Oracle

Getting Started with NoSQL Database Key/Value Java Driver

12c Release 1

Library Version 12.1.4.3

ORACLE
NOSQL DATABASE

Legal Notice

Published 17-Feb-2017

Table of Contents

Preface

There are two different APIs that can be used to write Oracle NoSQL Database (Oracle NoSQL Database) applications: the original Key/Value API, and the Table API. In addition, the Key/Value API is available in Java and C. The Table API is available in Java, C, node.js (Javascript), and Python. This document describes how to write Oracle NoSQL Database applications using the Key/Value API in Java.

Note

Most application developers should use one of the Table drivers because the Table API offers important features not found in the Key/Value API. The Key/Value API will no longer be enhanced in future releases of Oracle NoSQL Database.

This document provides the concepts surrounding Oracle NoSQL Database, data schema considerations, as well as introductory programming examples.

This document is aimed at the software engineer responsible for writing an Oracle NoSQL Database application.

Conventions Used in This Book

The following typographical conventions are used within in this manual:

Class names are represented in `monospaced font`, as are `method names`. For example: "The `KVStoreConfig()` constructor returns a `KVStoreConfig` class object."

Variable or non-literal text is presented in *italics*. For example: "Go to your *KVHOME* directory."

Program examples are displayed in a `monospaced font` on a shaded background. For example:

```
import oracle.kv.KVStore;
import oracle.kv.KVStoreConfig;

...

KVStoreConfig kconfig = new KVStoreConfig("exampleStore",
    "node1.example.org:5088, node2.example.org:4129");
KVStore kvstore = null;
```

In some situations, programming examples are updated from one chapter to the next. When this occurs, the new code is presented in **`monospaced bold`** font. For example:

```
import oracle.kv.KVStore;
import oracle.kv.KVStoreConfig;
import oracle.kv.KVStoreFactory;

...
```

```
KVStoreConfig kconfig = new KVStoreConfig("exampleStore",
    "node1.example.org:5088, node2.example.org:4129");
KVStore kvstore = null;

try {
    kvstore = KVStoreFactory.getStore(kconfig);
} catch (FaultException fe) {
    // Some internal error occurred. Either abort your application
    // or retry the operation.
}
```

Note

Finally, notes of special interest are represented using a note block such as this.

Chapter 1. Developing for Oracle NoSQL Database

You access the data in the Oracle NoSQL Database KVStore using Java drivers that are provided for the product. In addition to the Java drivers, several other drivers are also available. They are:

1. Java Table Driver
2. C Table Driver
3. C Key/Value Driver
4. Python Table Driver
5. node.js Table Driver

Note

New users should use one of the Table drivers unless they require a feature only available in the Key/Value API (such as Large Object support). The Key/Value API will no longer be enhanced in future releases of Oracle NoSQL Database.

The Java and C Key/Value driver provides access to store data using key/value pairs. All other drivers provide access using tables. Also, the Java Key/Value driver provides Large Object (LOB) support that as of this release does not appear in the other drivers. However, users of the Java Tables driver can access the LOB API, even though the LOB API is accessed using the Key/Value interface.

Users of the Table drivers are able to create and use secondary indexing. The Java and C Key/Value drivers do not provide this support.

To work, the C Table, Python Table, and node.js Table drivers require use of a proxy server which translates network activity between the driver and the Oracle NoSQL Database store. The proxy is written in Java, and can run on any machine that is network accessible by both your client code and the Oracle NoSQL Database store. However, for performance and security reasons, Oracle recommends that you run the proxy on the same local host as your driver, and that the proxy be used in a 1:1 configuration with your drivers (that is, each instance of the proxy should be used with just a single driver instance).

Regardless of the driver you decide to use, the provided classes and methods allow you to write data to the store, retrieve it, and delete it. You use these APIs to define consistency and durability guarantees. It is also possible to execute a sequence of store operations atomically so that all the operations succeed, or none of them do.

The rest of this book introduces the Java APIs that you use to access the store, and the concepts that go along with them.

The KVStore Handle

In order to perform store access of any kind, you must obtain a `KVStore` handle. You obtain a KVStore handle by using the `KVStoreFactory.getStore()` method.

When you get a KVStore handle, you must provide a KVStoreConfig object. This object identifies important properties about the store that you are accessing. We describe the KVStoreConfig class next in this chapter, but at a minimum you must use this class to identify:

- The name of the store. The name provided here must be identical to the name used when the store was installed.
- The network contact information for one or more helper hosts. These are the network name and port information for nodes currently belonging to the store. Multiple nodes can be identified using an array of strings. You can use one or many. Many does not hurt. The downside of using one is that the chosen host may be temporarily down, so it is a good idea to use more than one.

In addition to the KVStoreConfig class object, you can also provide a PasswordCredentials class object to KVStoreFactory.getStore(). You do this if you are using a store that has been configured to require authentication. See Using the Authentication APIs (page 3) for more information.

For a store that does not require authentication, you obtain a store handle like this:

```
package kvstore.basicExample;

import oracle.kv.KVStore;
import oracle.kv.KVStoreConfig;
import oracle.kv.KVStoreFactory;

...

String[] hhosts = {"n1.example.org:5088", "n2.example.org:4129"};
KVStoreConfig kconfig = new KVStoreConfig("exampleStore", hhosts);
KVStore kvstore = KVStoreFactory.getStore(kconfig);
```

The KVStoreConfig Class

The KVStoreConfig class is used to describe properties about a KVStore handle. Most of the properties are optional; those that are required are provided when you construct a class instance.

The properties that you can provide using KVStoreConfig are:

- Consistency

  Consistency is a property that describes how likely it is that a record read from a replica node is identical to the same record stored on a master node. For more information, see Consistency Guarantees (page 85).

- Durability

  Durability is a property that describes how likely it is that a write operation performed on the master node will not be lost if the master node is lost or is shut down abnormally. For more information, see Durability Guarantees (page 92).

- Helper Hosts

  Helper hosts are hostname/port pairs that identify where nodes within the store can be contacted. Multiple hosts can be identified using an array of strings. Typically an application developer will obtain these hostname/port pairs from the store's deployer and/or administrator. For example:

  ```
  String[] hhosts = {"n1.example.org:3333", "n2.example.org:3333"};
  ```

- Request Timeout

  Configures the amount of time the `KVStore` handle will wait for an operation to complete before it times out.

- Store name

  Identifies the name of the store.

- Password credentials and optionally a reauthentication handler

  See the next section on authentication.

Using the Authentication APIs

Oracle NoSQL Database can be installed such that your client code does not have to authenticate to the store. (For the sake of clarity, most of the examples in this book do not perform authentication.) However, if you want your store to operate in a secure manner, you can require authentication. Note that doing so will result in a performance cost due to the overhead of using SSL and authentication. While best practice is for a production store to require authentication over SSL, some sites that are performance sensitive may want to forgo that level of security.

Authentication involves sending username/password credentials to the store at the time a store handle is acquired.

A store that is configured to support authentication is automatically configured to communicate with clients using SSL in order to ensure privacy of the authentication and other sensitive information. When SSL is used, SSL certificates need to be installed on the machines where your client code runs in order to validate that the store that is being accessed is trustworthy.

Be aware that you can authenticate to the store in several different ways. You can use Kerberos, or you can specify a `LoginCredentials` implementation instance to `KVStoreFactory.getStore()`. (Oracle NoSQL Database provides the `PasswordCredentials` class as a `LoginCredentials` implementation.) If you use Kerberos, you can either use security properties understood by Oracle NoSQL Database to provide necessary Kerberos information, or you can use the Java Authentication and Authorization Service (JAAS) programming framework.

For information on using `LoginCredentials`, see Authentication using LoginCredentials (page 5). For information on using Kerberos, see Authentication using Kerberos (page 9). For

information on using JAAS with Kerberos, see Authentication using Kerberos and JAAS (page 11).

Configuring a store for authentication is described in the *Oracle NoSQL Database Security Guide*.

Configuring SSL

If you are using a secure store, then all communications between your client code and the store is transported over SSL, including authentication credentials. You must therefore configure your client code to use SSL. To do this, you identify where the SSL certificate data is, and you also separately indicate that the SSL transport is to be used.

Identifying the Trust Store

When an Oracle NoSQL Database store is configured to use the SSL transport, a series of security files are generated using a security configuration tool. One of these files is the `client.trust` file, which must be copied to any machine running Oracle NoSQL Database client code.

For information on using the security configuration tool, see the *Oracle NoSQL Database Security Guide*.

Your code must be told where the `client.trust` file can be found because it contains the certificates necessary to establish an SSL connection with the store. You indicate where this file is physically located on your machine using the `oracle.kv.ssl.trustStore` property. There are two ways to set this property:

1. Identify the location of the trust store by using a `Properties` object to set the `oracle.kv.ssl.trustStore` property. You then use `KVStoreConfig.setSecurityProperties()` to pass the `Properties` object to your `KVStore` handle.

   When you use this method, you use `KVSecurityConstants.SSL_TRUSTSTORE_FILE_PROPERTY` as the property name.

2. Use the `oracle.kv.security` property to refer to a properties file, such as the `client.trust` file. In that file, set the `oracle.kv.ssl.trustStore` property.

Setting the SSL Transport Property

In addition to identifying the location of the `client.trust` file, you must also tell your client code to use the SSL transport. You do this by setting the `oracle.kv.transport` property. There are two ways to set this property:

1. Identify the location of the trust store by using a `Properties` object to set the `oracle.kv.transport` property. You then use `KVStoreConfig.setSecurityProperties()` to pass the `Properties` object to your `KVStore` handle.

   When you use this method, you use `KVSecurityConstants.TRANSPORT_PROPERTY` as the property name, and `KVSecurityConstants.SSL_TRANSPORT_NAME` as the property value.

2. Use the `oracle.kv.security` property to refer to a properties file, such as the `client.trust` file. In that file, set the `oracle.kv.transport` property.

Authentication using LoginCredentials

You can authenticate to the store by specifying a `LoginCredentials` implementation instance to `KVStoreFactory.getStore()`. Oracle NoSQL Database provides the `PasswordCredentials` class as a `LoginCredentials` implementation. If your store requires SSL to be used as the transport, configure that prior to performing the authentication. (See the previous section for details.)

Your code should be prepared to handle a failed authentication attempt. `KVStoreFactory.getStore()` will throw `AuthenticationFailure` in the event of a failed authentication attempt. You can catch that exception and handle the problem there.

The following is a simple example of obtaining a store handle for a secured store. The SSL transport is used in this example.

```
import java.util.Properties;

import oracle.kv.AuthenticationFailure;
import oracle.kv.PasswordCredentials;
import oracle.kv.KVSecurityConstants;
import oracle.kv.KVStoreConfig;
import oracle.kv.KVStoreFactory;



KVStore store = null;
try {
    /*
     * storeName, hostName, port, username, and password are all
     * strings that would come from somewhere else in your
     * application.
     */
    KVStoreConfig kconfig =
        new KVStoreConfig(storeName, hostName + ":" + port);

    /* Set the required security properties */
    Properties secProps = new Properties();
    secProps.setProperty(KVSecurityConstants.TRANSPORT_PROPERTY,
                         KVSecurityConstants.SSL_TRANSPORT_NAME);
    secProps.setProperty
        (KVSecurityConstants.SSL_TRUSTSTORE_FILE_PROPERTY,
        "/home/kv/client.trust");
    kconfig.setSecurityProperties(secProps);

    store =
        KVStoreFactory.getStore(kconfig,
          new PasswordCredentials(username,
```

```
                                    password.toCharArray(),
                                    null /* ReauthenticateHandler */));
} catch (AuthenticationFailureException afe) {
    /*
     * Could potentially retry the login, possibly with different
     * credentials, but in this simple example, we just fail the
     * attempt.
     */
    System.out.println("authentication failed!");
    return;
}
```

Another way to handle the login is to place your authentication credentials in a flat text file that contains all the necessary properties for authentication. In order for this to work, a password store must have been configured for your Oracle NoSQL Database store. (See the *Oracle NoSQL Database Security Guide* for information on setting up password stores).

For example, suppose your store has been configured to use a password file password store and it is contained in a file called `login.pwd`. In that case, you might create a login properties file called `login.txt` that looks like this:

```
oracle.kv.auth.username=clientUID1
oracle.kv.auth.pwdfile.file=/home/nosql/login.pwd
oracle.kv.transport=ssl
oracle.kv.ssl.trustStore=/home/nosql/client.trust
```

In this case, you can perform authentication in the following way:

```
import oracle.kv.AuthenticationFailure;
import oracle.kv.PasswordCredentials;
import oracle.kv.KVStoreConfig;
import oracle.kv.KVStoreFactory;

/* the client gets login credentials from the login.txt file */
/* can be set on command line as well */
System.setProperty("oracle.kv.security", "/home/nosql/login.txt");

KVStore store = null;
try {
    /*
     * storeName, hostName, port are all strings that would come
     * from somewhere else in your application.
     *
     * Notice that we do not pass in any login credentials.
     * All of that information comes from login.txt
     */
    myStoreHandle =
        KVStoreFactory.getStore(
            new KVStoreConfig(storeName, hostName + ":" + port))
} catch (AuthenticationFailureException afe) {
    /*
     * Could potentially retry the login, possibly with different
```

```
     * credentials, but in this simple example, we just fail the
     * attempt.
     */
    System.out.println("authentication failed!")
    return;
}
```

Renewing Expired Login Credentials

It is possible for an authentication session to expire. This can happen for several reasons. One is that the store's administrator has configured the store to not allow session extension and the session has timed out. These properties are configured using `sessionExtendAllow` and `sessionTimeout`. See the *Oracle NoSQL Database Security Guide* for information on these properties.

Reauthentication might also be required if some kind of a major disruption has occurred to the store which caused the authentication session to become invalidated. This is a pathological condition which you should not see with any kind of frequency in a production store. Stores which are installed in labs might exhibit this condition more, especially if the stores are frequently restarted.

An application can encounter an expired authentication session at any point in its lifetime, so robust code that must remain running should always be written to respond to authentication session expirations.

When an authentication session expires, by default the method which is attempting store access will throw `AuthenticationRequiredException`. Upon seeing this, your code needs to reauthenticate to the store, and then retry the failed operation.

You can manually reauthenticate to the store by using the `KVStore.login()` method. This method requires you to provide the login credentials via a `LoginCredentials` class instance (such as `PasswordCredentials`):

```
try {
    ...
    /* Store access code happens here */
    ...
} catch (AuthenticationRequiredException are) {
    /*
     * myStoreHandle is a KVStore class instance.
     *
     * pwCreds is a PasswordCredentials class instance, obtained
     * from somewhere else in your code.
     */
    myStoreHandle.login(pwCreds);
}
```

Note that this is not required if you use the `oracle.kv.auth.username` and `oracle.kv.auth.pwdfile.file` properties, as shown in the previous section. In that case, your Oracle NoSQL Database client code will automatically and silently reauthenticate your client using the values specified by those properties.

A third option is to create a `ReauthenticationHandler` class implementation that performs your reauthentication for you. This option is only necessary if you provided a `LoginCredentials` implementation instance (that is, `PasswordCredentials`) in a call to `KVStoreFactory.getStore()`, and you want to avoid a subsequent need to retry operations by catching `AuthenticationRequiredException`.

A truly robust example of a `ReauthenticationHandler` implementation is beyond the scope of this manual (it would be driven by highly unique requirements that are unlikely to be appropriate for your site). Still, in the interest of completeness, the following shows a very simple and not very elegant implementation of `ReauthenticationHandler`:

```
package kvstore.basicExample

import oracle.kv.ReauthenticationHandler;
import oracle.kv.PasswordCredentials;

public class MyReauthHandler implements ReauthenticationHandler {
    public void reauthenticate(KVStore reauthStore) {
        /*
         * The code to obtain the username and password strings would
         * go here. This should be consistent with the code to perform
         * simple authentication for your client.
         */
        PasswordCredentials cred = new PasswordCredentials(username,
            password.toCharArray());

        reauthStore.login(cred);
    }
}
```

You would then supply a `MyReauthHandler` instance when you obtain your store handle:

```
import java.util.Properties;

import oracle.kv.AuthenticationFailure;
import oracle.kv.PasswordCredentials;
import oracle.kv.KVSecurityConstants;
import oracle.kv.KVStoreConfig;
import oracle.kv.KVStoreFactory;

import kvstore.basicExample.MyReauthHandler;

KVStore store = null;
try {
    /*
     * storeName, hostName, port, username, and password are all
     * strings that would come from somewhere else in your
     * application. The code you use to obtain your username
     * and password should be consistent with the code used to
     * obtain that information in MyReauthHandler.
     */
```

```
    KVStoreConfig kconfig =
        new KVStoreConfig(storeName, hostName + ":" + port);

    /* Set the required security properties */
    Properties secProps = new Properties();
    secProps.setProperty(KVSecurityConstants.TRANSPORT_PROPERTY,
                         KVSecurityConstants.SSL_TRANSPORT_NAME);
    secProps.setProperty
        (KVSecurityConstants.SSL_TRUSTSTORE_FILE_PROPERTY,
        "/home/kv/client.trust");
    kconfig.setSecurityProperties(secProps);

    store =
        KVStoreFactory.getStore(kconfig,
            new PasswordCredentials(username,
                                    password.toCharArray()));
            new MyReauthHandler());
} catch (AuthenticationFailureException afe) {
    /*
     * Could potentially retry the login, possibly with different
     * credentials, but in this simple example, we just fail the
     * attempt.
     */
    System.out.println("authentication failed!")
    return;
}
```

Authentication using Kerberos

You can authenticate to the store by using Kerberos. To do this, you must already have installed Kerberos and obtained the necessary login and service information. See the *Oracle NoSQL Database Security Guide* for details.

The following is a simple example of obtaining a store handle for a secured store, and using Kerberos to authenticate. Information specific to Kerberos, such as the Kerberos user name, is specified using `KVSecurityConstants` that are set as properties to the `KVStoreConfig` instance which is used to create the store handle.

```
import java.util.Properties;

import oracle.kv.KVSecurityConstants;
import oracle.kv.KVStore;
import oracle.kv.KVStoreConfig;
import oracle.kv.KVStoreFactory;

KVStore store = null;
/*
 * storeName, hostName, port, username, and password are all
 * strings that would come from somewhere else in your
 * application.
```

```
 */
KVStoreConfig kconfig =
    new KVStoreConfig(storeName, hostName + ":" + port);

/* Set the required security properties */
Properties secProps = new Properties();

/* Set the user name */
secProps.setProperty(KVSecurityConstants.AUTH_USERNAME_PROPERTY,
                     "krbuser");

/* Use Kerberos */
secProps.setProperty(KVSecurityConstants.AUTH_EXT_MECH_PROPERTY,
                      "kerberos");

/* Set SSL for the wire level encryption */
secProps.setProperty(KVSecurityConstants.TRANSPORT_PROPERTY,
                     KVSecurityConstants.SSL_TRANSPORT_NAME);

/* Set the location of the public trust file for SSL */
secProps.setProperty
    (KVSecurityConstants.SSL_TRUSTSTORE_FILE_PROPERTY,
    "/home/kv/client.trust");

/* Set the service principal associated with the helper host */
final String servicesDesc =
        "localhost:oraclenosql/localhost@EXAMPLE.COM";
secProps.setProperty(
        KVSecurityConstants.AUTH_KRB_SERVICES_PROPERTY,
        servicesDesc);

/*
 * Set the default realm name to permit using a short name for the
 * user principal
 */
secProps.setProperty(KVSecurityConstants.AUTH_KRB_REALM_PROPERTY,
                     "EXAMPLE.COM");

/* Specify the client keytab file location */
secProps.setProperty(KVSecurityConstants.AUTH_KRB_KEYTAB_PROPERTY,
                     "/tmp/krbuser.keytab");

kconfig.setSecurityProperties(secProps);

store = KVStoreFactory.getStore(kconfig);
```

Authentication using Kerberos and JAAS

You can authenticate to the store by using Kerberos and the Java Authentication and Authorization Service (JAAS) login API. To do this, you must already have installed Kerberos and obtained the necessary login and service information. See the *Oracle NoSQL Database Security Guide* for details.

The following is a simple example of obtaining a store handle for a secured store, and using Kerberos with JAAS to authenticate.

To use JAAS, you create a configuration file that contains required Kerberos configuration information. For example, the following could be placed in the file named `jaas.config`:

```
oraclenosql {
  com.sun.security.auth.module.Krb5LoginModule required
  principal="krbuser"
  useKeyTab="true"
  keyTab="/tmp/krbuser.keytab";
};
```

To identify this file to your application, set the Java property `java.security.auth.login.config` using the `-D` option when you run your application.

Beyond that, you use `KVSecurityConstants` to specify necessary properties, such as the SSL transport. You can also specify necessary Kerberos properties, such as the Kerberos user name, using `KVSecurityConstants`, or you can use the `KerberosCredentials` class to do this.

```
import java.security.PrivilegedActionException;
import java.security.PrivilegedExceptionAction;
import java.util.Properties;

import javax.security.auth.Subject;
import javax.security.auth.login.LoginContext;
import javax.security.auth.login.LoginException;

import oracle.kv.KerberosCredentials;
import oracle.kv.KVSecurityConstants;
import oracle.kv.KVStore;
import oracle.kv.KVStoreConfig;
import oracle.kv.KVStoreFactory;

/*
 * storeName, hostName, port, username, and password are all
 * strings that would come from somewhere else in your
 * application.
 */
final KVStoreConfig kconfig =
    new KVStoreConfig(storeName, hostName + ":" + port);

/* Set the required security properties */
Properties secProps = new Properties();
```

```
/* Set SSL for the wire level encryption */
secProps.setProperty(KVSecurityConstants.TRANSPORT_PROPERTY,
                     KVSecurityConstants.SSL_TRANSPORT_NAME);

/* Set the location of the public trust file for SSL */
secProps.setProperty
    (KVSecurityConstants.SSL_TRUSTSTORE_FILE_PROPERTY,
    "/home/kv/client.trust");

/* Use Kerberos */
secProps.setProperty(KVSecurityConstants.AUTH_EXT_MECH_PROPERTY,
                     "kerberos");

/* Set Kerberos properties */
final Properties krbProperties = new Properties();

/* Set the service principal associated with the helper host */
final String servicesPpal =
    "localhost:oraclenosql/localhost@EXAMPLE.COM";
krbProperties.setProperty(KVSecurityConstants.AUTH_KRB_SERVICES_PROPERTY,
                          hostName + ":" + servicesPpal);

/* Set default realm name, because the short name
 * for the user principal is used.
 */
krbProperties.setProperty(KVSecurityConstants.AUTH_KRB_REALM_PROPERTY,
                          "EXAMPLE.COM");

/* Specify Kerberos principal */
final KerberosCredentials krbCreds =
    new KerberosCredentials("krbuser", krbProperties);

try {
    /* Get a login context */
    final Subject subj = new Subject();
    final LoginContext lc = new LoginContext("oraclenosql", subj);

    /* Attempt to log in */
    lc.login();

    /* Get the store using the credentials specified in the subject */
    kconfig.setSecurityProperties(secProps);

    store = Subject.doAs(
        subj, new PrivilegedExceptionAction<KVStore>() {
            @Override
            public KVStore run() throws Exception {
                return KVStoreFactory.getStore(kconfig, krbCreds, null);
            }
```

```
        });
} catch (LoginException le) {
    // LoginException handling goes here
} catch (PrivilegedActionException pae) {
    // PrivilegedActionException handling goes here
} catch (Exception e) {
    // General Exception handling goes here
}
```

Unauthorized Access

Clients which must authenticate to a store are granted some amount of access to the store. This could range from a limited set of privileges to full, complete access. The amount of access is defined by the roles and privileges granted to the authenticating user. Therefore, a call to the Oracle NoSQL Database API could fail due to not having the authorization to perform the operation. When this happens, `UnauthorizedException` will be thrown.

