



Getting your FIX: Developing an add-on for the Financial Information eXchange Protocol

Josef Kuepker
Staff Security Specialist | Splunk
josef@splunk.com

splunk>

.conf19

Forward-Looking Statements

During the course of this presentation, we may make forward-looking statements regarding future events or the expected performance of the company. We caution you that such statements reflect our current expectations and estimates based on factors currently known to us and that actual events or results could differ materially. For important factors that may cause actual results to differ from those contained in our forward-looking statements, please review our filings with the SEC.

The forward-looking statements made in this presentation are being made as of the time and date of its live presentation. If reviewed after its live presentation, this presentation may not contain current or accurate information. We do not assume any obligation to update any forward-looking statements we may make. In addition, any information about our roadmap outlines our general product direction and is subject to change at any time without notice. It is for informational purposes only and shall not be incorporated into any contract or other commitment. Splunk undertakes no obligation either to develop the features or functionality described or to include any such feature or functionality in a future release.

Splunk, Splunk>, Listen to Your Data, The Engine for Machine Data, Splunk Cloud, Splunk Light and SPL are trademarks and registered trademarks of Splunk Inc. in the United States and other countries. All other brand names, product names, or trademarks belong to their respective owners. © 2019 Splunk Inc. All rights reserved.

**Making machine data
accessible, usable and
valuable to everyone.**

Agenda

What are we here for?

1. What is FIX
2. Getting FIX Data Into Splunk
3. Developing the Add-on
4. FIX Use Cases

What Is FIX

A quick primer on what the FIX protocol is used for and what it looks like.

.conf19

splunk>



“The Financial Information eXchange (FIX) protocol is an electronic communications protocol initiated in 1992 for international real-time exchange of information related to securities transactions and markets.”

ref: https://en.wikipedia.org/wiki/Financial_Information_eXchange

What is FIX

Some quick metrics about FIX

**100
s**

Institutions
using FIX
Protocol

**100
0s**

Symbols
(Stocks,
Commodities,
etc.) Traded

**10M
+**

Daily Unique
Trades

**1.5
B+**

Daily Shares
Traded

**\$100
B+**

Daily
Dollar Value
of Trades

What Does FIX Look Like?

Multiple formats

We are covering this today

- ▶ Originally a “tagvalue” format
 - A numeric tag with associated value
 - “Tagvalues” separated by Ascii Start of Header (SOH [0x01]) character
 - Sometime the SOH is replaced for readability

We are not coving these today

- ▶ Evolved into other formats
 - XML
 - JSON
 - Google Protocol Buffer
 - Simple Binary Encoding

What Does FIX Look Like

Some examples



Traditional FIX message:

```
8=FIX.4.29=6535=A49=SERVER56=CLIENT34=17752=20090107-18:15:1698=0108=3010=062
```

Some tools display as:

```
8=FIX.4.2•9=65•35=A•49=SERVER•56=CLIENT•34=177•52=20090107-18:15:16•98=0•108=30•10=062•
```

Simplified FIX message representation:

```
8=FIX.4.2|9=65|35=A|49=SERVER|56=CLIENT|34=177|52=20090107-18:15:16|98=0|108=30|10=062|
```

How Do I Get FIX Data Into Splunk?

Some ways are better
than others

- ▶ Flat files
 - Slow, not recommended except in limited use cases
- ▶ Syslog
 - Better, but the volume of data may lead to issues
- ▶ Kafka
 - Good, fast and can handle high capacity (generally)
- ▶ DSP
 - Great, similar to Kafka in speed and capacity, but adds many other capabilities

Developing the FIX Add-on

.conf19
splunk>



Where did I start?

