Standard Product Description

Science Tools Corporation

PRODUCT NAME: **The BigSur System™, Version 3.5** **SPD 1.1.0.3.5**

Description

The BigSur System™ (BigSur) was designed from best-practice principals to manage an entire scientific enterprise. It is poly-disciplinary and may join disparate scientific disciplines in one single, fully distributed computing environment. It provides for "end-to-end" management, from original source collection of data through automation of all processing workflows to derived data products, publishing and sharing of results, to visualization and other end-uses of resulting data products. BigSur is designed to handle configurations which are multi-site, multi-organizational, and with participants that are involved in multiple groups and projects distributed throughout. BigSur serves as a scientific notebook, includes precise and detailed tracking of all data and processing, computational resources, and what was shared with whom, when, how, and on whose authorization, with robust and extensible, high-granularity security. BigSur manages meta-data related to people in a way that prevents "shadows" of individual people from being accidentally confused as multiple people. BigSur is designed to learn and absorbs - rather than embeds - technical standards and data-types related to the assets it manages so that concurrent versions may be managed simultaneously. BigSur separates out "name space" so that "collisions" that would otherwise prevent concurrent running of different versions of various software on a single system do not occur.

BigSur includes a turn-key “grid” (or “cloud”) computing system providing a heterogeneous, fully distributed, high-performance, scientific-oriented computing environment with superior meta-data features. BigSur’s high-semantic relationship management of meta-data, including data objects, compute functions / processes, deployed code, computers, people, and more, and high-granularity, extensible security features provide scientific defensibility of results through full lineage of derived data products, full management of process source, object, and executable components for full context of all processing, data safety and robust tracking of data.

BigSur utilizes a relational database-centric client/server model operating across heterogeneous platforms. BigSur is comprised of a database schema called STDB, a set of “applications” that act upon that schema, a set of additional “helper” utilities, an Application Programming Interface (API), and related documentation.

This release is a maintenance release, with modest additions, for version 3.4 which in 2012 introduced a new "Dual Mode" capability in which separate read and write database connections can be established for the ultimate in performance. In such installations, one connection is used for all insert, update, and delete operations, while a different one is used for reading activities. This permits The BigSur System™ to be among the very first database API product offerings - if not the very first - to be able to take full advantage of the new in PostgreSql 9.2 (open source database) feature called "cascading binary replication" in which one "master" database feeds many "slaves". Science Tools has provided this support with FCS (First Customer Ship) exactly one week from the release of the PostgreSql 9.2 (September 11, 2012) which includes cascading binary replication, in a part of our on-going commitment to help our customers remain at the leading-edge of computing performance.

The BigSur Roadmap outlines release 4.0, the next major release, as the first to include an application-server design that can optionally be used, along side the current architecture. Starting with version 3.2 the API was altered to help ensure that very little modification, if any, will be required to upgrade to version 4.0. Both the older database-centric client/server orientated architecture and the newer application server architecture will be supported concurrently in version 4.0.

Key Benefits

Scientific "Notebook" & Scientific Defensibility

- Manages scientific objects (data), processes (code and executables), and the relationships between these including all meta-data about these
- Learns about researcher's scientific data types and processes, and the relationships and structures into which they belong
- Fully distributed architecture - object meta-data, object data, scientific processes and functions, and authorized users may be anywhere
- Robust knowledge about data-types - tracks access methods, visualization and processing tools

- Complete scientific object lineage, including full processing history and relationships between data-sets and objects
- Provides for collection
- Eliminates "shadows" of people in which one person appears to be many by tracking the relationships between actual people and the account names used on participating computers.
- Our Geo extension conforms to the U.S. Federally mandated Federal Geographic Data Committee (FGDC) Geospatial meta-data standard. Other extensions may be available or developed; only GeoDb is available for public release.

Workflow System

- Provides for archetypal process definitions
- Workflow steps can be computing steps, human-interaction steps, or logical (some say virtual) steps that represent some action.
- Workflows can be themselves named and encapsulated, thus simplifying representations of larger workflows and easing the burdens on making requests therefrom.

