BaseSpace® Clarity LIMS Reporting Module

Analyze lab data and uncover and resolve issues.

Introduction

The newest scientific technologies give genomics labs much greater insight into the human genome. Because these technologies are relatively new, they can be difficult to implement and fine tune. To maximize investment, instruments must run at peak performance and be used effectively. To remain competitive, labs must offer a fast turnaround and low costs using these technologies. And to provide the most accurate sample results possible, labs must maintain a high level of quality. Optimizing every workflow and procedure in the lab can be one way to reduce costs and improve turnaround time, while producing accurate results. But new methods and the hectic day to day in a lab make it difficult to keep pace, let alone get ahead and optimize.

Like most industries, genomics laboratories are turning to reporting and analytic tools to identify unknown issues, discover bottlenecks, or simply confirm intuition about how something is working (or not) in the lab.1 Yet, a recent study from Forrester Research suggests that most organizations, including laboratories, analyze only a small portion of their data, missing opportunities to improve operations that aren’t immediately apparent.2 Most of these organizations indicate that accessing their data and knowing where to start are the biggest challenges to launching reporting and analytic initiatives.

BaseSpace Clarity LIMS Reporting Module helps labs overcome these challenges. The Reporting Module is a ready-to-use application that enables labs to analyze metrics, identify trends, and share analytic reports. Preconfigured data sources and reports enable labs to work with the data immediately, eliminating the need for scripting or database query knowledge. Further, labs with database query knowledge that would like to construct their own data sources can do so with an add-on to the module: Reporting Module Pro.

This document describes the Reporting Module and the technical aspects of implementing it within an organization. Sections include:

- Overview of the BaseSpace Clarity LIMS Reporting Module
- Architecture
- The extract transform load (ETL) Service
- Preconfigured Data Sources
- Preconfigured Reporting Module Reports
- Licensing for the Reporting Module
- System Requirements
- Support and Documentation

Overview of the BaseSpace Clarity LIMS Reporting Module

The BaseSpace Clarity LIMS Reporting Module is built on the Tableau reporting platform, which is based on the philosophy that everybody should be able to see and understand data. This philosophy has been applied in a laboratory setting, helping scientists visualize, understand, and share data for more informed decisions about how to run the lab.

Traditionally, many of the limitations surrounding reporting involve the need for database or structured query language (SQL) knowledge. This barrier is removed by delivering preconfigured data sources that do this work for you. Data sources enable connection to the BaseSpace Clarity LIMS Database using different data views. These views present data for a specific purpose and combine data from 1 or more database tables for visualization or building and editing reports.

Similarly, the module comes with a collection of preconfigured reports for examining and identifying issues in the lab. Reports such as Turnaround Time can help labs identify how long certain milestones are taking to reach. Similarly, labs can see all history for a sample in Sample History or the Audit Trail reports.

Finally, the Reporting Module makes it easy to see the data available in the data sources, drag and drop that information into a report, and share that report with other members of the lab. The graphical and intuitive interface enables creation of new reports quickly, without having to go to the SQL layer or ask developers for help writing scripts.

Reporting Module Pro

The Reporting Module was designed to accommodate most labs interested in developing a comprehensive reporting and analytics program. For labs that already have a reporting or business analytics team, BaseSpace Clarity LIMS Reporting Module Pro is available as an add-on to the Reporting Module and allows access to additional features. These include creating data sources, generating dashboards, and accessing advanced calculations and analytics (Table 1).

Hosted and On-Premise Options

For the Reporting Module and Pro add-on, there are hosted and on-premise instances. For security reasons, the Reporting Module must be implemented the same way as your BaseSpace Clarity LIMS System. For example, an on-premise version of the LIMS, requires the on-premise version of reporting. Similarly, a hosted version of the LIMS dictates that a hosted version of the Reporting Module is needed.

Reporting Module Architecture

The BaseSpace Clarity LIMS Reporting Module uses a simple architecture in which data are extracted from the BaseSpace Clarity LIMS Database on its server and published to the Windows Tableau Server. When the ETL service runs, it extracts data from the BaseSpace Clarity LIMS Database and stores them in a separate database, the data warehouse. A configuration file connects the data warehouse to the Windows Server on which Tableau resides. Tableau then extracts the data from the data warehouse for users to visualize in reports (Figure 1).
It is recommended that Reporting Module users set up an empty database on their existing BaseSpace Clarity LIMS Server so that when the ETL Service runs, it’s accessing the same server and not having to go through the VPN tunnel or other remote connection. With Reporting Module Pro, users also receive a license for a desktop application that connects to the BaseSpace Clarity LIMS Server and the Tableau Server.