Examining the a FIX Dictionary

Tag	Field Name	XML Name	Data Type	Union Datatype	Description	Added	Depr.	Enums from tag
35	MsgType	@MsgTyp	String		Defines message type ALWAYS THIRD FIELD IN MESSAGE. (Always unencrypted) Note: A "U" as the first character in the MsgType field (i.e. U, U2, etc) indicates that the message format is privately defined between the sender and receiver. *** Note the use of lower case letters ***	Added FIX.2.7		
					0 = Heartbeat [Heartbeat] <i>The heartbeat monitors the status of the communication link and identifies when the last of a string of messages was not received.</i>			
					1 = TestRequest [TestRequest] <i>The test request message forces a heartbeat from the opposing application. The test request message checks sequence numbers or verifies communication line status. The opposite application responds to the Test Request with a Heartbeat containing the TestReqID.</i>			
					2 = ResendRequest [ResendRequest] <i>The resend request is sent by the receiving application to initiate the retransmission of messages. This function is utilized if a sequence number gap is detected, if the receiving application lost a message, or as a function of the initialization process.</i>			
					3 = Reject [Reject] <i>The reject message should be issued when a message is received but cannot be properly processed due to a session-level rule violation. An example of when a reject may be appropriate would be the receipt of a message with invalid basic data which successfully passes de-encryption, CheckSum and BodyLength checks.</i>			
					4 = SequenceReset [SequenceReset] <i>The sequence reset message is used by the sending application to reset the incoming sequence number on the opposing side.</i>			
					5 = Logout [Logout] <i>The logout message initiates or confirms the termination of a FIX session. Disconnection without the exchange of logout messages should be interpreted as an abnormal condition.</i>			
					6 = IOI [IOI] <i>Indication of interest messages are used to market merchandise which the broker is buying or selling in either a proprietary or agency capacity. The indications can be time bound with a specific expiration value. Indications are distributed with the understanding that other firms may react to the message first and that the merchandise may no longer be available due to prior trade. Indication messages can be transmitted in various transaction types; NEW, CANCEL, and REPLACE. All message types other than NEW modify the state of the message identified in IOIRefID.</i>			

So, there is a dictionary?

Definitions and more

Field or Component	Field Name	FIXML name	Req'd	Comments	Depr.
Component	StandardHeader	BaseHeader	✓	MsgType = D	
11	CIOrdID	@ID	✓	Unique identifier of the order as assigned by institution or by the intermediary (CIV term, not a hub/service bureau) with closest association with the investor.	
2422	OrderRequestID	@OrdReqID			
526	SecondaryCIOrdID	@ID2			
583	CIOrdLinkID	@LnkID			
Component	Parties	Pty		This is party information related to the submitter of the request.	
Component	TargetParties	TgtPty		Identifies parties not directly associated with or owning the order, who are to be informed to effect processing of the order.	
229	TradeOriginationDate	@OrigDt			
75	TradeDate	@TrdDt			
1	Account	@Acct			
660	AcctIDSource	@AcctIDSrc			
581	AccountType	@AcctTyp		Type of account associated with the order (Origin)	
589	DayBookingInst	@DayBkngInst			
590	BookingUnit	@BkngUnit			
591	PreallocMethod	@PreallocMeth			
70	AllocID	@AllocID		Used to assign an overall allocation id to the block of preallocations	
Component	PreAllocGrp	PreAll		Number of repeating groups for pre-trade allocation	
63	SettlType	@SettlTyp		For NDFs either SettlType or SettlDate should be specified.	
64	SettlDate	@SettlDt		Takes precedence over SettlType value and conditionally required/omitted for specific SettlType values. For NDFs either SettlType or SettlDate should be specified.	
544	CashMargin	@CshMgn			
635	ClearingFeeIndicator	@ClrFeeInd			
21	HandlInst	@HandlInst			
18	ExecInst	@ExecInst		Can contain multiple instructions, space delimited. If OrdType=P, exactly one of the following values (ExecInst = L, R, M, P, O, T, W, a, d) must be specified.	
1805	AuctionInstruction	@AuctInst			
110	MinQty	@MinQty			

What about Schema?

Yep, that's there too.