See the *Oracle NoSQL Database Security Guide* for information on how to define roles and privileges for users.

When `UnauthorizedException` is seen, the operation should not be retried. Instead, the operation should either be abandoned entirely, or your code could attempt to reauthenticate using different credentials that would have the required permissions necessary to perform the operation. Note that a client can log out of a store using `KVStore.logout()`. How your code logs back in is determined by how your store is configured for access, as described in the previous sections.

```
// Open a store handle, and perform authentication as you do
// as described earlier in this section.

...


try {
    // When you attempt some operation (such as a put or delete)
    // to a secure store, you should catch UnauthorizedException
    // in case the user credentials you are using do not have the
    // privileges necessary to perform the operation.
} catch (UnauthorizedException ue) {
    /*
     * When you see this, either abandon the operation entirely,
     * or log out and log back in with credentials that might
     * have the proper permissions for the operation.
     */
    System.out.println("authorization failed!")
    return;
}
```

Chapter 2. Introduction to Oracle KVLite

KVLite is a single-node, single shard store. It usually runs in a single process and is used to develop and test client applications. KVLite is installed when you install Oracle NoSQL Database.

Starting KVLite

You start KVLite by using the `kvlite` utility, which can be found in `KVHOME/lib/kvstore.jar`. If you use this utility without any command line options, then KVLite will run with the following default values:

- The store name is `kvstore`.
- The hostname is the local machine.
- The registry port is `5000`.
- The directory where Oracle NoSQL Database data is placed (known as KVROOT) is `./kvroot`.
- The administration process is turned on.
- Security is turned on.

This means that any processes that you want to communicate with KVLite can only connect to it on the local host (127.0.0.1) using port 5000. If you want to communicate with KVLite from some machine other than the local machine, then you must start it using non-default values. The command line options are described later in this chapter.

For example:

```
> java -Xmx256m -Xms256m -jar KVHOME/lib/kvstore.jar kvlite
```

Note

To avoid using too much heap space, you should specify -Xmx and -Xms flags for Java when running administrative and utility commands.

When KVLite has started successfully, it writes one of two statements to stdout, depending on whether it created a new store or is opening an existing store (the following assumes security is enabled):

```
Generated password for user admin: DG38I@zg*6]m
User login file: ./kvroot/security/user.security
Created new kvlite store with args:
-root ./kvroot -store <kvstore> -host localhost -port 5000
-secure-config enable
```

or

```
Opened existing kvlite store with config:
-root ./kvroot -store <kvstore name> -host <localhost> -port 5000
```

```
-secure-config enable
```

where <kvstore name> is the name of the store and <localhost> is the name of the local host. It takes about 10 - 60 seconds before this message is issued, depending on the speed of your machine.

Note that you will not get the command line prompt back until you stop KVLite.

Stopping and Restarting KVLite

To stop KVLite, use ^C from within the shell where KVLite is running.

To restart the process, simply run the `kvlite` utility without any command line options. Do this even if you provided non-standard options when you first started KVLite. This is because KVLite remembers information such as the port value and the store name in between run times. You cannot change these values by using the command line options.

If you want to start over with different options than you initially specified, delete the KVROOT directory (`./kvroot`, by default), and then re-run the `kvlite` utility with whatever options you desire. Alternatively, specify the `-root` command line option, making sure to specify a location other than your original KVROOT directory, as well as any other command line options that you want to change.

Verifying the Installation

There are several things you can do to verify your installation, and ensure that KVLite is running:

- Start another shell and run:

  ```
  jps -m
  ```

  The output should show KVLite (and possibly other things as well, depending on what you have running on your machine).

- Run the `kvclient` test application:

  1. cd KVHOME
  2. java -Xmx256m -Xms256m -jar lib/kvclient.jar

  This should write the release to stdout:

  ```
  12cR1.M.N.O...
  ```

- Compile and run the example program:

  1. cd KVHOME
  2. Compile the example:

     ```
     javac -g -cp lib/kvclient.jar:examples examples/hello/*.java
     ```

3. Run the example using all default parameters:

```
java -Xmx256m -Xms256m \
-cp lib/kvclient.jar:examples hello.HelloBigDataWorld
```

Or run it using non-default parameters, if you started KVLite using non-default values:

```
java -Xmx256m -Xms256m \
-cp lib/kvclient.jar:examples hello.HelloBigDataWorld \
    -host <hostname> -port <hostport> -store <kvstore name>
```

kvlite Utility Command Line Parameter Options

This section describes the command line options that you can use with the `kvlite` utility.

Note that you can only specify these options the first time KVLite is started. Most of the parameter values specified here are recorded in the KVHOME directory, and will be used when you restart the KVLite process regardless of what you provide as command line options. If you want to change your initial values, either delete your KVHOME directory before starting KVLite again, or specify the `-root` option (with a different KVHOME location than you initially used) when you provide the new values.

- `-help`

  Print a brief usage message, and exit.

- `-host <hostname>`

  Identifies the name of the host on which KVLite is running. Use this option ONLY if you are creating a new store.

  If you want to access this instance of KVLite from remote machines, supply the local host's real hostname. Otherwise, specify `localhost` for this option.

- `-noadmin`

  If this option is not specified, the administration user interface is started.

- `-port <port>`

  Identifies the port on which KVLite is listening for client connections. Use this option ONLY if you are creating a new store.

- `-root <path>`

  Identifies the path to the Oracle NoSQL Database home directory. This is the location where the store's database files are contained. The directory identified here must exist. If the appropriate database files do not exist at the location identified by the option, they are created for you.

- `-secure-config <enable|disable>`

If enabled, causes security to be enabled for the store. This means all clients connecting to the store must present security credentials. Security is enabled by default.

- `-store <storename>`

  Identifies the name of a new store. Use this option ONLY if you are creating a new store.

  See Using the Authentication APIs (page 3) for information on configuring your client code to connect to a secure store.

Chapter 3. Record Design Considerations

Oracle NoSQL Database KVStores offer storage of key-value pairs. Each such pair can be thought of as a single record in a database, where the key is used to locate the value. Both the key and the value are application-defined, given some loose restrictions imposed by Oracle NoSQL Database.

Every key in the KVStore is a list of strings. All keys must have one or more major components. Keys can also optionally have one or more minor components.

The value portion of the record can be simply a byte array, or it can use Avro to identify its schema. (See Avro Schemas (page 43) for more information.) The value portion can be as simple or complex as you want it to be.

Note

Avro is deprecated. If you want a fixed schema to define the value portion of a record, it is better to use the Table API. That API offers advantages that the Key/Value API with Avro does not — such as secondary indexes.

As a very simple example, suppose you wanted your store to contain information about people. You might then decide to do this:

- Key major: email address.
- Key minor: various properties, such as the user's street address, phone number, photograph, and name.
- Value: Avro-defined information related to the combination of major and minor key components. So, for example, the value for an email address plus a street address might be multiple fields related to street number, street name, city, and so forth.

This is a very simple example of what you might choose to store in Oracle NoSQL Database. However, from a performance point of view, this example might not be the best way for you to organize your data. How you design both your keys and your values can have important performance implications.

The remainder of this chapter describes the performance issues surrounding Oracle NoSQL Database schema design.

Keys

Oracle NoSQL Database organizes records using keys. All records have one or more major key components and, optionally, one or more minor key components. If minor key components are in use, the combination of the major and minor components uniquely identifies a single record in the store.

Keys are spread evenly using a hash across partitions based on the key's major component(s). Every key must have at least one major component, but you can optionally use a list of major components. This means that records that share the same combination of major key components are guaranteed to be in the same partition, which means they can be efficiently

queried. In addition, records with identical major key components can be operated upon using multiple operations but under a single atomic operation.

Remember that major key components are used to identify which partition contains a record, and that every partition is stored in a single shard. This means that major key components are used to identify which shard stores a given record. The combination of the major key components, plus the data access operation that you want performed is used to identify which node within the shard will service the request. Be aware that you cannot control which physical machine, or even which shard, will be used to store any given piece of data. That is all decided for you by the KV driver.

However, the fact that records are placed on the same physical node based on their major key components means that keys which share major key components can be queried efficiently in a single operation. This is because, conceptually, you are operating on a single physical database when you operate on keys stored together in a single partition. (In reality, a single shard uses multiple physical databases, but that level of complexity is hidden from you when interacting with the store.)

Remember that every partition is placed in a single shard, and that your store will have multiple shards. This is good, because it improves both read and write throughput performance. But in order to take full advantage of that performance enhancement, you need at least as many different major key components as you have partitions. In other words, do not create all your records under a single major key component, or even under a small number of major key components, because doing so will create performance bottle necks as the number of records in your store grow large.

Minor key components also offer performance improvements if used correctly, but in order to understand how you need to understand performance issues surrounding the value portion of your records. We will discuss those issues a little later in this chapter.

What is a Key Component?

A key component is a Java String. Issues of comparison can be answered by examining how Java Strings are compared using your preferred encoding.

Because it is a String, a key component can be anything you want it to be. Typically, some naming scheme is adopted for the application so as to logically organize records.

It helps to think of key components as being locations in a file system path. You can write out a record's components as if they were a file system path delimited by a forward slash ("/"). For example, suppose you used multiple major components to identify a record, and one such record using the following major components: "Smith", and "Bob." Another record might use "Smith" and "Patricia". And a third might use "Wong", and "Bill". Then the major components for those records could be written as:

```
/Smith/Bob
/Smith/Patricia
/Wong/Bill
```

Further, suppose you had different kinds of information about each user that you want to store. Then the different types of information could be further identified using minor

components such as "birthdate", "image", "phonenumber", "userID", and so forth. The minor portion of a key component is separated by the major components by a special slash-hyphen-slash delimiter (`/-/`).

By separating keys into major and minor key components, we could potentially store and operate upon the following records. Those that share a common major component can be operated upon in a single atomic operation:

```
/Smith/Bob/-/birthdate
/Smith/Bob/-/phonenumber
/Smith/Bob/-/image
/Smith/Bob/-/userID
/Smith/Patricia/-/birthdate
/Smith/Patricia/-/phonenumber
/Smith/Patricia/-/image
/Smith/Patricia/-/userID
/Wong/Bill/-/birthdate
/Wong/Bill/-/phonenumber
/Wong/Bill/-/image
/Wong/Bill/-/userID
```

Note that the above keys might not represent the most efficient way to organize your data. We discuss this issue in the next section.

Values

Records in the store are organized as key-value pairs. The *value* is the data that you want to store, manage and retrieve.

In simple cases, values can be organized as a byte array. (If more complexity is required, you should use the Table API.) If so, the mapping of the arrays to data structures (serialization and deserialization) is left entirely to the application.

There are no restrictions on the size of your values. However, you should consider your store's performance when deciding how large you are willing to allow your individual records to become. As is the case with any data storage scheme, the larger your record, the longer it takes to read the information from storage, and to write the information to storage. If your values become so large that they impact store read/write performance, or are even too large to fit into your memory cache (or even your Java heap space) then you should consider storing your values using Oracle NoSQL Database's large object support. See the Using the Large Object API (page 112) introduction for details.

On the other hand, every record carries with it some amount of overhead. Also, as the number of your records grows very large, search times may begin to be adversely affected. As a result, choosing to store an extremely large number of very small records can also harm your store's performance.

Therefore, when designing your store's content, you must find the appropriate balance between a small number of very large records and a large number of very small records. You should also consider how frequently any given piece of information will be accessed.

For example, suppose your store contains information about users, where each user is identified by their email address. There is a set of information that you want to maintain about each user. Some of this information is small in size, and some of it is large. Some of it you expect will be frequently accessed, while other information is infrequently accessed.

Small properties are:

- name
- gender
- address
- phone number

Large properties are:

- image file
- public key 1
- public key 2
- recorded voice greeting

There are several possible ways you can organize this data. How you should do it depends on your data access patterns.

Note

The following example discusses the use of Avro, which is deprecated. While it continues to be possible for you to use Avro to manage your data, the Table API is the better solution.

For example, suppose your application requires you to read and write all of the properties identified above every time you access a record. (This is unlikely, but it does represent the simplest case.) In that event, you might create a single Avro schema that represents each of the properties you maintain for the users in your application. You can then trivially organize your records using only major key components so that, for example, all of the data for user Bob Smith can be accessed using the key `/Smith/Bob`.

However, the chances are good that your application will not require you to access *all* of the properties for a user's record every time you access that record. While it is possible that you will always need to read all of the properties every time you perform a user look up, it is likely that on updates you will operate only on some properties.

Given this, it is useful to consider how frequently data will be accessed, and its size. Large, infrequently accessed properties should use a key other than that used by the frequently accessed properties. The different keys for these large properties can share major key components, while differing in their minor key components. However, if you are using large object support for your large properties, then these must be under a major key that is different from the major key you use for the other properties you are storing.

At the same time, there is overhead involved with every key your store contains, so you do not want to create a key for every possible user property. For this reason, if you have a lot of small properties, you might want to organize them all under a single key even if only some of them are likely to be updated or read for any given operation.

For example, for the properties identified above, suppose the application requires:

- all of the small properties to always be used whenever the user's record is accessed.
- all of the large properties to be read for simple user look ups.
- on user record updates, the public keys are always updated (written) at the same time.
- The image file and recorded voice greeting can be updated independently of everything else.

In this case, you might store user properties using four keys per user. Each key shares the same major components, and differs in its minor component, in the following way:

1. /*surname*/*familiar name*/-/contact

   The value for this key is a Avro record that contains all of the small user properties (name, phone number, address, and so forth).

2. /*surname*/*familiar name*/-/publickeys

   The value for this key is an Avro record that contains the user's public keys. These are always read and written at the same time, so it makes sense to organize them under one key.

3. /image.lob/-/*surname*/*familiar name*

   The value for this key is an image file, saved using Oracle NoSQL Database's large object support.

4. /audio.lob/-/voicegreeting/*surname*/*familiar name*

   The value for this key is an mp3 file, also saved using the large object support.

Any data organized under different keys which differ only in the minor key component allows you to read and update the various properties all at once using a single atomic operation, which gives you full ACID support for user record updates. At the same time, your application does not have to be reading and writing large properties (image files, voice recordings, and so forth) unless it is absolutely necessary. When it is necessary to read or write these large objects, you can use the Oracle NoSQL Database stream interface which is optimized for that kind of traffic.

Chapter 4. Writing and Deleting Records

This chapter discusses two different write operations: putting records into the store, and then deleting them.

Write Exceptions

There are many exceptions that you should handle whenever you perform a write operation to the store. Some of the more common exceptions are described here. For simple cases where you use default policies or are not using a secure store, you can probably avoid explicitly handling these. However, as your code complexity increases, so too will the desirability of explicitly managing these exceptions.

The first of these is `DurabilityException`. This exception indicates that the operation cannot be completed because the durability policy cannot be met. For more information, see Durability Guarantees (page 92).

The second is `RequestTimeoutException`. This simply means that the operation could not be completed within the amount of time provided by the store's timeout property. This probably indicates an overloaded system. Perhaps your network is experiencing a slowdown, or your store's nodes are overloaded with too many operations (especially write operations) coming in too short of a period of time.

To handle a `RequestTimeoutException`, you could simply log the error and move on, or you could pause for a short period of time and then retry the operation. You could also retry the operation, but use a longer timeout value. (There is a version of the `TableAPI.put()` method that allows you to specify a timeout value for that specific operation.)

You can also receive an `IllegalArgumentException`, which will be thrown if a `Row` that you are writing to the store does not have a primary key or is otherwise invalid.

You can also receive a general `FaultException`, which indicates that some exception occurred which is neither a problem with durability nor a problem with the request timeout. Your only recourse here is to either log the error and move along, or retry the operation.

Finally, if you are using a secure store that requires authentication, you can receive `AuthenticationFailureException` or `AuthenticationRequiredException` if you do not provide the proper authentication credentials. When using a secure store, you can also see `UnauthorizedException`, which means you are attempting an operation for which the authenticated user does not have the proper permissions.

Writing Records to the Store

Creating a new record in the store and updating an existing record are usually identical operations (although methods exist that work only if the record is being updated, or only if it is being created — these are described a little later in this section). In both cases, you simply write a record to the store that uses the appropriate key. If a record with that key does not currently exist in the store, then the record is created for you. If a record exists that does use the specified key, then that record is updated with the information that you are writing to the store.

In order to put an ordinary record into the store:

1. Construct a key, making sure to specify all of the key's major and minor path components. For information on major and minor path components, see Record Design Considerations (page 18).
2. Construct a value. This is the actual data that you want to put into the store.
3. Use one of the `KVStore` class's put methods to put the record to the store.

The following is a trivial example of writing a record to the store. It assumes that the `KVStore` handle has already been created. For the sake of simplicity, this example trivially serializes a string to use as the value for the put operation.

```
package kvstore.basicExample;

...

import oracle.kv.Key;
import oracle.kv.Value;
import java.util.ArrayList;

...

ArrayList<String> majorComponents = new ArrayList<String>();
ArrayList<String> minorComponents = new ArrayList<String>();

...

// Define the major and minor path components for the key
majorComponents.add("Smith");
majorComponents.add("Bob");

minorComponents.add("phonenumber");

// Create the key
Key myKey = Key.createKey(majorComponents, minorComponents);

String data = "408 555 5555";

// Create the value. Notice that we serialize the contents of the
// String object when we create the value.
Value myValue = Value.createValue(data.getBytes());

// Now put the record. Note that we do not show the creation of the
// kvstore handle here.

kvstore.put(myKey, myValue);
```

You can also load key/value pairs supplied by special purpose streams into the store. For more information, see Bulk Put Operations (page 25).

Other put Operations

Beyond the very simple usage of the `KVStore.put()` method illustrated above, there are three other important put operations that you can use:

- `KVStore.putIfAbsent()`

  This method will only put the record if the key DOES NOT current exist in the store. That is, this method is successful only if it results in a *create* operation.

- `KVStore.putIfPresent()`

  This method will only put the record if the key already exists in the store. That is, this method is only successful if it results in an *update* operation.

- `KVStore.putIfVersion()`

  This method will put the record only if the value matches the supplied version information. For more information, see Using Versions (page 83).

Bulk Put Operations

Bulk put operations allow you to load records supplied by special purpose streams into the store.

The bulk loading of the entries is optimized to make efficient use of hardware resources. As a result, this operation can achieve much higher throughput when compared with single put APIs.

The behavior of the bulk put operation with respect to duplicate entries contained in different streams is thus undefined. If the duplicate entries are just present in a single stream, then the first entry will be inserted (if it is not already present) and the second entry and subsequent entries will result in the invocation of `EntryStream.keyExists(E)` method. If duplicates exist across streams, then the first entry to win the race is inserted and subsequent duplicates will result in `EntryStream.keyExists(E)` being invoked on them.

To use bulk put, use one of the `KVStore.put()` methods that provide bulk put. These accept a set of streams to bulk load data.

When using these methods, you can also optionally specify a `BulkWriteOptions` class instance which allows you to specify the durability, timeout, and timeout unit to configure the bulk put operation.

For example, suppose you are loading 1000 key/value pairs with 3 input streams:

```
    import java.util.ArrayList;
    import java.util.List;
    import java.util.concurrent.atomic.AtomicLong;
    import oracle.kv.BulkWriteOptions;
    import oracle.kv.EntryStream;
    import oracle.kv.FaultException;
    import oracle.kv.KVStore;
    import oracle.kv.KVStoreConfig;
```

```
    import oracle.kv.KVStoreFactory;
    import oracle.kv.Key;
    import oracle.kv.KeyValue;
    import oracle.kv.Value;

    ...

    // KVStore handle creation is omitted for brevity

    ...
    Integer streamParallelism = 3;
    Integer perShardParallelism = 3;
    Integer heapPercent = 30;
    // In this case, sets the amount of key/value pairs to load
    int nLoad = 1000;

    BulkWriteOptions bulkWriteOptions =
                               new BulkWriteOptions(null, 0, null);
    // Set the number of streams. The default is 1 stream.
    bulkWriteOptions.setStreamParallelism(streamParallelism);
    // Set the number of writer threads per shard.
    // The default is 3 writer threads.
    bulkWriteOptions.setPerShardParallelism(perShardParallelism);
    // Set the percentage of max memory used for bulk put.
    // The default is 40 percent.
    bulkWriteOptions.setBulkHeapPercent(heapPercent);

    final List<EntryStream<KeyValue>> streams =
    new ArrayList<EntryStream<KeyValue>>(streamParallelism);
    final int num = (nLoad + (streamParallelism - 1)) /streamParallelism;
    for (int i = 0; i < streamParallelism; i++) {
     final int min = num * i;
     final int max = Math.min((min + num) , nLoad);
     streams.add(new LoadKVStream(i, min, max));
    }

    store.put(streams, bulkWriteOptions);

    long total = 0;
    long keyExists = 0;
    for (EntryStream<KeyValue> stream: streams) {
     total += ((LoadKVStream)stream).getCount();
     keyExists += ((LoadKVStream)stream).getKeyExistsCount();
    }
    final String fmt = "Loaded %,d records, %,d pre-existing.";
    System.err.println(String.format(fmt, total, keyExists));
}
```

You should implement the stream interface that supplies the data to be batched and loaded into the store. Entries are supplied by a list of `EntryStream` instances. Each stream is read sequentially, that is, each `EntryStream.getNext()` is allowed to finish before the next operation is issued. The load operation typically reads from these streams in parallel as determined by `BulkWriteOptions.getStreamParallelism()`.

```
private class LoadKVStream implements EntryStream<KeyValue> {

    private final String name;
    private final long index;
    private final long max;
    private final long min;
    private long id;
    private long count;
    private final AtomicLong keyExistsCount;

    LoadKVStream(String name, long index, long min, long max) {
        this.index = index;
        this.max = max;
        this.min = min;
        this.name = name;
        id = min;
        count = 0;
        keyExistsCount = new AtomicLong();
    }

    @Override
    public String name() {
        return name + "-" + index + ": " + min + "~" + max;
    }

    @Override
    public KeyValue getNext() {
        if (id++ == max) {
            return null;
        }
        Key key = Key.fromString("/bulk/" + id);
        Value value = Value.createValue(("value"+ id).getBytes());
        KeyValue kv = new KeyValue(key, value);
        count++;
        return kv;
    }

    @Override
    public void completed() {
        System.err.println(name() + " completed, loaded: " + count);
    }

    @Override
    public void keyExists(KeyValue entry) {
```

```
            keyExistsCount.incrementAndGet();
        }

        @Override
        public void catchException
        (RuntimeException exception, KeyValue entry) {
            System.err.println(name() + " catch exception: " +
                               exception.getMessage() + ": " +
                               entry.toString());
            throw exception;
        }

        public long getCount() {
            return count;
        }

        public long getKeyExistsCount() {
            return keyExistsCount.get();
        }
}
```

Deleting Records from the Store

You delete a single record from the store using the `KVStore.delete()` method. Records are deleted based on a key. You can also require a record to match a specified version before it will be deleted. To do this, use the `KVStore.deleteIfVersion()` method. Versions are described in Using Versions (page 83).

When you delete a record, you must handle the same exceptions as occur when you perform any write operation on the store. See Write Exceptions (page 23) for a high-level description of these exceptions.

```
package kvstore.basicExample;

...

import oracle.kv.Key;
import java.util.ArrayList;

...

ArrayList<String> majorComponents = new ArrayList<String>();
ArrayList<String> minorComponents = new ArrayList<String>();

...

// Define the major and minor path components for the key
majorComponents.add("Smith");
majorComponents.add("Bob");
```

```
minorComponents.add("phonenumber");

// Create the key
Key myKey = Key.createKey(majorComponents, minorComponents);

// Now delete the record. Note that we do not show the creation of the
// kvstore handle here.

kvstore.delete(myKey);
```

Using multiDelete()

You can delete multiple records at once, so long as they all share the same major path components. Note that you must provide a complete major path component. You can omit minor path components, or even provide partial path components.