ETL Service
BaseSpace Clarity LIMS uses ETL to access data in the LIMS Database for visualization and reporting. This ETL Service reads data from the LIMS Database, reformats it into a more reporting-friendly format, and loads the data into the data warehouse (Figure 2). Although most ETLs are fairly standard, this customized version more efficiently handles user-defined fields, or those custom fields set up by LIMS users.

Along with the ability to control the timing of the ETLs, users can disable an ETL completely. Users may want to do this if the lab does not use a particular feature or functionality. For example, if you have not enabled the audit trail on BaseSpace Clarity LIMS System, it might make sense to turn off the clarity_auditing_etl.

NOTE: Aggressive ETL or extract approaches can have an impact on the performance of the BaseSpace Clarity LIMS System.

Table 1: Comparison of BaseSpace Clarity LIMS Reporting Module and Reporting Module Pro

<table>
<thead>
<tr>
<th>Feature</th>
<th>BaseSpace Clarity LIMS Reporting Module</th>
<th>Reporting Module Pro</th>
</tr>
</thead>
<tbody>
<tr>
<td>Create/Edit New Reports</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Create/Edit Data Sources</td>
<td>Access to preconfigured data sources only</td>
<td>X</td>
</tr>
<tr>
<td>Create Dashboards</td>
<td>View and interact with preconfigured dashboards only</td>
<td>X</td>
</tr>
<tr>
<td>Advanced Calculations and Analytics</td>
<td>Access to standard calculations and analytics only</td>
<td>X</td>
</tr>
</tbody>
</table>

Figure 1: A High-Level View of the BaseSpace Clarity LIMS Reporting Architecture

Manipulating the ETL Service
The BaseSpace Clarity LIMS Reporting Module, requires installation of the ETL Service on the LIMS Server, or have the BaseSpace Clarity LIMS team install it for you. After it is installed, with access to the server, users can start, stop, get status, and log files for the service.

The ETL Service also uses an etl-config.properties file, which specifies the connection between the LIMS Database and data warehouse and the frequency with which ETL jobs will run (Table 2).

Figure 2: How the ETL Service Interacts with the LIMS Database and the Reporting Database Warehouse
For example, sample UDFs are populated in the views with the following naming convention, where # is the index of the view:
clarity_smp_udf_pivot#_view

Each view contains up to 500 columns with the first column being the submitted_sample_id and the remaining 499 columns the values for the sample UDFs. The number of columns in each view depends on the number of UDFs that are defined for samples.

If 499 or less UDFs are defined for samples, then one pivot view, clarity_smp_udf_pivot1_view, will contain all of the sample UDFs.

• If more than 499 UDFs are defined for samples, then a second pivot view, clarity_smp_udf_pivot2_view, will contain additional sample UDFs.
• As the number of sample UDFs increases, additional pivot views will be populated.

How the aliases are generated

The pivot views for all the UDF ETLs have been updated to have unique human-readable column names, each ending with the ID of the UDF the column represents. The mapping tables have also been updated to include a Udfid column. Putting UDFs in a report-friendly format requires using the UDF name as a column header in the data warehouse. Therefore, column headers are limited to a certain length and cannot contain special characters.

The aliases are generated by the following algorithm:
• Replace % with ‘pct’
• Replace & with ‘and’
• Replace all whitespace with an underscore character (‘_’)
• Remove all remaining non-word characters
• If the alias starts with a digit, add the ‘udf_’ prefix
• If the alias starts with an underscore character (‘_’), add the ‘udf’ prefix

Truncate the alias such that the Udfid can be added as a suffix with an underscore character (‘_’), and the entire alias remains less than or equal to the maximum allowable column name size for the reporting database.

Preconfigured Data Sources

The BaseSpace Clarity LIMS Reporting Module uses predesigned, reusable data sources to connect to the database tables. These data sources are built with data views, which present data for a specific purpose and can combine data from 1 or more database tables.

Views are how the data are depicted at the database level. Data sources are how they are combined or used alone to appear in the Reporting Module.

Because BaseSpace Clarity LIMS Reporting uses predesigned data sources, users do not need SQL skills to edit and create reports; users can simply work with the data in the predesigned data sources.

