# 100 Fivetran Facts

# A Quick Reference Guide/FAQ

| | |
|---|---|
| Authors | David Millman, Fivetran |
| Version | 1.0 |
| Date | Sep 1, 2023 |
| Description | Quick Reference Guide for Technical Questions for Fivetran General Knowledge and help in Certification Tests |

1. For most connectors, Fivetran can be summarized in three numbers:
    a. 15mins - the development time it takes to configure a channel/pipeline once networking/access issues are resolved
    b. 100% - Fivetran recommends moving 100% of the data from the transactional database/system of record to the analytic system to optimize the downstream process of AI/ML etc.
    c. 0 - Fivetran pipelines automatically manage schema drift, type changes etc. so are self healing, ensuring that maintenance time can be closer to 0% of project cost
2. Fivetran has 3 basic deployment architectures, a pure SaaS model for connecting to SaaS applications, Files and databases such as those found in RDS (Relational Database Service) instances.
    a. A second Hybrid approach is designed for moving data from larger databases, by accessing their binary transaction log from an agent.
    b. The third option is using installed Software called the LDP (Local Data Processing) hub and associated agents.
3. Loading files from an S3 bucket requires permissions for the file to be read by Fivetran, by making the folder public or similar and no installation as the Fivetran.com infrastructure can read the file.
4. All connectors provide a historical load function and the ability to detect changes, propagating the changes. There are however some API objects that do not allow a programmatic way to detect changes, so they may initiate a full load each time. These cases are very rare and dependent on the API provider.
5. To load CSV files, Datatypes for the table will be detected automatically and managed, e.g. if the number 2 is found a numeric type will be used and row number and file name are added to each row.
6. Teleport can be used on the following type of database deployments, RDS databases (i.e. RDBMS as a service), a database deployed in a customer's environment behind a firewall, a database installed on a bare metal machine and an oracle database deployed in a RAC configuration
7. If using an RDS instance as a source, consider reading changes from a read replica, this allows the provider to replicate to another instance, but also ensures that queries do not impede the base databases' performance.
8. For a Teleport connection, the following is true, an agent does not need to be installed, no database changes are required, beyond giving the user read access, in most cases a Primary Key is required
9. Teleport is only available on Fivetran.com, it is NOT available to LDP
10. Teleport detects changes, by issuing a set of SQL statements to determine what rows have changed
11. Of the four types of database connector that Fivetran provides, Teleport is the easiest to setup
12. Oracle, SQL Server, MySQL and PostgreSQL all offer Teleport options
13. No Teleport does not compress the data as it is moved from the source to target. Although it uses a compressed image to determine changes.
14. The minimum available sync time on Teleport connector is 1 minute, although 5 minutes is recommended to allow previous syncs to finish, especially on larger databases
15. Teleport is designed for SQL database, NOT noSQL ones like MongoDB
16. Teleport does NOT require Change Tables, or Log Miner to be configured. When using these options, the appropriate connector should be configured as these are supported in Fivetran.com