To delete multiple records at once, use the `KVStore.multiDelete()` method.

For example:

```
package kvstore.basicExample;

...

import oracle.kv.Key;
import java.util.ArrayList;

...

ArrayList<String> majorComponents = new ArrayList<String>();

...

// Define the major and minor path components for the key
majorComponents.add("Smith");
majorComponents.add("Bob");

// Create the key
Key myKey = Key.createKey(majorComponents);

// Now delete the record. Note that we do not show the creation of the
// kvstore handle here.

kvstore.multiDelete(myKey, null, null);
```

Chapter 5. Reading Records

There are several ways to retrieve records from the store. You can:

1. Retrieve a single record at a time using `KVStore.get()`.
2. Retrieve records that share a complete set of major components using either `KVStore.multiGet()` or `KVStore.multiGetIterator()`.
3. Retrieve records that share a partial set of major components using `KVStore.storeIterator()`.
4. Retrieve and process records from each shard in parallel using a single key as the retrieval criteria. Use one of the `KVStore.storeIterator()` or `KVStore.storeKeysIterator()` methods that provide parallel scans.
5. Retrieve and process records from each shard in parallel using a sequence of keys as the retrieval criteria. Use one of the `KVStore.storeIterator()` or `KVStore.storeKeysIterator()` methods that provide bulk retrievals.

Each of these are described in the following sections.

Read Exceptions

One of three exceptions can occur when you attempt a read operation in the store. The first of these is `ConsistencyException`. This exception indicates that the operation cannot be completed because the consistency policy cannot be met. For more information, see Consistency Guarantees (page 85).

The second exception is `RequestTimeoutException`. This means that the operation could not be completed within the amount of time provided by the store's timeout property. This probably indicates a store that is attempting to service too many read requests all at once. Remember that your data is partitioned across the shards in your store, with the partitioning occurring based on your shard keys. If you designed your keys such that a large number of read requests are occurring against a single key, you could see request timeouts even if some of the shards in your store are idle.

A request timeout could also be indicative of a network problem that is causing the network to be slow or even completely unresponsive.

To handle a `RequestTimeoutException`, you could simply log the error and move on, or you could pause for a short period of time and then retry the operation. You could also retry the operation, but use a longer timeout value.

You can also receive an `IllegalArgumentException`, which will be thrown if a `Row` that you are writing to the store does not have a primary key or is otherwise invalid.

You can also receive a general `FaultException`, which indicates that some exception occurred which is neither a problem with consistency nor a problem with the request timeout. Your only recourse here is to either log the error and move along, or retry the operation.

Finally, if you are using a secure store that requires authentication, you can receive `AuthenticationFailureException` or `AuthenticationRequiredException` if you do not provide the proper authentication credentials. When using a secure store, you can also see `UnauthorizedException`, which means you are attempting an operation for which the authenticated user does not have the proper permissions.

Retrieving a Single Record

To retrieve a record from the store, use the `KVStore.get()` method. This method returns a `ValueVersion` object. Use `ValueVersion.getValue()` to return the `Value` object associated with the key. It is then up to your application to turn the `Value`'s byte array into a useful form. Normally, this will require the use of an Avro binding. See Avro Bindings (page 58) for details.

For example, in Writing Records to the Store (page 23) we showed a trivial example of storing a key-value pair to the store, where the value was a simple String. The following trivial example shows how to retrieve that record. (Again, this is *not* how your code should deserialize data, because this example does not use Avro to manage the value's schema.)

```
package kvstore.basicExample;

...

import oracle.kv.Key;
import oracle.kv.Value;
import oracle.kv.ValueVersion;
import java.util.ArrayList;

...

ArrayList<String> majorComponents = new ArrayList<String>();
ArrayList<String> minorComponents = new ArrayList<String>();

...

// Define the major and minor path components for the key
majorComponents.add("Smith");
majorComponents.add("Bob");

minorComponents.add("phonenumber");

// Create the key
Key myKey = Key.createKey(majorComponents, minorComponents);

// Now retrieve the record. Note that we do not show the creation of
// the kvstore handle here.

ValueVersion vv = kvstore.get(myKey);
Value v = vv.getValue();
String data = new String(v.getValue());
```

Using multiGet()

`KVStore.multiGet()` allows you to retrieve multiple records at once, so long as they all share the same major path components. The major path components that you provide must represent a *complete* set of components.

Use `KVStore.multiGet()` only if your retrieval set will fit entirely in memory.

For example, suppose you use the following keys:

```
/Hats/-/baseball
/Hats/-/baseball/longbill
/Hats/-/baseball/longbill/blue
/Hats/-/baseball/longbill/red
/Hats/-/baseball/shortbill
/Hats/-/baseball/shortbill/blue
/Hats/-/baseball/shortbill/red
/Hats/-/western
/Hats/-/western/felt
/Hats/-/western/felt/black
/Hats/-/western/felt/gray
/Hat/-/swestern/leather
/Hat/-/swestern/leather/black
/Hat/-/swestern/leather/gray
```

Then you can retrieve all of the records that use the major key component `Hats` as follows:

```
package kvstore.basicExample;

...

import oracle.kv.ConsistencyException;
import oracle.kv.Key;
import oracle.kv.RequestTimeoutException;
import oracle.kv.Value;
import oracle.kv.ValueVersion;

import java.util.ArrayList;
import java.util.Iterator;
import java.util.SortedMap;
import java.util.Map;

...

ArrayList<String> majorComponents = new ArrayList<String>();

...

// Define the major and minor path components for the key
majorComponents.add("Hats");
```

```
// Create the retrieval key
Key myKey = Key.createKey(majorComponents);

// Now retrieve the records. Note that we do not show the creation of
// the kvstore handle here.

SortedMap<Key, ValueVersion> myRecords = null;

try {
    myRecords = kvstore.multiGet(myKey, null, null);
} catch (ConsistencyException ce) {
    // The consistency guarantee was not met
} catch (RequestTimeoutException re) {
    // The operation was not completed within the
    // timeout value
}
```

You can then iterate over the resulting sorted map as follows:

```
for (Map.Entry<Key, ValueVersion> entry : myRecords.entrySet()) {
    ValueVersion vv = entry.getValue();
    Value v = vv.getValue();
    // Do some work with the Value here
}
```

Using multiGetIterator()

If you believe your return set will be so large that it cannot fit into memory, use `KVStore.multiGetIterator()` instead of `KVStore.multiGet()`.

`KVStore.multiGetIterator()` allows you to perform an ordered traversal of a set of keys, as defined by a key and, optionally, a key range. Use this method if you believe your return set will not fit into memory, or if you believe the return set will be so large that it might strain your network resources.

`KVStore.multiGetIterator()` does not return the entire set of records all at once. Instead, it batches the fetching of key-value pairs in the iterator, to minimize the number of network round trips, while not monopolizing the available bandwidth.

Note that this method does not result in a transactional operation. Because the retrieval is batched, the return set can change over the course of the entire retrieval operation. As a result, you lose the atomicity of the operation when you use this method.

This method provides for an ordered traversal of records that share the same major path components. The major path components that you provide must represent a *complete* set of components.

To use this method, you must provide:

- A traversal direction.

- The suggested number of keys to fetch during each network round trip. If you provide a value of 0, an internally determined default is used.
- The key whose child pairs are to be fetched.

Note that there are other possible parameters that you can provide, but this above list represents the minimum information required to use this method.

For example, suppose the following is representative of the keys that you use:

```
/Hats/-/baseball
/Hats/-/baseball/longbill
/Hats/-/baseball/longbill/blue
/Hats/-/baseball/longbill/red
/Hats/-/baseball/shortbill
/Hats/-/baseball/shortbill/blue
/Hats/-/baseball/shortbill/red
/Hats/-/western
/Hats/-/western/felt
/Hats/-/western/felt/black
/Hats/-/western/felt/gray
/Hats/-/western/leather
/Hats/-/western/leather/black
/Hats/-/western/leather/gray
```

Then you can retrieve all of the records that use the major key component `Hats` as follows:

```
package kvstore.basicExample;

...

import oracle.kv.Direction;
import oracle.kv.Key;
import oracle.kv.Value;
import oracle.kv.KeyValueVersion;

import java.util.ArrayList;
import java.util.Iterator;

...

ArrayList<String> majorComponents = new ArrayList<String>();

...

// Define the major and minor path components for the key
majorComponents.add("Hats");

// Create the retrieval key
Key myKey = Key.createKey(majorComponents);
```

```
// Now retrieve the records. Note that we do not show the creation of
// the kvstore handle here.


Iterator<KeyValueVersion> i =
    kvstore.multiGetIterator(Direction.FORWARD, 0,
                             myKey, null, null);
while (i.hasNext()) {
    Value v = i.next().getValue();
    // Do some work with the Value here
}
```

Using storeIterator()

If you want to retrieve all the records that match only some of the major key components, use `KVStore.storeIterator()`. Using this method, you can iterate over all of the records in the store, or over all of the records that match a partial set of major components.

`KVStore.storeIterator()` does not return the entire set of records all at once. Instead, it batches the fetching of key-value pairs in the iterator, to minimize the number of network round trips, while not monopolizing the available bandwidth. Also, the records returned by this method are in unsorted order.

Note that this method does not result in a single atomic operation. Because the retrieval is batched, the return set can change over the course of the entire retrieval operation. As a result, you lose the atomicity of the operation when you use this method.

This method provides for an unsorted traversal of records in the store. If you do not provide a key, then this method will iterate over all of the records in the store. If you do provide a key, you must provide only a subset of the major key components used by your records. The key that you provide must NOT include any minor key components.

To use this method, at a minimum you must specify:

- A traversal direction.
- The suggested number of keys to fetch during each network round trip. If you provide a value of 0, an internally determined default is used.

This minimum list would iterate over all keys in the store. You can also iterate over all the descendants of a specified parent key. Key ranges may also be supplied.

This method performs single-threaded retrieval of records if `StoreIteratorConfig.setMaxConcurrentRequests` is anything other than 1. You might be able to achieve better performance by using parallel scans, which uses multiple threads to retrieve data from the store. See Parallel Scans (page 38) for more information.

For example, suppose you are storing user records that use keys like this:

```
/Smith/Bob/-/birthdate
/Smith/Bob/-/phonenumber
/Smith/Bob/-/image
```

```
/Smith/Bob/-/userID
/Smith/Patricia/-/birthdate
/Smith/Patricia/-/phonenumber
/Smith/Patricia/-/image
/Smith/Patricia/-/userID
/Smith/Richard/-/birthdate
/Smith/Richard/-/phonenumber
/Smith/Richard/-/image
/Smith/Richard/-/userID
/Wong/Bill/-/birthdate
/Wong/Bill/-/phonenumber
/Wong/Bill/-/image
/Wong/Bill/-/userID
```

Then in the simplest case, you can retrieve all of the records for all users whose surname is 'Smith' as follows:

```
package kvstore.basicExample;

...

import oracle.kv.Direction;
import oracle.kv.Key;
import oracle.kv.Value;
import oracle.kv.KeyValueVersion;
import java.util.ArrayList;
import java.util.Iterator;

...

ArrayList<String> majorComponents = new ArrayList<String>();

...

// Define the major and minor path components for the key
majorComponents.add("Smith");

// Create the retrieval key
Key myKey = Key.createKey(majorComponents);

// Now retrieve the records. Note that we do not show the creation of
// the kvstore handle here.


Iterator <KeyValueVersion>i =
    kvstore.storeIterator(Direction.UNORDERED, 0,
                          myKey, null, null);
while (i.hasNext()) {
    Value v = i.next().getValue();
    // Do some work with the Value here
```

```
}
```

Specifying Subranges

When performing multi-key operations in the store, you can specify a range of records to operate upon. You do this using the `KeyRange` class. This class defines a range of `String` values for the key components immediately following a key that is used in a multiple get operation.

For example, suppose you were using the following keys:

```
/Smith/Bob/-/birthdate
/Smith/Bobphone/-/number
/Smith/Bob/-/image
/Smith/Bob/-/userID
/Smith/Patricia/-/birthdate
/Smith/Patricia/-/phonenumber
/Smith/Patricia/-/image
/Smith/Patricia/-/userID
/Smith/Richard/-/birthdate
/Smith/Richard/-/phonenumber
/Smith/Richard/-/image
/Smith/Richard/-/userID
/Wong/Bill/-/birthdate
/Wong/Bill/-/phonenumber
/Wong/Bill/-/image
/Wong/Bill/-/userID
```

Given this, you could perform operations for all the records related to users Bob Smith and Patricia Smith by constructing a `KeyRange`. When you do this, you must identify the key components that defines the upper and lower bounds of the range. You must also identify if the key components that you provide are inclusive or exclusive.

In this case, we will define the start of the key range using the string "Bob" and the end of the key range to be "Patricia". Both ends of the key range will be inclusive.

```
package kvstore.basicExample;

...

import oracle.kv.KeyRange;

...

KeyRange kr = new KeyRange("Bob", true, "Patricia", true);
```

You then use the `KeyRange` instance when you perform your multi-key operation. For example, suppose you you were retrieving records from your store using `KVStore.storeIterator()`:

```
package kvstore.basicExample;

...
```

```
import oracle.kv.Direction;
import oracle.kv.Key;
import oracle.kv.Value;
import oracle.kv.KeyRange;
import oracle.kv.KeyValueVersion;

import java.util.ArrayList;
import java.util.Iterator;

...

ArrayList<String> majorComponents = new ArrayList<String>();

...

// Define the major and minor path components for the key
majorComponents.add("Smith");

// Create the retrieval key
Key myKey = Key.createKey(majorComponents);

KeyRange kr = new KeyRange("Bob", true, "Patricia", true);

// Now retrieve the records. Note that we do not show the creation of
// the kvstore handle here.


Iterator<KeyValueVersion> i =
    kvstore.storeIterator(Direction.UNORDERED, 0,
                          myKey, kr, null);
while (i.hasNext()) {
    Value v = i.next().getValue();
    // Do some work with the Value here
}
```

Parallel Scans

Until now the reads that we have discussed in this chapter are single-threaded. Reads are performed one shard at a time, in sequence, until all the desired records are retrieved. This has obvious performance implications if you are retrieving a large number of records that span multiple shards. However, you can speed up the read performance by using parallel scans.

That is, suppose you have a keyspace that looks like this:

```
/trades/<timestamp>/<symbol>/-/: <price>;<qty>
```

If you want to locate all trades for ORCL which are more than 10k shares, you would have to scan all the records under `/trades` (this part could be done with a key prefix restriction) and examine each record. You would use the `storeIterator()` call to perform this search. The

single-threaded `storeIterator()` retrieves records from each shard consecutively (that is, all records from shard 1, then all from shard 2, etc.).

Parallel Scan retrieves the records from each shard in parallel and allows the client to receive and process them in parallel. You can specify how many threads to use to perform the retrieval. If more threads are specified on the client side, then the user can expect better retrieval performance — until processor or network resources are saturated.

To specify that a parallel scan is to be performed, you use `StoreIteratorConfig` to identify the maximum number of client-side threads to be used for the scan. You can also set the number of results per request, and the maximum number of result batches that the Oracle NoSQL Database client can hold before the scan pauses. You provide this to `StoreIteratorConfig`, and then pass that instance to the overloaded form of `KVStore.storeIterator()` which accepts it. This creates a `ParallelScanIterator`. instance which you use to perform the parallel scan.

For example, to retrieve all of the records in the store using 5 threads in parallel, you would do this:

```
package kvstore.basicExample;

...

import oracle.kv.Consistency;
import oracle.kv.Direction;
import oracle.kv.ParallelScanIterator;
import oracle.kv.StoreIteratorConfig;

...
    /*
     *

    /*
     * Use multi-threading for this store iteration and limit the number
     * of threads (degree of parallelism) to 5.
     */
    final StoreIteratorConfig sc = new StoreIteratorConfig().
        setMaxConcurrentRequests(5);
    ParallelScanIterator<KeyValueVersion> iter = kvstore.storeIterator
        (Direction.UNORDERED,
         0,
         null /* parentKey */,
         null /* subRange */,
         null /* Depth */,
         Consistency.NONE,
         0 /* timeout */,
         null /* timeoutUnit */,
         sc, /* New Arg: StoreIteratorConfig */);

    try {
```

```
        while (iter.hasNext()) {
            KeyValueVersion kvv = iter.next();
            ...
        }
    } finally {
        if (iter != null) {
            iter.close();
        }
    }
```

Bulk Get Operations

Bulk get operations allow you to retrieve and process records from each shard in parallel, like a parallel scan, but using a set of keys instead of a single key as retrieval criteria.

A bulk get operation does not return the entire set of KV pairs all at once. Instead, it batches the fetching of KV pairs in the iterator, to minimize the number of network round trips, while not monopolizing the available bandwidth. Batches are fetched in parallel across multiple Replication Nodes. If more threads are specified on the client side, then the user can expect better retrieval performance – until processor or network resources are saturated.

To use bulk get, use one of the `KVStore.storeIterator()` or `KVStore.storeKeysIterator()` methods that provide bulk retrievals. These accept a set of keys instead of a single key as the retrieval criteria. The set is provided using either an `Iterator<Key>` or `List<Iterator<Key>>` value.

The methods retrieve the KV pairs or keys matching the keys supplied by the iterator(s).

Note

If the iterator yields duplicate keys, the `KeyValueVersion` associated with the duplicate keys will be returned at least once and potentially multiple times.

The supplied parent key must contain the complete major key path. The minor key path may be omitted or may be a partial path.

When using these methods, you can also optionally specify:

- The `Depth` parameter to specify how many children of the parent key to return.
- The `KeyRange` parameter to specify a range of records to operate on.
- `MaxConcurrentRequests` using a `StoreIteratorConfig` class instance to configure the number of threads used to perform the bulk get operation.

Note

If `MaxConcurrentRequests` is not specified, a default value is calculated based on the available hardware.

For example, suppose you use the following keys:

```
/Hats/-/baseball
/Hats/-/baseball/longbill
/Hats/-/baseball/longbill/blue
/Hats/-/baseball/longbill/red
/Hats/-/baseball/shortbill
/Hats/-/baseball/shortbill/blue
/Hats/-/baseball/shortbill/red
/Hats/-/western
/Hats/-/western/felt
/Hats/-/western/felt/black
/Pants/-/western/felt/gray
/Pants/-/baseball/cotton
/Pants/-/baseball/cotton/blue
/Pants/-/baseball/cotton/red
/Shoes/-/baseball/
/Shoes/-/baseball/blue
/Shoes/-/baseball/red
```

If you want to locate all the Hats and Pants used for baseball, using nine threads in parallel, you can retrieve all of the records that use the major key component `Hats` and minor key component `baseball` as well as the records that use the major key component `Pants` and minor key component `baseball` as follows:

```
package kvstore.basicExample;

...
import java.util.ArrayList;
import java.util.List;
import java.util.concurrent.TimeUnit;
import oracle.kv.Consistency;
import oracle.kv.Key;
import oracle.kv.StoreIteratorConfig;
import oracle.kv.ParallelScanIterator;
import oracle.kv.KeyValueVersion;

...

// KVStore handle creation is omitted for brevity

...

// Create the retrieval keys
Key myKey = Key.createKey("Hats","baseball");
Key myOtherKey = Key.createKey("Pants","baseball");


// Use multi-threading for this store iteration and limit the number of
// threads (degree of parallelism) to 9.
final StoreIteratorConfig storeIteratorConfig =
      new StoreIteratorConfig().setMaxConcurrentRequests(9);
```

```
// Create the parent key iterator.
List<Key> searchKeys = new ArrayList<Key>();

// Add the retrieval keys to the list.
searchKeys.add(myKey);
searchKeys.add(myOtherKey);

final ParallelScanIterator<KeyValueVersion> iterator =
      kvstore.storeIterator(searchKeys.iterator(),
                            0, //BatchSize
                            null, //SubRange
                            null, //Depth
                            Consistency.NONE_REQUIRED,
                            0, //Timeout
                            null,
                            storeIteratorConfig);

// Now retrieve the records.
try {
    while (iterator.hasNext()) {
    KeyValueVersion kvv = iterator.next();
    // Do some work with the Value here
    }
} finally {
   if (iterator != null) {
   iterator.close();
   }
}
```

Chapter 6. Avro Schemas

Avro is used to define the data schema for a record's value. This schema describes the fields allowed in the value, along with their data types.

Note

Avro is deprecated. If you want a fixed schema to define the value portion of a record, it is better to use the Table API. That API offers advantages that the Key/Value API with Avro does not — such as secondary indexes.

You apply a schema to the value portion of an Oracle NoSQL Database record using Avro bindings. These bindings are used to serialize values before writing them, and to deserialize values after reading them. The usage of these bindings requires your applications to use the Avro data format, which means that each stored value is associated with a schema.

The use of Avro schemas allows serialized values to be stored in a very space-efficient binary format. Each value is stored without any metadata other than a small internal schema identifier, between 1 and 4 bytes in size. One such reference is stored per key-value pair. In this way, the serialized Avro data format is always associated with the schema used to serialize it, with minimal overhead. This association is made transparently to the application, and the internal schema identifier is managed by the bindings supplied by the AvroCatalog class. The application never sees or uses the internal identifier directly.

The Avro API is the result of an open source project provided by the Apache Software Foundation. It is formally described here: http://avro.apache.org.

In addition, Avro makes use of the Jackson APIs for parsing JSON. This is likely to be of interest to you if you are integrating Oracle NoSQL Database with a JSON-based system. Jackson is formally described here: http://wiki.fasterxml.com/JacksonHome.

Creating Avro Schemas

An Avro schema is created using JSON format. JSON is short for *JavaScript Object Notation*, and it is a lightweight, text-based data interchange format that is intended to be easy for humans to read and write. JSON is described in a great many places, both on the web and in after-market documentation. However, it is formally described in the IETF's RFC 4627, which can be found at http://www.ietf.org/rfc/rfc4627.txt?number=4627.

To describe an Avro schema, you create a JSON record which identifies the schema, like this:

```
{
     "type": "record",
     "namespace": "com.example",
     "name": "FullName",
     "fields": [
       { "name": "first", "type": "string" },
       { "name": "last", "type": "string" }
     ]
}
```

The above example is a JSON record which describes schema that might be used by the value portion of a key-value pair in the store. It describes a schema for a person's full name.

Notice that for the record, there are four fields:

- `type`

  Identifies the JSON field type. For Avro schemas, this must always be `record` when it is specified at the schema's top level. The type `record` means that there will be multiple fields defined.

- `namespace`

  This identifies the namespace in which the object lives. Essentially, this is meant to be a URI that has meaning to you and your organization. It is used to differentiate one schema type from another should they share the same name.

- `name`

  This is the schema name which, when combined with the namespace, uniquely identifies the schema within the store. In the above example, the fully qualified name for the schema is `com.example.FullName`.

- `fields`

  This is the actual schema definition. It defines what fields are contained in the value, and the data type for each field. A field can be a simple data type, such as an integer or a string, or it can be complex data. We describe this in more detail, below.

  Note that schema field names must begin with [A-Za-z_], and subsequently contain only [A-Za-z0-9_].

To use the schema, you must define it in a flat text file, and then add the schema to your store using the appropriate command line call. You must also somehow provide it to your code. The schema that your code is using must correspond to the schema that has been added to your store.

The remainder of this chapter describes schemas and how to add them to your store. For a description of how to use schemas in your code, see Avro Bindings (page 58).

Avro Schema Definitions

Avro schema definitions are JSON records. Because it is a record, it can define multiple fields which are organized in a JSON array. Each such field identifies the field's name as well as its type. The type can be something simple, like an integer, or something complex, like another record.

