Proposed CAT Solution

SIFMA CAT Bidder Summit

July 29th, 2014

Manoj Narang
CEO and Managing Partner
Mike Beller
CTO and Managing Partner
Arzhang Kamarei
Managing Partner

Presented by Thesys Technologies, LLC at the SIFMA CAT Bidder Summit on July 29th, 2014.

Company Background

- Tradeworx was founded in 1999 as a financial technology company
  - Thesys is the technology affiliate established in 2009
- Highly accomplished in twin pillars of financial technology
  - Low Latency / High Performance Systems
  - Big Data and Analytics
- Superior Technology and Proprietary Middleware
  - Low cost solutions, highly efficient technology, low support burden
- Proven track record of delivering high-profile projects on budget / on time
- Trusted industry name: relationships with major financial institutions
  - Provider of the MIDAS solution to the U.S. Securities and Exchange Commission

Past Experience

- **HIGH PERFORMANCE TRADING**
  - Ultra-fast data feeds, OMS, simulation / market replay, microwave network
  - Managed solution accounting for 5% of the US Equity market's average daily volume
- **RISK GATEWAYS**
  - SEC-compliant pre-trade risk checks with median latency of 3 microseconds
- **DATA FEEDS**
  - High-performance data feeds using both software and hardware based solutions
  - Self-collect, normalize and timestamp direct feeds of major equities / futures exchanges
- **MIDAS**
  - Petabyte-scale data and analytics platform with over 100 SEC users
- **EXCHANGE TECHNOLOGY**
  - Fastest commercially available matching engine with 99% matches in less than 45us
- **ALGO TESTING**
  - Only commercially available paper trading system to quantify the economic impact of latency on trading strategies

Management Profile

Manoj Narang serves as **CEO and Managing Partner** of Thesys Technologies, LLC, and is responsible for the company's corporate and product strategy. An internationally recognized expert on algorithmic trading and market structure, Mr. Narang has taken a leading role in educating the public about these complex issues. Mr. Narang was named one of the "Seven Most Influential Players in High Frequency Trading" by CNBC.com, and in January 2014 was recognized by Institutional Investor magazine as one of the top financial technology executives in the world. Mr. Narang graduated from MIT in 1991 with a degree in Mathematics and Computer Science. Prior to founding Tradeworx, he spent several years working in technology and quantitative trading positions at major Wall Street firms including Citibank, Credit Suisse, and Goldman Sachs.

Mike Beller serves as **CTO and Managing Partner** of Thesys Technologies, LLC, and is responsible for all technology and network infrastructure for its trading technology and big data businesses. Mr. Beller has over 28 years' experience in the finance, telecom and information technology industries, with the last 15 years focused on financial technology, trading systems, and big data technology. Prior to Thesys, Mike spent 14 years as the CTO of Tradeworx, Inc, where he designed and implemented ultra high performance trading platforms and analytics that launched Tradeworx as a major player in algorithmic trading. Mike received his BS in Electrical Engineering from Cornell University, and his MS in Electrical Engineering from Columbia University.

Arzhang Kamarei, CFA, is a **Managing Partner** of Thesys Technologies, LLC, responsible for business development. Mr. Kamarei has over 13 years' experience in the financial industry, with 7 years in hedge funds. Prior to joining Thesys, he was a Vice President in International Equity Trading at Goldman Sachs, where he spent three years making markets in Canadian Equities for US and Canadian asset managers. He is a CFA Charterholder and received his BA in Psychology from Harvard.

The Vision

"Last year, the SEC put in motion two initiatives that will dramatically increase the quality and quantity of the data we receive and improve our understanding of the way today's markets function... The first of them is called MIDAS."

"MIDAS is becoming the world's greatest data sandbox. And the staff is planning to use it to make the SEC a leader in its use of market data. ... MIDAS will enable us to examine the fundamental mechanics of today's high-speed markets in a way that has never been done at the agency."

"Important as it is, though, MIDAS is just the beginning."

"Last year, the Commission voted to ... establish a market-wide consolidated audit trail (CAT), a system to significantly enhance the SEC's ability to monitor and analyze trading activity."

"CAT has the potential to expand our ability to analyze market structure and behavior even further than MIDAS while improving our ability to reconstruct and pinpoint the source of trading irregularities and investigate suspicious activity."

"I can't overestimate the importance of CAT... I believe this is an extraordinary initiative and — if properly constructed — the consolidated audit trail will be the backbone for our future surveillance and policy making efforts. And it will set the stage for expansion to additional instruments and markets as well. My long-term vision is a consolidated audit trail that spans products, markets and the globe."

- Excerpts from a speech by former SEC Chairman Elisse Walter in Feb 2013 entitled *"Harnessing Tomorrow's Technology for Today's Investors and Markets"* (emphasis added)

The Vision

Key Consideration #1: CAT must be simple to conform with (easier than OATS)

1. **Simple Reporting Format**
2. **Easy Access and Interactive Interfaces**
3. **Simulator to pre-validate submissions**
4. **Powerful tools to organize, visualize and correct errors**

Key Consideration #2: CAT should be useful in the end

1. **Aggregate Statistics**
2. **Direct Feeds, Order Book Reconstruction and Visualization**
3. **Fast and Powerful Queries**
4. **Real Time Reporting and Analytics**

The Vision

Key Consideration #3: CAT costs and financing are critical

1. **Realistic, appropriate, and reasonable**

Key Consideration #4: CAT needs proprietary tech, off the shelf is not sufficient

1. **Proprietary data stores, formats and processes**
2. **Custom Deployment for a custom built solution**
3. **Security Techniques / Innovations**

The Vision

Key Consideration #5: Characteristics of a winning CAT bidder

1. **Knows how to collect and manage a petabyte scale financial database**
2. **Understands how the data will eventually be used (analytics)**
3. **Understands security and its implications on performance**
4. **Experience with legacy reporting regimes (such as OATS): understands burden / shortcomings from a Reporter's point of view**

CAT Should Be Easy

Key Consideration #1: CAT must be simple to conform with (easier than OATS)

1) INPUT - Simple Reporting Format (no new format)

- Send us pre-existing trading logs and mapping files
- Thesys CAT accepts 50+ standard existing industry formats (e.g., ITCH)

2) CONNECTIVITY - Simple Access and Interactive Interfaces

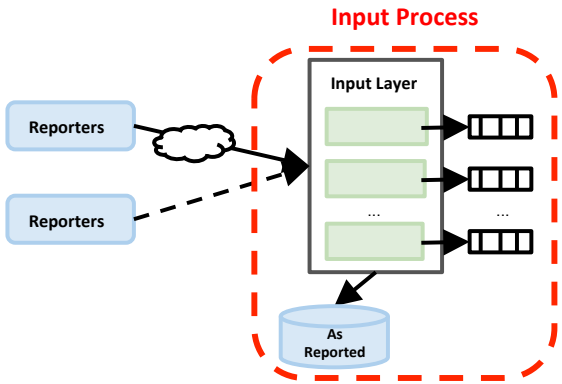
- Secure (free) VPN access for small reporters; dedicated lines for large
- Variety of interactive user interfaces

3) VERIFICATION - Simulator to pre-validate submissions & mappings

- Helps Reporters create log mappings during Onboarding Phase
- Real time status / feedback on submissions overnight after Go Live

4) CORRECTIONS - Powerful tools to organize, visualize and correct errors

1) EASE OF INPUTS: No new reporting format; send us existing logs & mappings.



REPORTER INPUT (e.g. FIX Log)

```
8=FIX.4.2|9=0198|35=D|43=N|34=00001202|
52=20130221-13:30:10.048|49=XYZQDARKPOOL03|50=0F|
56=DARKPOOL7|38=1300|54=5|60=20130221-13:30:10.048|
11=A01700027704|55=ZVZ|40=P|59=3|44=3.13|1=XYZQ|
47=A|114=N|18=M|10=163|
```

DATA VALIDATION CHECKS

Sanity checks and automated processes to catch anomalies or errors in data → *Cross sectional and time series checks to flag instances with high probability of being erroneous* → *Flagged instances manually checked as part of QA / QC process*

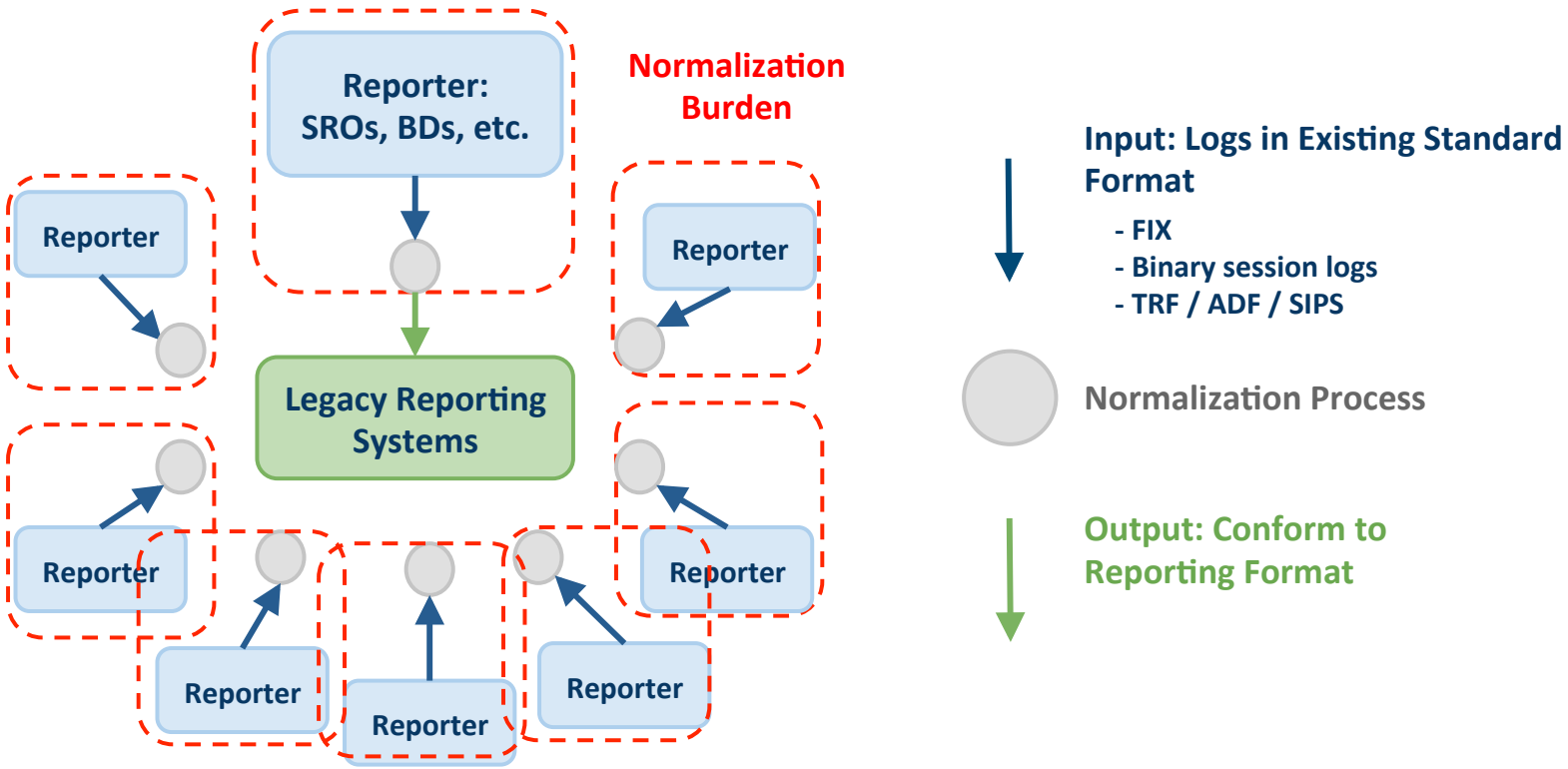
Pristine Audit Trail: No Deletions / Edits

