

CLAIMS

1. A method, comprising:
performing, by one or more computing devices:
determining, for a particular data stream comprising a sequence of a plurality of data records, a set of nodes configurable by one or more control components to perform stream management operations based on one or more policies including a stream partitioning policy for partitioning the particular data stream that comprises the sequence of the plurality of data records into a plurality of partitions, and wherein an initial number of the nodes to be instantiated is based at least on the stream partitioning policy;
providing data records in response to record retrieval requests received via one or more programmatic record retrieval interfaces, wherein the one or more programmatic record retrieval interfaces include a first retrieval interface enabling a non-sequential access pattern and a second retrieval interface enabling a sequential access pattern, and wherein a billing rate associated with a use of the first retrieval interface differs from a billing rate associated with a use of the second retrieval interface; and
generating a client billing amount associated with the particular data stream based at least in part on respective usage count metrics of the plurality of record retrieval interfaces.
2. The method as recited in claim 1, further comprising performing, by the one or more computing devices:
assigning, in accordance with the partitioning policy, a particular data record of the particular data stream to a first partition of the plurality of partitions of the particular data stream, based at least in part on a key associated with the particular data record, wherein the key is indicated by a write request corresponding to the particular data record.

3. The method as recited in claim 1, further comprising performing, by the one or more computing devices:

selecting, based at least in part on a partition to which a particular data record is assigned, one or more of (a) a particular node of a record ingestion subsystem responsible for accepting the particular data record, (b) a particular node of a record storage subsystem responsible for storing at least one copy of the particular data record, and (c) a particular node of a retrieval subsystem responsible for obtaining the particular data record from the record storage subsystem in response to a read request.

4. The method as recited in claim 1, further comprising performing, by the one or more computing devices:

receiving data records submitted via one or more programmatic record submission interfaces, wherein the one or more programmatic record submission interfaces include a first submission interface supporting in-line submission of data records and a second submission interface enabling submission of data records by reference to one of: (a) an object address at a storage service implemented by a provider network (b) a universal record locator (c) a database record.

5. The method as recited in claim 1, further comprising performing, by the one or more computing devices:

removing, from a particular node of a data storage subsystem, a particular data record based at least in part on one of: (a) a data de-duplication window configured for the particular data stream, (b) a data archival policy associated with the particular data stream, (c) a data retention policy indicated by a client, (d) a client request to remove the particular data record, or (e) an indication that the particular data record has been processed by one or more data consumers.

6. The method as recited in claim 1, wherein at least one subsystem of: (a) a record ingestion subsystem, (b) a record storage subsystem, or (c) a record retrieval subsystem

configured for the particular data stream comprises a plurality of nodes configured by the one or more control components as members of a redundancy group, wherein the redundancy group comprises: (a) one or more primary nodes assigned to perform operations on a set of data records of the stream and (b) one or more non-primary nodes configured to assume a primary role in response to one or more triggering events.

7. The method as recited in claim 6, wherein a particular primary node of the one or more primary nodes is instantiated at a particular data center, and wherein a particular non-primary node of the one or more non-primary nodes is instantiated at a different data center.

8. The method as recited in claim 6, further comprising performing, by the one or more computing devices:

detecting a triggering event of the one or more triggering events; and

notifying a particular non-primary node of the one or more non-primary nodes to assume the primary role.

9. The method as recited in claim 1, wherein the one or more control components comprise a plurality of control nodes configured as a redundancy group, comprising a primary control node configured to respond to requests for control-plane operations on one or more data streams including the particular data stream, and at least one non-primary control node configured to assume a primary role in response to one or more triggering events.

10. The method as recited in claim 1, further comprising performing, by the one or more computing devices:

storing, at a network-accessible database of a provider network, metadata of the

particular data stream including a partition map generated in accordance with the stream partitioning policy; and

accessing, from a record retrieval subsystem, the metadata to respond to a particular record retrieval request.

11. The method as recited in claim 1, further comprising performing, by the one or more computing devices:

generating, corresponding to a particular data record of the data stream, a particular sequence number indicative of an order in which the particular data record was received at a record ingestion subsystem relative to other data records of the partition to which the particular data record belongs;

storing, by a record storage subsystem, a plurality of data records including the particular data record in an order based at least in part on respective timestamps generated for the data records; and

in response to receiving a read request invoking the first retrieval interface with the particular sequence number as a parameter, retrieving the particular data record from the record storage subsystem.

12. A system including one or more processors and one or more memories, the one or more memories including program instructions that when executed on one or more processors implement a control node of a multi-tenant stream management service, wherein the control node is configured to:

receive a request to initialize a particular data stream to be comprised of a plurality of data records generated by one or more data producers;

determine, based at least in part on a partitioning policy for partitioning the particular data stream that comprises the sequence of the plurality of data records into a plurality of partitions, one or more parameters to be used to configure one or more subsystems for the particular data stream, including a record retrieval subsystem, wherein the one or more parameters comprise an initial number of nodes to be instantiated in the record retrieval subsystem;

identify one or more resources to be used for a particular node of the record retrieval subsystem, wherein the particular node is to be configured to implement a plurality of programmatic record retrieval interfaces, including a first retrieval interface enabling a non-sequential access pattern and a second retrieval interface enabling a sequential access pattern; and

configure the particular node of the record retrieval subsystem using the one or more resources.