For example, the following trivial Avro schema definition can be used for a value that contains just someone's age:

```
{
```

```
    "type" : "record",
    "name" : "userInfo",
    "namespace" : "my.example",
    "fields" : [{"name" : "age", "type" : "int"}]
}
```

Of course, if your data storage needs are this simple, you can just use a byte-array to store the integer in the store. (Although this is not considered best practice.)

Notice in the previous example that the top-level type for the schema definition is of type `record`, even though we are defining a single-field schema. Oracle NoSQL Database requires you to use `record` for the top-level type, even if you only need one field.

Also, it is best-practice to define default values for the fields in your schema. While this is optional, should you ever decide to change your schema, it can save you a lot of trouble. To define a default value, use the `default` attribute:

```
{
    "type" : "record",
    "name" : "userInfo",
    "namespace" : "my.example",
    "fields" : [{"name" : "age", "type" : "int", "default" : -1}]
}
```

You almost certainly will not be using single-field definitions. To add multiple fields, specify an array in the `fields` field. For example:

```
{
    "type" : "record",
    "name" : "userInfo",
    "namespace" : "my.example",
    "fields" : [{"name" : "username",
                 "type" : "string",
                 "default" : "NONE"},

                 {"name" : "age",
                 "type" : "int",
                 "default" : -1},

                 {"name" : "phone",
                 "type" : "string",
                 "default" : "NONE"},

                 {"name" : "housenum",
                 "type" : "string",
                 "default" : "NONE"},

                 {"name" : "street",
                 "type" : "string",
                 "default" : "NONE"},
```

```
                {"name" : "city",
                "type" : "string",
                "default" : "NONE"},

                {"name" : "state_province",
                "type" : "string",
                "default" : "NONE"},

                {"name" : "country",
                "type" : "string",
                "default" : "NONE"},

                {"name" : "zip",
                "type" : "string",
                "default" : "NONE"}]
}
```

The above schema definition provides a lot of information. However, simple as it is, you could add some more structure to it by using an embedded record:

```
{
    "type" : "record",
    "name" : "userInfo",
    "namespace" : "my.example",
    "fields" : [{"name" : "username",
                 "type" : "string",
                 "default" : "NONE"},

                {"name" : "age",
                 "type" : "int",
                 "default" : -1},

                 {"name" : "phone",
                  "type" : "string",
                  "default" : "NONE"},

                 {"name" : "housenum",
                  "type" : "string",
                  "default" : "NONE"},

                  {"name" : "address",
                   "type" : {
                         "type" : "record",
                         "name" : "mailing_address",
                         "fields" : [
                            {"name" : "street",
                             "type" : "string",
                             "default" : "NONE"},

                            {"name" : "city",
```

```
                                "type" : "string",
                                "default" : "NONE"},

                               {"name" : "state_prov",
                                "type" : "string",
                                "default" : "NONE"},

                               {"name" : "country",
                                "type" : "string",
                                "default" : "NONE"},

                               {"name" : "zip",
                                "type" : "string",
                                "default" : "NONE"}
                             ]},
                             "default" : {}
                 }
    ]
}
```

Note

It is unlikely that you will need just one record definition for your entire store. Probably you will have more than one type of record. You handle this by providing each of your record definitions individually in separate files. Your code must then be written to handle the different record definitions. We will discuss how to do that later in this chapter.

Primitive Data Types

In the previous Avro schema examples, we have only shown strings and integers. The complete list of primitive types which Avro supports are:

- `null`

  No value.

- `boolean`

  A binary value.

- `int`

  A 32-bit signed integer.

- `long`

  A 64-bit signed integer.

- `float`

A single precision (32 bit) IEEE 754 floating-point number.

- `double`

  A double precision (64-bit) IEEE 754 floating-point number.

- `bytes`

  A sequence of 8-bit unsigned bytes.

- `string`

  A Unicode character sequence.

These primitive types do not have any specified attributes. Primitive type names are also defined type names. For example, the schema "string" is equivalent to:

```
{"type" : "string"}
```

Complex Data Types

Beyond the primitive data types described in the previous section, Avro also supports six complex data types: Records, Enums, Arrays, Maps, Unions, and Fixed. They are described in this section.

record

A record represents an encapsulation of attributes that, all combined, describe a single thing. The attributes that an Avro record supports are:

- name

  This is the record's name, and it is required. It is meant to identify the thing that the record describes. For example: `PersonInformation` or `Automobiles` or `Hats` or `BankDeposit`.

  Note that record names must begin with [A-Za-z_], and subsequently contain only [A-Za-z0-9_].

- namespace

  A namespace is an optional attribute that uniquely identifies the record. It is optional, but it should be used when there is a chance that the record's name will collide with another record's name. For example, suppose you have a record that describes an employee. However, you might have several different types of employees: full-time, part time, and contractors. So you might then create all three types of records with the name `EmployeeInfo`, but then with namespaces such as `FullTime`, `PartTime` and `Contractor`. The fully qualified name for the records used to describe full time employees would then be `FullTime.EmployeeInfo`.

  Alternatively, if your store contains information for many different organizations, you might want to use a namespace that identifies the organization used by the record

so as to avoid collisions in the record names. In this case, you could end up with fully qualified records with names such as `My.Company.Manufacturing.EmployeeInfo` and `My.Company.Sales.EmployeeInfo`.

- doc

  This optional attribute simply provides documentation about the record. It is parsed and stored with the schema, and is available from the Schema object using the Avro API, but it is not used during serialization.

- aliases

  This optional attribute provides a JSON array of strings that are alternative names for the record. Note that there is no such thing as a rename operation for JSON schema. So if you want to refer to a schema by a name other than what you initially defined in the `name` attribute, use an alias.

- type

  A required attribute that is either the keyword `record`, or an embedded JSON schema definition. If this attribute is for the top-level schema definition, `record` must be used.

- fields

  A required attribute that provides a JSON array which lists all of the fields in the schema. Each field must provide a name and a type attribute. Each field may provide doc, order, aliases and default attributes:

  - The name, type, doc and aliases attributes are used in the exact same way as described earlier in this section.

    As is the case with record names, field names must begin with [A-Za-z_], and subsequently contain only [A-Za-z0-9_].

  - The order attribute is optional, and it is ignored by Oracle NoSQL Database. For applications (other than Oracle NoSQL Database) that honor it, this attribute describes how this field impacts sort ordering of this record. Valid values are `ascending`, `descending`, or `ignore`. For more information on how this works, see http://avro.apache.org/docs/current/spec.html#order.

  - The default attribute is optional, but highly recommended in order to support schema evolution. It provides a default value for the field that is used only for the purposes of schema evolution. Use of the default attribute does not mean that you can fail to initialize the field when creating a new value object; all fields must be initialized regardless of whether the default attribute is present.

    Schema evolution is described in Schema Evolution (page 51).

    Permitted values for the default attribute depend on the field's type. Default values for unions depend on the first field in the union. Default values for bytes and fixed fields are JSON strings.

Enum

Enums are enumerated types, and it supports the following attributes

- name

  A required attribute that provides the name for the enum. This name must begin with [A-Za-z_], and subsequently contain only [A-Za-z0-9_].

- namespace

  An optional attribute that qualifies the enum's name attribute.

- aliases

  An optional attribute that provides a JSON array of alternative names for the enum.

- doc

  An optional attribute that provides a comment string for the enum.

- symbols

  A required attribute that provides the enum's symbols as an array of names. These symbols must begin with [A-Za-z_], and subsequently contain only [A-Za-z0-9_].

For example:

```
{ "type" : "enum",
  "name" : "Colors",
  "namespace" : "palette",
  "doc" : "Colors supported by the palette.",
  "symbols" : ["WHITE", "BLUE", "GREEN", "RED", "BLACK"]}
```

Arrays

Defines an array field. It only supports the items attribute, which is required. The items attribute identifies the type of the items in the array:

```
{"type" : "array", "items" : "string"}
```

Maps

A map is an associative array, or dictionary, that organizes data as key-value pairs. The key for an Avro map must be a string. Avro maps supports only one attribute: values. This attribute is required and it defines the type for the value portion of the map.

```
{"type" : "map", "values" : "int"}
```

Unions

A union is used to indicate that a field may have more than one type. They are represented as JSON arrays.

For example, suppose you had a field that could be either a string or null. Then the union is represented as:

```
["string", "null"]
```

You might use this in the following way:

```
{
     "type": "record",
     "namespace": "com.example",
     "name": "FullName",
     "fields": [
       { "name": "first", "type": ["string", "null"] },
       { "name": "last", "type": "string", "default" : "Doe" }
     ]
}
```

Fixed

A fixed type is used to declare a fixed-sized field that can be used for storing binary data. It has two required attributes: the field's name, and the size in 1-byte quantities.

For example, to define a fixed field that is one megabyte in size:

```
{"type" : "fixed" , "name" : "bdata", "size" : 1048576}
```

Using Avro Schemas

Once you have defined your schema, you make use of it in your Oracle NoSQL Database application in the following way:

1. Add the schema to your store. See Managing Avro Schema in the Store (page 55) for information on how to do this.
2. Identify the schema to your application.
3. Serialize and/or deserialize Oracle NoSQL Database values which use the Avro data format. You use Avro bindings to perform the serialization functions. There are different bindings available to you, each of which offers pluses and negatives. We will describe the different bindings later in this section.

Other than that, the mechanisms you use to read/write/delete records in the store do not change just because you are using the Avro data format with your values. Avro affects your code only where you manage your values.

The following sections describe the bindings that you use to serialize and deserialize your data. The binding that you use defines how you provide your schema to the store.

Schema Evolution

Schema evolution is the term used for how the store behaves when Avro schema is changed after data has been written to the store using an older version of that schema. To change

an existing schema, you update the schema as stored in its flat-text file, then add the new schema to the store using the `ddl add-schema` command with the `-evolve` flag.

For example, if a middle name property is added to the FullName schema, it might be stored in a file named `schema2.avsc`, and then added to the store using the `ddl add-schema` command.

Note that when you change schema, the new field must be given a default value. This prevents errors when clients using an old version of the schema create new values that will be missing the new field:

```
{
   "type": "record",
   "namespace": "com.example",
   "name": "FullName",
   "fields": [
     { "name": "first", "type": "string" },
     { "name": "middle", "type": "string", "default": "" },
     { "name": "last", "type": "string" }
   ]
 }
```

These are the modifications you can safely perform to your schema without any concerns:

- A field with a default value is added.
- A field that was previously defined with a default value is removed.
- A field's doc attribute is changed, added or removed.
- A field's order attribute is changed, added or removed.
- A field's default value is added, or changed.
- Field or type aliases are added, or removed.
- A non-union type may be changed to a union that contains only the original type, or vice-versa.

Beyond these kind of changes, there are unsafe changes that you can do which will either cause the schema to be rejected when you attempt to add it to the store, or which can be performed so long as you are careful about how you go about upgrading clients which use the schema. These type of issues are identified when you try to modify (evolve) schema that is currently enabled in the store. See Changing Schema (page 56) for details.

Rules for Changing Schema

There are a few rules you need to remember if you are modifying schema that is already in use in your store:

1. For best results, always provide a default value for the fields in your schema. This makes it possible to delete fields later on if you decide it is necessary. *If you do not provide a default value for a field, you cannot delete that field from your schema.*

2. You cannot change a field's data type. If you have decided that a field should be some data type other than what it was originally created using, then add a whole new field to your schema that uses the appropriate data type.

3. When adding a field to your schema, you must provide a default value for the field.

4. You cannot rename an existing field. However, if you want to access the field by some name other than what it was originally created using, add and use aliases for the field.

Writer and Reader Schema

When a schema is changed, multiple versions of the schema will exist and be maintained by the store. The version of the schema used to serialize a value, before writing it to the store, is called the *writer schema*. The writer schema is specified by the application when creating a binding. It is associated with the value when calling the binding's `AvroBinding.toValue()` method to serialize the data. This writer schema is associated internally with every stored value.

The *reader schema* is used to deserialize a value after reading it from the store. Like the writer schema, the reader schema is specified by the client application when creating a binding. It is used to deserialize the data when calling the binding's `AvroBinding.toObject()` method, after reading a value from the store.

How Schema Evolution Works

Schema evolution is the automatic transformation of Avro schema. This transformation is between the version of the schema that the client is using (its local copy), and what is currently contained in the store. When the local copy of the schema is not identical to the schema used to write the value (that is, when the reader schema is different from the writer schema), this data transformation is performed. When the reader schema matches the schema used to write the value, no transformation is necessary.

Schema evolution is applied only during deserialization. If the reader schema is different from the value's writer schema, then the value is automatically modified during deserialization to conform to the reader schema. To do this, default values are used.

There are two cases to consider when using schema evolution: when you add a field and when you delete a field. Schema evolution takes care of both scenarios, so long as you originally assigned default values to the fields that were deleted, and assigned default values to the fields that were added.

Adding Fields

Suppose you had the following schema:

```
{
    "type" : "record",
    "name" : "userInfo",
    "namespace" : "my.example",
    "fields" : [{"name" : "name", "type" : "string", "default" : ""}]
}
```

In version 2 of the schema, you add a field:

```
{
    "type" : "record",
    "name" : "userInfo",
    "namespace" : "my.example",
    "fields" : [{"name" : "name", "type" : "string", "default" : ""},
                {"name" : "age", "type" : "int" , "default" : -1}]

}
```

In this scenario, a client that is using the new schema can deserialize a value that uses the old schema, even though the age field will be missing from the value. Upon deserialization, the value retrieved from the store will be automatically transformed such that the age field is contained in the value. The age field will be set to the default value, which is -1 in this case.

The reverse also works. A client that is using the old version of the schema attempts can deserialize a value that was written using the new version of the schema. In this case, the value retrieved from the store contains the age field, which from the client perspective is unexpected. So upon deserialization the age field is automatically removed from the retrieved object.

This has ramifications if you change your schema, and then have clients concurrently running that are using different schema versions. This scenario is not unusual in a large, distributed system of the type that Oracle NoSQL Database supports.

In this scenario, you might see fields revert to their default value, even though no client has explicitly touched those fields. This can happen in the following way:

1. Client v.2 creates a my.example.userInfo record, and sets the age field to 38. Then it writes that value to the store. Client v.2 is using schema version 2.

2. Client v.1 reads the record. It is using version 1 of the schema, so the age field is automatically removed from the value during deserialization.

   Client v.1 modifies the name field and then writes the record back to the store. When it does this, the age field is missing from the value that it writes to the store.

3. Client v.2 reads the record again. Because the age field is missing from the record (because Client v.1 last wrote it), the age field is set to the default value, which is -1. This means that the value of the age field has reverted to the default, even though no client explicitly modified it.

Deleting Fields

Field deletion works largely the same way as field addition, with the same concern for field values automatically reverting to the default. Suppose you had the following trivial schema:

```
{
    "type" : "record",
    "name" : "userInfo",
    "namespace" : "my.example",
```

```
    "fields" : [{"name" : "name", "type" : "string", "default" : ""},
                {"name" : "age", "type" : "int" , "default" : -1}]

}
```

In version 2 of the schema, you delete the age field:

```
{
    "type" : "record",
    "name" : "userInfo",
    "namespace" : "my.example",
    "fields" : [{"name" : "name", "type" : "string", "default" : ""}]
}
```

In this scenario, a client that is using the new schema can deserialize a value that uses the old schema, even though the age field is contained in that value. In this case, the age field is silently removed from the value during deserialization.

Further, a client that is using the old version of the schema attempts can deserialize a value that uses the new version of the schema. In this case, the value retrieved from the store does not contain the age field. So upon deserialization, the age field is automatically inserted into the schema (because the reader schema requires it) and the default value is used for the newly inserted field.

As with adding fields, this has ramifications if you change your schema, and then have clients concurrently running that are using different schema versions.

1. Client v.1 creates a my.example.userInfo record, and sets the age field to 38. Then it writes that value to the store. Client v.1 is using schema version 1.

2. Client v.2 reads the record. It is using version 2 of the schema, so it is not expecting the age field. As a result, the age field is automatically stripped from the value during deserialization.

   Client v.2 modifies the name field and then writes the record back to the store. When it does this, the age field is missing from the value that it writes to the store.

3. Client v.1 reads the record again. Because the age field is missing from the record (because Client v.2 last wrote it), the age field is automatically inserted into the value, using the default of -1. This means that the value of the age field has reverted to the default, even though no client explicitly modified it.

Managing Avro Schema in the Store

This section describes how to add, change, disable and enable, and show the Avro schema in your store.

Adding Schema

Avro schema is defined in a flat-text file, and then added to the store using the command line interface. For example, suppose you have schema defined in a file called `my_schema.avsc`.

Then (assuming your store is running) you start your command line interface and add the schema like this:

```
> java -Xmx256m -Xms256m \
-jar <kvhome>/lib/kvstore.jar runadmin -port <port> -host <host> \
-security USER/security/admin.security

kv-> ddl add-schema -file my_schema.avsc
```

Note

This assumes that you must have followed the steps as mentioned in Configuring Security with Remote Access section in the *Oracle NoSQL Database Administrator's Guide*.

Note that when adding schema to the store, some error checking is performed to ensure that the schema is correctly formed. Errors are problems that must be addressed before the schema can be added to the store. Warnings are problems that should be addressed, but are not so serious that the CLI refuses to add the schema. However, to add schema with Warnings, you must use the `-force` switch.

As of this release, the only Error that can be produced is if a field's default value does not conform to the field's type. That is, if the schema provides an integer as the default value where a string is required.

As of this release, the only Warning that can be produced is if the schema does not provide a default value for every field in the schema. Default values are required if you ever want to change (evolve) the schema. But in all other cases, the lack of a default value does no harm.

Changing Schema

To change (evolve) existing schema, use the `-evolve` flag:

```
kv-> ddl add-schema -file my_schema.avsc -evolve
```

Note that when changing schema in the store, some error checking is performed to ensure that schema evolution can be performed correctly. This error checking consists of comparing the new schema to all currently enabled versions of that schema.

This error checking can result in either Errors or Warnings. Errors are fatal problems that must be addressed before the modified schema can be added to the store. Errors represent situations where data written with an old version of the schema cannot be read by clients using a new version of the schema.

Possible errors are:

- A field is added without a default value.
- The size of a fixed type is changed.
- An enum symbol is removed.

- A union type is removed or, equivalently, a union type is changed to a non-union type and the new type is not the sole type in the old union.
- A change to a field's type (specifically to a different type name) is considered an error except when it is a type promotion, as defined by the Avro spec. And even a type promotion is a warning; see below. Another exception is changing from a non-union to a union; see below.

Warnings are problems that can be avoided using a two-phase upgrade process. In a two-phase upgrade, all clients begin using the schema only for reading in phase I (the old schema is still used for writing), and then use the new schema for both reading and writing in phase II. Phase II may not be begun until phase I is complete; that is, no client may use the new schema for writing until all clients are using it for reading.

Possible Warnings are:

- A field is deleted in the new schema when it does not contain a default value in the old schema.
- An enum symbol is added.
- A union type is added or, equivalently, a non-union type is changed to a union that includes the original type and additional types.
- A field's type is promoted, as defined by the Avro spec. Type promotions are: int to long, float or double; long to float or double; float to double.

Disabling and Enabling Schema

You cannot delete schema, but you can disable it:

```
kv-> ddl disable-schema -name avro.MyInfo.1
```

To enable schema that has been disabled:

```
kv-> ddl enable-schema -name avro.MyInfo.1
```

Showing Schema

To see all the schemas currently enabled in your store:

```
kv-> show schemas
```

To see all schemas, including those which are currently disabled:

```
kv-> show schemas -disabled
```

Chapter 7. Avro Bindings

Once you have defined your schema (as described in Avro Schemas (page 43)), you make use of it in your Oracle NoSQL Database application in the following way:

Note

Avro is deprecated. If you want a fixed schema to define the value portion of a record, it is better to use the Table API. That API offers advantages that the Key/Value API with Avro does not — such as secondary indexes.

1. Add the schema to your store. See Managing Avro Schema in the Store (page 55) for information on how to do this.
2. Identify the schema to your application.
3. Serialize and/or deserialize Oracle NoSQL Database values which use the Avro data format. You use Avro bindings to perform the serialization functions. There are different bindings available to you, each of which offers pluses and negatives.

Other than that, the mechanisms you use to read/write/delete records in the store do not change just because you are using the Avro data format with your values. Avro affects your code only where you manage your values.

The following sections describe the bindings that you use to serialize and deserialize your data. The binding that you use defines how you provide your schema to the store.

Avro Bindings Overview

There are four bindings you can use to serialize and deserialize store values that use the Avro data format. Each binding has strengths and weaknesses. The following sections go into three of the four in some detail. But, briefly, the four bindings are:

- Generic

  A generic binding is a general-purpose binding. Generic bindings identify fields to be read and written by supplying a simple string that names the field. This allows you to use the binding in a somewhat generic way, but it suffers from a lack of type safety at application compile time.

  If you are just starting out with Avro, and have no initial reason to prefer one of the other bindings, then you should start with the generic binding.

  Generic bindings are described in Generic Binding (page 59).

- Specific

  Specific bindings make use of classes that are generated from your schema specifications using an Ant tool. The generated classes allow you to manage the fields in your schema

using getter and setter methods. As a programming methodology, specific schemas might represent a familiar programming style, depending on your past experiences.

Unlike generic and JSON bindings, specific bindings require you to know in advance of compile time what all of your store schemas will be. This is because specific bindings make use of classes that are generated from your schema specifications using an Avro tool that can be called from Ant, and so there is no way for your application to dynamically discover the set of all schemas in use in the store.

Specific bindings are described in Specific Binding (page 68).

- JSON

  JSON bindings behave similarly to generic bindings, but offer less support for Avro data types. JSON bindings are most useful if you are integrating your Oracle NoSQL Database client code into a system that is JSON-oriented, because they expose JSON objects.

  JSON bindings are described in JSON Bindings (page 72).

- Raw

  Raw bindings is an advanced feature that allows you to use low-level Avro APIs to serialize and deserialize objects. Raw bindings should be used when the other three built-in bindings will not work for you, for whatever reason. Usage of Raw bindings requires a good understanding of the Avro API set. As a result, their usage is beyond the scope of this manual.

Note that you can read using one binding and write using another one. The Avro data format (which is what is actually used for the data placed into the store) is binding independent. Put another way, the bindings know how to read and write the Avro data format; the data format itself has no knowledge of the bindings.

Generic Binding

Generic bindings provide the widest support for the Avro data types. They are also flexible in that your application does not need to know the entire set of schemas in use in the store at compile time. This provides you good flexibility if your store has a constantly expanding set of schema.

The downside to generic bindings is that they do not provide compile-time type safety. Generic bindings identify fields using a string (as opposed to getter and setter methods provided by specific bindings), so it is not possible for the compiler to know, for example, whether you are using an integer where a real is expected.

Generic binding uses `AvroCatalog.getGenericBinding()` for a single schema binding, and uses `AvroCatalog.getGenericMultiBinding()` when using multiple schemas.

Using a Single Generic Schema Binding

```
{
    "type": "record",
```

```
    "name": "PersonInformation",
    "namespace": "avro",
    "fields": {"name": "ID", "type": "int"}
}
```

Further, suppose you placed that schema in a file named `PersonSchema.avsc`.