As Reported

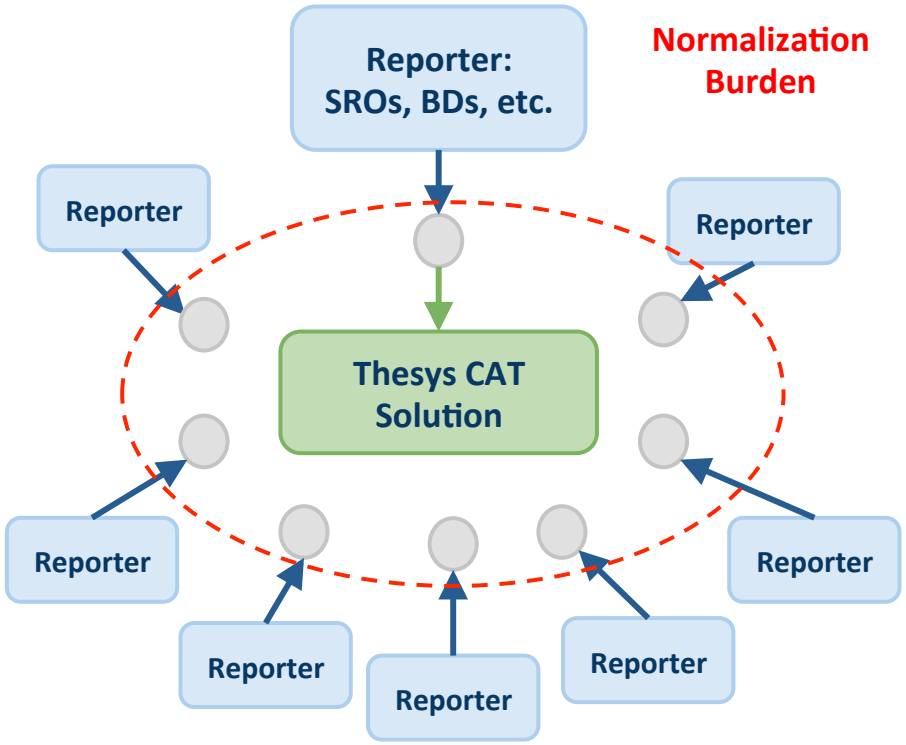
FEEDBACK / CORRECTIONS

Simple GUI to view feedback / submit corrections

Reporter Input Detail



Aggregate Reporter Conversions



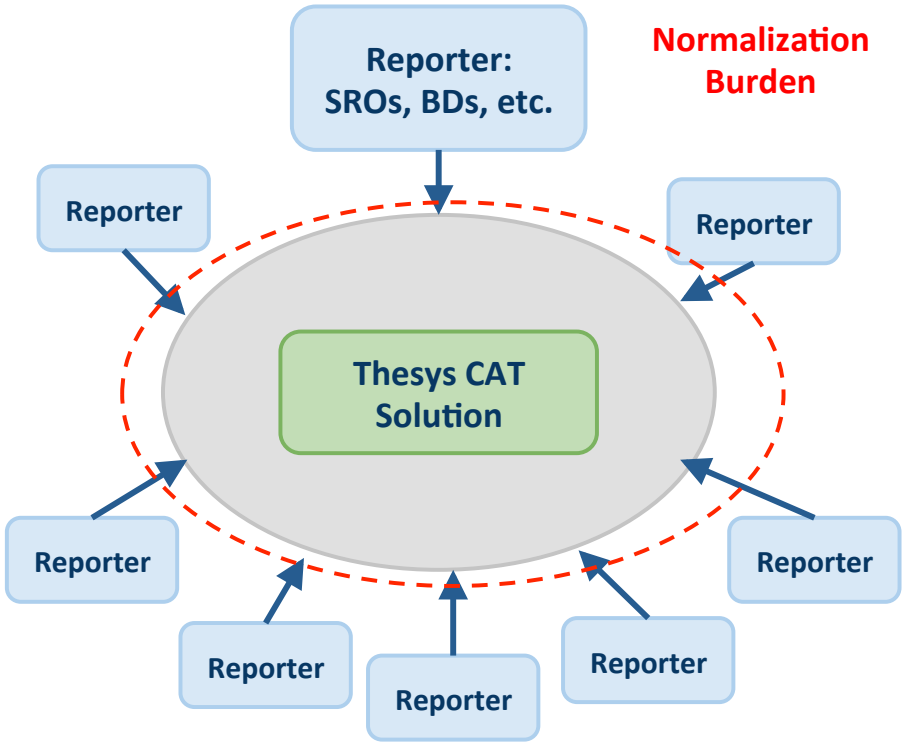
Input: Logs in Existing Standard Format

- FIX
- Binary session logs
- TRF / ADF / SIPS

Normalization Process

Output: Conform to Reporting Format

Reduce Reporting Burden



Normalization Burden

"Just send us your logs & mappings"

Input: Logs in Existing Standard Format

- FIX
- Binary session logs
- TRF / ADF / SIPS

Normalization Process

Output: Conform to Reporting Format

Order & Trade Data Supports This

Table: Data Types & Sources for CAT Equities

DATA TYPE	SOURCE	PRODUCT	DESCRIPTION	REPORTING FORMAT
Customer / Account	**BDs**	Equities / Options	Full submission of customer and account information (full submission will occur at initial CAT ingestion and weekly on non-peak hours)	**- Manual Input (Web Interface)** **- CSV File**
Order Data	**BDs**	Equities	CAT reportable orders and all related order events receives and generated by the BD	**- FIX** **- Exchange Native Binary Protocols** **- CSV File** **- FINRA ADF Input Format**
…	**…**	…	…	**…**
Trade Data	**FINRA**	Equities	All transactions reported to a FINRA TRF	**TRF format.**
Corporate Actions	**Listing Market**	Equities	Corporate events that affect the underlying instrument	**- Manual Input (Web Interface)** **- CSV File**
…	**…**	…	…	**…**
Quotes	**SIPs**	Equities	All CQS and UQDF data, including all quotes, appended NBBOs and admin messages (e.g. indications of market open / close, halts / resumes and circuit breakers)	**- CQS / UQDF native formats.**
Sponsored Access	**SROs**	Equities / Options	DMA relationships	**- Manual Input (Web Interface)** **- CSV File**
…	**…**	…	…	**…**
Trade Data	**SROs**	Equities / Options	Direct market data feeds	**- FIX** **- Exchange Native Binary Protocols**

Widespread Exchange Format Accepted

Table: Examples of Exchange-Proprietary Formats Accepted by Thesys CAT (Already Possessed by Reporters)

EXCHANGE	PROTOCOL	EXCHANGE	PROTOCOL	EXCHANGE	PROTOCOL
NASDAQ	OUCH 4.2	NASDAQ PSX	OUCH 4.2 for PSX	NYSE	NYSE CCG FIX
	OUCH 4.1		OUCH 4.1 for PSX		NYSE UTP
	RASHport 1.1		RASHport 1.1 for PSX	NYSE/ARCA	NYSE Arca FIX Specification
	INET FIX Version 4.2		RASHport 1.1 for PSX		ArcaDirect API Specification 4.0
	FIX Lite (Flite)		INET FIX Version 4.2 for PSX		ArcaDirect API Specification 4.1
	SUMO FIX		FIX Lite (Flite) for PSX	EDGA & EDGX	Edge XPRS Specification
	CTCI	BATS	BZX BOE		Direct Edge FIX Specifications
	QIX		BYX BOE	…	…
NASDAQ BX	OUCH 4.2 for BX		BZX FIX		
	OUCH 4.1 for BX		BYX FIX		
	RASHport 1.1 for BX	CBOE	CBOE FIX		
	RASHport 1.1 for BX	NSX	NSX FIX		
	INET FIX Version 4.2 for BX				
	FIX Lite (Flite) for BX				

Inputs Already In-House

How many records in this log file?

What order handling scenario?

SAMPLE FIX LOG SUBMISSION

```
8=FIX.4.2|9=0198|35=D|43=N|34=00001202|52=20130221-13:30:10.048|
49=XYZQDARKPOOL03|50=0F|56=DARKPOOL7|38=1300|54=5|
60=20130221-13:30:10.048|11=A01700027704|55=ZVZ|40=P|59=3|44=3.13|
1=XYZQ|47=A|114=N|18=M|10=163|
8=FIX.4.2|9=0198|35=D|43=N|34=00001202|52=20130221-13:30:10.199|
49=XYZQDARKPOOL02|50=0B|56=DARKPOOL7|38=800|54=5|
60=20130221-13:30:10.199|11=A01700027731|55=ZVZ|40=P|59=3|44=3.13|
1=XYZQ|47=A|114=N|10=180|
8=FIX.4.2|9=0196|35=D|43=N|34=00001202|52=20130221-13:30:10.235|
49=XYZQDARKPOOL01|50=0A|56=DARKPOOL7|38=100|54=5|
60=20130221-13:30:10.235|11=A01700027737|55=ZVZ|40=2|59=3|44=3.13|
1=XYZQ|47=A|114=N|10=019|
8=FIX.4.2|9=0198|35=D|43=N|34=00001204|52=20130221-13:30:28.081|
49=XYZQDARKPOOL04|50=0F|56=DARKPOOL7|38=8000|54=5|
60=20130221-13:30:28.081|11=A01700028526|55=ZVXQ|40=P|59=3|44=4.05|
1=XYZQ|47=A|114=N|10=180|
```

SAMPLE META FILE SUBMISSION

```
Reporter: B00123
NumRecords: 10321569
Format: FIX
Mode: Target
AccountLocator:
   Method: FixTag
   Tag: 50
   Mapping:
       0A: 12131121
       0B: 37212211
       0F: 37212212
```

Who's reporting the log?