13. The system as recited in claim 12, wherein the one or more subsystems include a record ingestion subsystem comprising one or more nodes configurable to receive data records of the particular data stream, and a record storage subsystem configurable to store data records of the stream at a selected set of storage locations.

14. The system as recited in claim 13, wherein at least one subsystem of the record ingestion subsystem, the record storage subsystem, and the record retrieval subsystem comprises a plurality of nodes configured as members of a redundancy group, wherein the redundancy group comprises: (a) one or more primary nodes assigned to perform operations on a set of data records of the stream and (b) one or more non-primary nodes configured to assume a primary role in response to one or more triggering events.

15. The system as recited in claim 13, wherein at least one node of one of: the record ingestion subsystem, the record storage subsystem or the record retrieval subsystem comprises a component of a compute instance of a virtualized computing service implemented at a provider network.

16. The system as recited in claim 13, wherein the one or more data producers comprise one or more of: (a) a logging component of a computing device, (b) a social media application, (c) a monitoring agent at a computing device, (d) a sensor device, (e) a mobile phone, or (f) a tablet computer.

17. The system as recited in claim 13, wherein at least one subsystem of the record ingestion subsystem, the record storage subsystem, and the record retrieval subsystem comprises a plurality of nodes configured as members of a redundancy group, wherein the redundancy group comprises: (a) one or more primary nodes assigned to perform operations on a set of data records of the stream and (b) one or more non-primary nodes configured to assume a primary

role in response to one or more triggering events, and wherein a particular primary node of the one or more primary nodes is instantiated at a particular data center, and wherein a particular non-primary node of the one or more non-primary nodes is instantiated at a different data center.

18. A system, comprising:

one or more computing devices configured to:

determine, by one or more control components of a multi-tenant stream management service, wherein the one or more control components are assigned to a particular data stream comprising a sequence of a plurality of data records generated by one or more data producers, a respective set of nodes of (a) a record ingestion subsystem, (b) a record storage subsystem, and (c) a record retrieval subsystem, wherein each subsystem of the record ingestion subsystem, the record storage subsystem, and the record retrieval subsystem comprises one or more nodes dynamically configurable by the one or more control components based on one or more policies including a partitioning policy for partitioning the particular data stream that comprises the sequence of the plurality of data records into a plurality of partitions, and wherein an initial number of the nodes to be instantiated in the record retrieval subsystem is based at least on the stream partitioning policy;

receive data records submitted via one or more programmatic record submission interfaces implemented at the record ingestion subsystem, wherein the one or more programmatic record submission interfaces include a first submission interface supporting in-line submission of data records and a second submission interface enabling submission of data records by reference to network addresses at which data is stored;

provide contents of data records in response to data record retrieval requests received via one or more programmatic record retrieval interfaces implemented at the record retrieval subsystem, wherein the one or more programmatic record retrieval interfaces include a first retrieval interface

enabling a non-sequential access pattern and a second retrieval interface enabling a sequential access pattern, and wherein a billing rate associated with a use of the first retrieval interface differs from a billing rate associated with a use of the second retrieval interface;

generate a client billing amount associated with the particular data stream based at least in part on respective usage count metrics of the plurality of record retrieval interfaces and the plurality of record submission interfaces.

19. The system as recited in claim 18, wherein the one or more computing devices are further configured to:

assign, in accordance with the partitioning policy, a particular data record of the particular data stream to a partition of the particular data stream, based at least in part on a key associated with the particular data record, wherein the key is indicated by a write request corresponding to the particular data record.

20. The system as recited in claim 18, wherein the one or more computing devices are further configured to:

select, based at least in part on a partition to which a particular data record is assigned, one or more of (a) a particular node of the record ingestion subsystem responsible for accepting the particular data record, (b) a particular node of the record storage subsystem responsible for storing at least one copy of the particular data record, and (c) a particular node of the retrieval subsystem responsible for obtaining the particular data record from the record storage subsystem in response to a read request.

21. The system as recited in claim 18, wherein at least one subsystem of the record ingestion subsystem, the record storage subsystem, and the record retrieval subsystem comprises a plurality of nodes configured by the one or more control components as members of a redundancy group, wherein the redundancy group comprises: (a) one or more primary nodes

assigned to perform operations on a set of data records and (b) one or more non-primary nodes configured to assume a primary role in response to a detection of one or more triggering events.

22. The system as recited in claim 18, wherein the one or more computing devices are further configured to:

- generate, corresponding to a particular data record, a particular sequence number indicative of an order in which the particular data record was received at the record ingestion subsystem relative to other data records of the partition to which the particular data record belongs;
- store, by the record storage subsystem, a plurality of data records including the particular data record in an order based at least in part on respective sequence numbers generated for the data records; and
- in response to a read request invoking the first retrieval interface with the particular sequence number as a parameter, retrieve the particular data record from the record storage subsystem.