In [4]:
```
df.head(5).T
```

Out[4]:

	0	1	2	3	4
export_time_ms	2023-05-18 19:58:37.999000+00:00	2023-05-18 19:58:37.999000+00:00	2023-05-18 19:58:38+00:00	2023-05-18 19:58:38.001000+00:00	2023-05-18 19:58:38.001000+00:00
import_time_ms	2023-05-19 19:49:19.778000+00:00	2023-05-19 19:53:56.175000+00:00	2023-05-19 19:57:12.603000+00:00	2023-05-19 19:55:01.374000+00:00	2023-05-19 19:57:34.540000+00:00
start_time_ns	2023-05-18 19:58:37.000006+00:00	2023-05-18 19:58:37.000077+00:00	2023-05-18 19:58:37.000099+00:00	2023-05-18 19:58:36.422655+00:00	2023-05-18 19:58:37.000144+00:00
end_time_ns	2023-05-18 19:58:37.000006+00:00	2023-05-18 19:58:37.000077+00:00	2023-05-18 19:58:37.000099+00:00	2023-05-18 19:58:37.000107+00:00	2023-05-18 19:58:37.000144+00:00
export_reason	idle	idle	idle	idle	idle
exporting_node	slac50n-ht1	slac50n-ht1	lbnl59-ht1	slac50n-ht1	lbnl59-ht2
router_name	slac50n-cr6	slac50n-cr6	lbnl59-cr6	slac50n-cr6	lbnl59-cr6
router_interface	1/1/c2/1	1/1/c2/1	2/1/c2/1	1/1/c2/1	2/1/c19/1
direction	out	out	out	out	out
vlan_id	1124	1124	765	1124	1126
sap_name	slac_se-58	slac_se-58	lbnl_se-408	slac_se-58	nersc_se-1122
sap_type	Layer 3 Virtual Interface	Layer 3 Virtual Interface	Layer 3 Virtual Interface	Layer 3 Virtual Interface	Layer 3 Virtual Interface
sap_routing_instance	Base	Base	Base	Base	Base
sap_bgp_policy_summary	Site	Site	Site	Site	Site
sap_organization_name	SLAC	SLAC	LBNL	SLAC	NERSC
asn_src	396982	43	2936	8075	400161
asn_dst	3671	3671	16	3671	2936
hash_fwd	15626446402169633286	5107403408903449222	963293011857245431	15436415860848499077	14668491204857827211
hash_rev	13739826294085640971	9276731171592102972	4285569469679602915	14915413259592191046	15798120477492198188
ip_version	4	4	4	4	4
ip_src	35.203.211.113	130.199.132.9	128.55.210.89	40.97.221.114	104.156.155.9
ip_src_bin	::ffff:23cb:d371	::ffff:82c7:8409	::ffff:8037:d259	::ffff:2861:dd72	::ffff:689c:9b09
ip_dst	134.79.17.146	134.79.35.10	128.3.7.159	198.129.119.119	128.55.240.152
ip_dst_bin	::ffff:864f:1192	::ffff:864f:230a	::ffff:8003:79f	::ffff:c681:7777	::ffff:8037:f098
ip_proto_num	6	17	6	6	6
l4_src_port	52990	64039	27017	443	48283
l4_dst_port	3087	53	62425	57962	9682
tcp_f_syn_ack_only	False	False	False	True	False
tcp_f_syn_only	True	False	False	False	True
tcp_f_cwr	False	False	False	False	False
tcp_f_ece	False	False	False	True	False
tcp_f_urg	False	False	False	False	False
tcp_f_ack	False	False	True	True	False
tcp_f_psh	False	False	True	True	False
tcp_f_rst	False	False	False	False	False
tcp_f_syn	True	False	False	True	True
tcp_f_fin	False	False	False	False	False
packets	1	1	1	33	1
bytes	44	83	952	16185	40
pkt_size_hist	[1, 0, 0, 0, 0, 0, 0, 0]	[0, 1, 0, 0, 0, 0, 0, 0]	[0, 0, 0, 0, 1, 0, 0, 0]	[5, 6, 1, 13, 1, 7, 0, 0]	[1, 0, 0, 0, 0, 0, 0, 0]
hash	10919528622545722641	14384134580495552194	5248862481536848346	11905085046731138507	12019867608640473783
esdb_name_src		BNL	NERSC		
esdb_name_dst	SLAC	SLAC	LBNL	SLAC	NERSC
caida_orgId_src	GOOGL-2-ARIN	BNL-ARIN	NERSC-Z-ARIN	MSFT-ARIN	AIRLL-ARIN
caida_orgId_dst	THELE-44-Z-ARIN	THELE-44-Z-ARIN	LBNL-ARIN	THELE-44-Z-ARIN	NERSC-Z-ARIN
caida_org_name_src	GOOGLE-CLOUD-PLATFORM	BNL-AS	NERSC	MICROSOFT-CORP-MSN-AS-BLOCK	ACDRESEARCH
caida_org_name_dst	SLAC	SLAC	LBL	SLAC	NERSC
caida_org_country_src	US	US	US	US	US
caida_org_country_dst	US	US	US	US	US
port_svc_src	unknown		unknown	https	unknown
port_svc_dst	unknown	domain	unknown	unknown	unknown
ip_proto	TCP	UDP	TCP	TCP	TCP

In [ ]:
```
```

Description of flow records:

The table above shows example data for 5 flow records from the ESnet High Touch Database. The full database contains roughly 1 Billion or more records captured every 24 hours.