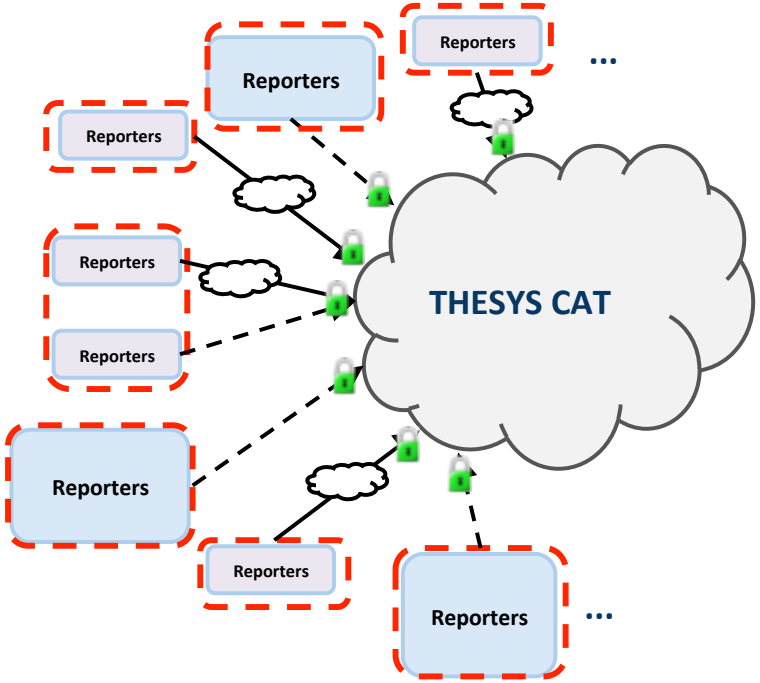
What's the format of the log?

Where are the Account IDs?

SAMPLE MAPPING SUBMISSION

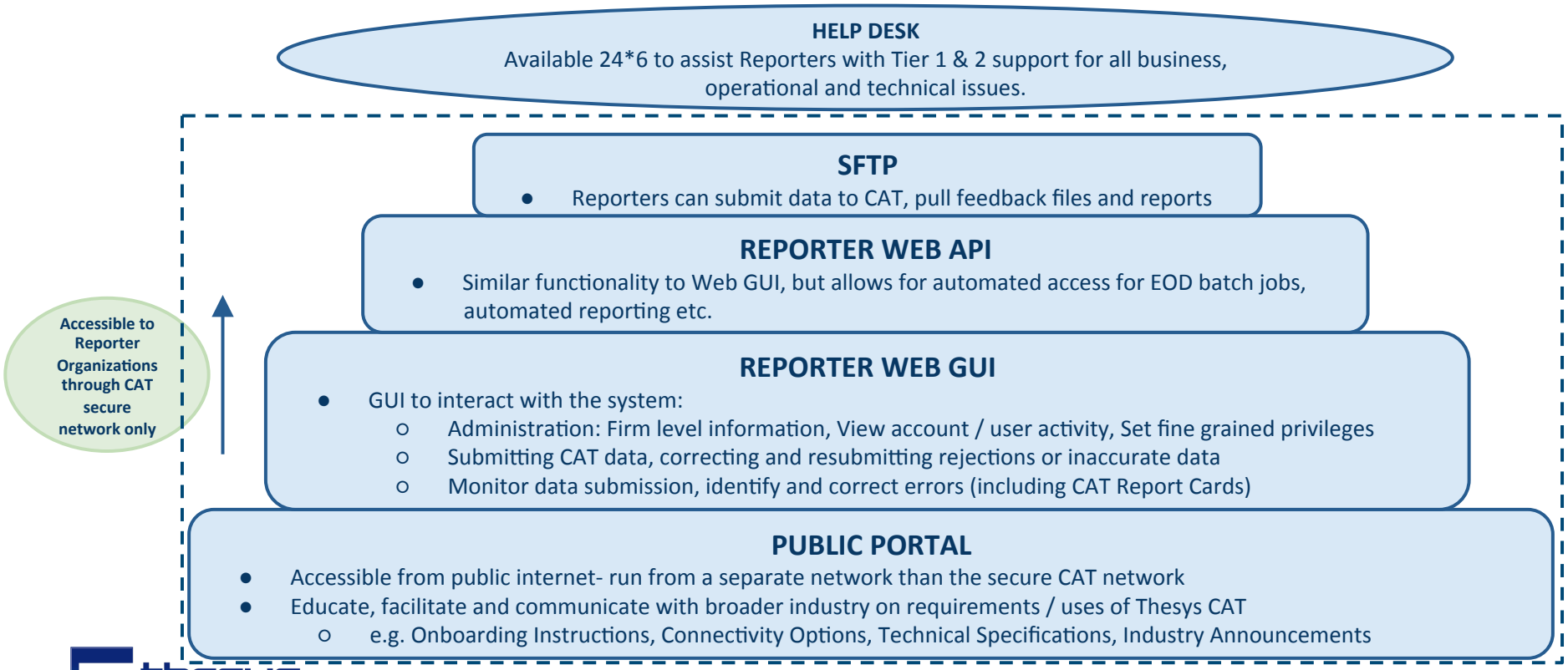
75628	91384	...	...
56482	34239	...	...
34157	94272	...	...
12304	43801	...	...
43092	32458	...	

2) EASE OF CONNECTIVITY: Secure and easy access with a variety of interactive interfaces - low burden on the industry, no hidden network costs.



- **2000+ Reporters** to access the system
- **Bulk of reporting:** 13 SROs + Large BDs
  - Dedicated 1-10 GB connections
    - Tools to monitor, log and audit connections
  - Integration with Extranet Providers e.g. SFTI, SAVVIS
- **Low burden on smaller BDs:** over the internet using a standard/free VPN client
- All communication is **encrypted** (SSL / HTTPS / SFTP), **logged** and **monitored**

Public & Private Interfaces



3) EASE OF VERIFICATION: Simulator to validate reporter inputs. Real time feedback overnight BEFORE final submissions are due.

Reporter ID	Order ID	Size	…	Status
R18423	O1243	100	…	Verified
R18423	O4093	175	…	Verified
R18423	O3420	10	…	Verified
R18423	O1275	55	…	Pending
R18423	O6793	130	…	Pending
R18423	O8932	5	…	Error
R18423	O5476	150	…	Error

Simulator Progress / Summary Statistics

Number of Records Submitted	1,232,000
Number of Records Verified	1,157,930
..	
Number of Errors	200

Progress Tracker

Click for Error Breakdown

Onboarding Phase

- Reporters test systems / mappings against simulator
- Receive real time feedback on data validation / breaks
- Submit OATS files to compare and find errors
- Seminars and 1 on 1 sessions with large Reporters to help create mappings / reuse existing systems

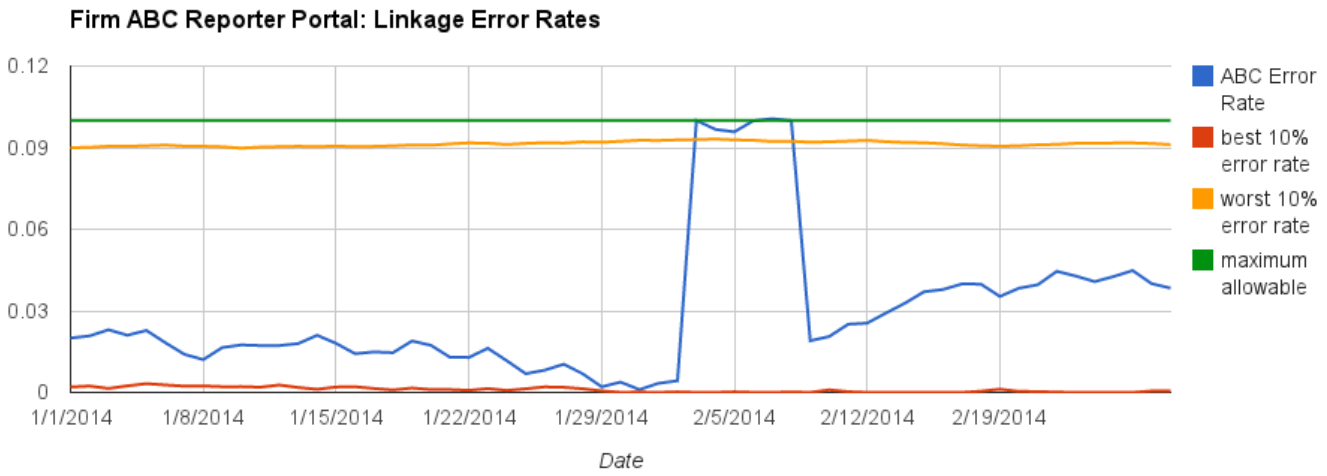
Ongoing / After Go Live

- Real time feedback on submissions after the close up until final submission at 8am on T+1
- Interactive updates / feedback / correction allowed overnight as lifecycles are reconstructed
- Simulator can be used to test change in formats / rules

4) EASE OF CORRECTIONS: Help organize, visualize and correct errors. Intelligent tools lower burden by identifying patterns / corrections across Reporters.

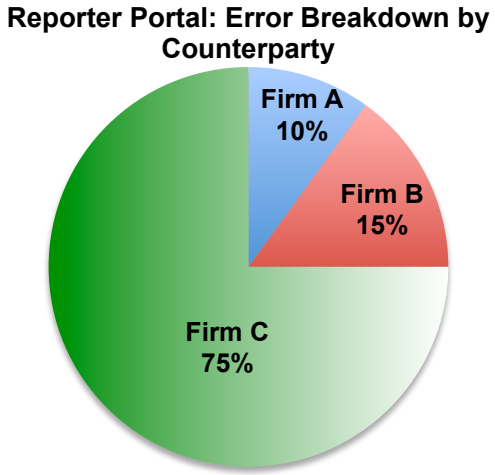
Interactive Report Cards

Compare your error rates with similar firms

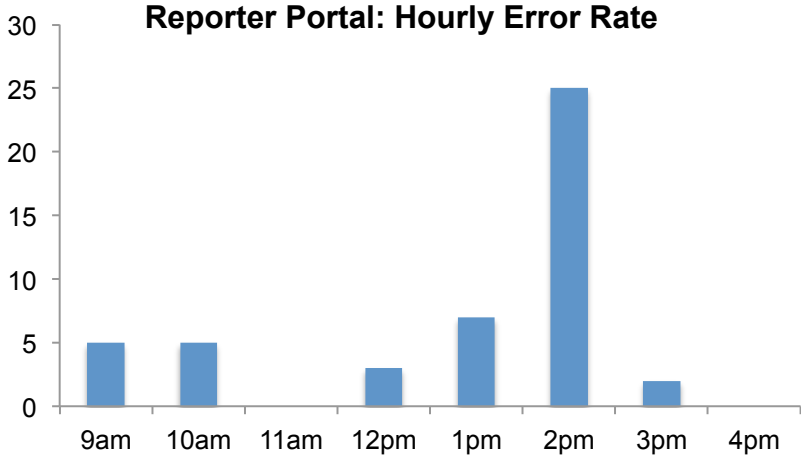


Tools for Catching Errors

Example #1:
Are errors related to orders sent to a specific firm?



Example #2:
Do errors start happening at a certain time of day?