17. With Analytic databases such as Redshift and Snowflake as a source, Teleport is used as the underlying data movement tool.
18. Soft Delete and History Mode are both available for Teleport
19. For some database applications such as EPIC where Composite Indexes are used, Teleport can be used to detect changes, as other techniques such as Change Tables are not available.
20. In complex applications such as EPIC Teleport maybe combined with other connectors such as HVA to move data.
21. Teleport is not recommended for Tables that are greater than 100MB
22. Teleport was acquired at the same time as HVR, but was from a different vendor
23. Teleport is normally used for smaller database sizes such as those less than 2TB and a smaller than 10 MB/Sec change rate
24. HVA uses the same Fivetran.com runtime, but with a local agent to optimize database reads.
25. The HVA/LDP agent compresses historical and change data, can read the active redo log and does all of this over a secure channel.
26. The LDP version of the agent allows multiple different types of reading change tables, such as Change Tables.
27. The HVA agent needs to be installed on the same machine, or very close (depending on the type of database) to the database server.
28. Only one Agent is required to deliver data to multiple destinations, this is important if the say source is potentially going to Google Big Query, Databricks and Snowflake at the same time.
29. The HVA agent is the same one used by HVR, so comes to Fivetran via acquisition, so it is well proven in high volume environments.
30. HVA only works on the Database server's transaction logs, for Change Tables etc, other connectors should be used.
31. As the agent resides on the same machine as the database server, 'localhost' is typically used as the hostname, or the private IP address.
32. HVA reads changes from the database's binary log and connects to the database to issue SQL statements for the historic data.
33. HVA cannot be used for RDS (Relational Database Service) instances as the transaction log is not available.
34. HVA achieves a performance increase over other technologies because it tails the transaction log, so never has to issue a SQL query for changes.
35. Connection between the HVA Agent and Fivetran.com infrastructure can be performed by SSH, Reverse SSH, Proxy Agent, Direct, Private Link and VPN. Availability will depend on the type of plan the customer has signed up for.
36. HVA can be used for SAP ECC and SAP/S4 databases, whether they be deployed on Oracle, SQL Server, Db2 and HANA.
37. HVA is available for Oracle, including RAC and ASM, SQL Server and DB2 for iSeries
38. HVA is built purely for Relational Databases, not SaaS applications, Files or any other type of data source.
39. HVA is available on the Enterprise plan.
40. For RDS instances, or where the database log is not available, Teleport and Database specific connectors with Log Miner and Change Tables.
41. In all but the most extreme outages, such as network connectivity issues, the agent will automatically continue propagating changes from the last successful update.
42. LDP stands for Local Data Processing and is a rebranding of the HVR Hub.
43. The LDP agent is 100% identical to the agent used in HVA
44. The LDP agent is used for the Capture cycle, it cannot be used for Integrate, but is NOT needed for integrate

45. The LDP hub has an inbuilt agent
46. Compare can identify individual records, by selecting the Diff Files on the compare
47. Multiple channels can be supported and often used with multiple concurrent channels.
48. The lowest possible latency is when the agent reads the active log. The archive log typically represents a 15 min plus delay on reading the changes.
49. Refresh is the process to perform an initial/historical load as well as repairing data that may have got corrupted or deleted in the target database.
50. LDP is not the optimal solution for RDS instance, it is better to use Fivetran.com.
51. New SaaS connectors should be built on Fivetran.com, to speed up this creation time a Lite Connector could be requested by the form.
52. LDP uses the agent to configure slicing, or reduce the amount of historic data bought across, this not available in HVA. HVA does do automatic slicing on some sources such as Oracle, but this is not configurable.
53. LDP is only for Databases, not SaaS solutions
54. LDP has two main processes.
    a. Capture reads the latest changes and pushes them to the Hub
    b. Integrate pushes data from the Hub to the Final Destination
55. LDP supports 1-1, 1-M, M-1 and M-M replication strategies
56. It is recommended for most cases that 1-1 is used. If there is a need to move to two destinations consider creating two 1-1 channels.
57. LDP supports Direct 1-1 Copy for transactional database replication, Soft Delete and Time key for analytics use-cases.
58. LDP's advantages over HVA are more configuration options and continuous data changes.
59. LDP requires either Business Critical and Private Deployment plans
60. All Fivetran products use MAR (Monthly Active Rows) as the pricing model
61. If a customer needs to add a new connector, nothing needs to be done assuming the current plan and MAR levels are not broken, Nothing needs to be done.
62. Fivetran gives flexibility to the user to define how they want Fivetran to handle schema changes (Allow All, Allow only new columns, Block All)
63. MAR resets every month.
64. Fivetran recommends the ELT methodology as it helps preserve raw data and helps in avoiding brittle pipelines which are generally caused by frequent schema changes, data type changes etc
65. Fivetran does do in-flight transformations but it is fully automatic i.e. automatic schema migrations, automatic data type changes, automatic normalizations, automatic deduplication etc
66. Fivetran's support for manual transformations is generally termed as 'push-down transformations'
67. Fivetran can handle a variety of database use cases, be it on-premises or on the cloud (AWS / GCP / Azure)
68. Fivetran supports both log-based and log-less methods of reading changes from databases to ensure we are able to provide CDC for a diverse variety of cases.
69. For log-based methods, Fivetran uses the native log-based utilities in databases to ensure authenticity and completeness. For example, for MySQL, Fivetran uses Binary Logs. For PostgreSQL Fivetran uses Write-Ahead Logs or WAL.
70. For log-less methods for databases Fivetran can also use native database properties and objects. For example, for PostgreSQL, Fivetran can use the XMIN column.
71. Fivetran also integrates heavily with Airflow for dependency management and advanced orchestrations
72. Fivetran is also able to support data warehouses as sources such as Snowflake and BigQuery for ease of data movement.
73. Fivetran is the only ELT vendor that is able to provide a data-delivery SLA

