

Bootstrap Capabilities	
Full table snapshot	When CDC process begins, take a full snapshot of the source table
Time range snapshot	Allow users to configure a start date/time from where to begin taking the snapshot
Parallel snapshot execution	Allow users to configure the number of parallel processes that concurrently snapshot a single source table
Snapshot parallelism	When snapshotting multiple tables, configure how many concurrent snapshots are allowed to run (may require multiple connections to the source DB)
Skip snapshot	Allow users to skip historical snapshot and only process new CDC events
Source Connector Ecosystem	
RDBMS	Key points: 1. Confirm that the vendor supports all the relational databases your organization requires. 2. Verify that the CDC solution is compatible with the versions of the RDBMS you are currently using or plan to use in the future.
	Common RDBMS platforms to look for include: 1. Oracle 2. Microsoft SQL Server 3. IBM Db2 4. MySQL 5. PostgreSQL 6. MariaDB 7. SQLite 8. Amazon RDS 9. Google Cloud SQL 10. Azure SQL Database
NoSQL Databases	Key points: 1. Confirm that the vendor supports all the NoSQL databases your organization requires. 2. Verify that the CDC solution is compatible with the versions of the NoSQL you are currently using or plan to use in the future.  Common NoSQL platforms to look for include: 1. MongoDB 2. Cassandra 3. Couchbase 4. DynamoDB 5. HBase 6. Redis 7. Elasticsearch 8. Neo4j 9. Amazon DocumentDB 10. Azure Cosmos DB
Mainframe	Support for various mainframe systems (e.g., IBM z/OS, Db2, IMS).
SAP	Does the vendor support the specific SAP modules you are using or plan to use? List all the SAP modules that you need support for, such as: SAP ERP (ECC) SAP S/4HANA SAP BW/4HANA SAP HANA. For SAP in particular drill into the specific methods available like trigger, Native DB based CDC or application extraction based.
Target Connector Ecosystem	
Event Stream platforms	Ability to propagate data changes to event streaming platforms hosted on a variety of platforms on on-premises, hybrid cloud or in public cloud.
	Common Event Streaming platforms to look for include:  - Apache Kafka - Confluent Kafka - Amazon Kinesis
Traditional Data Warehouses	Replication of data changes continuously to traditional on-premises data warehouses.  Traditional Data warehouses: - Oracle data warehouse (on-premises)- MS SQL data warehouse (on-premises)
Cloud Data Warehouses	Replication of data changes continuously to modern data warehouses or storage engineered as a cloud service or as a cloud-enabled repository.  - Snowflake, Big Query, Databricks with Delta Lake, AWS S3
NoSQL Databases	Propagation of data changes to NoSQL-type data structures such as graph, table-style, document store and key-value DBMSs
Write Operation	Allow users to define whether the target write operation should be a MERGE or APPEND
Essential Features	
Latency	Performance numbers are critical; speed is key in CDC tools. Ensure the solution offers low-latency data replication to meet real-time requirements.
Scalability	Check the vendor's performance metrics and key indicators. The solution should scale horizontally and vertically to handle increasing data volumes and transaction rates. Ask for benchmarks if they have one
HA Support	Assess the tool's high availability (HA) support. Is it active-active or active-passive? How is data consistency managed? Verify seamless failover capabilities by testing with a primary server switch-off scenario.
Zero data loss capability	Understand how the tool handles failures and errors to ensure zero data loss and recovery as desired by your organization.
Backlog Processing	Evaluate the tool's ability to capture historical changes or replay changes from a specific point in time, ensuring comprehensive data capture and recovery capabilities.
Point-in-time Recovery	Ensure the solution supports point-in-time recovery to reconstruct the data state at any given moment, aiding in data recovery and integrity.
Data Management and Transformation	
Data Transformation (In-Flight)	Check the breadth of transformation capabilities that the tool provides. While most ELT patterns focus on just moving data, transformations for operational purposes can be valuable.