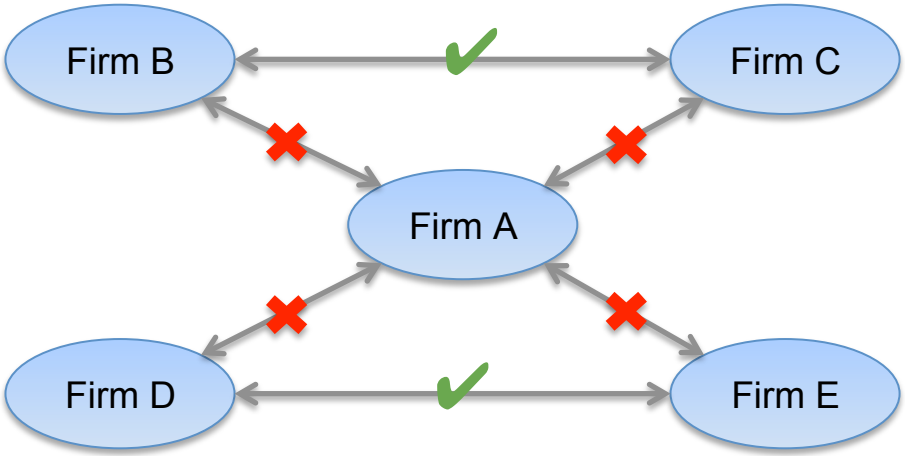
"Wisdom of Crowds" in Error Correction

Example #1: Reporter uses an old symbol.

Thesys CAT identifies invalid symbols, detects a pattern (reporter is using an old symbol), and suggests the change with the implied change in error rate.

Example #2: Linkage Errors

Thesys CAT analyzes the network of breaks / errors and identifies counterparties with a high probability of having issues.



CAT Should Be Effective

Key Consideration #2: CAT must be useful in the end (no "double taxation")

1) Aggregate Statistics (Included in Thesys CAT)
- Necessary for most basic questions

2) Direct Feeds, Order Book Reconstruction & Visualization (Included in Thesys CAT)
- Necessary for more pressing questions

3) Fast and Powerful Queries (Included in Thesys CAT)
- Reduce the cost of usage and compliance

4) Real Time Reporting and Analytics (Compatible with Thesys CAT)
- Reduce future costs by better technology

What's Missing: Aggregate Statistics

Sample Studies leveraging Aggregate Statistics (not in RFP):

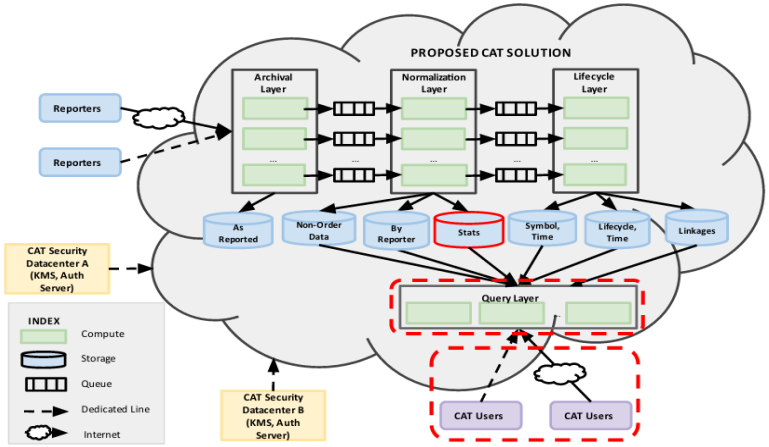
Example #1: Rank Reporters by trading efficiency (total volume / total messages) or cancellation rates over the last year.

Example #2: Rank Stocks by cancellation rate in the last minute of trading over the last two years.

Example #3: Rank Reporters by volume traded within a millisecond of an earnings announcement.

1) EFFECTIVNESS OF AGGREGATE STATISTICS: 100's of configurable pre-calculated statistics to allow SROs / SEC to query years of data and rapidly locate anomalies.

- Pre calculated 100s of statistics on a daily basis as the data is ingested into the system
- Large-scale analytics and statistics on the full data set
- Daily or even minute-by-minute or shorter basis



- Breakdowns along a **number of dimensions:** *Symbol, Underlying Symbol, Reporter, Customer Account etc.*
- **For Variables such as** *number of orders, cancels, executions, shares / contracts traded etc.*
- Highly efficient internal language can create **transformations on the statistics** e.g. moving averages

- Completely **configurable by SRO Users**- define their own statistics
- **Rapidly locate anomalies** and investigate by **drilling down** on time / event of interest
- Queries span **years of information** and **returned in seconds**

What's Missing: Direct Feeds

Sample Studies that REQUIRE Direct Feeds and Market Reconstruction (dropped from the NMS plan):

Example #1: Analyze and validate data involving a Reporter submitting prevailing market prices such as Best Execution Reports.

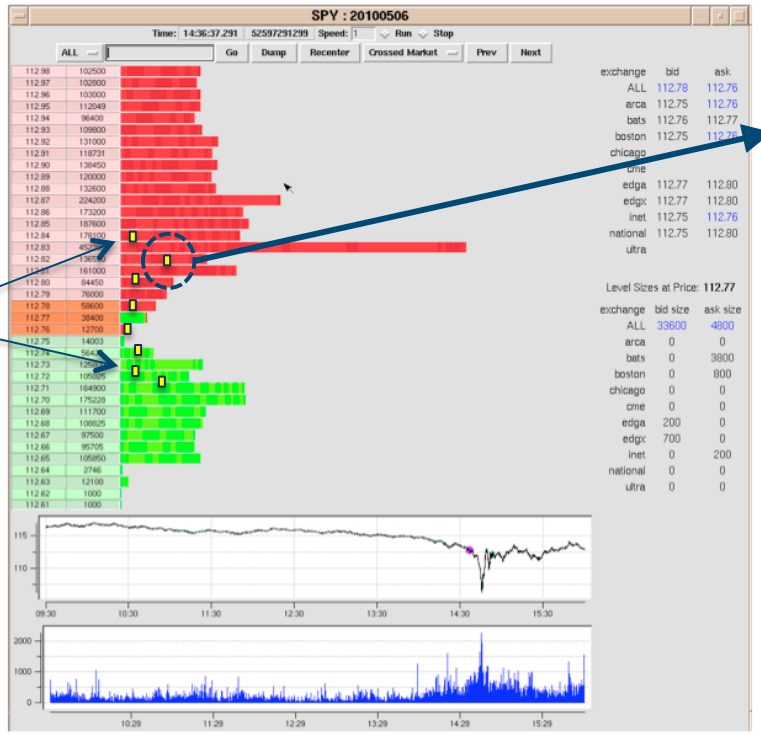
Example #2: Examine any latency sensitive trading or routing phenomena.

Example #3: Investigate concerns around manipulative trading practices.

2) EFFECTIVENESS OF DIRECT FEEDS/MARKET RECONSTRUCTION: Full depth order books and drill down to order lifecycle; All pertinent policy questions addressed.

Identify all Orders related to a specific Reporter. Drill down into any Orders of interest.

Notice that the orders have high time priority and are at multiple levels- probably an automated market maker.

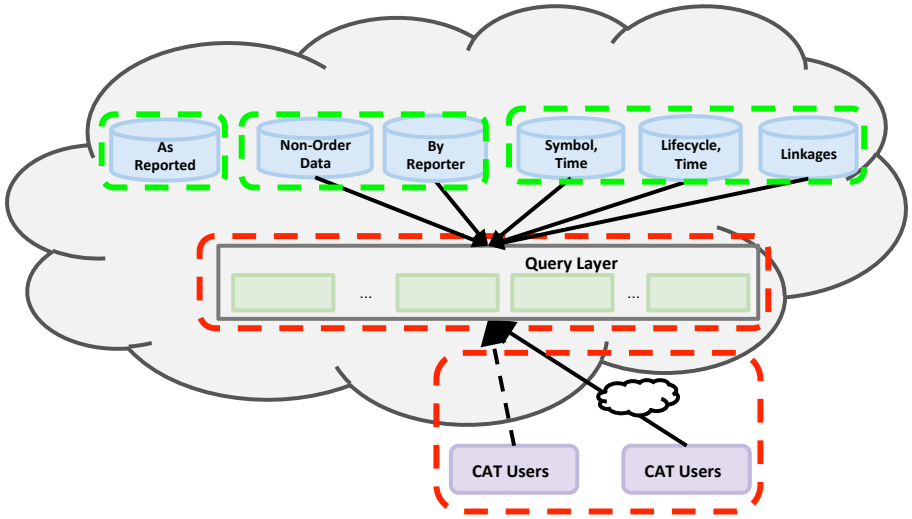


ORDER TYPE: NEW ORDER
REPORTER ID: BD-001
COUNTERPARTY ID: BD-034
SYMBOL..
PRICE:..
SIDE:..

O-45 → O-44 → O-12

- **Complete view of direct and regulatory order books (full depth)**
- **Ability to point and click drill down to any individual order in the book and query its lifecycle**

3) EFFECTIVENESS OF FAST AND POWERFUL QUERIES: Thesys CAT goes above and beyond the RFP requirements to provide query results in seconds, not minutes.



Simple query- query requiring reference to just one ordering of the dataset

Complex query- query that involves querying multiple orderings of the data

- **2000** simple queries per minute
- **250** complex queries a minute
- < **10 sec** for simple queries that return **tens of thousands** to a few million records
- < **1 min** for Complex queries that require joins

Highly scalable and low cost

Integrates with standard analytics packages e.g. Excel, MATLAB, R, or Python / Numpy

CAT Should Be Cheap

Key Consideration #3: CAT costs are critical and should be kept to a minimum

KEEPING COST LOW: Thesys CAT is a cost effective, scalable solution that aims to save the industry over $1bn in build, operating and compliance costs.

THESYS COST MODEL

Build	Year 1	Year 2	Year 3	Year 4	Year 5
$30M	$27.5 M	$27.5 M	$27.5 M	$27.5 M	$27.5 M
Thesys Total 5 year cost: $167.5 M (Vs. ~5bn in SEC Section 31 fees)					

- Fixed Cost
- No Variable Component
- Incentivized to minimize errors and issues
- No-Build Fee Option

100% Equally Weighted by Firm ⟷ **100% Allocated by CAT Msg Volume**

Total Annual Cost	27.5 M
# of Reporters	~ 2000
Annual Reporter Cost	~ $14,000

Example: Large Reporter creating 1% of CAT messages

By Equally Weighted	~ $14,000
By CAT Message Volume	~ $275,000

Total Annual Cost	27.5 M
# of CAT Messages	60 bn
Message Cost	~ $0.000002

CAT Should Be Highly Customized

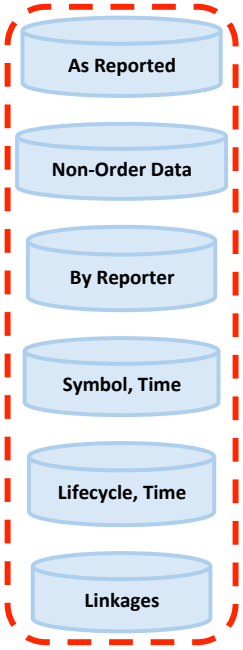
Key Consideration #4: CAT needs proprietary tech, off the shelf is not sufficient

1) **Proprietary data stores, formats and processes**
   - Designed, built and tested proprietary highly efficient data storage system
   - Verified performance (seconds, not minutes!) using CAT-scale test system
2) **Custom Deployment for a custom built solution**
   - Flexible architecture supports Cloud, Data Center & Hybrid deployments
3) **Security Techniques / Innovations- the key to a viable CAT**
   - Unique approach integrates access, authentication and key management
   - Unparalleled control over data access- fine-grained privileges
   - Vetted by a panel of world-renowned security experts

1) Proprietary Data Stores: Reviewed traditional Big Data technologies- off the shelf is not sufficient!

