Introduction

This document describes the basic requirements for installing **SMRT Analysis v2.3.0** on a customer system.

SMRT Analysis is designed to be installed on a wide variety of 64-bit Linux distributions **without** requiring root access. The installation is all within a **single directory tree**, and includes embedded Apache Tomcat and MySQL servers, as well as the whole set of analysis tools and the libraries to support them all.

The input and output data, as well as the temporary data used by the system, may be located on local or network storage. There are two directories (or softlinks) created in the root directory during the initial installation to point to this: `userdata` and `tmpdir`.

When running on a cluster, **all** machines in the queue **must** be able to see:

- The install directory
- Any remotely-mounted directories used for the data at the same paths as those used on the server where the tools are installed.

What's New in v2.3.0?

Bundled SMRT Portal Application Database

Pre-built binaries for MySQL Server are bundled with SMRT Analysis v2.2.0 and later. This provides a stand-alone, isolated environment for the SMRT Portal application database, free of external system dependencies.

This architecture:

- Allows for a more seamless installation and upgrade process
- Provides an additional measure of data security with automated schema backups.

After installation or upgrade, the default behavior is to embed the database. **All** data from the external database is migrated to the bundled server.

Bundled Software

SMRT Analysis includes the following third-party software packages bundled within the application; the application should **not** depend on what is already deployed on the system.

- Apache Tomcat™ 7.0.23
- Celera® Assembler 8.1
- Docutils 0.8.1
- GMAP (2014-08-04)
- HMMER 3.1b1 (May 2013)
- Java™ SE Runtime Environment (build 1.7.0_02-b13)
- Mono 3.0.7
- MySQL® 5.1.73

- Perl v5.8.8
- Python® 2.7.3
- SAMtools 0.1.17
- Scala 2.9.0 RC3

Note: GATK and associated executables are **no longer** included.

Release Notes

See the document **SMRT Analysis Release Notes (v2.3.0)** for changes and known issues.

- You can find the latest version of this document on the Pacific Biosciences web site.

SMRT Analysis Installation Environment Assumptions

- A 64-bit Linux machine with a `libc` version greater than 2.5.
- Installing as the **same** non-root user (`SMRT_USER`) that will be used to run the system.
- If upgrading, the **same** user with the same root directory (`SMRT_ROOT`) as used previously.
- The `SMRT_USER` has **full** permissions in the file system in the `SMRT_ROOT` directory and in all linked directories for `tmpdir` and `userdata`, and can lock files in these file systems. (Common problems include NFS setup problems, ACLs, and so on.)
- When running in distributed mode, all other nodes have the **same path** for `SMRT_ROOT` and for all linked directories.
- During the installation, no other daemons are running on the same ports chosen for the Tomcat and MySQL servers.

Post-Installation Assumptions

- No changes to `SMRT_USER`, `SMRT_ROOT`, the local hostname and/or to the IP that was used to install the system.

Installation Instruction Conventions

In the commands below, several environment variables are used and must be defined, **or** you must run the installation with the proper values substituted for these variables. These variables merely clarify where you have options during the installation process; they are **not** required during or after the install.

- `SMRT_ROOT`: Full path to the directory in which SMRT Analysis is or will be installed.
- `SMRT_USER`: The user name of the user running the installation and SMRT Analysis daemons and tools; or the user name of the user in previous installs. This should be the same value as `$(id -nu)`.
- `SMRT_GROUP`: The default group name for `SMRT_USER`. This should be the same value as `$(id -ng)`.

Below are typical values for `SMRT_ROOT`, `SMRT_USER`, and `SMRT_GROUP`. Although you can use different values, we recommend defaulting to the following conventions for most installations:

```
SMRT_ROOT=/opt/smrtanalysis/
SMRT_USER=smrtanalysis
SMRT_GROUP=smrtanalysis
```

Getting Started

Note: This section contains a summary of the commands used for a **quick installation**.

- Use these commands **only** if you are familiar with the installation/upgrade process.

- Proceed to "Installation Guide" on page 4 for the more detailed procedure.