Then to use that schema, first add it to your store using the `ddl add-schema` command:

```
> java -Xmx256m -Xms256m \
-jar <kvhome>/kvstore.jar runadmin -port <port> \
-host <host>
kv-> ddl add-schema -file PersonSchema.avsc
```

In your Oracle NoSQL Database client code, you must make the schema available to the code. One way to do this is to read the schema directly from the file where you created it:

```
package avro;

import java.io.File;
import org.apache.avro.generic.GenericData;
import org.apache.avro.generic.GenericRecord;
import org.apache.avro.Schema;

import oracle.kv.avro.AvroCatalog;
import oracle.kv.avro.GenericAvroBinding;

...

final Schema.Parser parser = new Schema.Parser();
parser.parse(new File("PersonSchema.avsc"));
```

Next, you need to make the schema available to your application:

```
final Schema personSchema =
    parser.getTypes().get("avro.PersonInformation");
```

Finally, before you can begin serializing and deserializing values that use the Avro data format, you must create a binding and then create an Avro record for that binding. In this example, we use the generic binding. But as we explain later in this chapter, there are other bindings and the generic binding may not necessarily be the best one for your purposes.

```
/**
 * Here, for the sake of brevity, we skip the necessary steps of
 * declaring and opening the store handle.
 */
final AvroCatalog catalog = store.getAvroCatalog();
final GenericAvroBinding binding =
    catalog.getGenericBinding(personSchema);
```

Once you have the binding, you need a way for your application to represent the fields in the schema, so that they can be read and written. You do this by creating an *Avro record*,

which is a data structure that allows you to read and/or write the fields in the schema. (Do not confuse an Avro record, which is a handle to a binary object containing data, to an Oracle NoSQL Database record, which is a single key-value pair contained in your store. An Oracle NoSQL Database record can contain a value that uses the Avro data format. An instance of the Avro data format, in turn, is managed within your client code using an Avro record.)

Because we are using the generic binding for this example, we will use the `GenericRecord` to manage the contents of the binding.

For example, assume we performed a store read, and now we want to examine the information stored with the Oracle NoSQL Database record.

```
/**
  * Assume a store read was performed here, and resulted in a
  * ValueVersion instance called 'vv'. Then, to deserialize
  * the value in the returned record:
  */
final GenericRecord member;
final int ID;
if (vv != null) {
    /* Deserialize the the value */
    member = binding.toObject(vv.getValue());
    /* Retrieve the contents of the ID field. Because we are
     * using a generic binding, we must type cast the retrieved
     * value.
     */
     ID = (Integer) member.get("ID");
}
```

If we want to write to a field (that is, we want to serialize some data), we use the record's `put()` method. As an example, suppose we wanted to create a brand new Avro object to be written to the store. Then:

```
final GenericRecord person = new GenericData.Record(personSchema);
final int ID = 100011;
person.put("ID", ID);

/**
  * To serialize this information so that it can be written to
  * the store, use GenericBinding.toValue() as the value for the
  * store put(). That is, assuming you already have a store handle
  * and a key:
  */
store.put(key, binding.toValue(person));
```

Using Multiple Generic Schema Bindings

It is unlikely that you will use only one schema with your application. In order to use more than one schema:

1. Specify each schema individually in separate files.

2. Add all these schemas to your store as described in Managing Avro Schema in the Store (page 55).
3. Use `HashMap` to organize your schemas, and then pass that to `AvroCatalog.getGenericMultiBinding()` in order to create your binding.

For example, suppose you had the following two schemas:

```
{
 "type": "record",
 "namespace": "avro",
 "name": "PersonInfo",
 "fields": [
   { "name": "first", "type": "string" },
   { "name": "last", "type": "string" },
   { "name": "age", "type": "int" }
 ]
}


{
 "type": "record",
 "namespace": "avro",
 "name": "AnimalInfo",
 "fields": [
   { "name": "species", "type": "string"},
   { "name": "name", "type": "string"},
   { "name": "age", "type": "int"}
 ]
}
```

Then put `Avro.PersonInfo` in a file (call it `PersonSchema.avsc`) and `Avro.AnimalInfo` in a second file (`AnimalSchema.avsc`). Add these schemas to your store using the command line interface.

At this point, you could simply create one binding for each schema that you are using, but that can quickly become awkward depending on how many schemas your code is using. Instead, create multiple schemas using `HashMap` and (in this case) `AvroCatalog.getGenericMultiBinding()`. To do this, first create a `HashMap` that you use to organize your schemas:

```
package avro;

import java.io.File;
import java.io.IOException;
import java.util.Arrays;
import java.util.HashMap;

import org.apache.avro.Schema;
import org.apache.avro.generic.GenericData;
import org.apache.avro.generic.GenericRecord;
```

```
...

import oracle.kv.ValueVersion;
import oracle.kv.avro.AvroCatalog;
import oracle.kv.avro.GenericAvroBinding;

...

HashMap<String, Schema> schemas = new HashMap<String, Schema>();
```

Then, parse each schema and add it to the `HashMap`:

```
final Schema.Parser parser = new Schema.Parser();

Schema personSchema = parser.parse(new File("PersonSchema.avsc"));
schemas.put(personSchema.getFullName(), personSchema);

Schema animalSchema = parser.parse(new File("AnimalSchema.avsc"));
schemas.put(animalSchema.getFullName(), animalSchema);
```

Then create your binding. You will only need one, because you are using a multi binding which is capable of using multiple schemas.

```
/*
 * Store creation is skipped for brevity
 */

catalog = store.getAvroCatalog();
binding = catalog.getGenericMultiBinding(schemas);
```

To use the binding, you call `toObject()` or `put()` in the same way as you would if you were using an ordinary single-schema binding. The multi-binding is capable of determining which schema you are using, and serializing/deserializing accordingly. For example, suppose you retrieve a record that uses the `Avro.AnimalInfo` schema. Then you can deserialize as if you are using a single-schema binding:

```
/*
 * Key creation and store retrieval skipped.
 * Assume we have retrieved a ValueVersion (vv1) that
 * contains an AnimalInfo value.
 */

final GenericRecord animalObject;
if (vv1 != null) {
    animalObject = binding.toObject(vv1.getValue());
    final String species = animalObject.get("species").toString();
    final String name = animalObject.get("name").toString();
    final int age = (Integer) animalObject.get("age");

    /* Do something with the data */
}
```

You can also create a new `Avro.PersonInfo` object for placement in the store using the same binding, like this:

```
final GenericRecord personObject =
    new GenericData.Record(personSchema);
personObject.put("name", "Sam Brown");
personObject.put("age", 34);

/*
 * Key creation and store handle creation skipped
 * for brevity's sake.
 */

store.put(aKey, binding.toValue(personObject));
```

Using Embedded Records

Suppose you have a schema that looks like this:

```
{
    "type" : "record",
    "name" : "hatInventory",
    "namespace" : "avro",
    "fields" : [{"name" : "sku", "type" : "string", "default" : ""},
                  {"name" : "description",
                     "type" : {
                         "type" : "record",
                         "name" : "hatInfo",
                         "fields" : [
                                     {"name" : "style",
                                      "type" : "string",
                                      "default" : ""},

                                     {"name" : "size",
                                      "type" : "string",
                                      "default" : ""},

                                     {"name" : "color",
                                      "type" : "string",
                                      "default" : ""},

                                     {"name" : "material",
                                      "type" : "string",
                                      "default" : ""}
                                ]}
                }
    ]
}
```

In order to address the fields in the embedded record `hatInfo`, you treat it as if it is a second piece of standalone schema. You only have to parse the schema file once. You then create two

schemas and two records, but only one binding. For example, to create a serialized object that uses this schema:

```
package avro;

import java.io.File;

import org.apache.avro.generic.GenericData;
import org.apache.avro.generic.GenericRecord;
import org.apache.avro.Schema;

import oracle.kv.KVStore;
import oracle.kv.Key;
import oracle.kv.ValueVersion;
import oracle.kv.avro.AvroCatalog;
import oracle.kv.avro.GenericAvroBinding;

...

// Parse our schema file
final Schema.Parser parser = new Schema.Parser();
try {
    parser.parse(new File("HatSchema.avsc"));
} catch (IOException io) {
    io.printStackTrace();
}


// Get two Schema objects. We need two because of the
// embedded record.
final Schema hatInventorySchema =
    parser.getTypes().get("avro.hatInventory");
final Schema hatInfoSchema =
    parser.getTypes().get("avro.hatInfo");

// Get two GenericRecords so we can manipulate both of
// the records in the schema
final GenericRecord hatRecord =
    new GenericData.Record(hatInventorySchema);
final GenericRecord hatInfoRecord =
    new GenericData.Record(hatInfoSchema);

// Now populate our records. Start with the
// embedded record.
hatInfoRecord.put("style", "western");
hatInfoRecord.put("size", "medium");
hatInfoRecord.put("color", "black");
hatInfoRecord.put("material", "leather");

// Now the top-level record. Notice that we
```

```
// set the embedded record as the value for the
// description field.
hatRecord.put("sku", "289163009");
hatRecord.put("description", hatInfoRecord);

// Now we need a binding. Only one is required,
// and we use the top-level schema to create it.
final AvroCatalog catalog = store.getAvroCatalog();
final GenericAvroBinding hatBinding =
    catalog.getGenericBinding(hatInventorySchema);

// Create a Key and write the value to the store.
final Key key = Key.createKey(Arrays.asList("hats", "0000000033"));
store.put(key, hatBinding.toValue(hatRecord));
```

On retrieval, you edit values of this type in the following way:

```
// Perform the retrieval
final ValueVersion vv = store.get(key);
if (vv != null) {
    // Deserialize the ValueVersion as normal
    GenericRecord hatR =
        new GenericData.Record(hatInventorySchema);
    hatR = hatBinding.toObject(vv.getValue());

    // To access the embedded record, create a GenericRecord
    // using the embedded record's schema. Then get the
    // embedded record from the field on the top-level
    // schema that contains it.
    GenericRecord hatInfoR =
        new GenericData.Record(hatInfoSchema);
    hatInfoR = (GenericRecord) hatR.get("description");

    // Finally, you can write to the top-level record and the
    // embedded record like this:

    // Modify a field on the embedded record:
    hatInfoR.put("style", "Fedora");

    // Modify the top-level record:
    hatR.put("sku", "300");
    hatR.put("description", hatInfoR);

    store.put(key, hatBinding.toValue(hatR)); }
```

Managing Generic Schemas Dynamically

A special use-case of generic bindings is that you do not necessarily know about all the schemas that will ever be used by your store at the time you write your code. That is, the use of a `HashMap` in the previous example is somewhat brittle if you are operating in an

environment with a constantly growing list of schemas. In that scenario, whenever you add to the schemas in use by your store, you potentially might need to rewrite your client code to add the new schemas to the `HashMap` used by your client. Otherwise, your code could retrieve a value that uses a schema which is unknown to your code. Depending on what your code is doing, this can cause you problems.

If this is a problem for you, you can avoid it by using `AvroCatalog.getCurrentSchemas()` with `AvroCatalog.getGenericMultiBinding()` so that you do not need to build a `HashMap` of all your schemas.

For example, in the previous example we showed client code that used two known schemas. We can change the previous example to use `getCurrentSchemas()` in the following way:

```
package avro;

import java.io.File;
import java.io.IOException;
import java.util.Arrays;

import org.apache.avro.Schema;
import org.apache.avro.generic.GenericData;
import org.apache.avro.generic.GenericRecord;

...

import oracle.kv.ValueVersion;
import oracle.kv.avro.AvroCatalog;
import oracle.kv.avro.GenericAvroBinding;

...

final Schema.Parser parser = new Schema.Parser();
Schema animalSchema = parser.parse(new File("AnimalSchema.avsc"));


/*
 * We skip creating a HashMap of Schemas because
 * it is not needed.
 */


/*
 * Store creation is skipped for brevity
 */

catalog = store.getAvroCatalog();
binding = catalog.getGenericMultiBinding(catalog.getCurrentSchemas());
```

If we then perform a read on the store, there is the possibility that we will retrieve an object that uses a schema which was not in use when our binding was created. (This

is particularly true for long-running client code). To handle this problem, we catch `SchemaNotAllowedException`.

```
/*
 * Key creation and store retrieval skipped.
 */

final GenericRecord animalObject;
if (vv1 != null) {
    try {
        animalObject = binding.toObject(vv1.getValue());
    } catch (SchemaNotAllowedException e) {
        // Take some action here. Potentially you could
        // recreate your binding, and then retry the
        // deserialization process
    }

    /*
     * Do something with the data. If your client code is
     * using more than one schema, you can identify which
     * schema the retrieved value is using by testing
     * the schema name. That is:
     *
     * String sName = animalObject.getSchema().getFullName()
     * if (sName.equals("avro.animalInfo")) { ... }
     */
}
```

Specific Binding

The `SpecificAvroBinding` interface represents values as instances of a generated Avro class which implements `SpecificRecord`. A single schema binding is created using `AvroCatalog.getSpecificBinding()`. A multiple schema binding is created using `AvroCatalog.getSpecificMultiBinding()`.

Avro-specific classes provide type safety and ease of use. Properties are accessed through getter/setter methods with parameters and return values of the correct type, as defined in the schema.

Further, when using Avro-specific classes, the client application does not need to be aware of the Avro schema at runtime. The generated class has an internal reference to its associated schema, and this schema is used by the binding, so the application does not need to explicitly supply a schema. Schemas are supplied at build time, when the source code for the specific class is generated.

The disadvantage to using specific bindings is that the set of specific classes is fixed at build time. Applications that wish to treat values generically, or dynamically based on the schema, should use a generic or JSON binding instead.

Generating Specific Avro Classes

When you use a specific binding, you provide your schema to your application using a generated class. You can do this using the `org.apache.avro.compiler.specific.SchemaTask` tool.

An example Ant file that uses `SchemaTask` can be found in `KVHOME/examples/avro/generate-specific.xml`. It automatically downloads all library dependencies required to generate the Avro class, and then creates generated classes for all schema found in the local directory. The example assumes your schema is contained in flat text files that use the `avsc` suffix.

For example, suppose you had a schema defined in a file named `my-schema.avsc`. Then to generate an Avro-specific class, place `generate-specific.xml` and `my-schema.avsc` in the same directory and run:

```
ant -f generate-specific.xml
```

After Ant downloads all the necessary library dependencies, it will create whatever generated classes are required by the contents of `my-schema.avsc`. The generated classes are named after the `name` field in the Avro schema. So if you had the (trivial) schema:

```
{
    "type": "record",
    "name": "MyInfoString",
    "namespace": "avro",
    "fields": [
        {"name": "ID", "type": "int"}
    ]
}
```

the generated class name would be `MyInfoString.java`. It contains getter and setter methods that you can use to access the fields in the schema. For example, a generated class for the above schema would contain the fields `MyInfoString.setID()` and `MyInfoString.getID()`.

Using Avro-specific Bindings

To use a schema encapsulated in a generated Avro-specific class, you first provide the schema to the store as described in Managing Avro Schema in the Store (page 55).

After that, you access the schema using the generated class. To do this, you need handles to the store, an Avro catalog, and a specific binding:

```
package avro;

import oracle.kv.KVStore;
import oracle.kv.ValueVersion;
import oracle.kv.avro.AvroCatalog;
import oracle.kv.avro.SpecificAvroBinding;
```

```
...


private final KVStore store;
private final AvroCatalog catalog;
private final SpecificAvroBinding<MyInfoString> binding;

...

/* store configuration and open omitted for brevity */

...

catalog = store.getAvroCatalog();
binding = catalog.getSpecificBinding(MyInfoString.class);
```

Having done that, you can use the binding to serialize and deserialize value objects that conform to the generated class' schema. For example, suppose you performed a store `get()` and retrieved a `ValueVersion`. Then you can access its information in the following way:

```
final MyInfoString member;
final int ID;

member = binding.toObject(valueVersion.getValue());

System.out.println("ID: " + member.getID());
```

To serialize the object for placement in the store, you use the generated class' setter method, as well as the binding's `toValue()` method:

```
member.setID(2045);
/* key creation omitted */
store.put(key, binding.toValue(member));
```

Using Multiple Avro-specific Bindings

If your client code needs to use multiple specific bindings, then you use `AvroCatalog.getSpecificMultiBinding()` to support the multiple schemas. For example, suppose your client code required the following two schemas:

```
{
 "type": "record",
 "namespace": "avro",
 "name": "PersonInfo",
 "fields": [
   { "name": "first", "type": "string" },
   { "name": "last", "type": "string" },
   { "name": "age", "type": "int" }
 ]
}
```

```
{
 "type": "record",
 "namespace": "avro",
 "name": "AnimalInfo",
 "fields": [
   { "name": "species", "type": "string"},
   { "name": "name", "type": "string"},
   { "name": "age", "type": "int"}
 ]
}
```

Then put `Avro.PersonInfo` in a file (call it `PersonSchema.avsc`) and `Avro.AnimalInfo` in a second file (`AnimalSchema.avsc`). Add these schemas to your store using the command line interface.

Next, generate your specific schema classes using:

```
ant -f generate-specific.xml
```

as discussed in Generating Specific Avro Classes (page 69). This results in two generated classes: `PersonInfo.java` and `AnimalInfo.java`. To use these classes in your client code, you can use a single binding, which you create using `AvroCatalog.getSpecificMultiBinding()`. Note that to do this, you must use `org.apache.avro.specific.SpecificRecord`:

```
package avro;

import java.util.Arrays;

import oracle.kv.Key;
import oracle.kv.ValueVersion;
import oracle.kv.avro.AvroCatalog;
import oracle.kv.avro.SpecificAvroBinding;

import org.apache.avro.specific.SpecificRecord;

...

private final KVStore store;
private final AvroCatalog catalog;
private final SpecificAvroBinding<SpecificRecord> binding;

/*
 * Store creation is skipped for the sake of brevity.
 */

...

catalog = store.getAvroCatalog();
binding = catalog.getSpecificMultiBinding();
```

You can then create `PersonInfo` and `AnimalInfo` objects, serialize them, and write them to the store, like this:

```
final Key key1 = Key.createKey(Arrays.asList("person", "11"));
final Key key2 = Key.createKey(Arrays.asList("animal", "11"));

final PersonInfo pi = new PersonInfo();
pi.setFirst("Jack");
pi.setLast("Smith");
pi.setAge(38);
store.put(key1, binding.toValue(pi));

final AnimalInfo ai = new AnimalInfo();
ai.setSpecies("Dog");
ai.setName("Bowzer");
ai.setAge(6);
store.put(key2, binding.toValue(ai));
```

Retrieval of the stored objects is performed with a normal store `get()`. However, to deserialize the retrieved objects you must identify the object's type. You can do this using either the Java `instanceof` operator, or by examining the schema's name, as follows:

```
final ValueVersion vv = store.get(someKey);
if (vv != null) {
    SpecificRecord sr = binding.toObject(vv.getValue());
    if (sr.getSchema().getFullName().equals("avro.PersonInfo")) {
        PersonInfo o = (PersonInfo) sr;
        /*
         * The object is now deserialized. You can access the object's
         * data using the specific class' getXXX() methods. For example,
         * o.getFirst().
         */
    } else {
        AnimalInfo o = (AnimalInfo) sr;
        /*
         * The object is now deserialized. You can access the object's
         * data using the specific class' getXXX() methods. For example,
         * o.getSpecies().
         */
    }
```

JSON Bindings

The `JsonAvroBinding` interface represents values as instances of `JsonRecord`. A single schema binding is created using `AvroCatalog.getJsonBinding()`. A multiple schema binding is created using `AvroCatalog.getJsonMultiBinding()`.

The most important reason to use a JSON binding is for interoperability with other components or external systems that use JSON objects. This is because JSON bindings expose JSON objects, which can be managed with the Jackson API; a popular API used to manage JSON records.

Like generic bindings, a JSON binding treats values generically. And, like generic bindings, the schemas used in the application need not be fixed at build time if the schemas are treated dynamically.

However, unlike `GenericRecord`, certain Avro data types are not represented conveniently using the JSON syntax:

- Avro `int` and `long` are both represented as a JSON `integer`.

  Note

  To avoid type conversion errors, it is safest to always define integer fields in the Avro schema using type `long` rather than `int` when using this binding.

- Avro `float` and `double` are both represented as a JSON `number`.

  Note

  Because Jackson represents floating point numbers as a Java `Double`, you must define floating point fields in the Avro schema using type `double` rather than `float` when using this binding.

- Avro `record` and `map` are represented as a JSON `object`.
- Avro `bytes` and `fixed` are both represented as a JSON `string`, using Unicode escape syntax.

  Note

  Characters greater than 0xFF are invalid because they cannot be translated to bytes without loss of information.

- An Avro `enum` is represented as a JSON `string`.
- Avro unions have a special JSON representation. .

For this reason, applications that use the JSON binding should limit the data types used in their Avro schemas, and should treat the above data types carefully.

Like `GenericRecord`, a JSON object does not provide type safety. It can be error prone, because fields are accessed by string name and so data types cannot be checked at compile time.

Using Avro JSON Bindings

To use schema encapsulated in a generated Avro-specific class, you first provide the schema to the store as described in Managing Avro Schema in the Store (page 55).

In your Oracle NoSQL Database client code, you must make the schema available to the code. Do this by reading the schema directly from the file where you created it. For example, suppose you had the following schema defined in a file named `my-schema.avsc`:

```
{
    "type": "record",
    "name": "MyInfo",
    "namespace": "avro",
    "fields": [
        {"name": "ID", "type": "int"}
    ]
}
```

Then to read that schema into your client code:

```
package avro;

import java.io.File;
import org.apache.avro.Schema;
import org.codehaus.jackson.JsonNode;
import org.codehaus.jackson.node.JsonNodeFactory;
import org.codehaus.jackson.node.ObjectNode;

import oracle.kv.avro.AvroCatalog;
import oracle.kv.avro.JsonAvroBinding;
import oracle.kv.avro.JsonRecord;

...

final Schema.Parser parser = new Schema.Parser();
parser.parse(new File("my-schema.avsc"));
final Schema myInfoSchema =
    parser.getTypes().get("avro.MyInfo");
```

Before you can begin serializing and deserializing values that use the Avro data format, you must create a JSON binding and then create an Avro record for that binding.

```
/**
 * For the sake of brevity, we skip the necessary steps of
 * declaring and opening the store handle.
 */
final AvroCatalog catalog = store.getAvroCatalog();
final JsonAvroBinding binding =
    catalog.getJsonBinding(myInfoSchema);
```

Once you have the binding, you need a way for your application to represent the fields in the schema, so that they can be read and written. You do this by creating a *JSON record* and from there an *Object Node*, which is a data structure that allows you to read and/or write the fields in the schema.

For example, assume we performed a store read, and now we want to examine the information stored with the Oracle NoSQL Database record.

```
/**
  * Assume a store read was performed here, and resulted in a
  * ValueVersion instance called 'vv'. Then, to deserialize
```

```
  * the value in the returned record:
  */
final JsonRecord record = binding.toObject(vv.getValue());
final ObjectNode member = (ObjectNode) record.getJsonNode();
/* Retrieve the contents of the ID field. Because we are
 * using a generic binding, we must type cast the retrieved
 * value.
 */
final int ID = member.get("ID").getIntValue();
```

If we want to write to a field (that is, we want to serialize some data), we use the member's `put()` method. As an example, suppose we wanted to create a brand new object to be written to the store. Then:

```
final ObjectNode member2 = JsonNodeFactory.instance.objectNode();
member2.put("ID", 100011);
final JsonRecord record2 = new JsonRecord(member2, myInfoSchema);
...
/**
  * Assume the creation of a store key here.
 */
...
store.put(key, binding.toValue(record2));
```

Using a JSON Binding with a JSON Record

The most common usage of the JSON binding is to store information that is already marked up with JSON. A typical example would be data collected from a web server in JSON format that you then want to place in the store.

Because Oracle NoSQL Database's Avro implementation relies on the Jackson API, which is a common API used to parse JSON records, everything you need is already in place to put a JSON record into the store.