System	Strengths	Weaknesses
Relational DBMS	Complex, flexible schema. ACID properties.	Poor scalability
Hadoop/MapReduce + HDFS/GFS	Highly scalable Organize many compute nodes	Slow response times Dedicated storage/compute ratios
HBASE/Cassandra/BigTable	Indexing added without the scalability issues of RDBMS	Generic approach adds large storage overhead
Apache SPARK/Google Dremel/ BigQuery/Redshift	Scalability of Hadoop with better response times	Optimized for specific analyses which don't map well to CAT (security and optimization)

Thesys CAT Data Stores



- **Proprietary highly efficient data storage system**
- **Leveraging the Cloud: designed on cloud based object stores**
- **Proprietary highly optimized data structure**
- **Order and quote data stored in multiple orderings**

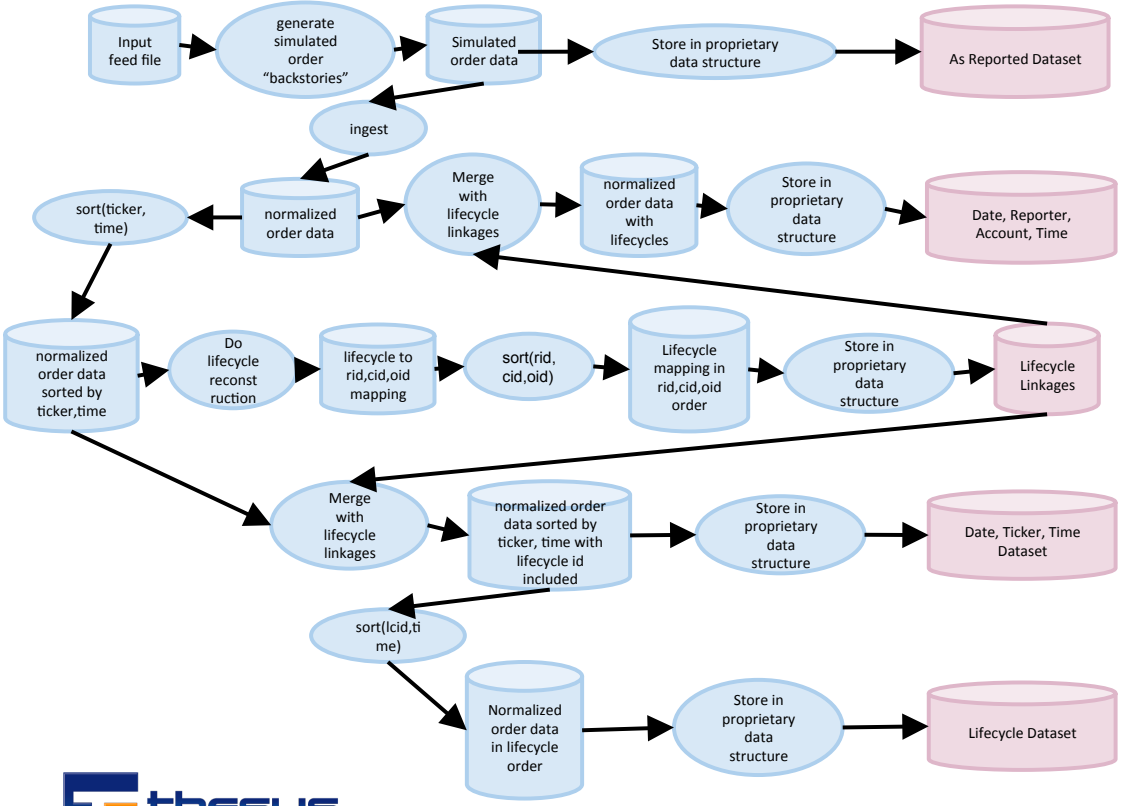
~~Off the Shelf Big Data Systems~~

Built Ground up for CAT

- Cost-Effective
- Arbitrarily Scalable
- Excellent Survivability

```
columns:
date,time,type,reporterid,counterpar
tyid,orderid,linkid,symbol,side,quan
tity,price,flags,lcid
20130221 48610048000 NEW B00123
A303123 A01700027704 - ZVZ S 1300
3.13 M LC000355217
```

Thesys built a CAT-scale proof of concept- tested and validated.



- **Generated enormous test database using our market data and extensive industry knowledge**
- **Billions of records of simulated order traffic- 10-12 bn records / day ~ 20% of CAT**
- **Implemented entre ingestion, normalization, lifecycle processing layers**
- **Tested and validated proprietary data store and distributed computing techniques in the cloud**
- **Ran and benchmarked various example queries- most queries run in a few seconds, far better than CAT requirements**

Simulated and benchmarked All Use Cases provided to Bidders in the RFP and conference calls.

REF #	USE CASES FROM RFP AND BIDDER CONFERENCES	PROCESSING WORK FLOW
PPT1	All orders, modifications, cancellations and executions in issue ABC submitted by all broker-dealer CAT Reporters and CAT Reporter Exchange (A) with an event timestamp between 15:14:28 and 15:17:22 for trade date May 16	{DATE | SYMBOL | TIME}, filter REPORTER ID
PPT2	All events in issue ABC and ABC Jan 50 Calls submitted by a specific CAT Reporter with an event timestamp between 9:59:30 and 10:01:30 for trade date Sept 20	{DATE | SYMBOL | TIME}, filter REPORTER ID
PPT3	All executions in issue ABC and all options class for ABC with a timestamp between 15:59:00 -16:01:30 for trade dates Aug 1 – Aug 5	{DATE | SYMBOL | TIME} for 4 trade dates
PPT4	All order events with an event timestamp between 9:59:30 and 10:01:30, submitted by a specific CAT Reporter in issue ABC for trade date Sept 20, along with associated order lifecycles (which may span additional dates)	{DATE | SYMBOL | TIME}, filter REPORTER ID, join {LIFECYCLE, TIME}
PPT5	All events in issue ABC submitted by broker-dealer and SRO CAT Reporters as being received from or routed to a specific CAT Reporter with an event timestamp between 15:14:28 and 15:17:22 for trade date Jan 11	{DATE | SYMBOL | TIME}, filter COUNTERPARTY ID
…	…	…

Sample Query Use Case

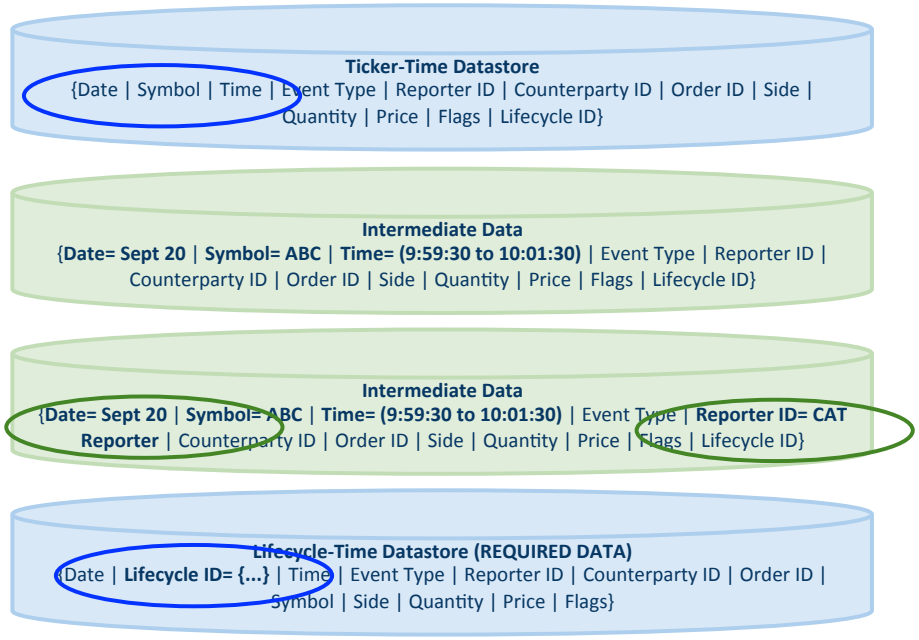
EXAMPLE 1: All order events with an event timestamp between 9:59:30 and 10:01:30, submitted by a specific CAT Reporter in issue ABC for trade date Sept 20, along with associated order lifecycles (which may span additional dates)

STEP 1: Filter Ticker-Time database for Symbol and Time range of interest

STEP 2: Filter resulting records for the Reporter ID of interest

STEP 3: Extract Required Lifecycles

STEP 4: Extract All Associated Lifecycles from Lifecycle-Time datastore



Simulation Results: Query results in seconds, not minutes!

EXAMPLE #1. All orders, modifications, cancellations and executions in issue AAPL submitted by all broker-dealer CAT Reporters and CAT Reporter Exchange (A) with an event timestamp between 15:14:28 and 15:17:22 for trade date May 16
(Here we used AAPL, a very large stock, instead of ABC from the original example, a representative one)

ROWS RETURNED: 45,271
TIME TO RUN: 1.5 seconds

EXAMPLE #2. All events in issue SPY submitted by broker-dealer and SRO CAT Reporters as being received from or routed to a specific CAT Reporter with an event timestamp between 15:14:28 and 15:17:22 for trade date Jan 11
(Here we use SPY, an extremely highly traded ETF, instead of ABC. We also have to select a specific "simulated" CAT reporter from our simulation data set.)