Download SMRT Analysis

Download SMRT Analysis from PacBio DevNet (**http://www.pacbiodevnet.com**):

```
wget https://s3.amazonaws.com/files.pacb.com/software/smrtanalysis/2.3.0/
smrtanalysis_2.3.0.140936.run
wget https://s3.amazonaws.com/files.pacb.com/software/smrtanalysis/2.3.0/smrtanalysis-
patch_2.3.0.140936.p5.run
```

Installation Summary

Set the installation environment variables. See "Installation Instruction Conventions" on page 2 for details on SMRT_* variables:

```
SMRT_ROOT=/opt/smrtanalysis/
SMRT_USER=smrtanalysis
SMRT_GROUP=smrtanalysis
```

Create the SMRT Analysis installation directory:
```
sudo mkdir $SMRT_ROOT
sudo chown $SMRT_USER:$SMRT_GROUP $SMRT_ROOT
```

Run the installation:
```
su -l $SMRT_USER
bash smrtanalysis_2.3.0.140936.run -p smrtanalysis-patch_2.3.0.140936.p5.run --rootdir
$SMRT_ROOT
```

Start the SMRT Analysis daemons:
```
$SMRT_ROOT/admin/bin/smrtportald-initd start
$SMRT_ROOT/admin/bin/kodosd start
```

Upgrade Summary

Set the installation environment variables. See "Installation Instruction Conventions" on page 2 for details on SMRT_* variables:

```
SMRT_ROOT=/opt/smrtanalysis/
SMRT_USER=smrtanalysis
```

Stop the SMRT Analysis daemons and run the upgrader:
```
su -l $SMRT_USER
$SMRT_ROOT/admin/bin/smrtportald-initd stop
$SMRT_ROOT/admin/bin/smrtupdater smrtanalysis_2.3.0.140936.run
```

Start the SMRT Analysis daemons:
```
$SMRT_ROOT/admin/bin/smrtportald-initd start
$SMRT_ROOT/admin/bin/kodosd start
```

Once SMRT Portal is installed, proceed to the following sections to complete setup:

1. "Set Up SMRT Portal" on page 6 (for new installations **only**.)
2. "Verify the Installation" on page 7 (for new installations **and** upgrades.)

Patch Summary

Set the installation environment variables. See "Installation Instruction Conventions" on page 2 for details on SMRT_* variables:

```
SMRT_ROOT=/opt/smrtanalysis/
SMRT_USER=smrtanalysis
```

Stop the SMRT Analysis daemons:
```
su -l $SMRT_USER
$SMRT_ROOT/admin/bin/smrtportald-initd stop
```

These two commands must be run as the SMRT_USER**. They perform the patch and restart the daemons:**

```
$SMRT_ROOT/admin/bin/smrtupdater <patchfile>
$SMRT_ROOT/admin/bin/smrtportald-initd start
```

Installation Guide

- For information on hardware and software requirements, see the document **SMRT® Analysis System Requirements**, available on the Pacific Biosciences web site.

Installation Details

Downloading SMRT Analysis

Download SMRT Analysis from PacBio DevNet (**http://www.pacbiodevnet.com**):

```
wget https://s3.amazonaws.com/files.pacb.com/software/smrtanalysis/2.3.0/
smrtanalysis_2.3.0.140936.run
wget https://s3.amazonaws.com/files.pacb.com/software/smrtanalysis/2.3.0/smrtanalysis-patch_2.
```

Create the SMRT Analysis User

We recommend that a system administrator create a special user called smrtanalysis, who belongs to the smrtanalysis group. This user will own **all** SMRT Analysis files, daemon processes, and SMRT Pipe jobs. Alternatively, the system may be installed and run by an individual user. Whatever user name you use, the instructions below assume that you are executing them as that user, and refer to that user and it's default group as SMRT_USER and SMRT_GROUP.

Create the Installation Path

The SMRT Analysis top-level directory, $SMRT_ROOT, can be **any** directory as long as the smrtanalysis user has read, write, and execute permissions in that directory. Historically, we referred to $SMRT_ROOT as /opt/smrtanalysis.

If the parent directory $SMRT_ROOT is **not** writable by the SMRT Analysis user, the $SMRT_ROOT directory **must** be pre-created with read/write/execute permissions for the SMRT Analysis user.

Run the Installer

The installation script attempts to discover inputs when possible, and performs the following configurations:

1. Confirms the valid non-root user that will own SMRT Pipe jobs and daemon processes.
2. Performs hardware, OS, and software prerequisite checks.
3. Identifies valid host names and IP addresses recognized by DNS.
4. Specifies the Tomcat web server **main** port and **shutdown** port numbers.
5. Creates and verifies symbolic links to the TMP and USERDATA directories.
6. Specifies MySQL server settings and initializes the SMRT Portal database.