A complete reference of each view exists in the user documentation for the BaseSpace Clarity LIMS Reporting Module.
Views
The BaseSpace Clarity LIMS Reporting data sources access 5 views: clarity_reporting_view, clarity_in_progress_view, clarity_dashboard_view, clarity_audit_trail_view, and clarity_reagents_view. Views are created in the BaseSpace Clarity LIMS data warehouse by the reporting service. For more detailed information, and descriptions of the columns each view includes, see the Reporting Data Model Documentation PDF.

clarity_reporting_view
The clarity_reporting_view extracts data from 24 separate tables. Each row represents a step that a sample has been through, and contains detailed sample, project, and client information. Includes information pertaining to:
- Submitted Sample
- Derived Sample
- Project
- Client
- Step
- Protocol
- Workflow
- Step Technician
- Container
- Instrument
- QC Flags
- Escalation

clarity_in_progress_view
The clarity_in_progress_view focuses specifically on the samples currently in progress in BaseSpace Clarity LIMS. Each row represents a derived sample in progress, along with the technician and status of the progress. Includes information pertaining to:
- Derived Samples in Progress
- Step Technician
- Step Status

clarity_dashboard_view
The clarity_dashboard_view contains data similar to that displayed in the BaseSpace Clarity LIMS Overview dashboard. Each row contains a submitted sample that is active in the LIMS. It includes information about the protocol and workflow in which each sample is currently queued, and the client and project the sample is associated with. Includes information pertaining to:
- Submitted Samples assigned to a workflow
- Workflow
- Protocol
- Client Account
- Client

clarity_audit_trail_view
Includes information pertaining to:
- Audit Events
- Audit Changes
- Event Types
- Change Types
- Application Name
- Application Type

clarity_reagents_view
Includes information pertaining to:
- Step
- Reagent Kit
- Reagent Lot
- Kit Supplier
- Lot Expiry Date

Data sources
Data sources are the reporting depictions of single or combinations of views. The data sources included in the Reporting Module are described in the documentation available (Figure 3).

ClarityReporting+Reagents Data Source
The ClarityReporting+Reagents data source forms the basis for most reports. The data source is created by combining the clarity_reporting_view and the clarity_reagents_view, both of which are defined within the LIMS Database.

This data source counts the number of steps completed by a sample. It is used for calculating turnaround time, client account information, reagents, and container details. This data source does not include UDFs, or audit trail information. The ClarityReporting+Reagents data source contains 3 prebuilt hierarchies (for workflow, project, and reagent). These allow the report user to drill down into the report details and examine the data at a more granular level. BaseSpace Clarity LIMS Reporting can connect directly to the ClarityReporting+Reagents data source without the need to make any joins or filter any data.

Figure 2: A Small Subset of Data Sources from BaseSpace Clarity LIMS Reporting Module
Data Sources Mapped from Views

Additional data sources are available as 1:1 mappings and can be used when creating reports and worksheets from multiple data sources and views. For detailed information on these data sources, see the Reporting Data Model Documentation PDF.

Refreshing Data in Data Sources

There are 2 major phases that occur when refreshing the data inside a data source. The first is the transfer of data from the LIMS Database to the data warehouse. By default, the ETL Service performs this transfer every night at midnight. The exception to this rule is the in-progress data and the dashboard data, which update every 5 minutes.

The second phase involves extracting the data from the data warehouse and transferring it to the BaseSpace Clarity LIMS Reporting Module. By default, this happens every weekday morning at 4:00 AM. Users with the proper privileges can modify these defaults. Also, you can refresh either phase on an as-needed basis with the BaseSpace Clarity LIMS Reporting Module.

Preconfigured Reports

The Reporting Module includes preconfigured reports that were designed to capture some of the most common data in which a lab would be interested. These reports can be used as they are or you can modify them to fit a lab's specific needs.

Sample History Report

The Sample History report summarizes information about a single sample or a small group of samples. This information is presented as a dashboard that compiles information from several independent reports. It includes information about the project, reagents, and steps. Users can filter the report by client account, project, and submitted sample name (Figure 4).

Escalations Report

The Escalations report focuses on samples escalated for manager review in Lab View. The report displays escalations based on project, technician, and timeline. This report is made up of 3 sections: Escalations by Project, Escalations by Technician, and Escalation Trends (Figure 4).

Reagents Report

The Reagents report focuses on reagent use and is helpful when troubleshooting problems related to sample quality. One of the possible causes of poor-quality results can be a bad reagent lot. With this report, users can quickly identify all samples that were exposed to a particular reagent lot number.

Users can also use this report to help manage reagent supply. The report indicates the number of times a particular reagent has been used, and how many active lot numbers are in the LIMS.