Field or Component	Field Name	FIXML name	Req'd	Comments
Component	StandardHeader	BaseHeader	✓	MsgType = D
11	CIOrdID	@ID	✓	Unique identifier of the order as assigned by institution or by the intermediary (CIV term, not a hub/service bureau) with closest association with the investor.
2422	OrderRequestID	@OrdReqID		
526	SecondaryCIOrdID	@ID2		
583	CIOrdLinkID	@LnkID		
Component(-)	Parties	Pty		This is party information related to the submitter of the request.
<i>Repeating Group 453</i>	NoPartyIDs			Repeating group below should contain unique combinations of PartyID, PartyIDSource, and PartyRole
448	PartyID	@ID		Required if NoPartyIDs(453) > 0. Identification of the party.
447	PartyIDSource	@Src		Required if NoPartyIDs(453) > 0. Used to identify classification source.
452	PartyRole	@R		Required if NoPartyIDs(453) > 0. Identifies the type of PartyID(448).
2376	PartyRoleQualifier	@Qual		
Component	PtysSubGrp	Sub		Repeating group of Party sub-identifiers.
<i>end Repeating Group</i>				
<i>end Component</i>				
Component(-)	TargetParties	TgtPty		Identifies parties not directly associated with or owning the order, who are to be informed to effect processing of the order.
<i>Repeating Group 1461</i>	NoTargetPartyIDs			Repeating group below should contain unique combinations of TargetPartyID, TargetPartyIDSource, and TargetPartyRole.
1462	TargetPartyID	@ID		Required if NoTargetPartyIDs(1461) > 0. Used to identify the party targeted for the action specified in the message.
1463	TargetPartyIDSource	@Src		Used to identify source of target party identifier.
1464	TargetPartyRole	@R		Used to identify the role of source party identifier.
1818	TargetPartyRoleQualifier	@Qual		Used to further qualify the role of the target party role.
Component	TargetPtysSubGrp	Sub		Repeating group of target party sub-identifiers.
<i>end Repeating Group</i>				
<i>end Component</i>				
229	TradeOriginationDate	@OrignDt		
75	TradeDate	@TrdDt		

Coding Configs

Using Python to create configs

- ▶ Python was used to create configs
 - Props
 - Transforms
- ▶ Generating Lookups as well
 - Lookups help translate the FIX values to human values
 - For OrdStatus = Order Status
 - Value 0 = New
 - Value 4 = Cancelled

```
def gen_alias(fpath, filename, outputProps):  
    url = "file://" + fpath + filename  
  
    page = urllib2.urlopen(url)  
    soup = BeautifulSoup(page.read(), convertEntities=BeautifulSoup.HTML_ENTITIES)  
    fieldname = soup.findAll('td', {'class': re.compile(r'(FldName)')})  
    tagname = fieldname[0].text  
    tag = re.search('tag(\d+)', filename)  
    foundtag = tag.group(1)  
    outputProps.write("FIELDALIAS-" + tagname + " = " + str(foundtag) + " AS " + tagname + "\n")  
    page.close()
```

A Closer Look At A Config Generator

```
import re
import urllib2

from BeautifulSoup import BeautifulSoup

def gen_alias(fpath, filename, outputProps):
    url = "file://" + fpath + filename

    page = urllib2.urlopen(url)
    soup = BeautifulSoup(page.read(), convertEntities=BeautifulSoup.HTML_ENTITIES)
    fieldname = soup.findAll('td', {'class': re.compile(r'(FldName)')})
    tagname = fieldname[0].text
    tag = re.search('tag(\d+)', filename)
    foundtag = tag.group(1)
    outputProps.write("FIELDALIAS-" + tagname + " = " + str(foundtag) + " AS " + tagname + "\n")
    page.close()

def gen_lookup(filename, outputProps):
    #for xy in filename:
    if re.match('(\d+)_(\^[_]+)_lookup.csv', filename):
        tag = re.search('(\d+)_(\^[_]+)_lookup', filename)
        foundtag = tag.group(1)
        foundname = tag.group(2)
        outputProps.write(
            "LOOKUP-" + foundname + " = " + foundtag + "_" + foundname + "_lookup" + " " + foundname + \
            " OUTPUTNEW " + foundname + "Description\n")
```

Challenges: Not All Was Well

What has issues and why?