74. For application ingestion use-cases, Fivetran has pre-built dbt models for building the final use-cases seamlessly
75. Fivetran can provide data-lineage graphs for a variety of dbt models showing lineage of how many tables are used for the dbt model and what the dependency is looking like.
76. Fivetran provides integrated scheduling for the dbt models so they can be executed as soon as the underlying connector has finished its sync so no manual involvement is required.
77. As part of complete schema replication, Fivetran replicates custom data whenever it exists and is accessible.Custom data includes custom objects, tables, and fields that you have configured in the source system to better suit your business needs. Custom objects are specific to your source, for example, custom Salesforce objects that match your business process. Custom tables are database tables that allow you to store information unique to your organization. Standard tables can also have custom fields, which are specific to an organization.There is no special action you need to take to make sure we replicate your custom data. It will happen automatically for the systems that enable it.
78. Fivetran can invoke other transformation engines, or downstream processes via web hooks.
79. As Fivetran.com is a SaaS tool, there are continuous improvements always being made, especially adding new connectors and improving the functionality of various database connectors, especially those implemented using HVA
80. LDP and HVR version 6+ are functionally identical.
81. There are installations of HVR 5.7 out there that partners should consider migrating to LDP/HVR v6.
82. Fivetran.com provides access to SAP RISE and all types of SAP deployment through a SAP connector that works only through the Netweaver interface.
    a. It requires a transport to be installed in SAP that will provide many of the advantages that the agent does in HVA/LDP.
83. Fivetran typically has four types of connector for databases, these being:
    a. Fivetran.com via Teleport
    b. Fivetran.com via native change data capture technique
    c. Fivetran.com via High Volume Agent (HVA)
    d. Fivetran LDP
84. Fivetran typically provides 3 types of connector for SAP
    a. Fivetran.com historic load via Netweaver, changes by Database Transaction Log
    b. Fivetran.com historic load via ODBC/JDBC and changes by Transaction Log
    c. LDP variations of a and b
85. For SAP databases, Fivetran takes care of unpacking and other related encodings that SAP performs within the database (Please clarify for specific details).
86. For many on-premise applications, the easiest way to integrate may be to use the relevant database connectors and change capture techniques, versus building an API based connector.
87. Fivetran's agent uses about 3-4% CPU when detecting changes, and predominantly is just moving a File pointer to detect changes
88. About 40-60% of what is found in a Database servers transaction log is dedicated to server state etc and so should not be propagated in CDC jobs. Fivetran's agent architecture ignores this state information, therefore freeing up network bandwidth over other solutions.
89. Fivetran's agent architecture compresses data by 10-20x, so further freeing up network bandwidth.
90. When starting a refresh or other job, Fivetran will automatically manage the pausing of any integrate process and continue this once the refresh job is complete.
91. Refresh jobs are completely free of charge.

92. A new MAR event occurs only when a new RowID/ObjectID is Created, Updated and Deleted in the source system and that object is propagated to the target.
93. MAR events are raised for C-UD (Create,-, Update and Delete) events in the source system, when the rowID/ObjectID has not had a previous event that month.
94. If a RowID or ObjectID is Created, Updated or Deleted 1000 times in a month, this will count as 1 MAR event.
95. If a RowID or ObjectID is Created, Updated or Deleted 1000 times over two months, this will count as 2 MAR events. One in month 1 and one in month 2.
96. Using Time Series as an option creates a series of events that can be used to replay changing aspects of a business.
97. Time Series data and other replication strategies can be augmented with dbt's incremental mode to minimize compute costs associated with materializing views in target cloud platforms, while still optimizing performance of created tables.
98. Fivetran's S3 destination is Iceberg compatible, allowing tools such as Amazon Athena to perform queries directly on the data without any extra processing
99. Fivetran's products are compatible with Databricks Unity catalog.
100. Fivetran supports DB2 on i-Series, zOS and LUW (Linux, Unix and Windows).