7. Distributed/non-distributed SMRT Pipe jobs.
8. Specifies Job Management System and related parameters for queues and parallel environments.

```
# see Installation Instruction Conventions for details about SMRT_* variables
# SMRT_ROOT is the directory where you want to install SMRT Analysis
sudo mkdir $SMRT_ROOT
# SMRT_USER and SMRT_GROUP are the user and group you are using to install SMRT Analysis
sudo chown $SMRT_USER:$SMRT_GROUP $SMRT_ROOT

su -l $SMRT_USER
bash smrtanalysis_2.3.0.140936.run --rootdir $SMRT_ROOT

$SMRT_ROOT/admin/bin/smrtportald-initd start
$SMRT_ROOT/admin/bin/kodosd start
```

If you cancelled out of the install prompt and want to rerun the script **without** extracting again, you can rerun using the `--no-extract` option:

```
bash smrtanalysis_2.3.0.140936.run --rootdir $SMRT_ROOT --no-extract
```

Apply Patches During Installation

If installing **after** a patch was released for the software, you can install **both** the software and the patch in one command using the `-p` option:

```
bash smrtanalysis_2.3.0.140936.run -p smrtanalysis-patch_2.3.0.140936.p5.run --rootdir
```

Set up Distributed Computing

Configuring Job Submission Templates

You configure distributed computing by editing three template files:

```
$SMRT_ROOT/current/analysis/etc/cluster/<JMS>/start.tmpl
$SMRT_ROOT/current/analysis/etc/cluster/<JMS>/interactive.tmpl
$SMRT_ROOT/current/analysis/etc/cluster/<JMS>/kill.tmpl
```

Specifying the SGE Job Management System

The install script will automatically discover the `queue name` and `parallel environment` name based on the SGE installed on your system. If you want to configure or add options to the `qsub` command, you must edit the `.tmp` files manually. For example, the default `interactive.tmpl` looks like the following:

```
qsub -pe smp ${NPROC} -S /bin/bash -V -q secondary -N ${JOB_ID} -o ${STDOUT_FILE} -e
${STDERR_FILE} ${EXTRAS} ${CMD}
```

If you are assembling large genomes (> 100 Mb) and wish to use the job distribution functionality within Celera Assembler, you **must** make sure the parallel environment is configured to use the `$pe_slots allocation` rule. For example, the `smp` parallel environment is configured as follows:

```
$ qconf -sp smp
pe_name            smp
slots              99999
user_lists         NONE
xuser_lists        NONE
start_proc_args    /bin/true
stop_proc_args     /bin/true
allocation_rule    $pe_slots
control_slaves     FALSE
job_is_first_task  TRUE
urgency_slots      min
accounting_summary FALSE
```

Specifying the PBS Job Management System

PBS does **not** have a `-sync` option, and the `interactive.tmpl` file runs a script named `qsw.py` to simulate the functionality. You must edit **both** `interactive.tmpl` and `start.tmpl`.

1. Change the queue name to one that exists on your system. (This is the `-q` option.)
2. Change the parallel environment to one that exists on your system. (This is the `-pe` option.)
3. Make sure that `interactive.tmpl` calls the `-PBS` option.

Specifying the LSF Job Management System

The equivalent SGE `-sync` option in LSF is `-K` and this should be provided with the `bsub` command in the `interactive.tmpl` file.

1. Change the queue name to one that exists on your system. (This is the `-q` option.)
2. Change the parallel environment to one that exists on your system. (This is the `-pe` option.)
3. Make sure that `interactive.tmpl` calls the `-K` option.

Specifying other Job Management Systems

1. Create a new directory `$SMRT_ROOT/current/analysis/etc/cluster/NEW_JMS`.
2. Edit `$SMRT_ROOT/current/analysis/etcsmrtpipe.rc`, and change the `CLUSTER_MANAGER` variable to `NEW_JMS`.
3. Once you have a new JMS directory specified, create and edit the `interactive.tmpl`, `start.tmpl`, and `kill.tmpl` files for your particular setup.