- ▶ Issue: **5704** Enabled Field Aliases Causes Search Latency
 - Field aliases make Splunk think the fields are mandatory
 - There are not 5000+ fields in each event, but Splunk thinks there are
 - Splunk keeps looking for the fields similar to Verbose Mode
 - Fix: Enable the Aliases that you need
 - Many of the FIX Tags are not used often
- ▶ Issue: **1818** individual lookups are a performance hit (especially automatic)
 - Resolution: I combined all the lookups into one
 - This limited me in how I can do the lookups
 - Looking into a custom command to do a high performance lookup

The Add-on

Tags and Values
Parsed

000 PM

8=FIX.4.1 9=138 35=8 34=7 49=EXEC 52=20180307-23:25:12 56=BANZAI 6=0 11=1352157912357 14=0 17=5 20=0 31=0 3

Event Actions ▾

Type	☒	Field	Value	Actions
Selected	☒	host ▾	fix_engine	▾
	☒	source ▾	fixOut.1.log	▾
	☒	sourcetype ▾	fix:msg	▾
Event	☐	10 ▾	252	▾
	☐	11 ▾	1352157912357	▾
	☐	14 ▾	0	▾
	☐	150 ▾	2	▾
	☐	151 ▾	0	▾
	☐	17 ▾	5	▾
	☐	20 ▾	0	▾
	☐	31 ▾	0	▾
	☐	32 ▾	0	▾
	☐	34 ▾	7	▾
	☐	35 ▾	8	▾
	☐	37 ▾	5	▾
	☐	38 ▾	10000	▾
	☐	39 ▾	0	▾
	☐	40 ▾	0	▾

The Add-on

Tags with Aliases

☐	AvgPx ▼	0	▼
☐	BeginString ▼	FIX.4.1	▼
☐	BodyLength ▼	138	▼
☐	Checksum ▼	252	▼
☐	ClOrdID ▼	1352157912357	▼
☐	CumQty ▼	0	▼
☐	ExecID ▼	5	▼
☐	ExecType ▼	2	▼
☐	LastPx ▼	0	▼
☐	LastQty ▼	0	▼
☐	LeavesQty ▼	0	▼
☐	MsgSeqNum ▼	7	▼
☐	MsgType ▼	8	▼
☐	OrdStatus ▼	0	▼
☐	OrderID ▼	5	▼
☐	OrderQty ▼	10000	▼
☐	SenderCompID ▼	EXEC	▼
☐	SendingTime ▼	20180307-23:25:12	▼
☐	Side ▼	1	▼

FIX Add-on Use Cases

What can we do with this data?



FIX Investigation Panels

Basic Investigations

FIX INVESTigator
Edit
Export ▾
...

Time Range
Info Level

All time ▾
☒ Limited
☐ Full
Hide Filters

Events

SendingTime ↕	MsgType ↕	MsgTypeDescription ↕	SenderComplID ↕	TargetComplID ↕	Text ↕	eventId ↕
20180907-20:00:44.655	A	Logon The logon message authenticates a user establishing a connection to a remote system. The logon message must be the first message sent by the application requesting to initiate a FIX session.	andpeggy_fx	aham_fx		main_2:22_jkuepker-mbp-12d13
20180907-20:00:44.568	A	Logon The logon message authenticates a user establishing a connection to a remote system. The logon message must be the first message sent by the application requesting to initiate a FIX session.	aham_fx	andpeggy_fx		main_2:16_jkuepker-mbp-12d13
20180907-04:05:01.021	5	Logout The logout message initiates or confirms the termination of a FIX session. Disconnection without the exchange of logout messages should be interpreted as an abnormal condition.	andpeggy_fx	aham_fx		main_2:12_jkuepker-mbp-12d13
20180907-04:05:01.021	5	Logout The logout message initiates or confirms the termination of a FIX session. Disconnection without the exchange of logout messages should be interpreted as an abnormal condition.	aham_fx	andpeggy_fx		main_2:8_jkuepker-mbp-12d13
20180308-12:05:06	A	Logon The logon message authenticates a user establishing a connection to a remote system. The logon message must be the first message sent by the application requesting to initiate a FIX session.	INWMGR	BRKR		main_0:157_jkuepker-mbp-12d13