The Key Fields are as follows:

start_time_ns : a hardware timestamp that identifies when the first packet of a 5-tuple flow is seen by the monitor. For TCP, this corresponds to the syn packet. start_time_ns is repeated for subsequent flow records for as long as a flow is active and being monitored. If a flow lasts for 60 seconds, it will produce 6 flow records, at 10 second intervals. Each with the same start_time_ns.

end_time_ns : a hardware based timestamp that identifies when the last packet in the current flow record is seen. For a long flow, this is very close to start_time_ns + 10s. However for a small flow that only has one or two packets, this time can be start_time_ns + a few microseconds.

Export reason : **idle** means that no packets were seen for this 5-tuple for a 1s period, and it is deemed expired or idle. **Active** means that packets have been seen for 10 consequtive seconds and a 10 second summary is being exported.

Exporting node : the hostname of the High Touch node in the ESnet topology
Router name : the name of the router, that the exporting node is connected to. Each router can have 1 or more HT Exporting nodes.

Router interface : the interface on the exporting router, that the HT packet was captured from.

Direction : **In** means the packet was entering the ESnet edge from another site or network. **Out** means the packet was leaving the ESnet edge towards another site or network.

vlan_id : Provides the VLAN ID that the packet was forwarded with. This allows the separation of IP name spaces, and provides Layer 2 context for the traffic.

sap_name : an identifier that describes a service end point in Nokia routers. This corresponds to different customer services that are being connected at the ESnet edge.

sap_type : L2 VPN , L3 VPN are all valid service types.

s**ap routing instance** : provides a name for an L3 VPN. For example “base” , “LHCONE” , etc.. these names are specific to the way ESnet manages overlay routing services.

s**ap bap policy summary** : another categorical name for sets of BGP policy for different VPNs.

sap organization name : a DOE site or entity that this overlap SAP was create for

asn_src : the ASN number determined by the global internet ASN assignments for the source ip of this packet.

asn_dst : the ASN number determined by the global Internet ASN assignments for the dest ip of the packet.

hash_fw : a convenience value. It is a single 64 bit integer that combines ip_src,ip_dst,l4_src_port,l4_dst_port and protocol. This allows for 5 tuple searches with a single integer lookup.

hash_rev : another convenience value. It is calculated like hash_fw, but the role of src, and dat is reversed. This allows for matching (hash_fwd == hash_rev) in order to find upstream and downstream flow pairs that correspond to a bi-directional TCP or UDP communication.

hash : hash_fwd + hash_rev. This makes it easy to search for flows, in a manner that is agnostic to client->server vs. server->client direction.

ip_version : 4 and 6 are popular.

ip_src : The source IPv4 or IPv6 address from the captured packets.

ip_dst : The destination IPv4 or IPv6 address from the captured packets.

l4_src_port : The source port number for either a UDP or TCP transport packet

l4_dst_port : The destination port number for either a UDP or TCP Transport packet

TCP Flags: A number of TCP flags are captured for each packet. A flow record corresponds to many packets for the same 5 tuple flow in a period of 1 or 10 seconds. It might have as many as 1 million packets summarized in a single record. As a result we need to accumulate the TCP flags, rather than report a unique set of flags for every packet. This is done using the "INCLUSIVE OR" operation. So if a flag is set in the database record, it means "one or more packets" had this flag set.

tcp_f_syn_only : This flag means we saw a packet where only the syn flag was set, and all other flags were not. This is useful for determining the beginning of the syn / ack sequence.

tcp_f_syn_ack_only : Only the syn_ack flags were set. This marks the first response to a tcp connection setup request.

tcp_f cwr , ece , urg , ack , syn , psh , rest and fin : are all consistent with the TCP protocol. Each flag in the flow record means that one or more packets had set the flag, but it does not tell us "which" packet set the flag.

packets : the total number of packets that are represented by the flow record. A flow record is unique to a 5 tuple flow, so this is the total number of packets for a single flow.

bytes: the exact number of bytes obtained by summing the total bytes in each packet represented by this flow record.

pkt_size_histogram : An array of 8 histogram bins which have a count of packet sizes based on the following boundaries:
[0-63 , 64-127 , 128-255 , 256-511 , 512-1023 , 1024-2047 , 2048-4096 , 4096+]

esdb_name_src / dst : an organization name based on ESnet's internal ESDB data base. This is null for organizations that are not ESnet customers.

caida_org_name_src / **dst** : an organization name based on determining the ASN number for an IP address, and obtaining the organization name that owns that particular ASN. This yields an answer for all registered prefixes, not just ESnet customers.

caida_org_country src / dst : the country in which the ASN is registered. This might or might not correspond to the geographic location of a host.

port_svc src / dst : The 16 bit integer from the packet header's port number, can be mapped to well known IP services. Such as http, ssh etc.. this field provides the service name as a convenience when it is well known.

ip_proto : String set to 'TCP' or 'UDP' for human readable protocol numbers.