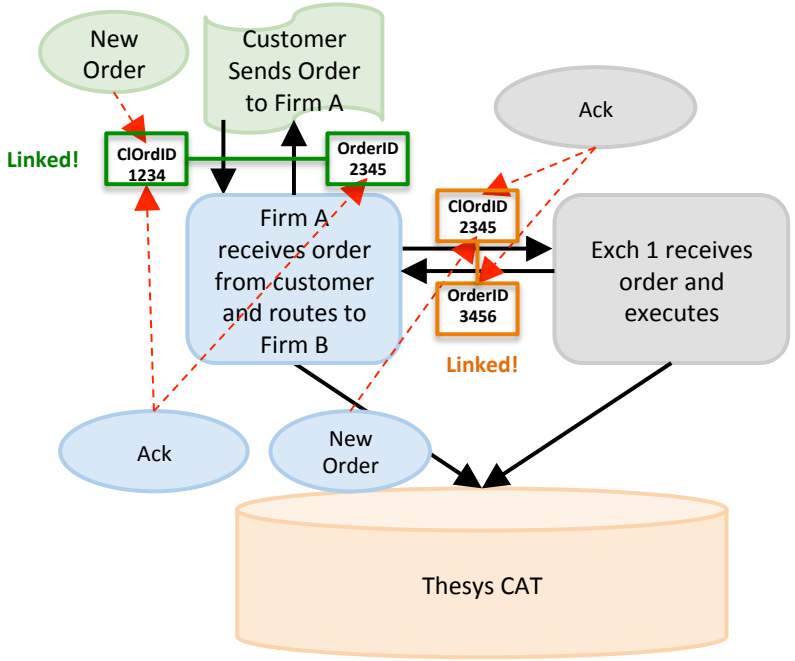
ROWS RETURNED: 5,777
TIME TO RUN: 1.7 seconds

EXAMPLE #3. All order events with an event timestamp between 9:59:30 and 10:01:30, submitted by a specific CAT Reporter in issue MSFT for trade date Sept 20, **along with associated order lifecycles** (which may span additional dates)
(Here we will use MSFT, a very large stock, instead of ABC, a representative one)

ROWS RETURNED: 195,754
TIME TO RUN: 7.9 seconds

1) Proprietary Data Formats and Processes: Tested and validated various Order Handling Scenarios provided to Bidders.

EXAMPLE ORDER HANDLING SCENARIO: DMA CUSTOMER TRADES ON EXCHANGE



FIRM A Submits:

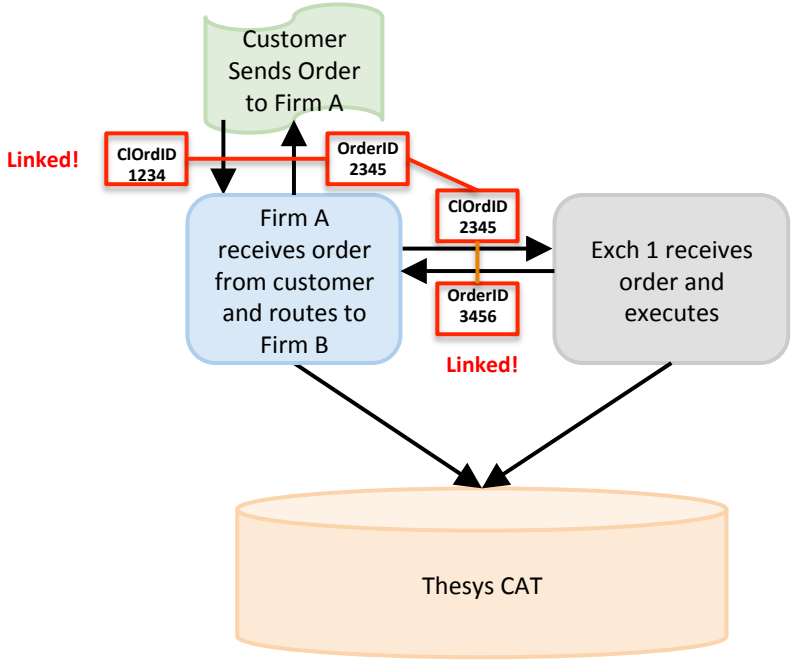
- **Customer and Account Info (EOD Updated)**
- **FIX logs for incoming order from customer (presumed to be trading electronically)**
- **FIX logs between Firm A and Firm B**

EXCH 1 Submits:

- **Native Binary logs between Firm A and Exch1**

1) Proprietary Data Formats and Processes: Tested and validated various Order Handling Scenarios provided to Bidders.

EXAMPLE ORDER HANDLING SCENARIO: DMA CUSTOMER TRADES ON EXCHANGE



DAISY CHAINING

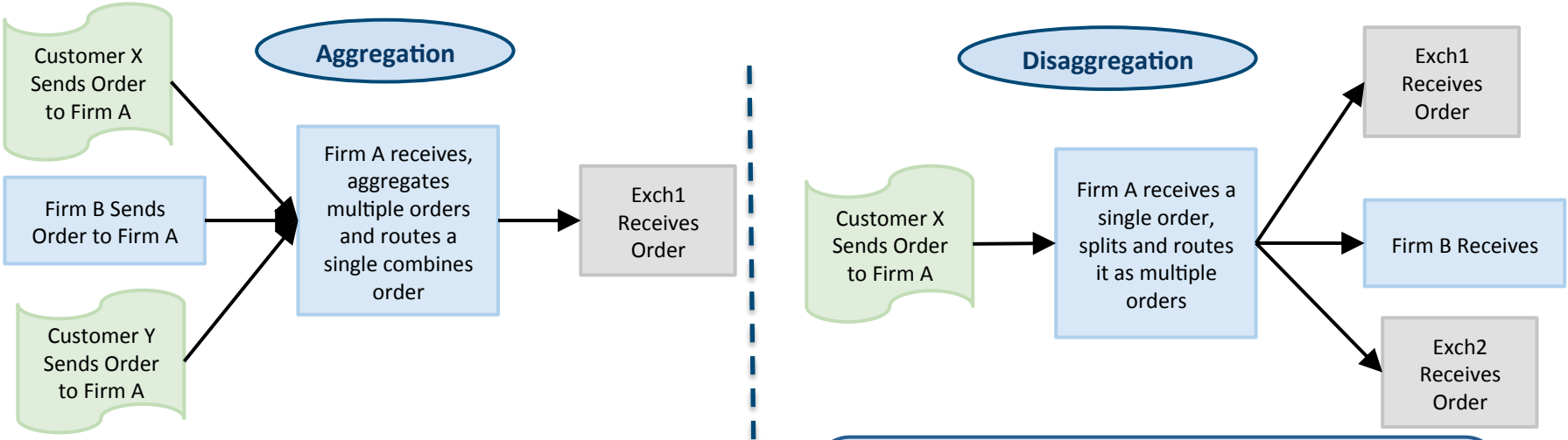
- 1234, 2345 (Customer/Firm A Order/Ack)
- 2345, 3456 (Firm A/Exch 1 Order/Ack)

FIX and Exchange Binary Format were shown as example, Logs could be in any format

Metafiles establish broker and account identifiers

Additional mapping required in some cases

Complex Order Handling Scenarios



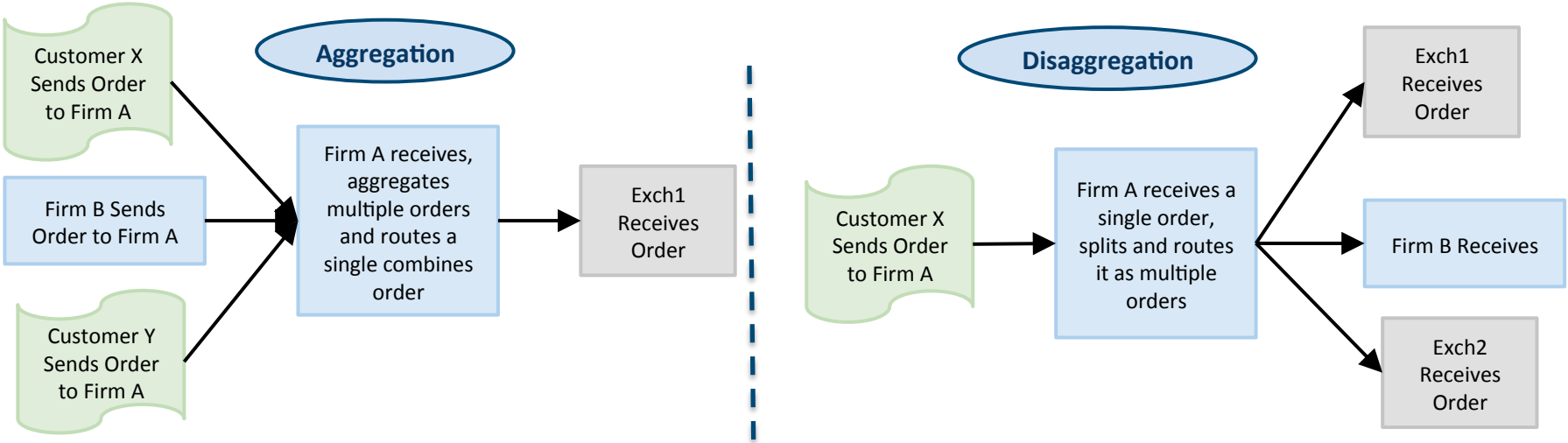
Some Common Aggregation Scenarios

- **Internalizer combines orders for efficient execution**
- **Post Trade Allocation**

Some Common Disaggregation Scenarios

- **Multi legged order**
- **Smart Order Router**
- **BD using an algo**

Thesys CAT Solution



- **Firm A sends a Master Order ID in the Acks to Customer X / Firm B & Customer Y**

OR

- **Firm A provides a CSV file with required mapping between incoming & aggregated orders**

- **Firm A uses Master Order ID in outgoing orders**

OR

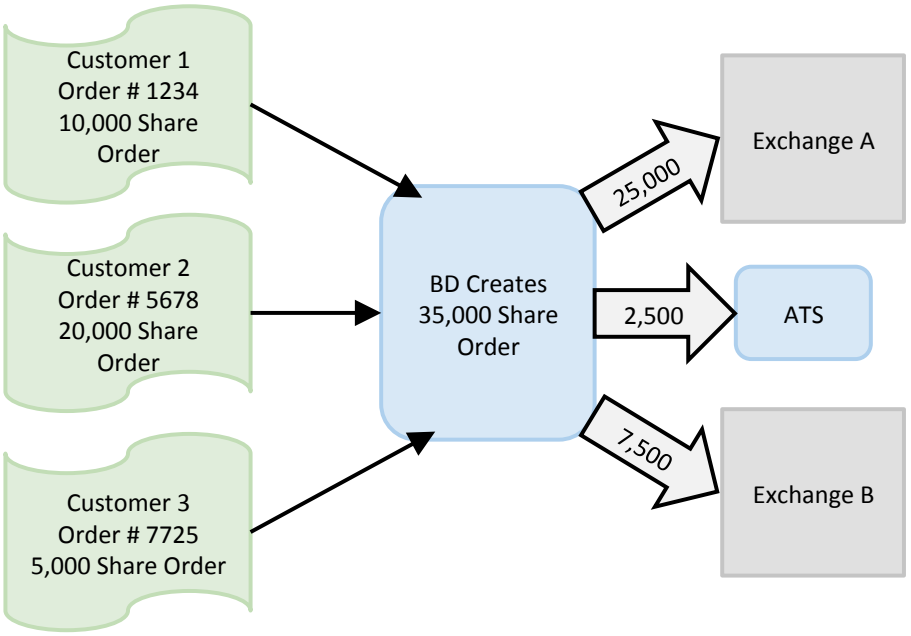
- **Firm A uses first n characters of the outgoing ID to represent a common parent order**

OR

- **Firm A reports a CSV file with the required mapping**

Example Complex Order Scenario

SCENARIO: ORDERS AGGREGATED WITH OTHER ORDERS FOR FURTHER ROUTING AND EXECUTION CAPTURING BOTH THE STREET SIDE EXECUTIONS FOR THE AGGREGATED ORDER AND THE FILLS TO EACH INDIVIDUAL CUSTOMER ORDER



1. BD reports logs facing customers and exchanges

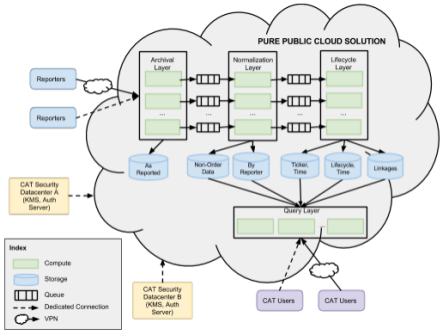
2. BD reports a mapping between Order ID's facing the customer, and outgoing orders facing

1234	99674	. . .	. . .
5678	99674	. . .	. . .
7725	99674	. . .	. . .
99674	7682	. . .	. . .
99674	8723	. . .	. . .
99674	2112	. . .	. . .