« Prev
1
2
3
4
5
6
7
Next »

Raw Events

Q
⏏
i
⌂
<1m ago

i	Time	Event
>	9/7/18 4:00:44.000 PM	8=FIX.4.4 9=112 35=A 49=andpeggy_fx 56=aham_fx 34=1 50=PJM 57=JCK 52=20180907-20:00:44.655 98=0 108=30 141=Y 10=201 host = fix.tarader.local source = logon_logout.sample sourcetype = fix:msg
>	9/7/18 4:00:44.000 PM	8=FIX.4.4 9=136 35=A 49=aham_fx 56=andpeggy_fx 34=1 52=20180907-20:00:44.568 98=0 108=30 141=Y 553=API_AHAM_BTHOMAS 554=XXX3t3! 10=19 host = fix.tarader.local source = logon_logout.sample sourcetype = fix:msg
>	9/7/18 12:05:01.000 AM	8=FIX.4.4 9=112 35=5 49=andpeggy_fx 56=aham_fx 34=0 52=20180907-04:05:01.021 10=157 host = fix.tarader.local source = logon_logout.sample sourcetype = fix:msg
>	9/7/18 12:05:01.000 AM	8=FIX.4.4 9=112 35=5 49=aham_fx 56=andpeggy_fx 34=0 52=20180907-04:05:01.021 10=157 host = fix.tarader.local source = logon_logout.sample sourcetype = fix:msg

FIX Investigation Panels

Friendly Parsed Data

FIX INVESTigator

Time Range

All time

Info Level

☒ Limited
 ☐ Full

Hide Filters

Edit

Export

...

Events

SendingTime	MsgType	MsgTypeDescription	SenderCompID	TargetCompID	Text	eventId
20180907-20:00:44.655	A	Logon The logon message authenticates a user establishing a connection to a remote system. The logon message must be the first message sent by the application requesting to initiate a FIX session.	andpeggy_fx	aham_fx		main_2:22_jkuepker-mbp-12d13
20180907-20:00:44.568	A	Logon The logon message authenticates a user establishing a connection to a remote system. The logon message must be the first message sent by the application requesting to initiate a FIX session.	aham_fx	andpeggy_fx		main_2:16_jkuepker-mbp-12d13
20180907-04:05:01.021	5	Logout The logout message initiates or confirms the termination of a FIX session. Disconnection without the exchange of logout messages should be interpreted as an abnormal condition.	andpeggy_fx	aham_fx		main_2:12_jkuepker-mbp-12d13
20180907-04:05:01.021	5	Logout The logout message initiates or confirms the termination of a FIX session. Disconnection without the exchange of logout messages should be interpreted as an abnormal condition.	aham_fx	andpeggy_fx		main_2:8_jkuepker-mbp-12d13
20180308-12:05:06	A	Logon The logon message	INVMGR	BRKR		main_0:157_jkuepker-mbp-12d13