Scientific (High-Performance) Processing and Automation

- Our Distributed Processing System (DPS) provides the core capability to schedule and automate scientific (or non-scientific) processing
- Turn-Key Grid - DPS is the world's first, and so far only turn-key grid computing system; It is possible to install and have an operating Grid Computing system processing your computational functions (Processes) in about an hour.
- No programming is required to use DPS, however with use of BigSur's DPS Process Templates and very modest programming, full and robust utilization of BigSur's features is also possible in only another hour or so.
- Processing may occur on any system, anywhere, so long as a connection to an appropriately licensed BigSur server system is available.
- The Demand-Engine provides processing capability whenever it is desired - "on demand."
- The Eager-Engine provides processing automatically, as soon as "parent" data objects become available
- Load-balancing and specialized compute resource control is easily managed from any client node.
- "Reaching through firewalls" to ensure security for all compute and storage resources is easily accomplished

Distributed Objects - Our meta-data provides an exceptionally robust distribution capability

- Objects may reside on any system, and in any form
- Objects may have data or not, as desired
- Object data may be in multiple parts or may be made up of set(s) of other objects
- Processing ("lineage") meta-data is fully distributed too - both local and remote inputs to processing are recorded in an identical fashion
- Copies of objects are also easily managed, as the canonical home location is known
- Automated transport - When called upon, the system automatically discovers where data objects are in the distributed environment, and with sufficient access privileges configured in advance, the system will transport data to environments that need it.
- Local Cache management helps reduce network transport costs through a mechanism that manages a local object cache.

Resource Discovery - Robust meta-data provides for easy searching of objects, functions, and relationships:

- Meta-data describes associations and relationships so discovering the data and processing algorithms of other researchers is easy, as is discovering what data were produced by what Processes (algorithms), etc.
- Special access handle generation creates a unique handle (name) for internal identification of objects which also provide for the ultimate in data access performance due to the ability to use exact-match fetches.
- Special associative mechanism permits easy and natural groupings of objects to fit researchers needs and greatly aiding searches
- For spatial data, R-Tree indexing provides for the worlds fastest geo-spatial searches (only available on certain RDBMSes)
- Natural and powerful typing system permits researchers to create their own data-types and object-hierarchies

Publishing System - Provides for publishing of results as they are generated

- Researchers may publish to any collaborative site by any means desired - email, FTP, etc., including other database systems.

- Publishing decisions may be based upon object type, process which ran, contact names, etc

Knowledge Capture - Beyond mere automation, BigSur provides for retaining knowledge when personnel move on, and for sharing knowledge between individuals by providing more than just computational features:

- Description, Purpose and Supplemental meta-data are available for every primary aspect of the entire enterprise
- Relationships are tracked (or are trackable) at every level, both technical and involving people.

Robust Security - BigSur's sensible, flexible, and extensible security system provides both high and low granularity security for all components, with defaults.

Multiple-Dimension Objects - Multi-Dimension Arrays (only available on certain RDBMSes)

- Our specialized MDA code allows as many descriptions of a single object as multiple dimension arrays may be defined thus improving access to the object by making natural descriptions available for selections of data
- Includes "snip" or "slice" functions for extracting subsets so that only minimal amounts of data need be moved through a network

Features

Please note that while BigSur's Client and Server installations are distinctive, Server installations also include a Client installation. Aside from what one could call "system internals", any distinction between Clients and Servers is effectively lost from the point of view of the feature set. Thus, we do not distinguish among installation types here.

(For more information on the distinctions between Client installations and Server installations, please see section "Ordering Options" below.)

Distributed Processing System - DPS

The following are the main components of DPS:

- **DaemonMaster™** - responsible for starting and stopping other Daemons on client hosts.
- **DemandEngine™** - responds to immediate demands for processing.
- **EagerEngine™** - assists in the production of standard data products that are made before anyone has requested (demanded) them.
- **FrontEndDaemon™** - provides a flexible means of connecting clients to servers for applications which cannot use the STDB API.

The following components also use DPS: (more information on each is provided below)

- **Archivist™** Client
- **ScienceMaster™** - provides a GUI for administrative access to BigSur installations (one instance comes with each STDB server)
- **ScienceInspector™** - provides non-administrator GUI access to BigSur installations. *NOTE: At present, ScienceInspector is only available on special request.*

Data and Meta-data Management

Object-Management

BigSur includes a full Object Management System which is very similar to but distinctive from the so-called Object-Oriented Model. Unlike OOM, BigSur's approach is "loosely-typed" wherein hierarchies may be developed, but there's no type enforcement, and functions or methods that operate on types are not compelled to handle "descendant" objects as would be expected in a true OOM system. This provides significant advantages in that the system doesn't break when inappropriate object definitions are made, with the modest cost that the user is responsible for ensuring the adherence to the OO model (via testing), if desired.