3. Exchange A, ATS, Exchange B report logs facing BD

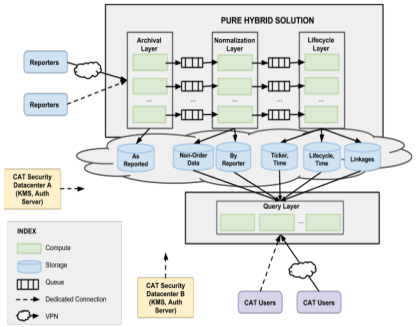
2) Deployment Options: Custom deployment for a custom built solution: Cloud, Private data center or a combination.

Public Cloud



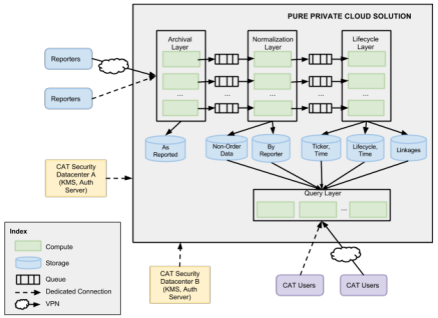
- **Early adopter** of the public cloud in 2007
- **Significant expertise** / experience
  - Petabyte-scale tick database
  - MIDAS deployed on AWS within 4 months
- **Public cloud partners:** Amazon Web Services and Google Cloud Platform
  - Showcased at global AWS events
  - Dedicated engineers

Hybrid Solution



Data collection, storage and analysis **systems across public cloud and private cloud domains**

Private Cloud



- **Past Experience:** High-performance trading platform
- Compute and storage infrastructure **colocated with every major exchange** in NJ and Chicago area
- Extensive **network mesh** connecting the various datacenters, with **microwave long haul** between NJ / Chicago

3) Security Techniques / Innovations- Securing access to the system and the data is critical to the viability of CAT- Importance of Security CAN'T be overstated!

CAT SECURITY GOAL

CAT must ensure submissions are valid, queries are authorized and data is protected from all threats- both within and outside the various organizations.

EXAMPLES OF SECURITY CONCERNS

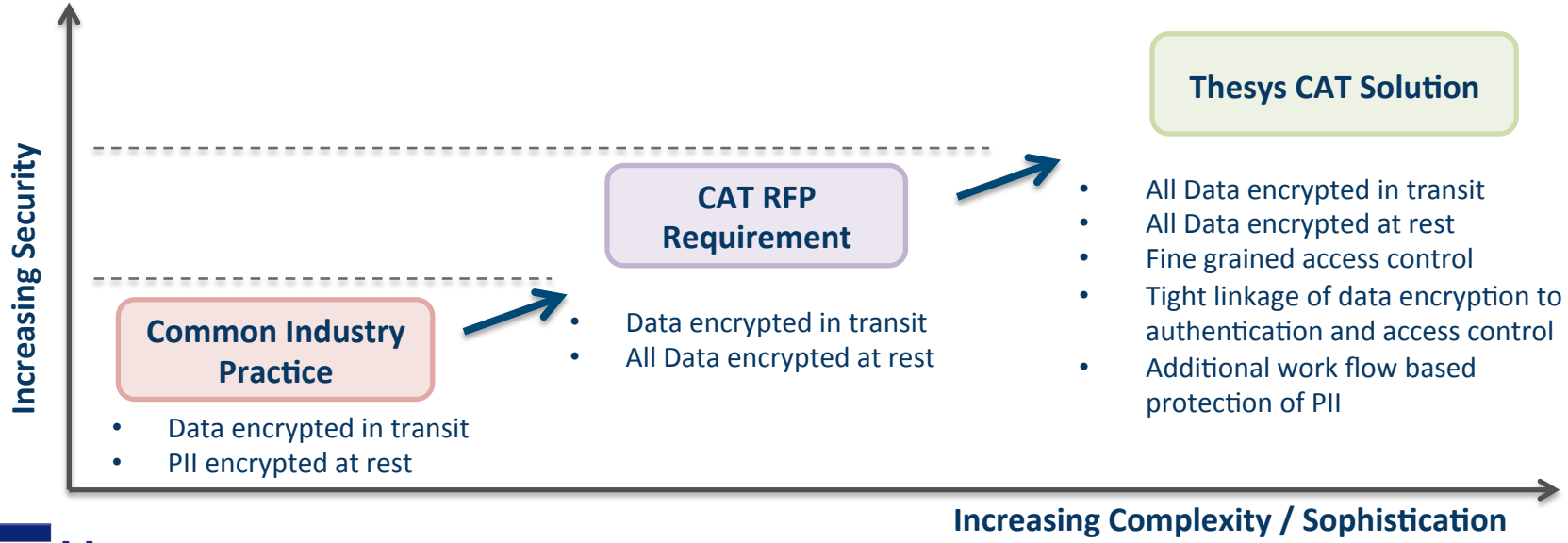
1. **Outside actors may attempt to break into the database**
2. **Errors or bugs in the system itself leak information to unauthorized users**

CAT SECURITY CHALLENGES

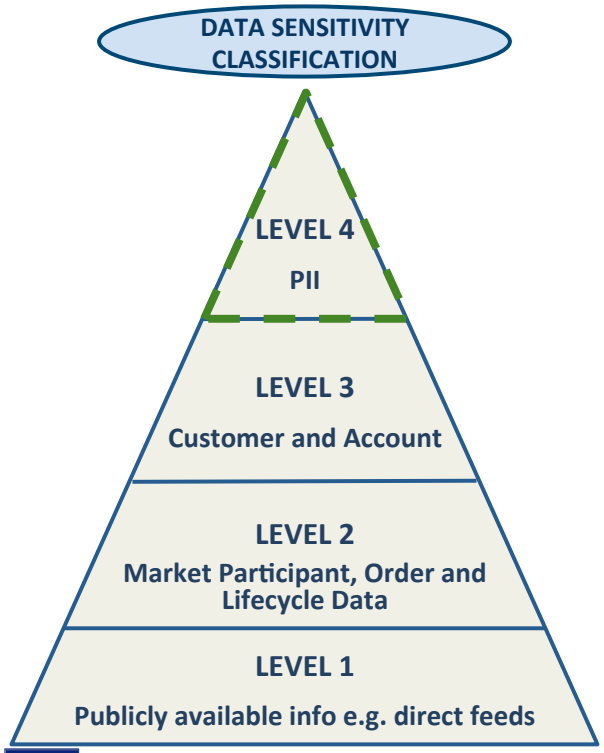
1. **High profile system (target)**
2. **Size of the database- petabyte scale**
3. **Number of Users- 2000+ Reporters, 3000+ Users**
4. **Sensitivity of Data- Personally Identifiable Information**

Thesys CAT Security is designed from the ground up and deeply integrated into each deployment scenario- state of the art is NOT good enough!

The Security Solution leverages our experience developing MIDAS (satisfies hundreds of NIST 800-53 requirements) and is vetted by a panel of world-renowned security experts in the areas of cryptography and information security.



ALL data is encrypted in transit and at rest- submissions are encrypted as well and require authentication. Additional security measures are taken to protect PII data.



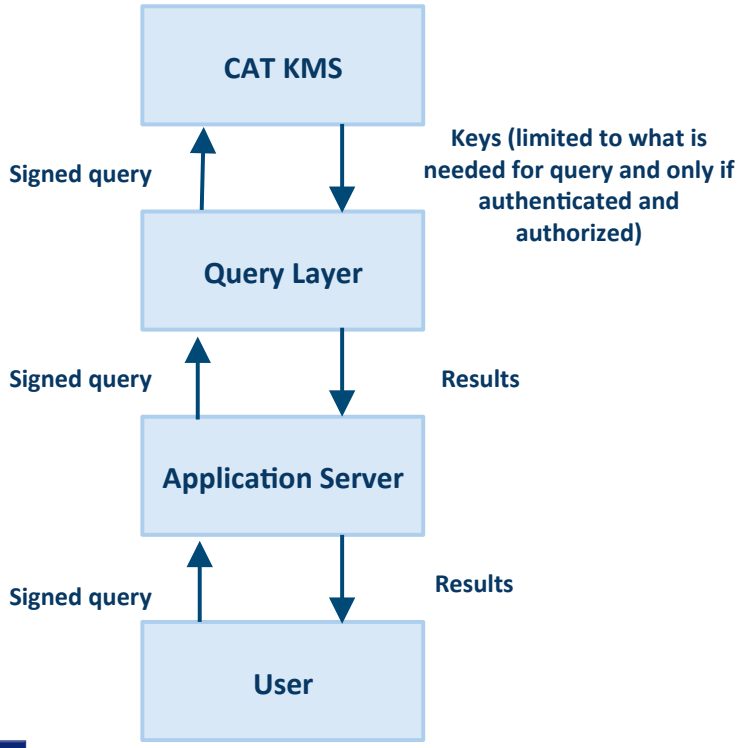
DATA ENCRYPTION (ALL SENSITIVITY LEVELS)

- Data in transit: all communication with Users / Reporters is always encrypted
- Data at rest: all data inside CAT is encrypted
  - Each dataset has a different set of keys, rotated on a daily basis
  - Keys assigned to "shard"- allows fine grained permissions

HANDLING PII

- Separate database from Customer Information and Account information
- Never included in regular CAT query results, access is restricted to privileged accounts pursuant to stringent review / approval
- "PII Access Requests" screen (part of CAT Online Query tool) is reviewed by SRP supervisory user and approved if appropriate
- All access to and usage of PII is monitored, logged and audited

Advanced key-agreement, encryption and authentication protocols minimize the probability of compromise even in the face of errors in application logic.



