Integration of Files with Alation Data Catalog

Information Asset has developed a solution to parse and transfer files of different file systems to Alation Data Catalog.
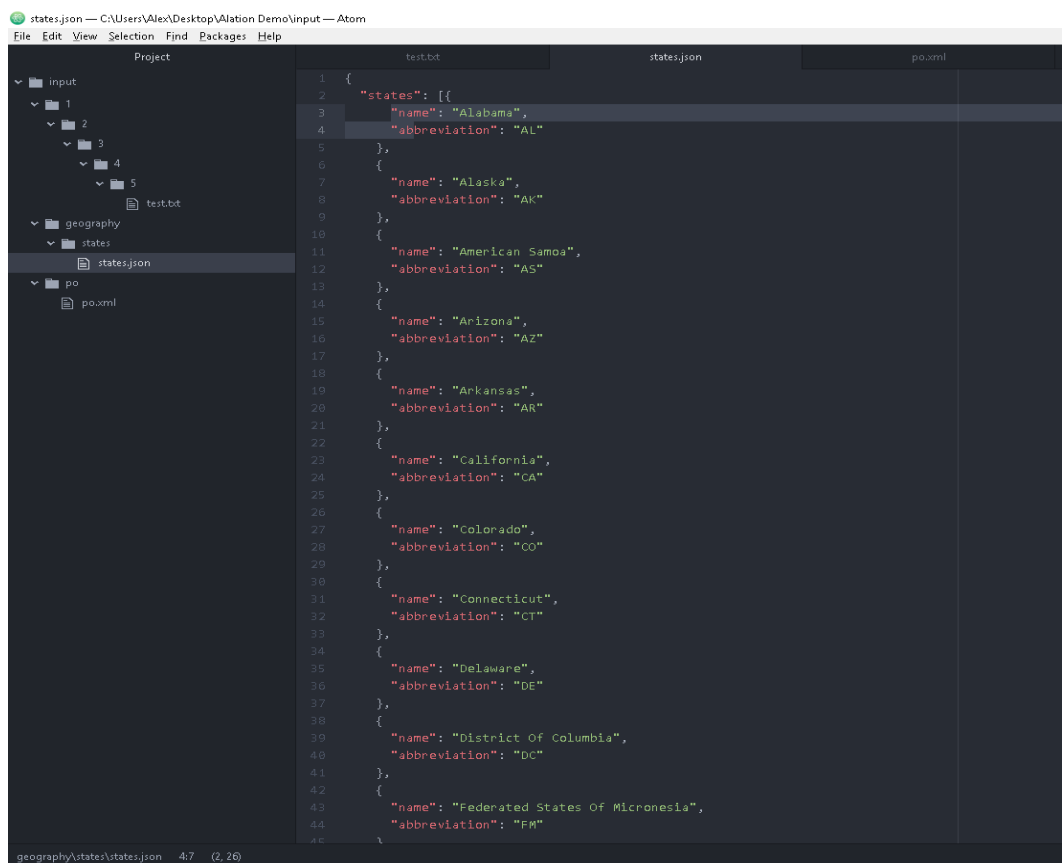
Alation Data Catalog

Alation Data Catalog is a market-leading data catalog built for collaboration. Alation empowers analysts to find, understand and share data for faster, more accurate insights. Alation's data catalog automatically captures the rich context of enterprise data and keeps it updated with human-machine learning collaboration.

Files

A file is used to store information. Without files, it would be difficult to understand where one piece of information stops and next begins. By separating the information and giving them names the information can be easily isolated and identified. There are different types of files such as JSON, XML, CSV, Parquet, and Avro so on.

Figure 1 shows the details of JSON file.

Figure 1: JSON file.