Anything can be an object, including fictitious objects that are merely conceptual ideas, such as a particular orbit of a satellite. Technically, an object is represented as meta-data, and any actual data for the object is then associated with that meta-data. Attributes of data-types are often extended within BigSur such that particular built-in types of objects have additional data or meta-data while others do not.

Many times users manage their scientific data wholly in "external data files," and BigSur has no problems with this whatsoever. However, occasionally it is desired to have a bit more direct access available. For these cases, through a feature called Extension Management, the object-extension mechanism described above extends full object support to additional, non-built-in object types such as those available through GeoDb, described elsewhere. By properly describing these additions to BigSur via Extension Management, other data-types, whether developed by customers or by us, gain the full power and capability of the features of BigSur. Thus, BigSur can be extended to efficiently handle every type of data in every type of circumstance.

Object-Data

Data that represents an Object may be of any form and may comprise sets of other Objects, or be managed in multiple parts. Conventions ("rules") ease management of complex Objects.

- There are no restrictions on the types of data that may be a part of an object or express that object, and indeed, there are multiple expressions of "type" to aid in the actual computational work of automation, including:
    - ObjectType - this is the overall, high-level label for an object, for example, an image.
    - DataType - this is the specific detailed type that the data is in, possibly called its format, for example a .gif image versus a .jpg image.
    - ValueType - this is most often used as an expression of how data is being transmitted, such as "filename", or "text."
- Data may be in "flat files" stored in the file system or may be in database tables in a database.
- A multi-part Object may contain multiple copies of the same data stored in different formats or each part may contain different contents, as appropriate to its type.
- An Object may be made up of other objects.

Schema and Extensions

All of the following either provide their own schema to the environment, or formally use Extension Management described elsewhere in this document:

- **MedDB™** schema extension to STDB - provides schema support for medical applications. *Note: MedDB™ has been withdrawn while it is being re-written.*
- **GeoDB™** schema extension to STDB - provides FGDC and SAIC Geographic Metadata Standard compliant schema extension for STDB.
- **The Publisher™** - provides automated movement of data and meta-data between BigSur server installations. (Requires DPS.)
- **The Archivist™** Server - provides schema extensions and executable utilities for managing data archives.
- **DbCryptor™** - provides a sophisticated encryption service enabling point-to-point, on-the-fly database connectivity. (Requires SSH)

The Publisher

The Publisher is an automation package that moves data objects and compute processes, under careful user control, to cooperating installations, either other BigSur installations or potentially any cooperating system for which a reasonable schema map can be identified.

A separate SPD is being developed for The Publisher. In the mean time, please ask your Science Tools representative for a full description.

The Archivist

The Archivist was originally developed to address the specific needs of large compute-resource providers in helping manage the work-space granted to various research efforts. Of particular interest was the identification of duplicates and managing their backup and restore without undue expenditure of resources.

It was subsequently recognized that larger scale optimizations in backup and restore could be realized from further development of this minimal-resource paradigm and a fully-fledged system emerged in which distinctions are made based on unique digital assets and their various, often multiple, locations within a file hierarchy. Key features include multiple-copy management, identification of the speeds of various storage media and identification (and use of) the most efficient mechanisms for restoration.

A separate SPD is being developed for The Archivist. In the mean time, please ask your Science Tools representative for a full description.

DbCryptor

DbCryptor is a product which helps provide database access in otherwise challenging and seemingly impossible security domain. It provides a way through firewalls and gateways in the very most strict environments by harnessing open-source, well trusted encryption technologies (SSL) in novel ways to route applications through to their database servers. When BigSur version 4.0 is released, DbCryptor will provide this same capability to connect applications with their application server.

A separate SPD is being developed for DbCryptor. In the mean time, please ask your Science Tools representative for a full description.

ScienceMaster

ScienceMaster is The BigSur System's™ administration graphical user interface (GUI) tool, and it can also serve as an application management environment for highly-technical end-users.

ScienceMaster includes over 77 frames (application windows) that provide a single-sign-on control capability over an entire BigSur environment. With it, virtually every aspect of BigSur administration can be performed from one user session.