This report is made up of 3 sections: Samples per Reagent, Reorder Reagents, and Reagents and Samples (Figure 4).

Sequencing Report

The Sequencing report provides lab managers with an understanding of sequencing runs completed, samples sequenced, and instruments used. It allows you to interrogate sequencing runs using filters for run date, container type, instrument type, and instrument name.

This report is made up of 3 independent reports: Sequencing Steps, Samples per Container, and Sequencing per Run Date (Figure 4).

Project Management Report

Information presented in the Project Management report is similar to the information shown in the Overview dashboard. However, this report lets users filter the data based on client account, project name, project client, and workflow assigned. The charts are color-coded by project. This report shows only open and active projects in your LIMS. This report is made up of 2 sections: Workflows by Project and Protocols by Project (Figure 4).

Figure 4: A Sampling of Preconfigured Reports
In Progress Report
The In Progress report allows you to see all steps in progress, samples in progress, and the technicians currently working on samples. This report is made up of 2 sections: Steps in Progress and Samples in Progress (Figure 5).

Steps Completed Report
The Steps Completed report displays the number of steps each sample has completed in the LIMS, per project. The report graphs the number of distinct step IDs associated with each project.

An increased number of steps may indicate that samples needed to be reworked or requeued. Highlighting these increases can help users to identify potentially problematic steps, protocols, and samples, and fix any inefficiencies (Figure 5).

Top Projects Report
The Top Projects report displays the top number of projects based on the number of submitted samples that have completed at least 1 step in the LIMS. The report graphs the number of distinct submitted sample IDs associated with each project.

This visualization allows lab managers and directors to identify the top projects in their labs, over a particular date range or for a particular client account (Figure 5).

Top Workflows Report
The Top Workflows report displays the top workflows based on the number of submitted samples assigned to the workflow that have completed at least 1 step in the LIMS. The report graphs the number of distinct submitted sample IDs associated with each workflow (Figure 5).

This visualization allows lab managers and directors to identify the top workflows in their labs, over a particular date range or for a particular client account.

Turnaround Time Report
Lab managers of a commercial or core lab need to track the time between sample arrival to sending results to clients for individual projects or samples, and for a batch of projects or samples. They may also need to calculate the average turnaround time of all projects or samples in a given time period (Figure 5).

The Turnaround Time by Client or Project graphs the average turnaround time associated with each project. The report includes automatic calculations and averages for turnaround time based on client account or project.

NOTE: The Turnaround Time Report is also available for examining the amount of time it takes to complete steps in the LIMS.

Audit Trail Report
The Audit Trail IT and lab managers need to show auditors everything that has occurred in the LIMS Database over a given time period, for a specific sample or for a particular database.

While the Audit Trail feature in the LIMS tracks this information, the Audit Trail- Events and Audit Trail- Change reports allows lab managers to interrogate the audit events and audit change tables based on many factors, at a granular level.

The Audit Trail - Events report tracks all events that occur in the LIMS.
The Audit Trail - Change report tracks all changes that occur in the LIMS (Figure 5).

Container View Report
The Container View report is based on the contents of a container. This view allows you to see the contents of a specific container, including the sample names and the date that derived samples in the container were created (Figure 5).

Figure 5: A Preconfigured Sampling of Reports
UDF Map Report

The UDF Map reports show how UDF names in the LIMS map to the field names used in the Reporting Module (Figure 5). The reports may be used to help decipher truncated UDF names. 7 reports are provided:

- Sample UDF Map
- Project UDF Map
- Account UDF Map
- Client UDF Map
- Container UDF Map
- Step UDF Map
- Output (Artifact) UDF Map

The following columns of information are included in each report:

- The UDF name used in the LIMS (column name: BaseSpace Clarity LIMS Name)
- The UDF alias created for reporting purposes (column name: Reporting Alias)
- The name of the udf_pivot_view that contains each UDF (column name: Pivottable)
- The unique ID for each UDF (column name: Udfid)

Editing and Creating Reports

Users with the appropriate permissions can create new reports, duplicate reports, and rename reports, directly from the user interface. In addition, there are a wide variety of components that users can modify to make sure that each report displays the data of interest.

Comprehensive documents for creating and editing reports can be found on Zendesk and accessed through the online help available from Tableau.

Licensing for the Reporting Module

The BaseSpace Clarity LIMS Reporting Module comes with 2 named licensed. Users and organizations can purchase additional licenses at any time. Reporting Module Pro comes with 1 additional license. Full documentation on accessing and logging on to the Reporting Module is included in the BaseSpace Clarity LIMS Reporting Module Documentation.