Figure 2 shows a JSON file in Alation.
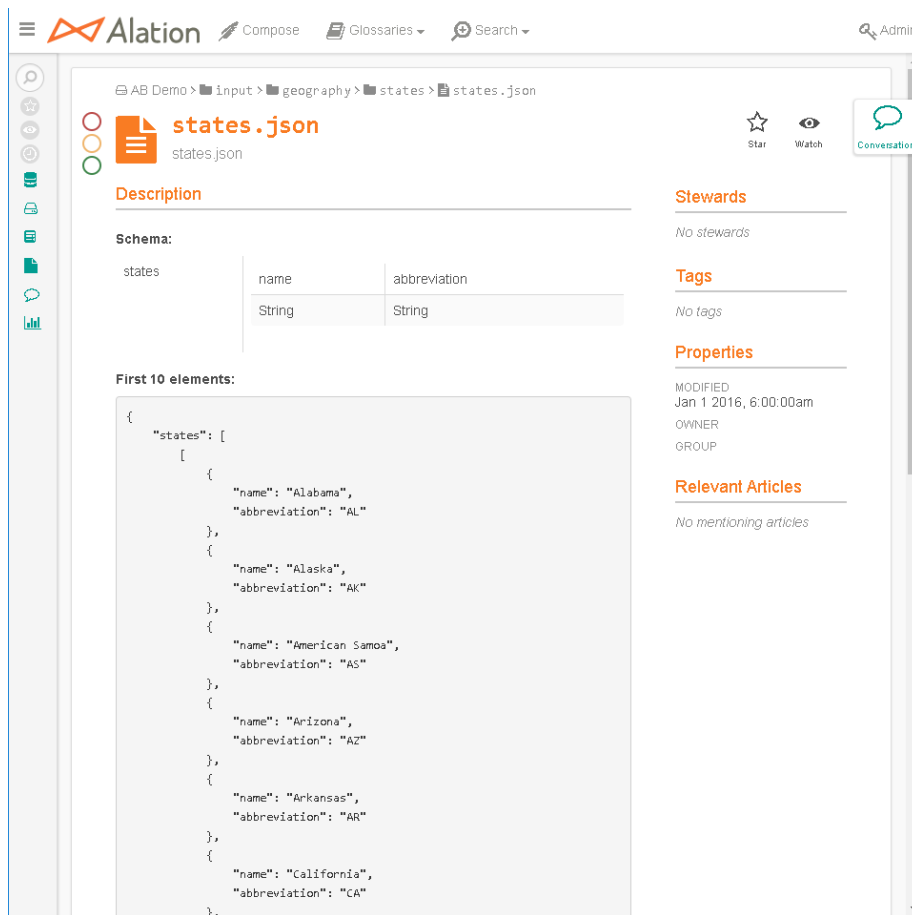
Figure 2: JSON file in Alation.

Figure 3 shows a CSV file.

Figure 3: CSV file.

Figure 4 shows the details of the CSV file in Alation.

Figure 4: CSV file in Alation.

Figure 5 shows the XML file.

Figure 5: XML file.

Figure 6 shows the details of an XML file in Alation.

Figure 6: XML file in Alation.

Figure 7 shows the Avro file.

{u'username': u'miguno', u'timestamp': 1366150681, u'tweet': u'Rock: Nerf paper, scissors is fine.'}
{u'username': u'BlizzardCS', u'timestamp': 1366154481, u'tweet': u'Works as intended. Terran is IMBA.'}

Figure 7: Avro file.

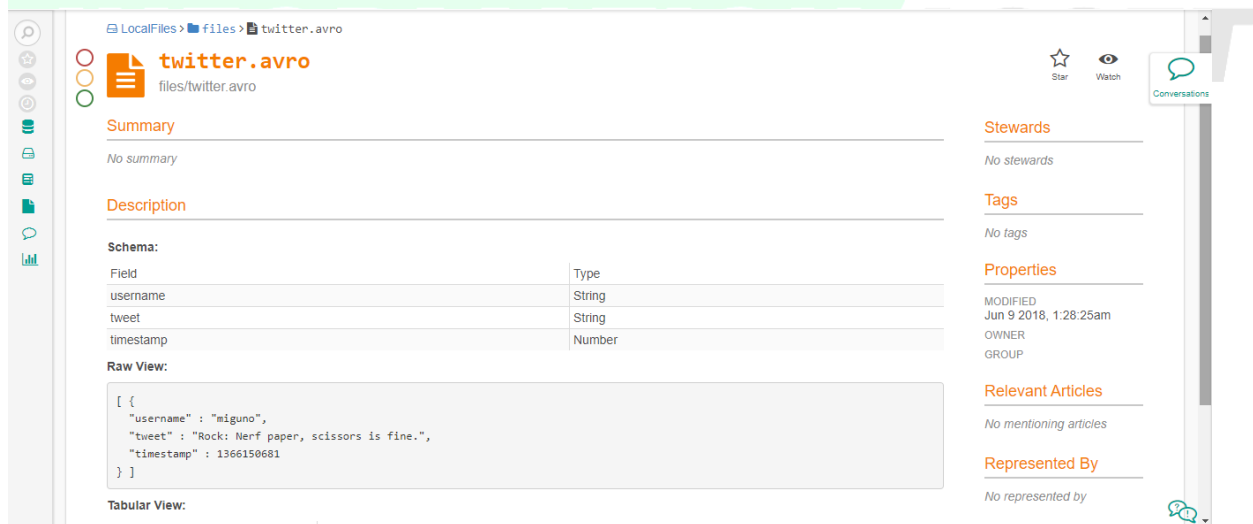Figure 8 shows the details of an Avro file in Alation.

