Full reference is available here: https://www1.ncdc.noaa.gov/pub/data/noaa/isd-format-document.pdf

Column Name	Description	Format Details	Example	Notes
STATION	Weather station unique identifier	6 digit USAF id followed by 5 digit WBAN id	12930099999	Only need USAF part to merge with the stations dataset
DATE	UTC DateTime	YYYY-MM-DD**T**HH:MM:SS Time: 00:00:00-23:00:00	2016-11-14T11:00:00	The date is separated from time by 'T'. We will need to convert arrival/departure times in the airlines dataset to UTC format as well.
CALL_SIGN	Call letters assigned to the weather station	A set of ASCII characters. 99999 if missing	TXKF	Can be used to extract airport codes as the last 3 characters of the call
WND	Summarized wind observation	{direction angle,quality code,type code,speed rate} where angle: 001-360 q-code: 0-9 t-code: [A,B,C,H,N,R,Q,T,V,9(if missing)] Speed: 0000-0900 (9999 if missing)	030,1,N,0010,1	We might only need the speed rate. Optionally, may exclude data with erroneous or suspect quality codes.
CIG	Sky condition observation (the height above	{ceiling height,quality code,determination code,CAVOC code}	00000,1,9,9	Might only need the ceiling height. Optionally, may exclude

	ground level of the lowest cloud)	Where height: 00000-22000 (meters), 99999 if missing q-code: 0-9 d-code: [A,B,C,D,E,M,P,R,S,U,V,W,9(if missing)] CAVOK: [N(no),Y(yes),9(missing)]		data with erroneous or suspect quality codes.
VIS	Visibility observation (horizontal distance at which an object can be seen and identified)	{distance,quality code,variability code, quality variability code} Where distance: 000000-160000 (meters), 999999 if missing q-code: 0-9 v-code: [N(not variable), V(variable), 9(missing)] q-vcode: 0-9	000000,1,9,9	Might only need the visibility distance. Optionally, may exclude data with erroneous or suspect quality codes.
TMP	Air temperature (degrees Celsius)	{temperature,quality code} Where temp: -0932 - +0618, (+9999 if missing) q-code: 0-9,A,C,I,M,P,R,U	+0028,1	Might only need the temperature. Optionally, may exclude data with erroneous or suspect quality codes.
DEW	Dew point temperature (degrees Celsius)	{dew,quality code} Where dew: -0982 - +0368, (+9999 if missing) q-code: 0-9,A,C,I,M,P,R,U	-0017,1	Might only need the dew point. Optionally, may exclude data with erroneous or suspect quality codes.

SLP	Atmospheric pressure relative to Mean Sea Level (Hectopascals)	{pressure,quality code} Where pressure: 08600-10900 (99999 if missing) q-code: 0-9	10341,1	Might only need the air pressure. Optionally, may exclude data with erroneous or suspect quality codes.
AA1, AA2	Liquid precipitation amount	{time period,depth,condition code,quality code} Where period: 00-98(hours), 99(missing) depth: 0000-9998(millimeters), 9999(missing) c-code: 1-8,E,I,J,9(missing) q-code: 0-9,A,C,I,M,P,R,U	01,0027,3,1	Many observations have 'null' values in one or both of the columns. When the data is available, we might aggregate it as average precipitation per hour and possibly bin it.
AJ1	Snow amount	{time period,depth,condition code,quality code} Where period: 00-98(hours), 99(missing) depth: 0000-1200(centimeters), 9999(missing) c-code: 1-6,E,9(missing) q-code: 0-9,A,C,I,M,P,R,U	24,0089,9,1	Many observations have 'null' values. When the data is available, we might aggregate it as average precipitation per hour and possibly bin it.
		Of the second half, I think: ● First element of GA2, GA3 ● First element of IA1 (ground condition) The rest of the columns seem to have bad data		

AY1-AY2	Past weather	Hard to interpret, likely not as useful	`AU,13,BR  ,5`	I don't think this variable will be useful

GA1-GA3	Amount of sky covered	{coverage code, coverage quality code, base height dimension, base height quality code, cloud type code, cloud type quality code}
		
		Where
		Coverage code: 00 (none) - 10 (partial obscuration); 99 missing
		Quality code: 0-9
		Base height dimension: -00400 to +35000 in meters (height relative to lowest surface of a cloud); +99999 = missing
		Base height quality code: 0-9
		Cloud type code: 00-23; 99 missing
		Cloud type quality code: 0-9
			`GA1: AW,18,SN  ,1`
			`GA2: 1,0,02,0,0,1,5`
			`GA3: 2,0,02,0,0,3,5`
				Using the coverage code (first element) as a categorical variable seems the most interesting
				GA1 seems misformatted

GE1	Sky condition	{convective cloud attribute, vertical datum attribute, base height upper range attribute, base height lower range attribute} Where Connective cloud attribute: 0-9 describing the cloud type (altocumulus…. etc.)	`62,5`	I think a more detailed version of information in GA*, for simplicity we can ignore.
GF1	Sky condition	{a lot} Where Total coverage code: 0-19; 99 missing	`92,5`	Using the total coverage code as a categorical variable seems the most interesting. This again

				seems misformatted.
IA1	Ground surface observation	{code, quality code} Where Code: 0-31; 99 missing (snow, wet, dry) Quality code: 0-3; 9 passed	`92,1`	Use code as a categorical variable
KA1-KA2	Extreme air temperature	{period quantity, code, air temperature, temperature quality code} Where Period quantity: 001 - 480 hours Air temperature: air temperature in degrees celsius * 10 (+0618 to -0932)	`04,5,+01676,5,99,9` `07,5,+00853,5,99,9`	Unsure. The description and the values don't match up here.
MA1	Atmospheric pressure	{altimeter setting rate, altimeter quality code, station pressure rate, station pressure quality code} Where Alimeter setting rate: 08635-10904 in hectopascals; 99999 is missing Station pressure rate: 04500-10900 in hectopascals; 99999 is missing	`07,5,+01067,5,99,9`	Unsure. Again, the description and the values don't match up here
MD1	Atmospheric pressure change	{a lot} Where Tendency code (first one): 0-8; 9 missing - describes whether atmospheric pressure is increasing,	`08,5,+03048,5,99,9`	The values don't seem to match up with the description

		decreasing… etc.		
MW1	Manual present weather condition	{manual atmospheric condition code, quality manual atmospheric condition code} Where Atmospheric condition code: 0-99 categorical (e.g. thunderstorm, rain… etc.)	`2,99,1,+01676,5,9`	Manual atmospheric condition code looks interesting. The values don't seem to match up with the description though
OC1	Wind gust	{speed rate, quality code} Where Speed rate: 0050 to 1100 in meters per second; 9999 is missing	`4,99,1,+03048,5,9`	Again, the values don't seem to match up with the description
OD1	Supplementary wind observation	{a lot} Where Type code (1): 1-6; the type of the observation Speed rate (3rd element): 0-2000 meters per second	`3,99,1,+01433,5,9`	Speed rate looks the most interesting, but again the values don't match up with the description
SA1	Sea surface temperature	{temperature, quality code}	`9,AGL ,+99999,+99999`	Again the values don't match up with the description
UA1	Wave measurement	{method code, wave period quantity, wave height dimension, quality code, sea state code, sea state code quality code}	`04,99,5,99,9,99,9,01676,1,99,9,99,9`	I don't think this variable will be useful

REM	Remarks	No key	10261,5,10067,5	I don’t think this variable will be useful
EQD	Element quality data	No key	01,5	I don’t think this variable will be useful