System Requirements

ETL Service Requirements

The prerequisites described must be in place before the BaseSpace Clarity LIMS Reporting ETL Service can be installed and used. If users have a hosted implementation of BaseSpace Clarity LIMS Reporting, the BaseSpace Clarity LIMS Support team will perform any steps required to meet these prerequisites.

Client and Server Systems Meet BaseSpace Clarity LIMS Requirements

Reporting is supported on BaseSpace Clarity LIMS versions 3.5.9, 4.0.12, 4.1.4, 4.2.4, and later.

Users need to make sure that your system meets the requirements described in the BaseSpace Clarity LIMS Technical Overview document for the version of BaseSpace Clarity LIMS (located in the Technical Requirements forum of the LIMS documentation).

NOTE: The BaseSpace Clarity LIMS Reporting Module supports Red Hat and CentOS operating systems only. The SUSE operating system is not supported.

An Empty Database/Schema is Available for the Reporting Data Warehouse

BaseSpace Clarity LIMS Systems with a PostgreSQL Database: Make sure that a new database has been created and configured on the server designated for reporting.

BaseSpace Clarity LIMS Systems with an Oracle Database: Make sure that an empty schema has been created and configured on your database server designated for reporting.

In addition to the new database/schema, a new database user is required. This user will be used by the ETL Service and will have full read/write access to the database/schema. Instructions for configuring user access and permissions are provided in the BaseSpace Clarity LIMS Reporting Installation Guide. The reporting integration populates this database/schema at run time.
The database/schema:

- Can be on the same server as the current Clarity LIMS Database.
- Must not be the same database/schema used by Clarity LIMS.

Requirements for BaseSpace Clarity Reporting Module and Pro

BaseSpace Clarity LIMS integrates with Tableau Server and Tableau Desktop to provide BaseSpace Clarity LIMS Reporting Module and BaseSpace Clarity LIMS Reporting Pro.

NOTE: The requirements listed in this section are based on Tableau Server requirements and recommendations. See Minimum Hardware Requirements and Recommendations for Tableau Server.

BaseSpace Clarity LIMS Reporting Module: Server Requirements and Recommendations

- Microsoft Windows Server 2008 SP2 with platform update, 2008 R2, 2012, or 2012 R2; Windows 7, 8, 8.1, or 10 on x86 or x64 chipsets
- 64-bit versions of Windows
- Minimum of Intel Pentium 4 or AMD Opteron processor
- 32-bit color depth recommended
- Internet Protocol version 4 (IPv4)
- Internet connection speed:
  - Minimum: 100 Mbps Ethernet network or faster
  - Recommended: 1 Gbps Ethernet network or faster
- No server application such as Internet Information Services (IIS) web server should be installed

Minimum

The minimum specifications listed are suggested for prototyping and testing of BaseSpace Clarity LIMS Reporting Module only. The installer checks for these minimum system requirements and will not proceed on computers with less than the hardware minimums.

64-bit computers

- 4 cores
- 8 GB system memory
- 15 GB minimum free disk space

Recommended

The minimum configuration recommended for production usage of BaseSpace Clarity LIMS Reporting Module is based on:

- Single computer
- 64-bit processor
- 8 physical cores, 2.0 GHz or higher CPU
- 32 GB system memory
- 200 GB minimum free disk space on a second drive

NOTE: Core count is based on ‘physical’ cores. Physical cores can represent actual server hardware or cores on a virtual machine (VM). Hyperthreading is ignored for the purposes of counting cores. For BaseSpace Clarity LIMS Reporting Module running on a 64-bit virtual machine, you will need a minimum of 4 physical cores.

BaseSpace Clarity LIMS Reporting Pro: System Requirements

Windows

- Microsoft Windows Vista SP2 or later (64-bit)
- Microsoft Server 2008 R2 or later (64-bit)
- Minimum of Intel Pentium 4 or AMD Opteron processor (SSE2 or later required)
- 2 GB memory
- Minimum 750 megabytes free disk space
- Internet Explorer 8 or later

Mac

- iMac/MacBook 2009 or later
- OS X 10.9 or later

Support and Documentation

Support is provided through the BaseSpace Clarity LIMS Team. Full documentation is available on Zendesk after licenses are acquired.

Learn More

To learn more about the BaseSpace Clarity LIMS Reporting Module, contact informatics@illumina.com.

References