ScienceMaster also includes an incredible "visualization" capability in which data objects of every type can be visualized, audible-ized, or otherwise observed in their original forms. It does this by providing an easy harnessing mechanism to other digital tools; if there's a way to "view" something, ScienceMaster can launch the appropriate technology, on the fly, so that results can be observed immediately.

A separate SPD is being developed for Science Master. In the mean time, please ask your Science Tools representative for a full description.

PolyglotSQL™ database dialect translation

PolyglotSQL™ provides both on-the-fly SQL dialect translation during live database connections and static translation of files containing either DDL (Data Definition Language - schema related) or DML (Data Manipulation Language) statements. 5 SQL dialects are presently supported, and virtually any database system that supports SQL can be configured.

SQL - the Standard Query Language, developed by IBM, and now an international standard, is used by nearly every database management system now in existence.

- **Dialect translation services** - BigSur translates any supported SQL dialect into any other supported dialect, on-the-fly including:
  - DB2
  - Informix
  - Ingres
  - Oracle
  - Postgres (AKA PostgreSQL)

  Only translatable features are translated.
- **Client-side Journaling** - this feature provides the ultimate in reliability, data-safety, portability, and auditability. When enabled, all update access to the server is logged on client-side files. The resulting logs may be replayed against any supported database, and they are automatically uploaded to the server system, when possible.
- **Automatic Re-Try** - Applications no longer have to worry about catching basic database access errors unrelated to their own SQL statements because BigSur will catch any exceptions, evaluate the exception to see if it is a permanently fatal error or if there may be value in re-trying the database access, and doing so if appropriate.
- **Automatic Reconnection**
- **Dual-Mode** provides for the division of read and write activity so that reads can happen on one connection while write activity happens on another. This permits the harnessing of extreme-performance database systems like PostgreSQL 9.2 with it's cascading binary replication, for the ultimate in performance. Writes are performed only on the master, while read access is performed on any copy.
- Helper programs that work with PolyglotSQL™:
  - **ConvertSchema** - this program does static conversion of SQL DDL statements from one dialect to another.
  - **ConvertSql** - This program does static conversion of SQL DML statements from one dialect to another.
  - **LoadJournal** - This program reads BigSur Client-Side Journaling log files and executes the statements contained therein against a BigSur Server installation.
  - **LoadSchema** - This program executes SQL DDL statements against a live database instance, converting as necessary.
  - **LoadSql** - This program executes SQL DML statements against a live database instance, converting as necessary.
  - **UnloadDb** - This program fetches the contents of a single BigSur Server table and writes it to an output file either in raw format or in SQL. NOTE: Output of an entire BigSur Server database is planned for a future release of UnloadDb but is not yet implemented.
  - **UnloadJournal** - This program reads a BigSur Server's journal that exists in the server's database and creates a file-based journal that can be read by LoadJournal.

Hardware Requirements

N/A

For data safety, use of Virtual Machine technology on server installations is not recommended due to data safety issues of supporting database systems on VMs.

Software Requirements

Licensee must provide an installation of one of the following supported Relational Database Management Systems, an installation of Oracle Corporation's JAVA version 1.4 or younger (either JRE or SDK as desired), and must ensure that a version compatible JDBC driver (provided by the RDBMS) is configured for BigSur use.

Supported RDBMSes:

DB2, Informix, Oracle, Postgres, Sybase (certification of Ingres is pending)

Ordering Information

When ordering, for each instance of The BigSur System™, please cite:

- The SPD number
- The number of bits (eg 32 or 64) of the CPU
- Operating system type
- The System Administrator's account name (case sensitive)
- The system's full host name, including fully qualified domain name
- The "MAC" address of any network hardware that is expected to remain attached to the system for its service life, or at least during the length of license.
- Whether the installation is client-only or a server

What options are desired from the Ordering Options listed below. Also, please see "Software Licensing" below for further information.

Ordering Options

Client Installations - BigSur System client installations have the following options:

- **Distributed Processing System**™ (DPS) Client
  - **DemandEngine™** - responds to immediate demands for processing.
  - **EagerEngine™** - assists in the production of standard data products that are made before anyone has requested (demanded) them.
  - **FrontEndDaemon™** - provides a flexible means of connecting clients to servers for applications which cannot use the STDB API.
- **Archivist™ Client**
- **DbCryptor™** - provides flexible and secure connections that can pierce firewalls and gateways in a secure, high performance manner.
- **ScienceMaster™** - provides a GUI for administrative access to BigSur installations (one instance comes with each STDB server)
- **ScienceInspector™** - provides non-administrator GUI access to BigSur installations.
- **Individual Client** - Required for each user when seat based licensing is in effect.
- **Individual Developer** - Provides Java API tracing, and enables application development features (per STDB instance).