Event Breakout

fixTag	fixTagName	fixTagValue	fixTagValueDescription	fixTagDescription
8	BeginString	FIX.4.4	-	Identifies beginning of new message and protocol version. ALWAYS FIRST FIELD IN MESSAGE. (Always unencrypted)Valid values:FIXT.1.1
9	BodyLength	112	-	Message length, in bytes, forward to the CheckSum field. ALWAYS SECOND FIELD IN MESSAGE. (Always unencrypted)
98	EncryptMethod	0	None / Other	Method of encryption.
108	HeartBtInt	30	-	Heartbeat interval (seconds)
34	MsgSeqNum	1	-	Integer message sequence number.
35	MsgType	A	Logon The logon message authenticates a user establishing a connection to a remote system. The logon message must be the first message sent by the application requesting to initiate a FIX session.	Defines message type ALWAYS THIRD FIELD IN MESSAGE. (Always unencrypted)Note: A "U" as the first character in the MsgType field (i.e. U, U2, etc) indicates that the message format is privately defined between the sender and receiver.*** Note the use of lower case letters ***
141	ResetSeqNumFlag	Y	Yes, reset sequence numbers	Indicates that both sides of the FIX session should reset sequence numbers.
49	SenderCompID	andpeggy_fx	-	Assigned value used to identify firm sending message.
50	SenderSubID	PJM	-	Assigned value used to identify specific message originator (desk, trader, etc.)
52	SendingTime	20180907-20:00:44.655	-	Time of message transmission (always expressed in UTC (Universal Time Coordinated, also known as "GMT")
56	TargetCompID	aham_fx	-	Assigned value used to identify receiving firm.

FIX Analyzer with Checksum Validation Command

Quickly Examine Individual FIX Messages

FIX Analyzer
Edit
Export ▾
...

Enter a raw FIX message in the box below.

FIX Message

8=FIX.4.1 9=153 35=8 34=4 49=EXEC 52=20180307-23:24:42 56=BANZAI 6=12.3 11=1352157882577 14=10000 17=2 20=0 31=12.3 32=10000 37=2 38=10000 39=2 54=1 55=MSFT 150=2 151=0 10=230

Submit
Hide Filters

FIX Checksum Validation

fixIsValid ▾

INVALID

FIX Message Analysis

fixTag ▾	fixTagName ▾	fixTagValue ▾	fixTagValueDescription ▾	fixTagDescription ▾
8	BeginString	FIX.4.1	-	Identifies beginning of new message and protocol version. ALWAYS FIRST FIELD IN MESSAGE. (Always unencrypted)Valid values:FIXT.1.1
9	BodyLength	153	-	Message length, in bytes, forward to the CheckSum field. ALWAYS SECOND FIELD IN MESSAGE. (Always unencrypted)
35	MsgType	8	ExecutionReport The execution report message is used to: 1. confirm the receipt of an order 2. confirm changes to an existing order (i.e. accept cancel and replace requests) 3. relay order status information 4. relay fill information on working orders 5. relay fill information on tradeable or restricted tradeable quotes 6. reject orders 7. report post-trade fees calculations associated with a trade	Defines message type ALWAYS THIRD FIELD IN MESSAGE. (Always unencrypted)Note: A "U" as the first character in the MsgType field (i.e. U, U2, etc) indicates that the message format is privately defined between the sender and receiver.*** Note the use of lower case letters ***
34	MsgSeqNum	4	-	Integer message sequence number.
49	SenderCompID	EXEC	-	Assigned value used to identify firm sending message.
52	SendingTime	20180307-23:24:42	-	Time of message transmission (always expressed in UTC (Universal Time Coordinated, also known as "GMT")

Other FIX Use Cases

What are some other

- ▶ Most popular symbol
- ▶ Trading anomalies
- ▶ Long running orders
- ▶ Top cancellation reasons
- ▶ Top reject reasons
- ▶ Client Metrics
- ▶ Invalid Checksums
- ▶ Unusual number of logins
- ▶ Infrastructure Errors
- ▶ Connectivity Issues
 - Drop in number of FIX events
- ▶ Correlation of FIX Trades with News Event

Key Takeaways

What did we cover?

1. FIX Protocol is a complex data format
2. FIX is used heavily in Financial Services
3. Building an Splunk Add-on for complex data may be faster with some code
4. There are a ton of high-value use cases for FIX data in Splunk



**Thank
You!**