Suppose you had a JSON record that looked like this:

```
{
    "name": {
        "first": "Percival",
        "last":  "Lowell"
    },
    "age": 156,
    "address": {
        "street": "Mars Hill Rd",
        "city": "Flagstaff",
        "state": "AZ",
        "zip": 86001
    }
}
```

To support records of this type, you need a schema definition:

```
{
    "type": "record",
    "name": "MemberInfo",
    "namespace": "avro",
    "fields": [
        {"name": "name", "type": {
            "type": "record",
            "name": "FullName",
            "fields": [
                {"name": "first", "type": "string"},
                {"name": "last", "type": "string"}
            ]
        }},
        {"name": "age", "type": "int"},
        {"name": "address", "type": {
            "type": "record",
            "name": "Address",
            "fields": [
                {"name": "street", "type": "string"},
                {"name": "city", "type": "string"},
                {"name": "state", "type": "string"},
                {"name": "zip", "type": "int"}
            ]
        }}
    ]
}
```

Assuming that you added that schema to a file called `MemberInfo.avsc`, you add it to the store using the CLI (using the `ddl add-schema` command), and then create your JSON binding in the same way as shown in the previous example.

(Note that one additional class is required for this example: `ObjectMapper`. In addition, we use `java.io.BufferedReader` and `java.io.FileReader` to support reading the JSON record from disk, which is something that you probably would not do in production client code.)

```
package avro;

import java.io.BufferedReader;
import java.io.File;
import org.apache.avro.Schema;
import org.codehaus.jackson.JsonNode;
import org.codehaus.jackson.map.ObjectMapper;
import org.codehaus.jackson.node.JsonNodeFactory;
import org.codehaus.jackson.node.ObjectNode;

import oracle.kv.avro.AvroCatalog;
import oracle.kv.avro.JsonAvroBinding;
import oracle.kv.avro.JsonRecord;

import oracle.kv.Key;
```

```
import oracle.kv.Value;

...

final Schema.Parser parser = new Schema.Parser();
parser.parse(new File("MemberInfo.avsc"));
final Schema memberInfoSchema =
    parser.getTypes().get("avro.MemberInfo");

...

/**
 * For the sake of brevity, we skip the necessary steps of
 * declaring and opening the store handle.
 */

...

final AvroCatalog catalog = store.getAvroCatalog();
final JsonAvroBinding binding =
    catalog.getJsonBinding(memberInfoSchema);
```

The next step is to read the JSON record into your application. Typically, this would occur over the network, but for simplicity, we show the record being read from a file on disk:

```
try {
    final String jsonText =
        readFile(new File("MemberInfoRecord.json"));
```

Next, we use the Jackson API to parse the text so as to create a JSON object:

```
    final ObjectMapper jsonMapper = new ObjectMapper();

    final JsonNode jsonObject = jsonMapper.readTree(jsonText);
```

Finally, we write the JSON object to a JSON record, which we then use to create an Oracle NoSQL Database `Value`.

```
    final JsonRecord jsonRecord =
        new JsonRecord(jsonObject, memberInfoSchema);

    final Value value = binding.toValue(jsonRecord);
```

The final step is to the write the object to the store in the usual way:

```
    store.put(Key.fromString("/any/old/key"), value);
} catch (IOException io) {
    io.printStackTrace();
}
```

To be complete, the code for the `readFile()` method used in this example is:

```
private String readFile(File f) throws IOException {
```

```
    final BufferedReader r = new BufferedReader(new FileReader(f));
    final StringBuilder buf = new StringBuilder(1000);
    String line;
    while ((line = r.readLine()) != null) {
        buf.append(line);
        buf.append("\n");
    }
    return buf.toString();
}
```

Chapter 8. Key Ranges and Depth for Multi-Key Operations

When performing multi-key operations (for example, `KVStore.multiGet()`, `KVStore.multiDelete()`, `KVStore.storeIterator()`), you can limit the records that are operated upon by specifying key ranges and depth. Key ranges allow you to identify a subset of keys to use out of a matching set. Depth allows you to specify how many children you want the multi-key operation to use.

Specifying Subranges

When performing multi-key operations in the store, you can specify a range of records to operate upon. You do this using the `KeyRange` class. This class defines a range of `String` values for the key components immediately following a key that is used in a multiple get operation.

For example, suppose you were using the following keys:

```
/Smith/Bob/-/birthdate
/Smith/Bob/-/phonenumber
/Smith/Bob/-/image
/Smith/Bob/-/userID
/Smith/Patricia/-/birthdate
/Smith/Patricia/-/phonenumber
/Smith/Patricia/-/image
/Smith/Patricia/-/userID
/Smith/Richard/-/birthdate
/Smith/Richard/-/phonenumber
/Smith/Richard/-/image
/Smith/Richard/-/userID
/Wong/Bill/-/birthdate
/Wong/Bill/-/phonenumber
/Wong/Bill/-/image
/Wong/Bill/-/userID
```

Given this, you could perform operations for all the records related to users Bob Smith and Patricia Smith by constructing a `KeyRange`. When you do this, you must identify the key components that define the upper and lower bounds of the range. You must also identify if the key components that you provide are inclusive or exclusive.

In this case, we will define the start of the key range using the string "Bob" and the end of the key range to be "Patricia". Both ends of the key range will be inclusive.

```
package kvstore.basicExample;

import oracle.kv.KeyRange;

...
```

```
KeyRange kr = new KeyRange("Bob", true, "Patricia", true);
```

You then use the `KeyRange` instance when you perform your multi-key operation. For example, suppose you were retrieving records from your store using `KVStore.storeIterator()`:

```
package kvstore.basicExample;

...

import oracle.kv.Direction;
import oracle.kv.Key;
import oracle.kv.Value;
import oracle.kv.KeyRange;
import oracle.kv.KeyValueVersion;
import oracle.kv.RequestTimeoutException;

import java.util.ArrayList;
import java.util.Iterator;

...

ArrayList<String> majorComponents = new ArrayList<String>();

...

// Define the major and minor path components for the key
majorComponents.add("Smith");

// Create the retrieval key
Key myKey = Key.createKey(majorComponents);

KeyRange kr = new KeyRange("Bob", true, "Patricia", true);

// Now retrieve the records. Note that we do not show the creation of
// the kvstore handle here.


try {
    Iterator<KeyValueVersion> i =
        kvstore.storeIterator(Direction.FORWARD, 0,
                              myKey, kr, null);
    while (i.hasNext()) {
        Value v = i.next().getValue();
        // Do some work with the Value here
    }
} catch (RequestTimeoutException re) {
    // The operation was not completed within the
    // timeout value
}
```

Specifying Depth

When performing multi-key operations in the store, you can specify a depth of records to operate upon. That is, you can indicate whether you want to operate upon:

- The specified key and all its children.
- The specified key and its most immediate children.
- Only the immediate children of the specified key. (The specified key is omitted.)
- All of the children of the specified key. (The specified key is omitted.)

By default, multi-key operations operate upon the specified key and all of its children. To limit the operation to something else, such as just the key's immediate children, specify `Depth.CHILDREN_ONLY` to the operation's `Depth` parameter.

For example, suppose you were using the following keys:

```
/Products/Hats/-/baseball
/Products/Hats/-/baseball/longbill
/Products/Hats/-/baseball/longbill/blue
/Products/Hats/-/baseball/longbill/red
/Products/Hats/-/baseball/shortbill
/Products/Hats/-/baseball/shortbill/blue
/Products/Hats/-/baseball/shortbill/red
/Products/Hats/-/western
/Products/Hats/-/western/felt
/Products/Hats/-/western/felt/black
/Products/Hats/-/western/felt/gray
/Products/Hats/-/western/leather
/Products/Hats/-/western/leather/black
/Products/Hats/-/western/leather/gray
```

Further, suppose you wanted to retrieve just these records:

```
/Products/Hats/-/baseball
/Products/Hats/-/western
```

Then you could do this using `KVStore.multiGet()` with the appropriate `Depth` argument.

```
package kvstore.basicExample;

...

import oracle.kv.Depth;
import oracle.kv.Key;
import oracle.kv.RequestTimeoutException;
import oracle.kv.Value;
import oracle.kv.ValueVersion;

import java.util.ArrayList;
```

```
import java.util.Iterator;
import java.util.SortedMap;

...

ArrayList<String> majorComponents = new ArrayList<String>();

...

// Define the major and minor path components for the key
majorComponents.add("Product");
majorComponents.add("Hats");

// Create the retrieval key
Key myKey = Key.createKey(majorComponents);

// Now retrieve the records. Note that we do not show the creation of
// the kvstore handle here.

try {
    SortedMap<Key, ValueVersion> myRecords = null;

    myRecords = kvstore.multiGet(myKey,
                                 null,
                                 Depth.CHILDREN_ONLY);
} catch (RequestTimeoutException re) {
    // The operation was not completed within the
    // timeout value
}
```

Chapter 9. Using Versions

When a record (that is, a key-value pair) is initially inserted in the store, and each time it is updated, it is assigned a unique version token. The version is always returned by the method that wrote to the store (for example, `KVStore.put()`). The version information is also returned by methods that retrieve records from the store.

There are two reasons why versions might be important.

1. When an update or delete is to be performed, it may be important to only perform the operation if the record's value has not changed. This is particularly interesting in an application where there can be multiple threads or processes simultaneously operating on the record. In this case, read the record, examining its version when you do so. You can then perform a put operation, but only allow the put to proceed if the version has not changed (this is often referred to as a *Compare and Set* (CAS) or *Read, Modify, Write* (RMW) operation). You use `KVStore.putIfVersion()` or `KVStore.deleteIfVersion()` to guarantee this.

2. When a client reads data that was previously written, it may be important to ensure that the Oracle NoSQL Database node servicing the read operation has been updated with the information previously written. This can be accomplished by passing the version of the previously written data as a consistency parameter to the read operation. For more information on using consistency, see Consistency Guarantees (page 85).

Versions are managed using the `Version` class. In some situations, it is returned as part of another encapsulating class, such as `KeyValueVersion` or `ValueVersion`.

The following code fragment retrieves a record, and then stores that record only if the version has not changed:

```
package kvstore.basicExample;

...

import oracle.kv.Key;
import oracle.kv.Value;
import oracle.kv.ValueVersion;
import java.util.ArrayList;

...

ArrayList<String> majorComponents = new ArrayList<String>();
ArrayList<String> minorComponents = new ArrayList<String>();

...

// Define the major and minor path components for the key
majorComponents.add("Smith");
majorComponents.add("Bob");
```

```
minorComponents.add("phonenumber");

// Create the key
Key myKey = Key.createKey(majorComponents, minorComponents);

// Now retrieve the record. Note that we do not show the creation of
// the kvstore handle here.

ValueVersion vv = kvstore.get(myKey);
Value value = vv.getValue();
Version version = vv.getVersion();

...

////////////////////////////////////////////////////////
////////////// Do work on the value here /////////////////
////////////////////////////////////////////////////////

...

// Put if the version is correct. Notice that here we examine
// the return code. If it is null, that means that the put was
// unsuccessful, probably because the record was changed elsewhere.
// In this case, you could retry the entire get/putIfVersion
// operation.
Version newVersion = kvstore.putIfVersion(myKey, value, version);
if (newVersion == null) {
    // Unsuccessful. Someone else probably modified the record.
}
```

Chapter 10. Consistency Guarantees

The KV store is built from some number of computers (generically referred to as *nodes*) that are working together using a network. All data in your store is first written to a master node. The master node then copies that data to other nodes in the store. Nodes which are not master nodes are referred to as *replicas*.

Because of the relatively slow performance of distributed systems, there can be a possibility that, at any given moment, a write operation that was performed on the master node will not yet have been performed on some other node in the store.

Consistency, then, is the policy describing whether it is possible for a record on Node A to be different from the same record on Node B.

When there is a high likelihood that a record stored on one node is identical to the same record stored on another node, we say that we have a *high consistency guarantee*. Likewise, a *low consistency guarantee* means that there is a good possibility that a record on one node differs in some way from the same record stored on another node.

You can control how high you want your consistency guarantee to be. Note that the trade-off in setting a high consistency guarantee is that your store's read performance might not be as high as if you use a low consistency guarantee.

There are several different forms of consistency guarantees that you can use. They are described in the following sections.

Note that by default, Oracle NoSQL Database uses the lowest possible consistency possible.

Specifying Consistency Policies

To specify a consistency policy, you use one of the static instances of the `Consistency` class, or one of its nested classes.

Once you have selected a consistency policy, you can put it to use in one of two ways. First, you can use it to define a default consistency policy using the `KVStoreConfig.setConsistency()` method. Specifying a consistency policy in this way means that all store operations will use that policy, unless they are overridden on an operation by operation basis.

The second way to use a consistency policy is to override the default policy using the `Consistency` parameter on the `KVStore` method that you are using to perform the store read operation.

The following example shows how to set a default consistency policy for the store. We will show the per-operation method of specifying consistency policies in the following sections.

```
package kvstore.basicExample;

import oracle.kv.Consistency;
import oracle.kv.KVStore;
import oracle.kv.KVStoreConfig;
import oracle.kv.KVStoreFactory;
```

```
...

KVStoreConfig kconfig = new KVStoreConfig("exampleStore",
    "node1.example.org:5088, node2.example.org:4129");

kconfig.setConsistency(Consistency.NONE_REQUIRED);

KVStore kvstore = KVStoreFactory.getStore(kconfig);
```

Using Simple Consistency

You can use static instances of the `Consistency` base class to specify certain rigid consistency guarantees. There are three such instances that you can use:

1. `Consistency.ABSOLUTE`

   Requires that the operation be serviced at the master node. In this way, the record(s) will always be consistent with the master.

   This is the strongest possible consistency guarantee that you can require, but it comes at the cost of servicing all read and write requests at the master node. If you direct all your traffic to the master node (which is just one machine for each partition), then you will not be distributing your read operations across your replicas. You also will slow your write operations because your master will be busy servicing read requests. For this reason, you should use this consistency guarantee sparingly.

2. `Consistency.NONE_REQUIRED`

   Allows the store operation to proceed regardless of the state of the replica relative to the master. This is the most relaxed consistency guarantee that you can require. It allows for the maximum possible store performance, but at the high possibility that your application will be operating on stale or out-of-date information.

3. `Consistency.NONE_REQUIRED_NO_MASTER`

   Requires read operations to be serviced on a replica; never the Master. When this policy is used, the read operation will not be performed if the only node available is the Master.

   Where possible, this consistency policy should be avoided in favor of the secondary zones feature.

For example, suppose you are performing a critical read operation that you know must absolutely have the most up-to-date data. Then do this:

```
package kvstore.basicExample;

...

import oracle.kv.Consistency;
import oracle.kv.ConsistencyException;
import oracle.kv.Key;
```

```
import oracle.kv.Value;
import oracle.kv.ValueVersion;

import java.util.ArrayList;

...

ArrayList<String> majorComponents = new ArrayList<String>();

...

// Define the major path components for the key
majorComponents.add("Smith");
majorComponents.add("Bob");

// Create the key
Key myKey = Key.createKey(majorComponents);

// Now retrieve the record. Note that we do not show the creation of
// the kvstore handle here.

try {
    ValueVersion vv = kvstore.get(myKey,
                          Consistency.ABSOLUTE,
                          0,     // Timeout parameter.
                                 // 0 means use the default.
                          null); // Timeout units. Null because
                                 // the Timeout is 0.

    Value v = vv.getValue();
    /*
     * From here, deserialize using your Avro binding.
     */
} catch (ConsistencyException ce) {
    // The consistency guarantee was not met
}
```

Using Time-Based Consistency

A time-based consistency policy describes the amount of time that a replica node is allowed to lag behind the master node. If the replica's data is more than the specified amount of time out-of-date relative to the master, then a `ConsistencyException` is thrown. In that event, you can either abandon the operation, retry it immediately, or pause and then retry it.

In order for this type of a consistency policy to be effective, the clocks on all the nodes in the store must be synchronized using a protocol such as NTP.

In order to specify a time-based consistency policy, you use the `Consistency.Time` class. The constructor for this class requires the following information:

- `permissibleLag`

  A `long` that describes the number of `TimeUnits` the replica is allowed to lag behind the master.

- `permissibleLagUnits`

  A `TimeUnit` that identifies the units used by `permissibleLag`. For example: `TimeUnit.MILLISECONDS`.

- `timeout`

  A `long` that describes how long the replica is permitted to wait in an attempt to meet the permissible lag limit. That is, if the replica cannot immediately meet the permissible lag requirement, then it will wait this amount of time to see if it is updated with the required data from the master. If the replica cannot meet the permissible lag requirement within the timeout period, a `ConsistencyException` is thrown.

- `timeoutUnit`

  A `TimeUnit` that identifies the units used by `timeout`. For example: `TimeUnit.SECONDS`.

The following sets a default time-based consistency policy of 2 seconds. The timeout is 4 seconds.

```
package kvstore.basicExample;

import oracle.kv.Consistency;
import oracle.kv.KVStore;
import oracle.kv.KVStoreConfig;
import oracle.kv.KVStoreFactory;

import java.util.concurrent.TimeUnit;

...

KVStoreConfig kconfig = new KVStoreConfig("exampleStore",
    "node1.example.org:5088, node2.example.org:4129");

Consistency.Time cpolicy =
        new Consistency.Time(2, TimeUnit.SECONDS,
                             4, TimeUnit.SECONDS);
kconfig.setConsistency(cpolicy);

KVStore kvstore = KVStoreFactory.getStore(kconfig);
```

Using Version-Based Consistency

Version-based consistency is used on a per-operation basis. It ensures that a read performed on a replica is at least as current as some previous write performed on the master.

An example of how this might be used is a web application that collects some information from a customer (such as her name). It then customizes all subsequent pages presented to the customer with her name. The storage of the customer's name is a write operation that can only be performed by the master node, while subsequent page creation is performed as a read-only operation that can occur at any node in the store.

Use of this consistency policy might require that version information be transferred between processes in your application.

To create a version-based consistency policy, use the `Consistency.Version` class. When you do this, you must provide the following information:

- `version`

  The `Version` that the read must match.

- `timeout`

  A `long` that describes how long the replica is permitted to wait in an attempt to meet the version requirement. That is, if the replica cannot immediately meet the version requirement, then it will wait this amount of time to see if it is updated with the required data from the master. If the replica cannot meet the requirement within the timeout period, a `ConsistencyException` is thrown.

- `timeoutUnit`

  A `TimeUnit` that identifies the units used by `timeout`. For example: `TimeUnit.SECONDS`.

For example, the following code performs a store write, collects the version information, then uses it to construct a version-based consistency policy. In this example, assume we are using a generic Avro binding to store some person information.

```
package kvstore.basicExample;

...

import oracle.kv.Key;
import oracle.kv.Value;
import oracle.kv.Version;
import java.util.ArrayList;

import org.apache.avro.Schema;
import oracle.kv.avro.GenericAvroBinding;
import oracle.kv.avro.GenericRecord;

...

ArrayList<String> majorComponents = new ArrayList<String>();

...
```

```
// Define the major path components for the key
majorComponents.add("Smith");
majorComponents.add("Bob");

// Create the key
Key myKey = Key.createKey(majorComponents);

...

// Binding and schema creation omitted

...

final GenericRecord person = new GenericData.Record(personSchema);
person.put("ID", 100011);
person.put("FamiliarName", "Bob");
person.put("Surname", "Smith");
person.put("PrimaryPhone", "408 555 5555");

Value myValue = binding.toValue(person);

// Now put the record. Note that we do not show the creation of the
// kvstore handle here.

Version matchVersion = kvstore.put(myKey, myValue);
```

At some other point in this application's code, or perhaps in another application entirely, we use the `matchVersion` captured above to create a version-based consistency policy.

```
package kvstore.basicExample;

...

import oracle.kv.Consistency;
import oracle.kv.ConsistencyException;
import oracle.kv.Key;
import oracle.kv.Value;
import oracle.kv.ValueVersion;
import oracle.kv.Version;

import org.apache.avro.Schema;
import oracle.kv.avro.GenericAvroBinding;
import oracle.kv.avro.GenericRecord;

import java.util.ArrayList;
import java.util.concurrent.TimeUnit;

...

ArrayList<String> majorComponents = new ArrayList<String>();
```

```
...

// Define the major path components for the key
majorComponents.add("Smith");
majorComponents.add("Bob");

// Create the key
Key myKey = Key.createKey(majorComponents);

// Create the consistency policy using the
// Version object we captured, above.
Consistency.Version versionConsistency =
        new Consistency.Version(matchVersion,
                                200,
                                TimeUnit.NANOSECONDS);

// Now retrieve the record. Note that we do not show the creation of
// the kvstore handle here.

try {
    ValueVersion vv = kvstore.get(myKey,
                          versionConsistency,
                          0,     // Timeout parameter.
                                 // 0 means use the default.
                          null); // Timeout units. Null because
                                 // the Timeout is 0.

    // Deserialize with our generic avro binding
    // (creation of this binding is not shown).

    final GenericRecord member = binding.toObject(vv.getValue());

    // Do work with the generic record here.
} catch (ConsistencyException ce) {
    // The consistency guarantee was not met
}
```

Chapter 11. Durability Guarantees

Writes are performed in the Oracle NoSQL Database store by performing the write operation (be it a creation, update, or delete operation) on a master node. As a part of performing the write operation, the master node will usually make sure that the operation has made it to stable storage before considering the operation complete.

The master node will also transmit the write operation to the replica nodes in its shard. It is possible to ask the master node to wait for acknowledgments from its replicas before considering the operation complete.

Note

If your store is configured such that secondary zones are in use, then write acknowledgements are never required for the replicas in the secondary zones. That is, write acknowledgements are only returned by replicas in primary zones. See the *Oracle NoSQL Database Administrator's Guide* for more information on zones.

The replicas, in turn, will not acknowledge the write operation until they have applied the operation to their own database.

A *durability guarantee,* then, is a policy which describes how strongly persistent your data is in the event of some kind of catastrophic failure within the store. (Examples of a catastrophic failure are power outages, disk crashes, physical memory corruption, or even fatal application programming errors.)

A high durability guarantee means that there is a very high probability that the write operation will be retained in the event of a catastrophic failure. A low durability guarantee means that the write is very unlikely to be retained in the event of a catastrophic failure.

The higher your durability guarantee, the slower your write-throughput will be in the store. This is because a high durability guarantee requires a great deal of disk and network activity.

Usually you want some kind of a durability guarantee, although if you have highly transient data that changes from run-time to run-time, you might want the lowest possible durability guarantee for that data.

Durability guarantees include two types of information: acknowledgment guarantees and synchronization guarantees. These two types of guarantees are described in the next sections. We then show how to set a durability guarantee.

Note that by default, Oracle NoSQL Database uses a low durability guarantee.

Setting Acknowledgment-Based Durability Policies

Whenever a master node performs a write operation (create, update or delete), it must send that operation to its various replica nodes. The replica nodes then apply the write operation(s) to their local databases so that the replicas are consistent relative to the master node.

Upon successfully applying write operations to their local databases, replicas in primary zones send an *acknowledgment message* back to the master node. This message simply says that the write operation was received and successfully applied to the replica's local database. Replicas in secondary zones do not send these acknowledgement messages.

Note

The exception to this are replicas in secondary zones, which will never acknowledge write operations. See the *Oracle NoSQL Database Administrator's Guide* for more information on zones.

An acknowledgment-based durability policy describes whether the master node will wait for these acknowledgments before considering the write operation to have completed successfully. You can require the master node to wait for no acknowledgments, acknowledgments from a simple majority of replica nodes in primary zones, or acknowledgments from all replica nodes in primary zones.

The more acknowledgments the master requires, the slower its write performance will be. Waiting for acknowledgments means waiting for a write message to travel from the master to the replicas, then for the write operation to be performed at the replica (this may mean disk I/O), then for an acknowledgment message to travel from the replica back to the master. From a computer application's point of view, this can all take a long time.

When setting an acknowledgment-based durability policy, you can require acknowledgment from:

- All replicas. That is, all of the replica nodes in the shard that reside in a primary zone. Remember that your store has more than one shard, so the master node is not waiting for acknowledgments from every machine in the store.
- No replicas. In this case, the master returns with normal status from the write operation as soon as it has met its synchronization-based durability policy. These are described in the next section.
- A simple majority of replicas in primary zones. That is, if the shard has 5 replica nodes residing in primary zones, then the master will wait for acknowledgments from 3 nodes.

