AIS Parser

- Prerequisite: Install gdal library/tool for GDB Parser
- - brew install gdal
- Prerequisite: command to Convert GDB into CSV
- - ogr2ogr -f CSV output.csv ./Zone1_2009_01.gdb -lco GEOMETRY=AS_XYZ
- Parse the 2015-2017 flat csv and transform data into Vessel, Broadcast, and Voyage files to be uniform with the GDB Exported Data
- Rename all the columns to a unified naming convention and select only the common attributes across all zones/years

In [94]:
```
# from IPython.display import Image, HTML
import os
import numpy as np
import pandas as pd
import datetime
import warnings

warnings.filterwarnings("ignore")           # Suppress Warning
```

In [95]:
```
# https://afdata.s3.us-gov-west-1.amazonaws.com/index.html#Scenario_Data/
# Prerequisite:
# Download raw data to local folder
# Unzip everything into ./csv/AIS/ Folder

WorkingFolder = '/data/vessel data/'
```

Process Flat AIS Data

In [151]:
```
OutputDir = WorkingFolder + "AIS_Processed/Zone01_2016_01/"

if not os.path.exists(OutputDir):
    os.mkdir(OutputDir)
```

In [152]:
```
ais = pd.read_csv(WorkingFolder + "AIS/AIS_2016_01_Zone01.csv", sep=",")
ais.head(3)
```

Out[152]:

	MMSI	BaseDateTime	LAT	LON	SOG	COG	Heading	VesselName	IMO
0	367303490	2016-01-01T00:08:23	51.86139	-176.63799	0.1	3.6	359.0	RESOLVE PIONEER	IMO7528843
1	367303490	2016-01-01T00:23:24	51.86144	-176.63798	0.0	3.6	359.0	RESOLVE PIONEER	IMO7528843
2	367303490	2016-01-01T00:47:22	51.86144	-176.63799	0.1	3.6	0.0	RESOLVE PIONEER	IMO7528843

In [153]:
```
ais.shape
```

Out[153]:
```
(21426, 16)
```

In [154]:
```
# ais.columns
```

In [155]:
```
ais.rename(columns={'MMSI':'mmsi_id', 'BaseDateTime':'date_time', 'LAT':'lat
ais.head(3)
```

Out[155]:

	mmsi_id	date_time	lat	lon	speed_over_ground	course_over_ground	heading
0	367303490	2016-01-01T00:08:23	51.86139	-176.63799	0.1	3.6	359.0
1	367303490	2016-01-01T00:23:24	51.86144	-176.63798	0.0	3.6	359.0
2	367303490	2016-01-01T00:47:22	51.86144	-176.63799	0.1	3.6	0.0

In [156]:
```
# Voyages got collapsed
ais['voyage_id'] = np.nan
```

In [157]:
```
# ais.columns
```

In [158]:
```
# Extract Broadcast Data
Broadcast = ais[['mmsi_id', 'date_time', 'lat', 'lon', 'speed_over_ground',
Broadcast.head()
```

Out[158]:

	mmsi_id	date_time	lat	lon	speed_over_ground	course_over_ground	voyage_id
0	367303490	2016-01-01T00:08:23	51.86139	-176.63799	0.1	3.6	NaN
1	367303490	2016-01-01T00:23:24	51.86144	-176.63798	0.0	3.6	NaN
2	367303490	2016-01-01T00:47:22	51.86144	-176.63799	0.1	3.6	NaN
3	367303490	2016-01-01T01:02:23	51.86144	-176.63791	0.0	3.6	NaN
4	367303490	2016-01-01T01:08:21	51.86143	-176.63791	0.0	3.6	NaN

In [159]:
```
# Extract Vessel Data
Vessel = ais[['mmsi_id', 'imo', 'call_sign', 'vessel_name', 'vessel_type','
Vessel.head()
```

Out[159]:

	mmsi_id	imo	call_sign	vessel_name	vessel_type	length	width
0	367303490	IMO7528843	WDD8846	RESOLVE PIONEER	1005.0	63.23	12.25
396	366947000	IMO8225101	WRC6707	NORTHERN GLACIER	1001.0	61.27	13.72
428	357058000	IMO9227601	3FWC9	NO.2 POHAH	70.0	120.00	16.00
523	413478230	IMO9608427	BOFD	DA XIN	1004.0	179.57	28.00
524	338626000	IMO8213225	WDG2215	ALASKA PROVIDER	1001.0	54.00	12.20

In [160]:
```
# Extract Voyage Data
Voyage = ais[['mmsi_id', 'voyage_id', 'draft', 'cargo']].drop_duplicates()
Voyage.head()
```

Out[160]:

	mmsi_id	voyage_id	draft	cargo
0	367303490	NaN	4.5	NaN
396	366947000	NaN	5.3	NaN
428	357058000	NaN	5.3	70.0
523	413478230	NaN	10.5	70.0
524	338626000	NaN	3.8	NaN

In [161]:
```
# Output result to csv
Broadcast.to_csv(OutputDir + "Broadcast.csv", index=False)
Vessel.to_csv(OutputDir + "Vessel.csv", index=False)
Voyage.to_csv(OutputDir + "Voyage.csv", index=False)
```

Process GDB Format

In [224]:
```
GDB_Folder = WorkingFolder + "AIS/Zone10_2014_01.gdb/"
OutputDir = WorkingFolder + "AIS_Processed/Zone10_2014_01/"

if not os.path.exists(OutputDir):
    os.mkdir(OutputDir)
```

In [225]:
```
Broadcast = pd.read_csv(GDB_Folder + "Broadcast.csv", sep=",")
Broadcast.head(3)
```

Out[225]:

	X	Y	Z	SOG	COG	Heading	ROT	BaseDateTime	Status	VoyageID	
0	-122.361145	47.581332	0	0.0	39.599998	511	128	2013/12/31 23:57:44	0	1	3
1	-123.990592	45.835737	0	6.7	355.399990	359	129	2013/12/31 23:57:44	15	2	3
2	-122.382117	47.631067	0	0.0	192.100010	180	0	2013/12/31 23:57:44	7	3	3

In [226]:
```
Broadcast.rename(columns={'MMSI':'mmsi_id', 'BaseDateTime':'date_time', 'Y'
Broadcast.head(3)
```

Out[226]:

	lon	lat	Z	speed_over_ground	course_over_ground	heading	ROT	date_time	s
0	-122.361145	47.581332	0	0.0	39.599998	511	128	2013/12/31 23:57:44	
1	-123.990592	45.835737	0	6.7	355.399990	359	129	2013/12/31 23:57:44	
2	-122.382117	47.631067	0	0.0	192.100010	180	0	2013/12/31 23:57:44	

In [227]:
```
# Broadcast.shape
```

In [228]:
```
# Extract Broadcast Data
Broadcast = Broadcast[['mmsi_id', 'date_time', 'lat', 'lon', 'speed_over_gr
Broadcast.head()
```

Out[228]:

	mmsi_id	date_time	lat	lon	speed_over_ground	course_over_ground	voyage_
0	366025993	2013/12/31 23:57:44	47.581332	-122.361145	0.0	39.599998	
1	367160890	2013/12/31 23:57:44	45.835737	-123.990592	6.7	355.399990	
2	366490600	2013/12/31 23:57:44	47.631067	-122.382117	0.0	192.100010	
3	338000406	2013/12/31 23:57:44	48.123443	-123.444115	0.0	14.200000	
4	367840001	2013/12/31 23:57:44	48.121267	-122.726412	11.4	55.400002	

In [229]:
```
# Broadcast.shape
```

In [230]:
```
Vessel = pd.read_csv(GDB_Folder + "Vessel.csv", sep=",")
Vessel.head(3)
```

Out[230]:

	MMSI	IMO	CallSign	Name	VesselType	Length	Width	DimensionComponents
0	235469970	NaN	NaN	NaN	71.0	365.0	52.0	145,220,26,26
1	367000504	NaN	NaN	NaN	60.0	22.0	7.0	10,12,4,3
2	367870800	NaN	NaN	NaN	32.0	35.0	12.0	5,30,6,6

In [231]:
```
Vessel.rename(columns={'MMSI':'mmsi_id', 'Name':'vessel_name', 'IMO':'imo',
Vessel.head(3)
```

Out[231]:

	mmsi_id	imo	call_sign	vessel_name	vessel_type	length	width	DimensionComponents
0	235469970	NaN	NaN	NaN	71.0	365.0	52.0	145,220,26,26
1	367000504	NaN	NaN	NaN	60.0	22.0	7.0	10,12,4,3
2	367870800	NaN	NaN	NaN	32.0	35.0	12.0	5,30,6,6

In [232]:
```
# Extract Vessel Data
Vessel = Vessel[['mmsi_id', 'imo', 'call_sign', 'vessel_name', 'vessel_type
Vessel.head()
```

Out[232]:

	mmsi_id	imo	call_sign	vessel_name	vessel_type	length	width
0	235469970	NaN	NaN	NaN	71.0	365.0	52.0
1	367000504	NaN	NaN	NaN	60.0	22.0	7.0
2	367870800	NaN	NaN	NaN	32.0	35.0	12.0
3	366281509	NaN	NaN	NaN	90.0	25.0	5.0
4	576110500	NaN	NaN	NaN	70.0	175.0	28.0

In [233]:
```
Voyage = pd.read_csv(GDB_Folder + "Voyage.csv", sep=",")
Voyage.head(3)
```

Out[233]:

	VoyageID	Destination	Cargo	Draught	ETA	StartTime	EndTime	MMSI
0	337	OAKLAND	71	120	2014/01/02 07:30:00	2014/01/01 00:00:00	2014/01/04 14:33:07	235469970
1	232	SFO ANCH 9	32	45	2014/12/30 18:00:00	2014/01/01 00:00:00	2014/01/31 15:38:20	367870800
2	113	TIANJIN,CHINA	70	93	2014/01/21 06:00:00	2014/01/01 00:00:04	NaN	576110500

In [234]:
```
Voyage.rename(columns={'MMSI':'mmsi_id', 'VoyageID':'voyage_id', 'Draught':
Voyage.head(3)
```

Out[234]:

	voyage_id	Destination	cargo	draft	ETA	StartTime	EndTime	mmsi_id
0	337	OAKLAND	71	120	2014/01/02 07:30:00	2014/01/01 00:00:00	2014/01/04 14:33:07	235469970
1	232	SFO ANCH 9	32	45	2014/12/30 18:00:00	2014/01/01 00:00:00	2014/01/31 15:38:20	367870800
2	113	TIANJIN,CHINA	70	93	2014/01/21 06:00:00	2014/01/01 00:00:04	NaN	576110500

In [235]:
```
# Extract Voyage Data
Voyage = Voyage[['mmsi_id', 'voyage_id', 'draft', 'cargo']]
Voyage.head()
```

Out[235]:

	mmsi_id	voyage_id	draft	cargo
0	235469970	337	120	71
1	367870800	232	45	32
2	576110500	113	93	70
3	316004579	306	32	52
4	477221200	519	70	70

In [236]:
```
# Output result to csv
Broadcast.to_csv(OutputDir + "Broadcast.csv", index=False)
Vessel.to_csv(OutputDir + "Vessel.csv", index=False)
Voyage.to_csv(OutputDir + "Voyage.csv", index=False)
```