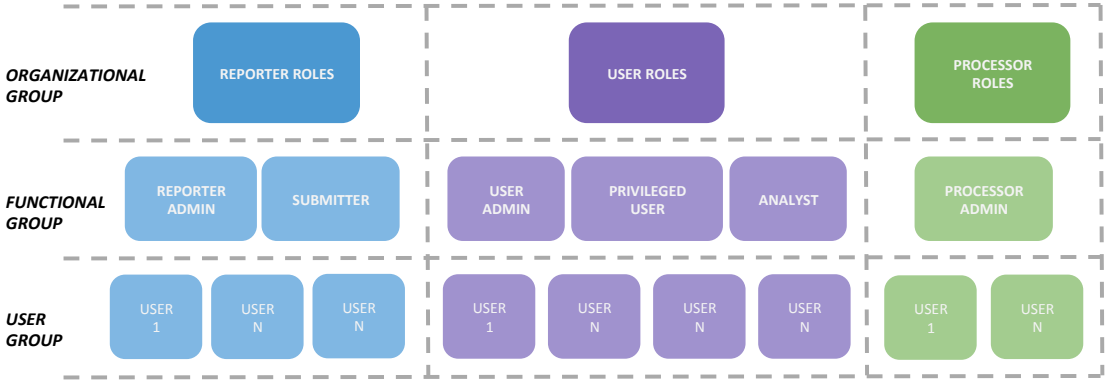
UNIQUE KEY MANAGEMENT APPROACH

- Fully integrating authentication, access control, and data encryption key management
- All users of CAT are issued security tokens
- Each query is signed by the users security tokens
- KMS returns only required keys, and only if query is authenticated as coming from a specific user

ADVANTAGES

- Much lower probability of compromise due to nefarious application
- Lower risk of intentional or accidental compromise due to errors in application logic

Unique encryption and key management approach allows unparalleled control over data access- fine-grained privileges to enforce separation of duties.

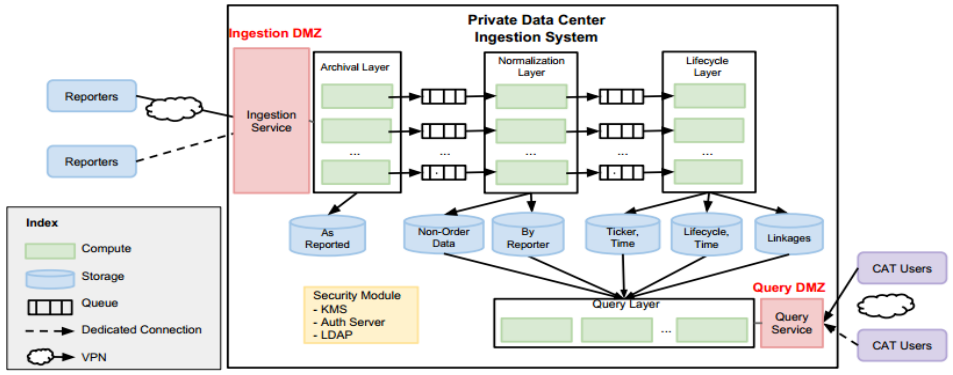


Example Reporter / User Admin Functions

- **Account Management**
  - Add / Modify Accounts
- **Set Fine-Grained Privileges**
  - Access to Data Sets (Level 1 / 2)
  - Actions (Submit / view)
  - Usage (# of queries)

Access, Authentication & Network Security

- **Username / password, MFA tokens**
- **VPN or dedicated line only**
  - **No Public Internet**
- **DMZ with Archival and Query layers**
- **All communication at the boundary is encrypted, logged and monitored**

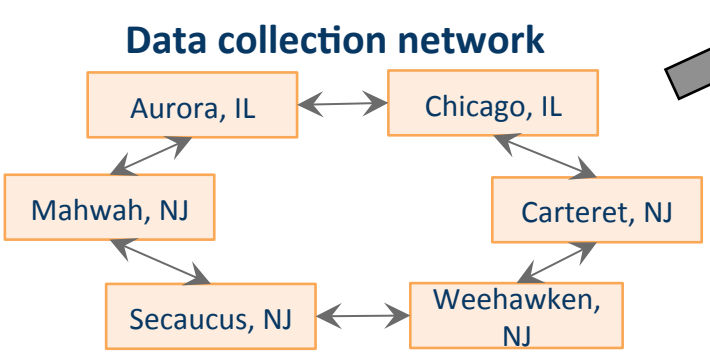


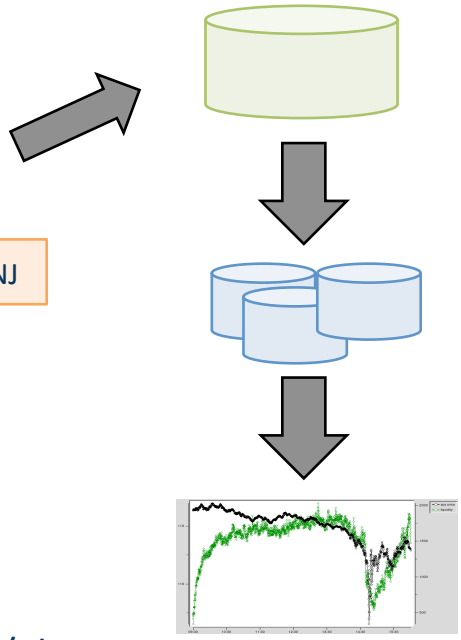
CAT Provider Should Have Domain Expertise

Key Consideration #5: Characteristics of a winning CAT bidder

1) Knows how to collect and manage a petabyte scale financial database
2) Understands how the data will eventually be used (analytics)
3) Understands security and its implications on performance
4) Experience with legacy reporting regimes (such as OATS): understands burden / shortcomings from a reporters point of view

Thesys CAT leverages years of expertise in collecting and analyzing petabytes of financial market data.





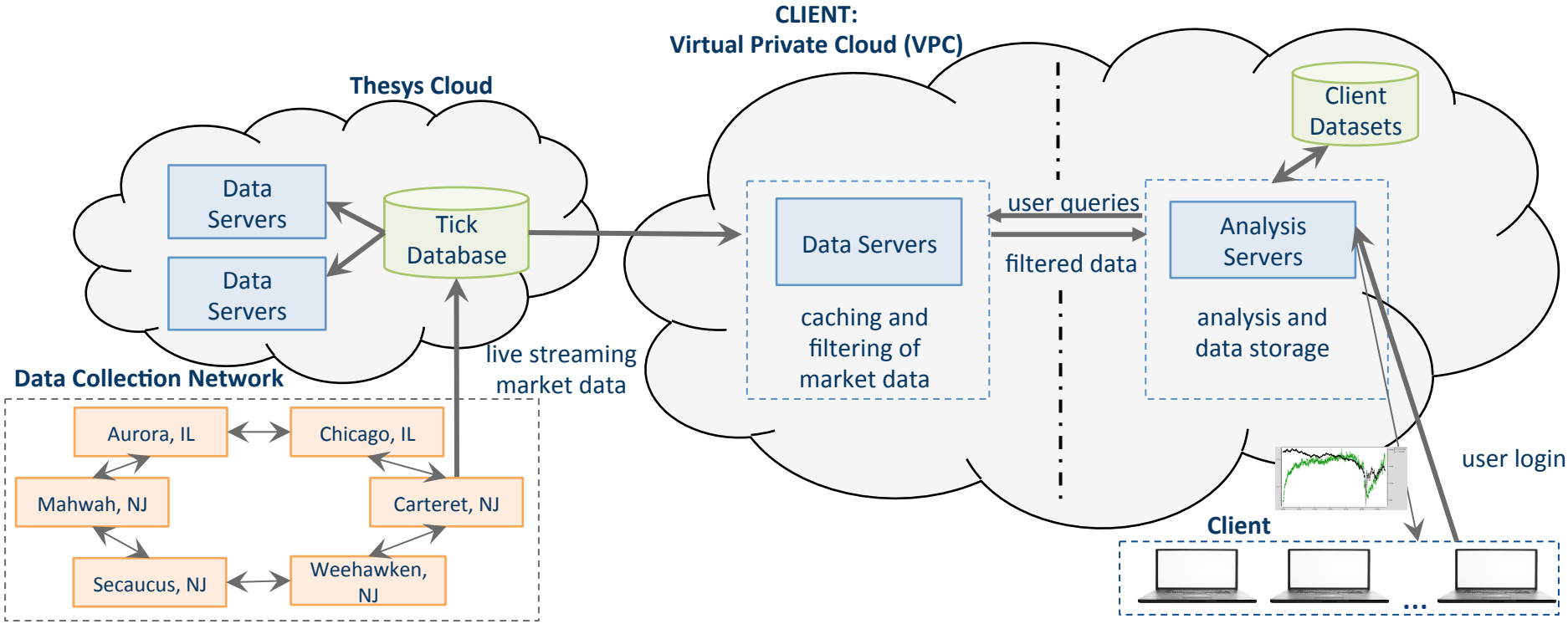
Capture
- Colocation
- Time Sync
- Networking

Storage
- Data Center
- AWS Cloud

Analysis
- Statistical tools
- Visualization Tools

- 6 colocated datacenters
- 15 different protocols
- Millions of messages / second
- 1 TB of market data messages / day
- 15 billion messages / day

MIDAS: SEC Custom Deployment



Thesys Competitive Advantage

METRIC	RFP REQUIREMENT / CURRENT SYSTEM	THESYS SOLUTION
LOW COST	• **Low cost:** $30 million to build, $27 million per year to operate • **Cost savings** arise from applying innovative solutions and leveraging pre-existing technology, not by cutting corners • **Cost savings** for industry due to simple conformance (e.g. FIX/Binary logs, OATS converter, etc.) • Dedicated **CAT-specific input network**	✔
HIGHLY SECURE	• **Battle-tested techniques** to secure access to the system and data at rest and in transit • **Proprietary security innovations** for CAT, including security protocols for authentication and access to data • **Panel of world renowned security experts** to vet the system • **Past organizational experience** and expertise with FISMA moderate qualifications and FedRAMP certification, via the SEC MIDAS project	✔

Thesys Competitive Advantage

METRIC	RFP REQUIREMENT / CURRENT SYSTEM	THESYS SOLUTION
EASY COMPLIANCE / LOW INDUSTRY BURDEN	• **Philosophy:** Avoid burdening reporters with duplicative tasks, wherever possible • **Connectivity:** Connect and interact in multiple ways - e.g. Web GUI, Web API, etc. • **Simple reporting format**, handful of data fields, reduced error rates and reduced fines • **Tools:** OATS-CAT converter, pre-submission validation tool; runs and debugs locally	✔
SUPERIOR TECH CHARACTERISTICS	• **Highly-optimized proprietary big data architecture** designed to enable ultra-efficient analytical operations • **High-Performance**, fast ingestion, processing, queries • Massively **Scalable** • **Flexible Deployment Scenarios**: Pure Public vs Pure Private vs Hybrid	✔

QUESTIONS?