Setting Synchronization-Based Durability Policies

Whenever a node performs a write operation, the node must know whether it should wait for the data to be written to stable storage before successfully returning from the operation.

As a part of performing a write operation, the data modification is first made to an in-memory cache. It is then written to the filesystem's data buffers. And, finally, the contents of the data buffers are synchronized to stable storage (typically, a hard drive).

You can control how much of this process the master node will wait to complete before it returns from the write operation with a normal status. There are three different levels of synchronization durability that you can require:

- NO_SYNC

The data is written to the host's in-memory cache, but the master node does not wait for the data to be written to the file system's data buffers, or for the data to be physically transferred to stable storage. This is the fastest, but least durable, synchronization policy.

- WRITE_NO_SYNC

  The data is written to the in-memory cache, and then written to the file system's data buffers, but the data is not necessarily transferred to stable storage before the operation completes normally.

- SYNC

  The data is written to the in-memory cache, then transferred to the file system's data buffers, and then synchronized to stable storage before the write operation completes normally. This is the slowest, but most durable, synchronization policy.

Notice that in all cases, the data is eventually written to stable storage (assuming some failure does not occur to prevent it). The only question is, how much of this process will be completed before the write operation returns and your application can proceed to its next operation.

Setting Durability Guarantees

To set a durability guarantee, use the `Durability` class. When you do this, you must provide three pieces of information:

- The acknowledgment policy.
- A synchronization policy at the master node.
- A synchronization policy at the replica nodes.

The combination of policies that you use is driven by how sensitive your application might be to potential data loss, and by your write performance requirements.

For example, the fastest possible write performance can be achieved through a durability policy that requires:

- No acknowledgments.
- NO_SYNC at the master.
- NO_SYNC at the replicas.

However, this durability policy also leaves your data with the greatest risk of loss due to application or machine failure between the time the operation returns and the time when the data is written to stable storage.

On the other hand, if you want the highest possible durability guarantee, you can use:

- All replicas must acknowledge the write operation.
- SYNC at the master.

- SYNC at the replicas.

Of course, this also results in the slowest possible write performance.

Most commonly, durability policies attempt to strike a balance between write performance and data durability guarantees. For example:

- Simple majority of replicas must acknowledge the write.
- SYNC at the master.
- NO_SYNC at the replicas.

Note that you can set a default durability policy for your `KVStore` handle, but you can also override the policy on a per-operation basis for those situations where some of your data need not be as durable (or needs to be MORE durable) than the default.

For example, suppose you want an intermediate durability policy for most of your data, but sometimes you have transient or easily re-created data whose durability really is not very important. Then you would do something like this:

First, set the default durability policy for the `KVStore` handle:

```
package kvstore.basicExample;

import oracle.kv.Durability;
import oracle.kv.KVStore;
import oracle.kv.KVStoreConfig;
import oracle.kv.KVStoreFactory;

...

KVStoreConfig kconfig = new KVStoreConfig("exampleStore",
    "node1.example.org:5088, node2.example.org:4129");

Durability defaultDurability =
    new Durability(Durability.SyncPolicy.SYNC,    // Master sync
                   Durability.SyncPolicy.NO_SYNC, // Replica sync
                   Durability.ReplicaAckPolicy.SIMPLE_MAJORITY);
kconfig.setDurability(defaultDurability);

KVStore kvstore = KVStoreFactory.getStore(kconfig);
```

In another part of your code, for some unusual write operations, you might then want to relax the durability guarantee so as to speed up the write performance for those specific write operations:

```
package kvstore.basicExample;

...

import oracle.kv.Durability;
import oracle.kv.DurabilityException;
import oracle.kv.Key;
```

```
import oracle.kv.RequestTimeoutException;
import oracle.kv.Value;
import java.util.ArrayList;

import org.apache.avro.Schema;
import oracle.kv.avro.GenericAvroBinding;
import oracle.kv.avro.GenericRecord;

...

ArrayList<String> majorComponents = new ArrayList<String>();

...

// Define the major and minor path components for the key
majorComponents.add("Smith");
majorComponents.add("Bob");

// Create the key
Key myKey = Key.createKey(majorComponents);

...

// Binding and schema creation omitted

...

final GenericRecord person = new GenericData.Record(personSchema);
person.put("ID", 100011);
person.put("FamiliarName", "Bob");
person.put("Surname", "Smith");
person.put("PrimaryPhone", "408 555 5555");

Value myValue = binding.toValue(person);

// Create the special durability policy
Durability durability =
    new Durability(Durability.SyncPolicy.NO_SYNC, // Master sync
                   Durability.SyncPolicy.NO_SYNC, // Replica sync
                   Durability.ReplicaAckPolicy.NONE);

// Now put the record. Note that we do not show the creation of the
// kvstore handle here.
try {
    kvstore.put(myKey, myValue,
                null,       // ReturnValueVersion is null because
                            // we aren't using it.
                durability, // The per-operation durability
                0,          // Use the default request timeout
```

```
                null);      // Use the default timeunit value
} catch (DurabilityException de) {
    // The durability guarantee was not met
} catch (RequestTimeoutException re) {
    // The operation was not completed within the
    // timeout value
}
```

Chapter 12. Executing a Sequence of Operations

You can execute a sequence of write operations as a single atomic unit so long as all the records that you are operating upon share the same major path components. By *atomic unit*, we mean all of the operations will execute successfully, or none of them will.

Also, the sequence is performed in isolation. This means that if you have a thread running a particularly long sequence, then another thread cannot intrude on the data in use by the sequence. The second thread will not be able to see any of the modifications made by the long-running sequence until the sequence is complete. The second thread also will not be able to modify any of the data in use by the long-running sequence.

Be aware that sequences only support write operations. You can perform puts and deletes, but you cannot retrieve data when using sequences.

When using a sequence of operations:

- All of the keys in use by the sequence must share the same major path components.
- Operations are placed into a list, but the operations are not necessarily executed in the order that they appear in the list. Instead, they are executed in an internally defined sequence that prevents deadlocks.
- You cannot create two or more operations that operate on the same key. Doing so results in an exception, and the entire operation is aborted.

The rest of this chapter shows how to use `OperationFactory` and `KVStore.execute()` to create and run a sequence of operations.

Sequence Errors

If any operation within the sequence experiences an error, then the entire operation is aborted. In this case, your data is left in the same state it would have been in if the sequence had never been run at all — no matter how much of the sequence was run before the error occurred.

Fundamentally, there are two reasons why a sequence might abort:

1. An internal operation results in an exception that is considered a fault. For example, the operation throws a `DurabilityException`. Also, if there is an internal failure due to message delivery or a networking error.
2. An individual operation returns normally but is unsuccessful as defined by the particular operation. (For example, you attempt to delete a key that does not exist). If this occurs AND you specified `true` for the `abortIfUnsuccessful` parameter when the operation was created using OperationFactory, then an `OperationExecutionException` is thrown. This exception contains information about the failed operation.

Creating a Sequence

You create a sequence by using the `OperationFactory` class to create `Operation` class instances, each of which represents exactly one operation in the store. You obtain an instance of `OperationFactory` by using `KVStore.getOperationFactory()`.

For example, suppose you are using the following keys:

```
/Products/Hats/-/baseball
/Products/Hats/-/baseball/longbill
/Products/Hats/-/baseball/longbill/blue
/Products/Hats/-/baseball/longbill/red
/Products/Hats/-/baseball/shortbill
/Products/Hats/-/baseball/shortbill/blue
/Products/Hats/-/baseball/shortbill/red
/Products/Hats/-/western
/Products/Hats/-/western/felt
/Products/Hats/-/western/felt/black
/Products/Hats/-/western/felt/gray
/Products/Hats/-/western/leather
/Products/Hats/-/western/leather/black
/Products/Hats/-/western/leather/gray
```

And further suppose each of the following records has some information (such as a price refresh date) that you want to update in such a fashion as to make sure that the information is consistent for all of the records:

```
/Products/Hats/-/western
/Products/Hats/-/western/felt
/Products/Hats/-/western/leather
```

Then you can create a sequence in the following way:

```
package kvstore.basicExample;

...

import oracle.kv.Key;
import oracle.kv.Value;
import oracle.kv.Operation;
import oracle.kv.OperationFactory;
import java.util.ArrayList;

import org.apache.avro.Schema;
import oracle.kv.avro.GenericAvroBinding;
import oracle.kv.avro.GenericRecord;

...

// Get the operation factory. Note that we do not show the
// creation of the kvstore handle here.
```

```
OperationFactory of = kvstore.getOperationFactory();

// We need a List to hold the operations in the
// sequence.
ArrayList<Operation> opList = new ArrayList<Operation>();

...

ArrayList<String> majorComponents = new ArrayList<String>();
ArrayList<String> minorComponents1 = new ArrayList<String>();
ArrayList<String> minorComponents2 = new ArrayList<String>();
ArrayList<String> minorComponents3 = new ArrayList<String>();

...

// Define the major and minor path components for our keys
majorComponents.add("Products");
majorComponents.add("Hats");

minorComponents1.add("western");
minorComponents2.add("western");
minorComponents2.add("felt");
minorComponents3.add("western");
minorComponents3.add("leather");

// Create the three keys that we will need
Key key1 = Key.createKey(majorComponents, minorComponents1);
Key key2 = Key.createKey(majorComponents, minorComponents2);
Key key3 = Key.createKey(majorComponents, minorComponents3);

...

// Binding and schema creation omitted

...

final GenericRecord hat1 = new GenericData.Record(hatSchema);
hat1.put("randomHatData", "someRandomData");
final Value value1 = binding.toValue(hat1);

final GenericRecord hat2 = new GenericData.Record(hatSchema);
hat2.put("randomHatData", "someMoreRandomData");
final Value value2 = binding.toValue(hat2);

final GenericRecord hat3 = new GenericData.Record(hatSchema);
hat3.put("randomHatData", "someMoreRandomData");
final Value value3 = binding.toValue(hat3);

...
```

```
// Here we would perform whatever actions we need to create
// our record values. We won't show how the values get created,
// but assume it results in three Value objects: value1, value2,
// and value3.

...

// Now create our list of operations for the key pairs
// key1/value1, key2/value2, and key3/value3. In this
// trivial example we will put store all three records
// in a single atomic operation.


opList.add(of.createPut(key1, value1));
opList.add(of.createPut(key2, value2));
opList.add(of.createPut(key3, value3));
```

Note in the above example that we create three unique keys that differ only in their minor path components. If the major path components were different for any of the three keys, we could not successfully execute the sequence.

Executing a Sequence

To execute the sequence we created in the previous section, use the `KVStore.execute()` method:

```
package kvstore.basicExample;

...

import oracle.kv.DurabilityException;
import oracle.kv.FaultException;
import oracle.kv.OperationExecutionException;
import oracle.kv.RequestTimeoutException;

...

try {
    kvstore.execute(opList);
} catch (OperationExecutionException oee) {
    // Some error occurred that prevented the sequence
    // from executing successfully. Use
    // oee.getFailedOperationIndex() to determine which
    // operation failed. Use oee.getFailedOperationResult()
    // to obtain an OperationResult object, which you can
    // use to troubleshoot the cause of the execution
    // exception.
} catch (DurabilityException de) {
    // The durability guarantee could not be met.
```

```
} catch (IllegalArgumentException iae) {
    // An operation in the list was null or empty.

    // Or at least one operation operates on a key
    // with a major path component that is different
    // than the others.

    // Or more than one operation uses the same key.
} catch (RequestTimeoutException rte) {
    // The operation was not completed inside of the
    // default request timeout limit.
} catch (FaultException fe) {
    // A generic error occurred
}
```

Note that if any of the above exceptions are thrown, then the entire sequence is aborted, and your data will be in the state it would have been in if you had never executed the sequence at all.

A richer form of `KVStore.execute()` is available. It allows you to specify:

- The list of operations.
- The durability guarantee that you want to use for this sequence. If you want to use the default durability guarantee, pass `null` for this parameter.
- A timeout value that identifies the upper bound on the time interval allowed for processing the entire sequence. If you provide `0`, the default request timeout value is used.
- A `TimeUnit` that identifies the units used by the timeout value. For example: `TimeUnit.MILLISECONDS`.

For an example of using `WriteOptions`, see Durability Guarantees (page 92).

Chapter 13. Index Views

Index views are a design pattern you use to create auxiliary records that are reflective of information contained in your primary records. There are many ways you can create index views. This document describes two of them.

> **Note**
>
> This article assumes that you are using Oracle NoSQL Database's Key/Value API and have read and understood the *Oracle NoSQL Database Getting Started with the Key/Value API* guide. If you have not read that manual, you should do so before reading this document.
>
> Users of the Tables API have a built-in indexing mechanism available, and so this article is not meant for them.

As described in Reading Records (page 30) records are generally retrieved from the store using their key major and minor paths. You can either retrieve a single record using its key, or you can retrieve multiple records using part of a major path and then iterate over the result.

For example, suppose your store contains records related to users. The key might contain user organization information and other identifying information such as a user ID. Each record's data, however, would likely contain additional details about people such as names, addresses, phone numbers, and so forth. While your application may frequently want to query a person by user ID (that is, by the information stored as a part of the key path), it may also on occasion want to locate people by, say, their name.

Rather than iterating through all of the records in your store, examining each in turn for a given person's name, you can instead create application-managed index views. There are multiple ways to implement index views, but in general they are simply key/value pairs where the key relates to some information within your primary record, and the value identifies the primary record where that information can be found.

That is, if you had a record which contained the name `Peter`, then the key for its index view would contain `Peter` and the value would contain the major and minor key paths to that record.

Using Traditional Key/Data Pairs

This method of creating index views is, intuitively, the way many developers familiar with key/value stores will think to implement views.

For a different approach to building index views, see Using Key-Only Records (page 105).

When you use traditional key/data pair records to build index views, you create records where:

- The record's key path is some information in your primary data records that you want to quickly query.

- The record's data is the key path to a record which has the information contained in the key path.

For example, suppose you had records that used the following schema:

```
{
    "type": "record",
    "name": "PrimaryDBValue",
    "namespace": "oracle.kv.indexView",
    "fields": [
        {"name": "name", "type": "string", "default": ""},
        {"name": "email", "type": "string", "default": ""},
        {"name": "phone", "type": "string", "default": ""},
        {"name": "date", "type": "string", "default": ""},
        {"name": "org", "type": "string", "default": ""},
        {"name": "cost", "type": "long", "default": 0}
    ]
}
```

Further, suppose these records are stored using the employee's unique identifier. For example, these records might use key paths which end with an employee unique identifier, like this:

```
/Corporate/people/10012
/Corporate/people/10013
/Corporate/people/10014
```

In this case, in order to find all people who belong to the organization called "Support," you would have to iterate over every record whose key begins with `/Corporate/people`, examine each in turn for the proper organization name, and construct a list of those people who belong to that organization. Depending on the number of people contained in your store, this could be a lengthy operation.

The alternative is to create an index view that is keyed by the organization name. For example:

```
/IndexView/People/Organization/Engineering
/IndexView/People/Organization/Sales
/IndexView/People/Organization/Support
```

There are two ways to handle the data portion of these records. One way is for each record to contain a list of keys corresponding to the people records belonging to that organization. That is, the key:

```
/IndexView/People/Organization/Support
```

would return a data item with was a list of major keys for all those people entries containing an 'org' of 'Support'. As an Avro schema, you would represent the data item in the following way:

```
{
    "type": "record",
    "name": "SecondaryDBValue",
```

```
    "namespace": "oracle.kv.indexView",
    "fields": [
        {"name": "arrays",
         "type": {"type" : "array", "items" : "string"},
         "default" : []}
  ]
}
```

While this approach will work for small-to-medium sized indexes, it ultimately suffers from an inability to scale. It would be far too easy to create a view whose list of primary keys is too large to be easily handled by your code. In fact, it could easily grow so large that it could not fit into available memory. Given the size of the datasets for which Oracle NoSQL Database is designed, this is a very real consideration.

A different approach would be to create index views where each record referred to one and only one primary record. That is, the data portion of the record contains a simple string representing the key path to a primary record. (You could also carry this information as an array of key path components.) However, you cannot duplicate keys in Oracle NoSQL Database, so in this case the key needs to somehow be unique, based on the information found in the primary record. As an example, you could create keys that contains both the organization name, as well as the user's UID:

```
/IndexView/People/Organization/Support/-/10012
```

refers to the primary record:

```
/Corporate/People/10012
```

Using Key-Only Records

Key-only index view records carry all of the record's information in the key; the data portion of the record is set to an empty value. In this scheme, each index view record represents a single pairing between the secondary key and the primary record key to which it refers. Because Oracle NoSQL Database is good at scaling up to large numbers of records, this eliminates the scalability problem described in the previous section.

Note

The following examples use fairly long key paths. This is done for the purpose of clarity. However, in general, shorter key paths are desirable and so the paths shown here should not be taken as advice for how to construct the keys for your records.

Essentially, key-only index view records carry the index view's key in the major portion of the key path, and the corresponding primary record's key in the minor portion of the key path. That is:

```
/Secondary/Key/Path/-/Primary/Key/Path
```

The minor path component here is the key path for a primary record. For example, building on the example presented in the previous section, this might be:

```
/Secondary/Key/Path/-/Corporate/people/10012
```

The major key path portion of the record needs to carry more information:

- Index key prefix

  This is simply a prefix value used to indicate that the record is an index view record. The prefix can be anything so long as it is unique within your store; for example, `IDX`.

- Index name

  This is used to differentiate this index view from other types of index views. You could use something fairly simple here that is indicative of the information indexed by this record, such as `EMPLOYEE_NAME` or `EMPLOYEE_LOCATION`. However, it is possible to carry more complex information if you set up your code correctly. We discuss this further in Complex Index Names (page 106).

- Field value(s)

  The remainder of the major key path is a sequence of one or more field values that are obtained from the associated primary record. This is the actual information that you are indexing.

  In the simplest case, this portion of the key contains only one field value; for example, an organization name if what you are doing is indexing all employees by organization. For example:

  ```
  /IDX/ORGANIZATION/Engineering/-/Corporate/people/10012
  ```

  is a view entry that indicates employee record 10012 belongs to the Engineering organization.

  However, this portion of the key path can contain multiple field values, which gives you multi-column views. An example of this is indexing by employee common and family name, both of which would be individual fields in the primary record:

  ```
  /IDX/EMP_NAME/Smith/Robert/-/Corporate/people/10012
  /IDX/EMP_NAME/Smith/Patricia/-/Corporate/people/40288
  /IDX/EMP_NAME/Smyth/Don/-/Corporate/people/7893
  ```

Complex Index Names

As described above, an index name can be a simple text label, especially if you have fairly simple indexing requirements. However, it is possible to carry more information about the view record in the index name. You can construct the index name so that it identifies:

- The Avro schema name used by the primary record.
- A list of the field names that this view is indexing. This information is useful for generalizing your Avro binding code, especially as the number of fields you are indexing grows large, and/or as the number of types of index views grows large.

One way to construct an index name that carries this information is to create a list object that holds all the information you want in your index name, then create a one-way hash of the

information using `java.security.MessageDigest`. Converting the list to a byte array can be accomplished using the `Key.createKey()` method. For example:

```
    /**
     * Construct and return an index name representing an index type.
     */
    private String getIndexName(String schemaName,
                                List<String> indexFieldNames) {

        MessageDigest md = null;
        try {
            /*
             * The implementation for digestCache is omitted
             * for brevity.
             */
            md = digestCache.get();
            List<String> minorPath = new ArrayList<String>();
            minorPath.add(schemaName);
            minorPath.addAll(indexFieldNames);
            byte[] bytes =
                Key.createKey("", minorPath).toString().getBytes();
            md.update(bytes);
            return new String(md.digest());
        } finally {
            digestCache.free(md);
        }
    }
```

This means that the information you are carrying in your index name is locked up in a one-way hash. There is no way to retrieve that information from the hash, so you need to store it somewhere. You need a separate set of records to record index view metadata.

Managing Index View Metadata

Index view metadata is information you want to record about each index type. Mostly, this is information you use to construct your index names (if you use complex index names). You can also record your index state as a part of your metadata.

You can collect your index view metadata as a series of key-only records. In this case, the keys are constructed like this:

```
/PREFIX/INDEX_NAME/-/SCHEMA_NAME/FNAME1/FNAME2/.../STATE
```

where:

- `PREFIX` is a unique identifier that you use to indicate this record is an index view metadata record. For example: `META`.
- `INDEX_NAME` is the name you are using for the type of index for which you are collecting metadata. If you are using a simple name (for example, `ORGANIZATION` or `EMP_NAME`), then use that. If you are using a hashed complex name, such as is described in the previous section, then use that here.

- SCHEMA_NAME is the name of the Avro schema used by the primary record. This must be the same schema name as you used to construct your complex index name.
- FNAME1, FNAME2, and so forth, are the primary record field names this view type is using. Again, these must be identical to the field names you used to construct your complex index name. They must also appear in the same order as the field values used to construct your index view record keys.
- STATE is the current state of the index type represented by this metadata record. Examples of view STATE are:
  - BUILDING to indicate that the index view is currently being built.
  - DELETING to indicate that the index view is currently being deleted.
  - READY to indicate that the index view is ready for use.

  These are just some suggestions. STATE can really indicate anything that is useful to your code. But in the example given here, your code would only use the view if its state was READY.

Using Index View Records and Metadata Together

Putting it all together, to create an index view that uses complex index names, you would:

1. Create the index name, using the schema and field names that you are working with.
2. Create the metadata record, as described in the previous section, setting its state to BUILDING.
3. Iterate over your store, creating a view record for each primary record that you want to index. Use the index name you created in step 1 as part of the view record's major key path. See Using Key-Only Records (page 105) and Complex Index Names (page 106) for more information.
4. When you are done creating the view, change the status for the metadata record to READY. (To do this, you delete the old record and create a new one.)

To use (read) index views, you:

1. Check the corresponding metadata record to make sure the index view is in a READY state. If it is not, you can abort the read, or pause until the state has changed to READY.
2. Iterate over the index view records that interest you for the search.
3. For each such record, use it to retrieve the corresponding primary record.
4. For each primary record, use the schema and field names, contained in the corresponding metadata record, along with your Avro binding, to serialize/deserialize the primary record's data.

To update an existing record, you:

1. Retrieve the primary record.
2. Retrieve the index view record.
3. Modify the primary record as needed.
4. Modify the index view record to reflect the changes to the primary record.
5. Check the status of the index view to ensure that it is in a READY state. If it is, then write the index view record back to the store.

   If the index view status is not READY, then either wait for the status to change to READY before writing the index view record, or fail the operation.
6. Write the modified primary record back to the store.

An example of performing all these operations is available in your Oracle NoSQL Database distribution. See Example (page 111) for details.

Key Size Consideration

The longer your keys, the more memory you are using at your nodes. Keys can therefore grow so large that they harm your system's overall read/write throughput due to an inability to maintain enough records in cache.

The key-only design pattern described here will probably result in very long keys. Whether those key sizes are so large that they cause you a performance problem is a function of how long your keys actually are, how many keys you need to manage, and how much memory is available on your nodes.

If your keys are so large that they will cause an I/O throughput issue, then you need to implement some other design approach.

General Index Views Considerations

While creating index views can vastly improve your stores read performance (depending on the size of your data set, and the kinds of questions you want to ask of it), there are some limitations of which you need to be aware.

Additional Write Activity

Maintaining an index view necessarily requires additional read and write activity over and above what is required just to maintain a primary record. Whether this additional activity will measurably affect your write throughput depends on the size of the dataset you are indexing, and the size of your views.