Figure 8: Avro file in Alation.

Figure 9 shows the Parquet file.

registration_dttm	id	first_name	last_name	email	gender	ip_address	cc	country	birthdate	salary	title
03-02-2016 13:36:39	1	Donald	Lewis	dlewis0@clic	Male	102.22.124.20		Indonesia	7/9/1972	140249.37	Senior Finc
03-02-2016 00:22:28	2	Walter	Collins	wcollins1@b	Male	247.28.26.93	3587726:	China			
03-02-2016 18:29:04	3	Michelle	Henderson	mhendersor	Female	193.68.146.150		France	1/15/1964	236219.26	Teacher
03-02-2016 13:42:19	4	Lori	Hudson	lhudson3@d		34.252.168.48	3568840	Russia	4/22/1988		Nuclear Po
03-02-2016 00:15:29	5	Howard	Miller	hmiller4@fer	Male	103.193.150.230	3583473:	France	11/26/1998	50210.02	Senior Edit
03-02-2016 10:49:07	6	Frances	Adams	fadams5@1:	Female	106.196.106.93		Russia	3/27/1997	82175.77	Account C
03-02-2016 19:44:12		Steven	Hanson	shanson6@c	Male	234.130.172.185	3550842	Indonesia		129582.61	
03-02-2016 08:11:34	8	Louis	Simmons	lsimmons7@	Male	18.69.80.15		China	6/1/1992	90744.86	Product Er
03-02-2016 23:56:51	9	Keith	Parker	kparker8@a	Male	108.205.40.64		Guadeloupe	12/30/1992	60618.9	Developer
03-02-2016 07:47:29	10	Wanda	Walker	wwalker9@k	Female	246.214.98.78	3539421	Portugal		137664.53	
03-02-2016 16:39:23	11	Kathryn	Weaver	kweavera@	Female	157.237.161.75	2014250	Sweden		117572.65	
03-02-2016 06:50:56	12	Philip	Ward	pwardb@sa	Male	77.140.225.69	2015080:	Greece	9/3/1984	238925.79	Human Re
03-02-2016 23:36:58	13	Evelyn	Harvey	eharveyc@t		254.174.154.7	3539535	China	5/15/1979		Software E
03-02-2016 07:33:24	14	Andrea	Lane	alaned@go	Female	192.253.116.192	5100174	Indonesia	1/19/1989	166778.42	Operator
03-02-2016 13:45:04	15	Bobby	Vasquez	bvasqueze@	Male	126.60.18.195	3581051	Philippines	1/25/1975	138184.83	Senior Edit
03-02-2016 21:58:10	16	Kenneth	Gibson	kgibsonf@sc	Male	91.153.142.170	5389947	Peru	11/3/1975	98614.53	Environme
03-02-2016 16:22:34	17	Emily	Hill	ehillg@hous	Female	109.107.174.205		Palestinian Ten	5/18/1956	218781.48	Executive
03-02-2016 23:20:49	18	Kelly	Fowler	kfowlerh@d	Female	147.58.88.116	3551741	Greece	6/11/1975	117249.56	Systems Ac
03-02-2016 18:28:46	19	Diana	Howell	dhowelli@sp	Female	21.240.75.42	4026635	Iran	7/7/1993	174844.52	Teacher
03-02-2016 02:23:26	20	Johnny	Collins	jcollinsj@go	Male	38.173.129.250	3723016	Afghanistan	7/28/1987	155908.69	Social Wor
03-02-2016 10:05:12	21	Frank	Bradley	fbradleyk@s	Male	186.9.38.46	4913033	Czech Republi		211051.83	

Figure 9: Parquet file.

Figure 10 shows the details of Parquet file in Alation.

Figure 10: Parquet file in Alation.

Integration of Files and Alation

Information Asset has developed a solution that enables a user to import JSON, CSV, XML, Avro, and Parquet file types into Alation Data Catalog. The data scientist can view the data structure along with sample data.

Next Steps

To learn more about this solution, please request a demo by contacting sales@information-asset.com or visit our website at www.information-asset.com.