ML/AI in flight	Determine if the tool can apply ML algorithms to incoming data in-flight before it is persisted. This can add real-time intelligence and insights to your data streams.
Schema Evolution	This one is a given and checking how the tool handles Schema evolution is critical. Evaluate how the tool handles schema evolution. Ensure it supports 1. Automatic table inclusion 2. Automatic table inclusion 3. Column name change/ deletion management without impacting the downstream pipeline
Automatic Schema Discovery	Ensure the tool can automatically discover the schema of the source and target databases upon registration, simplifying setup and maintenance.
Data Quality	Today with the increased focus on the trustability of the data, can the tool perform Data quality check on the streamed data inflight or once it has landed seamlessly
Data Validation	Verify if there are automated validation processes on the source and target sides to ensure proper reconciliation and validation. Is there a report or dashboard provided
Strong ordering	Enforce and ensure original CDC event ordering per table
Logical deletes	Allow users to configure if delete operations should be performed as hard deletes (permanently deleting the target row) or soft/logical deletes (mark a row as deleted)
Data type mapping	Automatically map source data types to appropriate target data types
Data type coarsening	Automatically coarse or cast compatible source to target data types.
Integration and Extendability	
Multi-Replicate Capability	The capability to read once and replicate to multiple data sources is crucial if you want to replicate to several targets. In today's modern architecture, combining operational and analytical patterns is key. You may want to sync your data to a translation store for real-time campaign needs while also taking the same data to your lakehouses.
API Support	Check how much of the process can be automated through APIs. Reviewing the API documentation will help you understand if the process can be fully codified.
Custom Connectors	Determine how the vendor supports custom connectors. Check if it is possible to support these via published SDKs or other means.
Deployment	
On-premises	Does the tool support on-premises deployment strategy
Cloud as a Service ( Managed)	Does the tool support managed cloud deployment strategy
Cloud (Self-hosted)	Does the tool support self-hosted deployment strategy
Vendor Support and Community	
Vendor Reputation	Track record and reputation of the vendor in the industry. Mentions of Analyst Reviews etc
Support Services	Quality and availability of technical support and professional services in the working hours of your company.
Community and Resources	Availability of user community, documentation, and training resources.
Skill-Set	While CDC tooling is probably the easiest to learn and become a master, expertise comes in handy while handling complex requirements. You may need to check the market availability for skilled resources.
Security and Monitoring	
Monitoring	Tools and facilities for monitoring and controlling runtime processes and the ability to monitor at query, database and server levels. A few useful monitoring capabilities are, 1. Snapshot data processing 2. Change Data processing
Maintoring failures and Data Quality Violations	Share detailed metrics indicating which one and whether an inline transformation failed to execute. Share detailed metrics indicating which one and whether a DQ expectation was violated and how many times.
Statistics	Collection of runtime statistics to determine usage and efficiency, as well as an application-style interface for visualization and evaluation
Is the solution available On-Premise	Depending on your technical needs and the availability of your sources systems on the premise you need to evaluate the tools that meet this requirement. In the absence of a pure play on-prem solution do they provide you with any Gateways/agents that can be installed on-premise. Also can it be made available serverless
Data encryption	At Rest/Motion, understand what types of protocols are being used
Access controls & Governance	Access controls & Governance
Data Sovereignty/Data Residency	Check if the vendor stores any data in flight for data consistency as this may be against your privacy requirements as set by your company.
Resiliency	Resilient to recover and resume (with transactional integrity) and capture the data changes from a source where it left off before it was shut down in the event of downtime (planned or unplanned)
Error handling	Predefined and customizable approaches for implementing standard error-handling processes
User Experience	
Ease of Use	Ability to configure all the configurations required through a GUI based tool
Measure User Experience	While it is not easy to quantify and can be very subjective, understanding how simple and easy it is to use the product based on limited exposure to the product can be measured.