Server Installations - BigSur System server installations have the following options in addition to all of the Client Installation options:

- **Distributed Processing System** (DPS) - this license permits DPS processes to connect to the server in question. Note that a client-side license is also required for all clients connecting to the server. A DPS server-side license includes a client license for the server system.
- **MedDB™** schema extension to STDB - provides schema support for medical applications. *Note that the MedDb schema extension is presently being re-written to include support for medical record management. There is presently neither fee for nor commitment to capabilities provided by MedDB.*
- **GeoDB™** schema extension to STDB - provides FGDC and SAIC Geographic Metadata Standard compliant schema extension for STDB.
- **The Publisher™** - provides automated movement of data and meta-data between BigSur server installations. (Requires DPS.)
- **The Archivist™ Server** - provides schema extensions and executable utilities for managing data archives.

- **DbCryptor™** - provides a sophisticated encryption service enabling point-to-point, on-the-fly database connectivity. (Requires SSH)
- The **mode** the server is permitted to operate in may be:
  - **node-locked** - Connections from other systems are not permitted.
  - **site** - All properly licensed computers from the same site are permitted to connect.
  - **domain** - All computers in the same network domain are permitted to connect.
  - **world-wide** - All computers are permitted to connect.

Software Licensing

Every system running BigSur must be licensed as either a client or server. Identification of such systems must include the system's full internet name, where it exists, internal network name when no externally accessible network name exists, and optionally its primary MAC address (especially useful for mobile computers).

Licenses options include:

- Mode: node-locked, domain, site, world-wide
- Seats: per human being

Server installations include a client license for the server system. Each instance of a BigSur server installation may be licensed to accept three forms of clients: “System”, indicating that only clients from the server host will be accepted, “Project”, which indicates that only clients from the licensed project scope will be accepted, or “Worldwide”, which indicates that there are no limits placed upon connectivity. A client installation may support any number of Individual Client computers, however, each needs to have a license.

License Management Facility

License management is integrated with the STJC utility and may also be enforced through the BigSur API.

Software Product Services

Technical Support

In consideration of the appropriate Technical Support Fee(s) paid by Customer and Customer’s agreement to meet the responsibilities set forth below, Science Tools shall provide to Customer Technical Support for Product as appropriate to the selected level of service. Available levels of service are presently “Basic Support”, “Enhanced Support” and “Premium Support”. Science Tools makes the following specific commitments in response to requests for Technical Support from Customer:

- Science Tools shall make best effort, for bona fide defect or problem report, to develop a software fix or workaround in a timely fashion;
- Science Tools shall assist Customer Contact in diagnosing errors and malfunctions which occur when the Developed Product is used by Licensee;
- Science Tools shall provide support services to Customer to attempt to correct diagnosed errors and malfunctions.

Basic Support

- Hours of service are between 8:00 a.m. and 5:00 p.m., Pacific Time, Monday through Friday, holidays excluded.
- Technical Support will be provided via email.
- Accept/acknowledge support request response time: within three (3) business days of receipt of request
- Report on Science Tools’ analysis and/or assessment, including options and estimated time for resolution: five (5) business days from acknowledgement of support request

Enhanced Support

- Hours of service are between 8:00 a.m. and 5:00 p.m., Pacific Time, Monday through Friday, holidays excluded.
- Technical Support will be provided via email or telephone.
- Accept/acknowledge support request response time: within one (1) business days of receipt of request
- Report on Science Tools’ analysis and/or assessment, including options and estimated time for resolution: three (3) business days from acknowledgement of support request

Premium Support

- Hours of service are “24 X 7 X 365” with best-effort four hour response time.
- for support calls identified by Customer as “urgent,” to provide a proposed resolution or response within one (1) business day.
- An assigned technical support person.

Customer responsibilities and acknowledgements:

- Customer shall designate one Customer Contact (“contact”) person and one alternate.

- Customer acknowledges that Technical Support is not a substitute for training or consulting services.

Product Updates

Update service is available for this product.

Standard Terms

License is "30 day net" from date of invoice with product ship following payment. Support fees, terms and conditions are provided in a separate document.