

WHAT IS CLAIMED IS:

1. A method, comprising:

locally storing fingerprints for data to a fingerprint dictionary, wherein the data comprises
5 a plurality of data units, and wherein each fingerprint in the fingerprint dictionary uniquely identifies a respective data unit in the data;

receiving, from a device via a network, one or more fingerprints each corresponding to a different data unit cached at the device;

searching the fingerprint dictionary for each of the one or more fingerprints received from
10 the device to determine if the fingerprint is in the fingerprint dictionary or is not in the fingerprint dictionary, wherein determining that a fingerprint is not in the fingerprint dictionary indicates a corresponding data unit to be uploaded;

sending, to the device via the network, an indication of one or more data units to be uploaded as determined by said searching the fingerprint dictionary; and

15 receiving, from the device via the network, the indicated one or more data units, wherein each received data unit corresponds to a fingerprint that is not in the fingerprint dictionary.

2. The method as recited in claim 1, wherein each fingerprint is a hash of a
20 respective data unit.

3. The method as recited in claim 2, wherein the hash is generated according to a cryptographically strong one-way hash function applied to the respective data unit.

25 4. The method as recited in claim 1, further comprising updating the fingerprint dictionary with the fingerprint corresponding to each data unit received from the device.

5. The method as recited in claim 1, wherein the data comprises a plurality of data blocks, wherein each data block comprises two or more of the plurality of data units.
30

6. The method as recited in claim 1, wherein the data is stored in a data store, wherein the method further comprises storing the one or more data units received from the device to the data store.

7. The method as recited in claim 6, wherein the method further comprises providing the data store as a virtualized data store to a plurality of customers via the network, wherein the data stored in the data store includes customer data for each of the plurality of customers, wherein the device corresponds to one of the plurality of customers, and wherein the device
5 locally caches at least a portion of the customer data of the respective customer from the data store.

8. The method as recited in claim 1, further comprising:

after said receiving the indicated one or more data units from the device, determining that
10 at least one fingerprint that said searching indicated was in the fingerprint dictionary has subsequently been deleted from the fingerprint dictionary;
sending, to the device via the network, an indication of at least one data unit corresponding to the at least one fingerprint; and
receiving, from the device via the network, the indicated at least one data unit.

9. A device, comprising:

at least one processor; and

a memory comprising program instructions, wherein the program instructions are executable by the at least one processor to:

20 generate fingerprints for a plurality of data units of locally cached data, wherein each fingerprint uniquely identifies a respective data unit in the locally cached data;

send the fingerprints to a remote data storage service via a network, wherein the remote data storage service maintains a primary data store of the data;

25 receive, from the remote data storage service via the network, an indication of one or more of the data units that are to be stored to the primary data store; and
send, via the network to the remote data storage service, the one or more data units for storage, by the remote data storage service, to the primary data store.

30 10. The device as recited in claim 9, wherein, to generate a fingerprint for a data unit, the program instructions are executable by the at least one processor to apply a hash function to the data unit.

11. The device as recited in claim 10, wherein the hash function is a cryptographically strong one-way hash function.

12. The device as recited in claim 9, wherein the locally cached data is cached as a plurality of data blocks, wherein the plurality of data units correspond to one of the plurality of data blocks.

13. The device as recited in claim 12, wherein the one of the plurality of data blocks is marked as dirty to indicate that the locally cached data block has been created or modified and thus requires uploading from the locally cached data to the primary data store.

14. The device as recited in claim 9, wherein the remote data storage service provides a virtualized data store to a plurality of customers via the network, wherein data stored in the virtualized data store includes customer data for each of the plurality of customers, wherein the device is associated with one of the plurality of customers, and wherein the program instructions are executable by the at least one processor to locally cache at least a portion of the customer data of the respective customer from the data store.

15. The device as recited in claim 14, wherein the device is communicatively coupled to a local network of a customer site, and wherein the program instructions are executable by the at least one processor to:

download and locally cache portions of the customer data from the remote data storage service; and

provide local access to the locally cached customer data to other processes on the local network including one or more virtual computing systems implemented on other devices on the local network.

16. The device as recited in claim 9, wherein the device is allocated a portion of bandwidth on a communications channel to the remote data storage service via the network, and wherein the program instructions are executable by the at least one processor to request allocation of additional bandwidth on the communications channel for uploading data units to the remote data storage service as needed, and to release bandwidth on the communications channel when not needed.