Start the SMRT Analysis Services

Start the MySQL and Tomcat Daemons

The following command starts **both** Tomcat and MySQL. Use this command to restart services.

```
$SMRT_ROOT/admin/bin/smrtportald-initd start
```

You can also individually control MySQL and Tomcat for troubleshooting purposes:

```
$SMRT_ROOT/admin/bin/mysqld start
$SMRT_ROOT/admin/bin/tomcatd start
```

You can check that the services are on or off using the `ps` command:

```
ps -ef | grep tomcat
ps -ef | grep mysql
```

Start the Kodos Daemon

```
$SMRT_ROOT/admin/bin/kodosd start
```

You can check that the services are on or off using the `ps` command:

```
ps -ef | grep kodos
```

Set Up SMRT Portal

1. Use a web browser to start SMRT Portal: `http://HOST:PORT/smrtportal`
2. Click **Register** at the top right.
3. Create a user named `administrator` (all lowercase). This user is special - it is the only user that does **not** require activation on creation.
4. Enter the user name `administrator`.

5. Enter an email address. All administrative emails, such as new user registrations, are sent to this address.
6. Enter, then confirm the password.
7. Select **Click Here** to access **Change Settings**.
8. To set up the mail server, enter the SMTP server information and click **Apply**. For email authentication, enter a user name and password. You can also enable Transport Layer Security.
9. To enable automated submission from a PacBio instrument, click **Add** under the Instrument Web Services URI field. Then, enter the following into the dialog box and click **OK**:

   ```
   http://INSTRUMENT_PAP01:8081
   ```

   – `INSTRUMENT_PAP01` is the IP address or name (`pap01`) of the instrument.
   – `8081` is the port for the instrument web service.

10. Select the new URI, then click **Test** to check if SMRT Portal can communicate with the instrument service.
11. (Optional) You can delete the pre-existing instrument entry by clicking **Remove**.

Verify the Installation

Create a test job in SMRT Portal using the provided lambda sequence data. This is data from a single SMRT Cell that has been down-sampled to reduce overall tarball size. If you are upgrading, this cell will already have been imported into your system, and you can skip to Step 10 below.

1. Open your web browser and clear the browser cache:
   – **Google Chrome**: Choose **More Tools > Clear browsing data**. Choose **the beginning of time** from the droplist, then check **Cached images and files** and click **Clear browsing data**.
   – **Internet Explorer**: Choose **Tools > Internet Options > General,** then under **Browsing history**, click **Delete.** Check **Temporary Internet files and website files,** then click **Delete.**
   – **Firefox®**: Choose **Options > Advanced**, then click the **Network** tab. In the Cached Web Content section, click **Clear Now**.
2. Refresh the current page by pressing **F5**.
3. Navigate to SMRT Portal at `http://HOST:PORT/smrtportal`, then log in.
4. Click **Design Job**.
5. Click **Import and Manage**.
6. Click **Import SMRT Cells**.
7. Click **Add**.
8. Enter `common/test/primary`, then click **OK**.
9. Select the new path and click **Scan**. You should get a dialog saying "One input was scanned."
10. Click **Design Job**.
11. Click **Create New**.
12. Enter a job name and comment.
13. Select the protocol `RS_Resequencing.1`.
14. Under **SMRT Cells Available**, select a lambda cell and click the right-arrow button.
15. Click **Save** on the bottom right, then click **Start**. The job should complete successfully.
16. Click the **SMRT View** button. SMRT View should open with tracks displayed, and the reads displayed in the Details panel.

Optional Configurations

Set Up the User Data Directory

The user data folder, `$SMRT_ROOT/userdata`, expands rapidly because it contains all jobs, references, and drop boxes. We recommend softlinking this folder to an **external** directory with more storage:

```
mv $SMRT_ROOT/userdata /path/to/NFS/mounted/offline_storage
ln -s /path/to/NFS/mounted/offline_storage $SMRT_ROOT/userdata
```

Upgrade Details

Supported Upgrade Path
- For SMRT Analysis v2.3.0, **only** upgrades directly from **v2.2.0** are supported.
- SMRT Analysis does **not** support upgrades from SMRT Analysis v2.0.1 or earlier. The recommended upgrade path is to incrementally upgrade to each version, that is:

```
1.4 -> 2.0.0 -> 2.0.1 -> 2.1.0 -> 2.2.0 ->2.3.0
```

Alternately, you may opt for a fresh installation of SMRT Analysis v2.3.0 and then manually import old SMRT Cells and jobs to preserve analysis history.