For small datasets and small views, this additional activity will not be noticeable. But as the size of the data to be indexed grows, and so your views themselves grow larger, you might

notice a reduction in throughput, particularly in write throughput. Given this, when planning your store size, be sure to consider overhead to maintain index views.

Non-Atomic Updates

Because index views are managed by the application, Oracle NoSQL Database cannot insure that operations performed on the primary record are atomic with operations performed on the corresponding view records. This means that it is possible to change your primary records, but have the corresponding operation(s) on your index view(s) fail thereby causing them to be out of sync with the primary data. The reverse can also happen, where the index view update operation is successful, but the update to the primary record fails.

Note that for some workloads, non-atomic updates to primary records and their index views can be desirable. This is sometimes the case for workloads that require the maximum read and write throughput possible. For these types of applications, consistency between the index view and the primary data is not nearly as important as overall performance.

That said, you should still make an attempt to determine whether your indexes are out of sync relative to your primary data, so that you can perform compensating transactions if your code detects a problem. You may also need to flag your index views as being in an unsafe state if some aspect of the update operations fail. The safest way (not necessarily the fastest way) to update a primary record for which you are maintaining an index view is:

1. Check whether your view is in a READY state. If it is, proceed with the update operation. If it is not, either pause and wait for the state to change, or abort the update entirely.
2. Update your primary record as necessary, but *do not write the results back to the store yet*.
3. Update your index view to be reflective of the changes you have made to the primary record.
4. Write the primary record to the store. If the write fails, perform a compensating transaction to fix the problem. Either retry the write operation with the updated record, or check to ensure that the record which is currently in the store is not corrupted or altered in any way.
5. If the update to the primary record succeeds, then write the changes to the index view to the store. If this succeeds, then you are done with your update.
6. If the update to the index view record fails, then immediately mark your index view as being in a non-READY state. How you do this depends on how you are storing index view state flags, but assuming you are using metadata records, that needs to be updated before you take steps to fix your index view.

A similar algorithm is required for the creation and deletion of primary records.

Of course, this means that before you perform a read with your index view, you need to check the view's state before you proceed. If the view's state is not READY, then you need to either pause until the state is READY, or you need to abandon the read entirely. In addition to this check, you also need to ensure that your index views are in a state that is consistent with the primary records. This is described next.

Decoupled Consistency

As described above, index views can be out of sync with your primary data due to some generic failure in the update operation. Your code needs to be robust enough to recognize when this has happened, and take corrective action (including rebuilding the index view, if necessary). A related, but temporary, problem is that for any given node, changes to your views may not have caught up to changes to your primary records due to replication delays. Note that it is also possible for views on the local node to have been updated when the corresponding primary data modifications have not yet arrived.

Again, for some workloads, it might not be critically important that your views are in sync with your primary data. However, if your workload is such that you need assurance your views accurately reflect your primary data, you need to make use of Oracle NoSQL Database's built-in consistency guarantees.

One way to do this is to use an absolute consistency guarantee for any reads that you perform using your views. But this can ultimately harm your read and write performance because absolute consistency requires the read operation to be performed on a master node. (See Using Simple Consistency (page 86) for details.) For this reason, you should use absolute consistency only when it is truly critical that your views are completely up-to-date relative to your primary data.

A better way to resolve the problem is to use a version-based consistency guarantee when using your index views. You will need to check the version information on both the primary data and the views when performing your reads in order to ensure that they are in sync. You may also need to create a way to transfer version information between different processes in your application, if one process is responsible for performing store writes and other processes are performing store reads. For details on using version-based consistency, see Using Version-Based Consistency (page 88).

Example

An example of creating and managing index views is included in the Oracle NoSQL Database distribution. It can be found here:

```
<KVROOT>/examples/secondaryindex
```

The example exposes a command line interface that allows you to create and delete index views, retrieve the primary data referred to by an index view record, and insert, delete, and update new primary records. The application is a very simple application that allows you to create views against customer billing records.

The example uses key-only index view records, with complex index names and associated metadata records. The code that is responsible for managing the views and associated metadata is contained in this class:

```
<KVROOT>/examples/secondaryindex/IndexViewService.java
```

Note that the example is one expression of the index view design pattern. Its operations may not be a match for the way your code operates, but it should serve as good design guide. Feel free to adapt, expand, or simplify the example code to match your own design needs and goals.

Appendix A. Using the Large Object API

Oracle NoSQL Database provides an interface you can use to read and write Large Objects (LOBs) such as audio and video files. As a general rule, any object greater than 1 MB is a good candidate for representation as a LOB. The LOB API permits access to large values, without having to materialize the value in its entirety by providing streaming APIs for reading and writing these objects.

A LOB is stored as a sequence of chunks whose sizes are optimized for the underlying storage system. The chunks constituting a LOB may not all be the same size. Individual chunk sizes are chosen automatically by the system based upon its knowledge of the underlying storage architecture and hardware. Splitting a LOB into chunks permits low latency operations across mixed workloads with values of varying sizes. The stream based APIs serve to insulate the application from the actual representation of the LOB in the underlying storage system.

The LOB interface makes use of text-only keys that can be used with either the Tables API or the Key/Value API. This document provides high-level concepts pertinent to the LOB interface, and then provides examples of using it with both the Tables and the Key/Value APIs.

LOB Keys

LOBs are stored and retrieved using `oracle.kv.Key` objects. Each such object contains a series of strings which represent a key path. Key paths are divided into two parts: major and (optionally) minor. The last key path component in a LOB key must end with a trailing suffix that is `.lob` by default.

The LOB suffix may be defined using the `KVStoreConfig.setLOBSuffix()` method.

Unlike other objects stored using `Key` objects (such as when using the Key/Value API), LOB data is not stored on shards driven by the major and minor key path components in the key path. Instead, LOB data uses a hidden keyspace, and its various chunks are distributed across partitions based on this keyspace, instead of based on the `Key` which you provide.

Also, be aware that the actual key you use for your LOBs is stored on a single partition based on its major/minor key path components, and the partition used for this storage is selected in the same way that the store partitions any data based on the key's major and minor key paths. However, assuming reasonably brief keys, this represents a small amount of data and it should not substantially affect your store's data sizing requirements.

The rest of this section provides a brief overview of Oracle NoSQL Database `Keys`, which may be of interest to users of the Table API. Key/Value API users should already be familiar with these concepts and so they can skip to the next section.

A `Key` object can be constructed in several different ways, depending on how you want to represent the key path components. Most commonly, arrays are used to represent the major and minor key path components, and these arrays are then provided to the `Key.createKey()` method. Users of the Table API may find the `Key.fromString()` method convenient because

it is easy to store a string representation of a key path in a table cell. Alternatively, Table API users can store key paths components as arrays in table cells, or construct the key path array using information found in table cells.

When represented as a string, key paths begin with a forward slash ('/), which is also used to delimit each path component. If a minor key path is used, then it is delimited from the major key path using a dash ('-). For example, a LOB key used for an image file might be:

```
/Records/People/-/Smith/Sam/Image.lob
```

Users of the Table API should take care to ensure that their LOB paths do not collide with the keys used internally by their tables. In general this is easy to do because the key paths used internally to store table data begin with a numerical representation of the table's name.

LOB Key Checks

All of the LOB APIs verify that the key used to access LOBs meets the trailing suffix requirement. If the trailing suffix requirement is not met, the LOB APIs throw an `IllegalArgumentException` exception. This requirement permits non-LOB methods to check for inadvertent modifications to LOB objects.

This is a summary of LOB-related key checks performed across all methods (LOB and non-LOB):

- When using the Key/Value API, all non-LOB write operations check for the absence of the LOB suffix as part of the other key validity checks. If the check fails (that is, the key contains the LOB suffix), the non-LOB API throws an `IllegalArgumentException`.
- When using the Key/Value API, all non-LOB read operations return the associated opaque value used internally to construct a LOB stream.
- All LOB write and read operations check for the presence of the LOB suffix. If the check fails it will result in an `IllegalArgumentException`.

LOB APIs

Due to their representation as a sequence of chunks, LOBs must be accessed exclusively using the LOB APIs. If you use a LOB key with the family of `KVStore.getXXX` Key/Value methods, you will receive a value that is internal to the KVS implementation and cannot be used directly by the application.

The LOB API is declared in the `KVLargeObject` interface. This is a superinterface of `KVStore`, so you create a `KVStore` handle in the usual way and then use that to access the LOB methods.

The LOB methods are:

- `KVLargeObject.deleteLOB()`

  Deletes a LOB from the store.

- `KVLargeObject.getLOB()`

  Retrieves a LOB from the store.

- `KVLargeObject.putLOB()`, `KVLargeObject.putLOBIfAbsent()`, and `KVLargeObject.putLOBIfPresent()`.

  Writes a LOB to the store.

LOB Operation Exceptions

The methods used to read and insert LOBs are not atomic. This relaxing of the atomicity requirement permits distribution of LOB chunks across the entire store. It is therefore the application's responsibility to coordinate concurrent operations on a LOB. The LOB implementation will make a good faith effort to detect concurrent modification of a LOB and throw `ConcurrentModificationException` when it detects conflicts, but there is no guarantee that the API will detect all such conflicts. The safe course of action upon encountering this exception is to delete the LOB and replace it with a new value after fixing the application level coordination issue that provoked the exception.

Failures during a LOB modification operation result in the creation of a partial LOB. The LOB value of a partial LOB is in some intermediate state, where it cannot be read by the application; attempts to `getLOB()` on a partial LOB will result in a `PartialLOBException`. A partial LOB resulting from an incomplete `putLOB()` operation can be repaired by retrying the operation. Or it can be deleted and a new key/value pair can be created in its place. A partial LOB resulting from an incomplete delete operation must have the delete retried. The documentation associated with individual LOB methods describes their behavior when invoked on partial LOBs in greater detail.

Key/Value LOB Example

The following example writes and then reads a LOB value using the Key/Value API. Notice that the object is never actually materialized within the application; instead, the value is streamed directly from the file system to the store. On reading from the store, this simple example merely counts the number of bytes retrieved.

Also, this example only provides bare-bones exception handling. In production code, you will probably want to do more than simply report the exceptions caught by this example.

```
package kvstore.lobExample;

import java.io.File;
import java.io.FileNotFoundException;
import java.io.IOException;
import java.io.FileInputStream;
import java.io.InputStream;
import java.util.Arrays;
import java.util.concurrent.TimeUnit;

import oracle.kv.Consistency;
import oracle.kv.Durability;
import oracle.kv.KVStore;
import oracle.kv.KVStoreConfig;
import oracle.kv.KVStoreFactory;
import oracle.kv.Key;
```

```
import oracle.kv.RequestTimeoutException;
import oracle.kv.Version;
import oracle.kv.lob.InputStreamVersion;
import oracle.kv.lob.PartialLOBException;

public class LOBKV {

    private String[] hhosts = {"localhost:5000"};

    public static void main(String args[]) {
        LOBKV lobkv = new LOBKV();
        lobkv.run(args);
        System.out.println("All done.");
    }
    private void run(String args[]) {

        KVStoreConfig kconfig = new KVStoreConfig("kvstore", hhosts);
        KVStore kvstore = KVStoreFactory.getStore(kconfig);

        // Use key "/test/lob/1.lob" to store the jpeg object.
        // Note that we are not using a minor key in this
        // example. As required, the final key component
        // uses a ".lob" suffix.
        final Key key =
            Key.createKey(Arrays.asList("test", "lob", "1.lob"));

        File lobFile = new File("test.jpg");
        try {
            FileInputStream fis  = new FileInputStream(lobFile);

            // The image file is streamed from the filesystem into
            // the store without materializing it within the
            // application. A medium level of durability is
            // used for this put operation. A timeout value
            // of 5 seconds is set in which each chunk of the LOB
            // must be written, or the operation fails with a
            // RequestTimeoutException.
            kvstore.putLOB(key, fis,
                    Durability.COMMIT_WRITE_NO_SYNC,
                    5, TimeUnit.SECONDS);


            // Now read the LOB. It is streamed from the store,
            // without materialization within the application code.
            // Here, we only count the number of bytes retrieved.
            //
            // We use the least stringent consistency policy
            // available for the read. Each LOB chunk must be read
            // within a 5 second window or a RequestTimeoutException
```

```
            // is thrown.
            InputStreamVersion istreamVersion =
                 kvstore.getLOB(key,
                             Consistency.NONE_REQUIRED,
                             5, TimeUnit.SECONDS);

             InputStream stream = istreamVersion.getInputStream();
             int byteCount = 0;
             while (stream.read() != -1) {
                 byteCount++;
             }
             System.out.println(byteCount);
        } catch (FileNotFoundException fnf) {
            System.err.println("Input file not found.");

            System.err.println("FileNotFoundException: " +
                    fnf.toString());
            fnf.printStackTrace();
            System.exit(-1);
        } catch (RequestTimeoutException rte) {
            System.err.println("A LOB chunk was either not read or");
            System.err.println("not written in the alloted time.");

            System.err.println("RequestTimeoutException: " +
                rte.toString());
            rte.printStackTrace();
            System.exit(-1);
        } catch (PartialLOBException ple) {
            System.err.println("Retrieval (getLOB()) only retrieved");
            System.err.println("a portion of the requested object.");

            System.err.println("PartialLOBException: " + ple.toString());
            ple.printStackTrace();
            System.exit(-1);
        } catch (IOException e) {
            System.err.println("IO Exception: " + e.toString());
            e.printStackTrace();
            System.exit(-1);
        }
    }

    protected LOBKV() {}
}
```

Table LOB Example

The following example writes and then reads a LOB value using the Table API. Notice that the object is never actually materialized within the application; instead, the value is streamed

directly from the file system to the store. On reading from the store, this simple example merely counts the number of bytes retrieved.

When you use LOBs with the Table API, you must still use a `Key` to identify the LOB object. In other words, you cannot directly store the LOB in a table row. Typically you will construct the `Key` using information stored in your table. For example, you can simply store the LOB's key as a text string in one of your table cells. Or you could store the key's values as an array in a table cell (or two arrays, if you are using minor key components). Finally, you can construct the key based on values retrieved from one or more cells in the row.

Also, this example only provides bare-bones exception handling. In production code, you will probably want to do more than simply report the exceptions caught by this example.

Before beginning, we must define and create the table in the store. The following table definition describes a very minimal table of user information. It then uses a child table to identify one or more image files associated with the user.

```
table create -name userTable
add-field -type STRING -name userid
add-field -type STRING -name familiarname
add-field -type STRING -name surname
primary-key -field userid -field familiarname -field surname
shard-key -field userid
exit
plan add-table -name userTable -wait

table create -name userTable.images
add-field -type STRING -name imageFileName
add-field -type STRING -name imageVersion
add-field -type STRING -name imageDescription
primary-key -field imageFileName
exit
plan add-table -name userTable.images -wait
```

Add the table definition to the store:

```
> java -Xmx256m -Xms256m \
-jar KVHOME/lib/kvstore.jar runadmin -host <hostName> \
-port <port> -store <storeName>
kv-> load -file createLOBTable.txt
Table userTable built.
Executed plan 5, waiting for completion...
Plan 5 ended successfully
Table userTable.images built.
Executed plan 6, waiting for completion...
Plan 6 ended successfully
```

Now we can write and read table data. In the following example, we create two users that between them have three associated images. First the table rows are created (written), and then the BLOB data is saved to the store. The example then iterates over the tables, showing relevant information, and along the way showing the images associated with each user. In this case, we limit the BLOB display to merely reporting on the BLOB's byte count.

```
package kvstore.lobExample;

import java.io.File;
import java.io.FileNotFoundException;
import java.io.IOException;
import java.io.FileInputStream;
import java.io.InputStream;
import java.util.Arrays;
import java.util.concurrent.TimeUnit;

import oracle.kv.Consistency;
import oracle.kv.Durability;
import oracle.kv.KVStore;
import oracle.kv.KVStoreConfig;
import oracle.kv.KVStoreFactory;
import oracle.kv.Key;
import oracle.kv.RequestTimeoutException;
import oracle.kv.Version;
import oracle.kv.lob.InputStreamVersion;
import oracle.kv.lob.PartialLOBException;

import oracle.kv.table.PrimaryKey;
import oracle.kv.table.Row;
import oracle.kv.table.Table;
import oracle.kv.table.TableAPI;
import oracle.kv.table.TableIterator;
import oracle.kv.table.MultiRowOptions;

public class LOBTable {

    private String[] hhosts = {"localhost:5000"};

    // Store handles
    private KVStoreConfig kconfig;
    private KVStore kvstore;

    // Table handles
    private TableAPI tableH;
    private Table userTable;
    private Table userImageTable;
    private Row row;

    private static String blobPfx = "blobpfx";
    private static String imgSfx = "userimage.lob";

    public static void main(String args[]) {
        LOBTable lobtable = new LOBTable();
        lobtable.run(args);
        System.out.println("All done.");
```

```
    }

    private void run(String args[]) {

        // Open a store handle
        kconfig = new KVStoreConfig("kvstore", hhosts);
        kvstore = KVStoreFactory.getStore(kconfig);
        tableH = kvstore.getTableAPI();


        // Obtain table handles
        userTable = tableH.getTable("userTable");
        userImageTable = tableH.getTable("userTable.images");

        // populate the tables, and load LOBs into the store
        addData();

        // retrieve tables, and retrieve LOBs from the store
        // and show some details about the tables and LOBs.
        retrieveData();
    }

    // Creates some table rows and loads images into the store
    private void addData() {

        // Load up a couple of rows in the user (parent) table.
        row = userTable.createRow();
        row.put("userid","m.beckstrom.3267");
        row.put("familiarname","Mary");
        row.put("surname","Beckstrom");
        tableH.put(row, null, null);

        row = userTable.createRow();
        row.put("userid","h.zwaska.9140");
        row.put("familiarname","Harry");
        row.put("surname","Zwaska");
        tableH.put(row, null, null);

        // Now populate each row's image (child) table
        // and stream in a BLOB as each row is created.
        row = userImageTable.createRow();
        row.put("userid","m.beckstrom.3267");
        row.put("imageFileName","IMG_2581.jpg");
        row.put("imageDescription","Highrise sunset");
        tableH.put(row, null, null);
        loadBlob("m.beckstrom.3267", "IMG_2581.jpg");

        row = userImageTable.createRow();
        row.put("userid","m.beckstrom.3267");
```

```
        row.put("imageFileName","IMG_2607.jpg");
        row.put("imageDescription","Swing set at dawn");
        tableH.put(row, null, null);
        loadBlob("m.beckstrom.3267", "IMG_2607.jpg");

        row = userImageTable.createRow();
        row.put("userid","h.zwaska.9140");
        row.put("imageFileName","mom1.jpg");
        row.put("imageDescription","Mom's 89th birthday");
        tableH.put(row, null, null);
        loadBlob("h.zwaska.9140", "mom1.jpg");
    }

    // Loads a blob of data into the store
    private void loadBlob(String userid, String filename) {

        // Construct the key.
        // userid and filename are information saved in the
        // table, so later we can recreate the key by table
        // examination. blobPfx is a constant that we use for
        // all BLOB data. imgSfx ends the key path with the
        // required suffix. We use a fixed constant partially
        // to meet the BLOB suffix requirement, but in a
        // larger system this could also be used to
        // differentiate the type of data contained in the
        // BLOB (image data versus an audio file, for example).

        final Key key = Key.createKey(
                Arrays.asList(blobPfx, userid, filename, imgSfx));

        File lobFile = new File(filename);
        try {
            FileInputStream fis  = new FileInputStream(lobFile);
            // The image file is streamed from the filesystem into
            // the store without materializing it within the
            // application. A medium level of durability is
            // used for this put operation. A timeout value
            // of 5 seconds is set in which each chunk of the LOB
            // must be written, or the operation fails with a
            // RequestTimeoutException.
            kvstore.putLOB(key, fis,
                    Durability.COMMIT_WRITE_NO_SYNC,
                    5, TimeUnit.SECONDS);
        } catch (FileNotFoundException fnf) {
            System.err.println("Input file not found.");

            System.err.println("FileNotFoundException: " +
                    fnf.toString());
            fnf.printStackTrace();
```

```
            System.exit(-1);
        } catch (RequestTimeoutException rte) {
            System.err.println("A LOB chunk was either not read or");
            System.err.println("not written in the alloted time.");

            System.err.println("RequestTimeoutException: " +
                rte.toString());
            rte.printStackTrace();
            System.exit(-1);
        } catch (IOException e) {
            System.err.println("IO Exception: " + e.toString());
            e.printStackTrace();
            System.exit(-1);
        }
    }

    // Retrieves the user (parent) table, as well as the images
    // (child) table, and then iterates over the user table,
    // displaying each row as it is retrieved.
    private void retrieveData() {

        PrimaryKey key = userTable.createPrimaryKey();
        // Iterate over every row in the user table including
        // images (child) table in the result set.
        MultiRowOptions mro = new MultiRowOptions(null, null,
                Arrays.asList(userImageTable));
        TableIterator<Row> iter =
                   tableH.tableIterator(key, mro, null);
        try {
            while (iter.hasNext()) {
                Row row = iter.next();
                displayRow(row);
            }
        } finally {
            iter.close();
        }
    }

    // Display a single table row. Tests to see if the
    // table row belongs to a user table or a user images
    // table, and then displays the row appropriately.
    private void displayRow(Row row) {
        if (row.getTable().equals(userTable)) {
            System.out.println("\nName: " +
                    row.get("familiarname").asString().get() +
                    " " +
                    row.get("surname").asString().get());
            System.out.println("UID: " +
                    row.get("userid").asString().get());
```

```
        } else if (row.getTable().equals(userImageTable)) {
            showBlob(row);
        }
    }

   // Retrieves and displays a BLOB of data. For this limited
    // example, the BLOB display is limited to simply reporting
    // on its size.
    private void showBlob(Row row) {
        // Build the blob key based on information stored in the
        // row, plus external constants.
        String userid = row.get("userid").asString().get();
        String filename  = row.get("imageFileName").asString().get();
        final Key key = Key.createKey(
                Arrays.asList(blobPfx, userid, filename, imgSfx));

        // Show supporting information about the file which we have
        // stored in the table row:
        System.out.println("\n\tFile: " + filename);
        System.out.println("\tDescription: " +
                row.get("imageDescription").asString().get());

        try {
            // Now read the LOB. It is streamed from the store,
            // without materialization within the application code.
            // Here, we only count the number of bytes retrieved.
            //
            // We use the least stringent consistency policy
            // available for the read. Each LOB chunk must be read
            // within a 5 second window or a RequestTimeoutException
            // is thrown.
            InputStreamVersion istreamVersion =
                 kvstore.getLOB(key,
                             Consistency.NONE_REQUIRED,
                             5, TimeUnit.SECONDS);

             InputStream stream = istreamVersion.getInputStream();
             int byteCount = 0;
             while (stream.read() != -1) {
                 byteCount++;
             }
             System.out.println("\tBLOB size: " + byteCount);

        } catch (RequestTimeoutException rte) {
            System.err.println("A LOB chunk was either not read or");
            System.err.println("not written in the alloted time.");

            System.err.println("RequestTimeoutException: " +
                rte.toString());
```

```
            rte.printStackTrace();
            System.exit(-1);
        } catch (PartialLOBException ple) {
            System.err.println("Retrieval (getLOB()) only retrieved");
            System.err.println("a portion of the requested object.");

            System.err.println("PartialLOBException: " + ple.toString());
            ple.printStackTrace();
            System.exit(-1);
        } catch (IOException e) {
            System.err.println("IO Exception: " + e.toString());
            e.printStackTrace();
            System.exit(-1);
        }
    }

    protected LOBTable() {}
}
```

Appendix B. Third Party Licenses

All of the third party licenses used by Oracle NoSQL Database are described in the `LICENSE.txt` file, which you can find in your KVHOME directory.