3	Implement code to take OPRA data from RAM and send to network sockets	2/06/08		
	2. Head End	2/06/08		Jose

2.1	Create a Decoder for OPRA FAST data	2/06/08	3/20/08	
2.2	Create a template for our data to send out	2/11/08	2/20/08	
2.3	Create UDP socket connection to receive data from CD			
2.4	Get timing metrics on UDP receiving speed and CODEC speed			
	3. Last Value Cache	2/06/08		<u> Jong Su</u>
3.1	Provide ability to use different hash algorithms	3/03/08	3/10/08	
	4. Data Distributor	2/06/08		Lance
4.1	Separate subscription from LVC	2/12/08	3/12/08	
4.2	Interface subscription with Head End	2/18/08	2/25/08	
	5. Client Application	2/06/08		Jong Su
5.1	Convert into c++.	2/10/08	2/17/08	
5.2	convert monitor application into c++	2/17/08	2/24/08	
5.3	implement hashing functions in c++ (superfast, CRC32, adler, tiger)	2/25/08	3/02/08	
5.4	plug hashing functions into LVC module			
			Baseline End Date	5/2/08

Research/Optimization team

	Week	
1	2/17/08 - 2/24/08	- In depth view of ADLER32 and CRC32
2	2/24/08 - 3/02/08	- Get code to decode OPRA data - Test hash functions using the metrics chosen(time &reliability) - Select a hash function
3	3/02/08 - 3/09/08	- Implementation of the chose hash function in to the LVC - Possibly, presentation on partitioning data sets
4	3/09/08 - 3/16/08	- presentation on Distributed and parallel processing - Hardware support
5	3/16/08 - 3/23/08	- Presentation on hardware support – necessary hardware structure? - Analysis of real workloads
6	3/23/08 - 3/30/08	- Presentation on analysis of workloads - Handling system failures
7	3/30/08 - 4/6/08	- presentation on handling system failures

IPRO/Web Design team

Week	Task	
02/17/08	Finalize project plan and Code of Ethics deliverables	- <u>Starting web development</u>
02/24/08	Midterm presentation creation	- <u>Establish basic webspace</u>
03/02/08	Collect materials for Midterm Report deliverable	- <u>Research options and design</u>
03/09/08	Finalize and reivise midterm report	- <u>Layout</u>
03/16/08	Web Development	- <u>Basic text content</u> - <u>Layout</u>
03/23/08		- <u>Basic text content</u>
03/30/08		- <u>Graphics</u>
04/06/08		- <u>Finalize and ready for updates from other team</u> _[wgy1] - <u>Final adjustments and updates</u>
04/13/08	Collect materials for IPRO Day and Final report	
04/20/08	IPRO day materials ready	Power point, poster, information booklet, CD
04/27/08	Presentation practice	

Detailed information available upon request

7. Individual Team Member Assignments

7.1 Overview

Name	Role	Major	Other
Khanh Duong	Master schedule maker	CPE	Web design experience
Jose Acuna-Rohter	Minute taker	CS	IPRO experience
Tarun Anupaju	Agenda maker, Time keeper	<u>CPE</u>	
Lance Cooper		<u>CS</u>	<u>C++</u> <u>Web design experience</u>
Martin Kolodziej		<u>EE</u>	Web design experience
Konstantin Roytman		<u>CS</u>	
Oluwaseun Shonubi	Team leader	<u>EE</u>	
Jing Kai Tan		<u>EE</u>	
Jong Su Yoon		<u>CS</u>	Project experience
Wai Gen Yee	Professor	Prof. es	
Ben Van Vliet	Advisor	Advisor	

7.2 Sub-teams

Development (programming) Team		
Developing the software and hardware for the system		
Name	Role	Responsibilities
Konstantin Roytman	Sub-team leader	Overall, data generator
Jong Su Yoon		Client application
Lance Cooper		Data distributor
Jose Acuna-Rohter		Head end

Research/Optimization team		
Look at the base system developed, research for solutions to improve and optimize the system for maximum reliability and performance.		
Name	Role	Responsibilities
Tarun Anupaju	Sub-team leader	Overall
Jing Kai Tan		Assist and research

IPRO/ Web design team		
Responsible for handling the creation of IPRO deliverables and the IPRO 313 Website		

Name	Role	Responsibilities
Khanh Duong	Sub-team leader	Overall
Oluwaseun Shonubi		Deliverable research
Martin Kolodziej		Web development

|

8. Designation of Roles

Team Leader:

Oluwaseun Shonubi

Minute Taker:

Jose Acuna-Rohter

In charge of recording decisions made during meetings including task assignments or changes under consideration.

Agenda Maker:

Tarun Anupaju

Responsible for creating an agenda for each team meeting. This provides structure to the meetings and offers a productive environment.

Time Keeper:

Tarun Anupaju

Responsible for making sure meetings go according to agendas.

Master Schedule Maker:

Khanh Duong

Responsible for collecting schedules from all the team members and developing a master schedule, which tells the team when the members are available and how to contact them.

References

J. Johnson, Trading at Light Speed. Technical Report, Tabb Group, March, 2007

P. Hsieh, Hash Functions. Web Document, Accessed 2/20/08.
<http://www.azillionmonkeys.com/qed/hash.html>