See **https://github.com/PacificBiosciences/SMRT-Analysis/wiki/Official-Documentation** for information about upgrading from earlier versions of SMRT Analysis.

Run the Upgrader
Upgrades are handled by the script `smrtupdater` located in `$SMRT_ROOT/admin/bin/smrtupdater`. The script performs the following:

1. Confirms the valid non-root user that will own SMRT Pipe jobs and daemon processes.
2. Checks for running services, and stops them if necessary.
3. Performs hardware, OS, and software prerequisite checks.
4. Transfers the computing configurations from the previous installation.
5. Performs a Reference Repository upgrade check.
6. Confirms and validates symbolic links to `TMP` and `USERDATA` directories.
7. Upgrades the MySQL Database.

The upgrade script does **not** port over protocols that were defined in previous versions of SMRT Analysis. This is because protocol files can vary a great deal between versions due to rapid code development and change. Please **recreate** any custom protocols you may have.

```
$SMRT_ROOT/admin/bin/smrtupdater smrtanalysis_2.3.0.140936.run
```

Applying Patches During Upgrade
```
$SMRT_ROOT/admin/bin/smrtupdater -p smrtanalysis-patch_2.3.0.140936.p5.run
smrtanalysis_2.3.0.140936.run
```

Start the SMRT Analysis Services

Start the MySQL and Tomcat Daemons
```
$SMRT_ROOT/admin/bin/smrtportald-initd start
```

Start the Kodos Daemon
```
$SMRT_ROOT/admin/bin/kodosd start
```

Known Installation Problems and Workarounds

Remote Storage Issues
We have seen some problems with the `mysql` portion of the installation due to the inability of the `mysql` scripts to change ownership (and possibly to change permissions) of files in `$SMRT_ROOT/userdata/runtime/tmp`. In each case, `userdata` was linked to remote NFS storage where the problem could be demonstrated with simple tests like creating a temporary file and running `chown` on it.

The **best** way to resolve this problem is to fix the storage issue, but you can use the following workaround instead:

```
SMRT_ROOT=<customer_specific>
# you can actually put these new directories anywhere on the head node
# local filesystem but these are shown as an example
SMRT_DB=$SMRT_ROOT/../smrtanalysis_db
SMRT_RTTMP=$SMRT_ROOT/../smrtanalysis_runtime_tmp
SAUSER=<smrtanalysis_user>
SAGRP=<smrtanalysis_group>
sudo mkdir $SMRT_DB; sudo chown $SAUSER:$SAGRP  $SMRT_DB
sudo mkdir $SMRT_RTTMP; sudo chown $SAUSER:$SAGRP  $SMRT_RTTMP

# replace old directories with links to these new ones
# note that this is safe to do only because this database
# directory is new with the 2.2 install and you have not
# yet finished a 2.2. install or used 2.2 yet
sudo rm -rf $SMRT_ROOT/userdata/database
sudo rm -rf $SMRT_ROOT/userdata/runtime/tmp

# then as the  <smrtanalysis_user>:
ln -s $SMRT_DB $SMRT_ROOT/userdata/database
ln -s $SMRT_RTTMP $SMRT_ROOT/userdata/runtime/tmp

#From there, you should be able to execute the install or upgrade as shown above.

# the following is probably not necessary due to the way that we resolve paths in our scripts,
# but this will cleanup broken links created during the install
rm $SMRT_ROOT/userdata/database/mysql/log
rm $SMRT_ROOT/userdata/database/mysql/runtime
ln -s $SMRT_ROOT/userdata/log $SMRT_ROOT/userdata/database/mysql/log
ln -s $SMRT_ROOT/userdata/runtime $SMRT_ROOT/userdata/database/mysql/runtime
```

ACL Problems

If you use ACLs in the SMRT_ROOT or any of the linked storage, you may have obscure install or execution problems if the smrtanalysis user does **not** have full permissions. Installations have failed due to the inability to copy a file with cp -a in some of the install scripts. If you suspect ACL related problems, try disabling them and retrying.

Advanced Deployment

Using Amazon Web Services

Users wishing to run SMRT Analysis in the cloud can use an Amazon Machine Image (AMI) with SMRT Analysis pre-installed. For details, see the document **Running SMRT® Analysis on Amazon**, available on the Pacific Biosciences web site.

P/N 100-376-500-07