WHAT IS CLAIMED IS:

1. A method, comprising:
locally storing, in a network-based virtualized data store of a virtualized data store provider, fingerprints for data to a fingerprint dictionary, wherein the data comprises a plurality of data units, and wherein each fingerprint in the fingerprint dictionary uniquely identifies a respective data unit in the data;
receiving, by the virtualized data store provider and from a customer device of a customer of a plurality of customers of the virtualized data store provider via a network, one or more fingerprints each corresponding to a different data unit cached at the device;
searching the fingerprint dictionary for each of the one or more fingerprints received from the device to determine if the fingerprint is in the fingerprint dictionary or is not in the fingerprint dictionary, wherein determining that a fingerprint is not in the fingerprint dictionary indicates a corresponding data unit to be uploaded;
sending, to the customer device via the network, an indication of one or more data units to be uploaded as determined by said searching the fingerprint dictionary; and
receiving, from the customer device via the network, the indicated one or more data units, wherein each received data unit corresponds to a fingerprint that is not in the fingerprint dictionary.
2. The method as recited in claim 1, wherein each fingerprint is a hash of a respective data unit.
3. The method as recited in claim 2, wherein the hash is generated according to a cryptographically strong one-way hash function applied to the respective data unit.
4. The method as recited in claim 1, further comprising updating the fingerprint dictionary with the fingerprint corresponding to each data unit received from the device.
5. The method as recited in claim 1, wherein the data comprises a plurality of data blocks, wherein each data block comprises two or more of the plurality of data units.

6. The method as recited in claim 1, wherein the method further comprises storing the one or more data units received from the customer device to the data store.

7. The method as recited in claim 6, wherein the data stored in the data store includes customer data for each of the plurality of customers, and wherein the device locally caches at least a portion of the customer data of the respective customer from the data store.

8. The method as recited in claim 1, further comprising:
after said receiving the indicated one or more data units from the device, determining that at least one fingerprint that said searching indicated was in the fingerprint dictionary has subsequently been deleted from the fingerprint dictionary;
sending, to the device via the network, an indication of at least one data unit corresponding to the at least one fingerprint; and
receiving, from the device via the network, the indicated at least one data unit.

9. A device, comprising:
at least one processor; and
a memory comprising program instructions, wherein the program instructions are executable by the at least one processor to:
generate fingerprints for a plurality of data units of locally cached data, wherein each fingerprint uniquely identifies a respective data unit in the locally cached data;
send the fingerprints to a remote virtualized data storage service via a network, wherein the remote data storage service maintains a network-based virtualized primary data store of [[the]] data, and wherein the remote virtualized data storage service provides a network-based virtualized primary data store to a plurality of customers via the network;
receive, from the remote virtualized data storage service via the network, an indication of one or more of the data units that are to be stored to the network-based virtualized primary data store; and
send, via the network to the remote virtualized data storage service, the one or more data units for storage, by the remote virtualized data storage service, to the network-based virtualized primary data store.

10. The device as recited in claim 9, wherein, to generate a fingerprint for a data unit, the program instructions are executable by the at least one processor to apply a hash function to the data unit.

11. The device as recited in claim 10, wherein the hash function is a cryptographically strong one-way hash function.

12. The device as recited in claim 9, wherein the locally cached data is cached as a plurality of data blocks, wherein the plurality of data units correspond to one of the plurality of data blocks.

13. The device as recited in claim 12, wherein the one of the plurality of data blocks is marked as dirty to indicate that the locally cached data block has been created or modified and thus requires uploading from the locally cached data to the primary data store.

14. The device as recited in claim 9, wherein data stored in the virtualized data store includes customer data for each of the plurality of customers, wherein the device is associated with one of the plurality of customers, and wherein the program instructions are executable by the at least one processor to locally cache at least a portion of the customer data of the respective customer from the data store.

15. The device as recited in claim 14, wherein the device is communicatively coupled to a local network of a customer site, and wherein the program instructions are executable by the at least one processor to:

download and locally cache portions of the customer data from the remote data storage service; and

provide local access to the locally cached customer data to other processes on the local network including one or more virtual computing systems implemented on other devices on the local network.

16. The device as recited in claim 9, wherein the device is allocated a portion of bandwidth on a communications channel to the remote data storage service via the network, and wherein the program instructions are executable by the at least one processor to request allocation of additional bandwidth on the communications channel for uploading data units to

the remote data storage service as needed, and to release bandwidth on the communications channel when not needed.