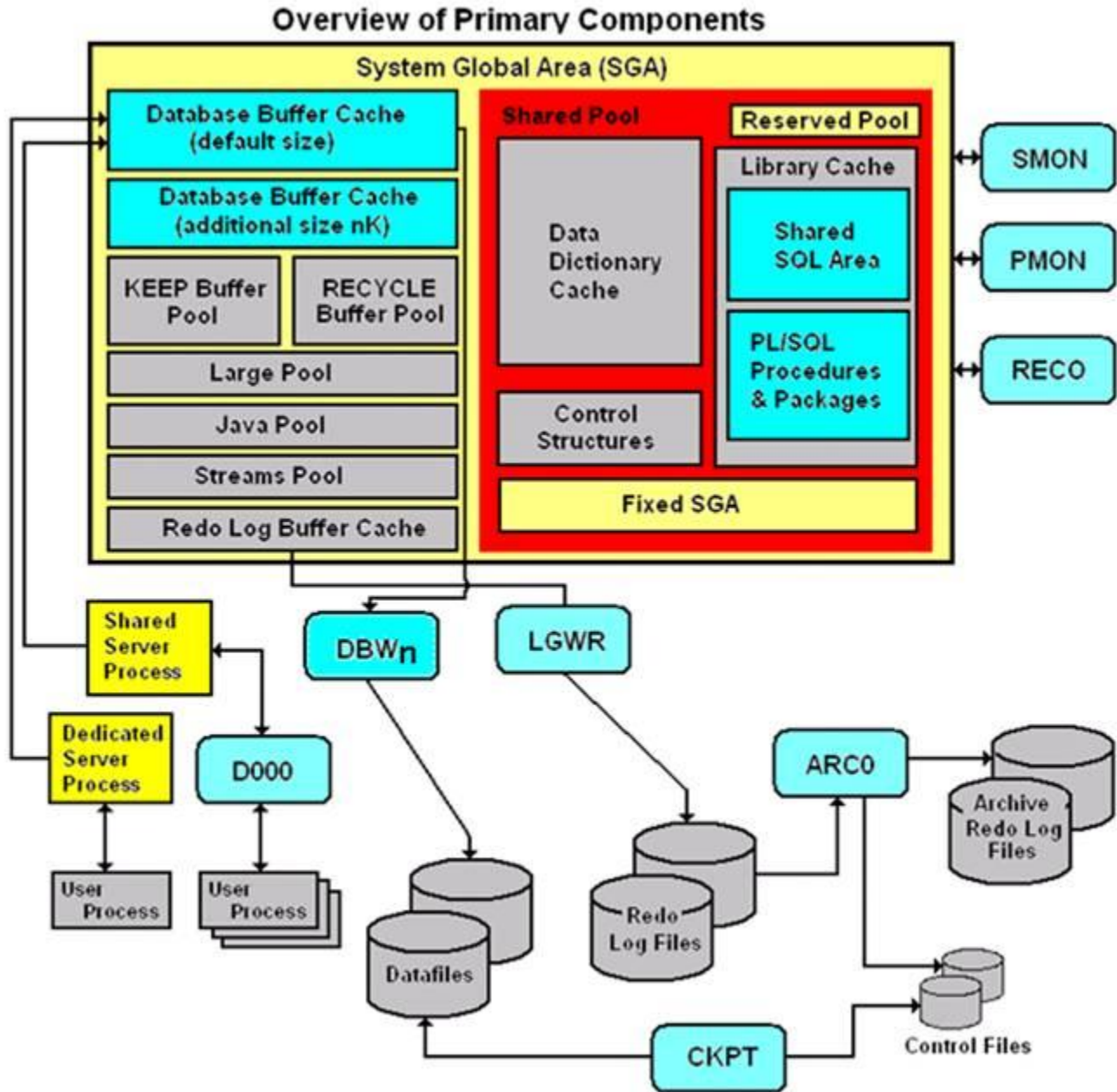


Objectives

These notes introduce the Oracle server architecture. The architecture includes physical components, memory components, processes, and logical structures.

Primary Architecture Components



The figure shown above details the Oracle architecture.

Oracle server: An Oracle server includes an **Oracle Instance** and an **Oracle database**.

- An Oracle database includes several different types of files: datafiles, control files, redo log files and archive redo log files. Oracle database also contains User and server processes.
- The database server must manage large amounts of data in a multi-user environment.
- The server must manage concurrent access to the same data.
- The server must deliver high performance. This generally means fast response times.

Oracle instance: An Oracle Instance consists of two different sets of components:

- The first component set is the set of **background processes** (PMON, SMON, RECO, DBW0, LGWR, CKPT, D000 and others).
 - These processes perform input/output and monitor other Oracle processes to provide good performance and database reliability.
- The second component set includes the **memory structures** that comprise the Oracle instance.
 - When an instance starts up, a memory structure called the System Global Area (SGA) is allocated.
 - At this point the background processes also start.
- An Oracle Instance provides access to one and only one Oracle database.

Oracle database: An Oracle database consists of files.

- Sometimes these are referred to as operating system files, but they are actually **database files** that store the database information that a firm or organization needs in order to operate.
- The **redo log files** are used to recover the database in the event of application program failures, instance failures and other minor failures.
- The **archived redo log files** are used to recover the database if a disk fails.
- Other files not shown in the figure include:
 - The required **parameter file** that is used to specify parameters for configuring an Oracle instance when it starts up.
 - The optional **password file** authenticates special users of the database – these are termed **privileged users** and include database administrators.
 - **Alert** and **Trace Log Files** – these files store information about errors and actions taken that affect the configuration of the database.

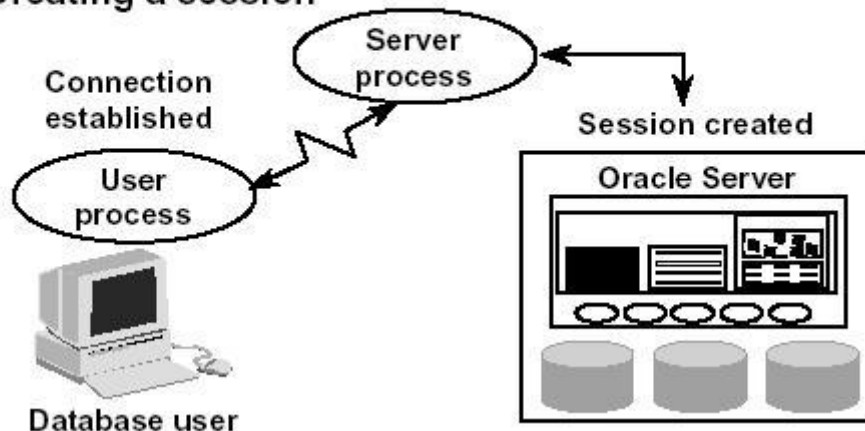
User and server processes: The processes shown in the figure are called **user** and **server processes**. These processes are used to manage the execution of SQL statements.

- A **Shared Server Process** can share memory and variable processing for multiple user processes.
- A **Dedicated Server Process** manages memory and variables for a single user process.

Connecting to an Oracle Instance – Creating a Session

Connecting to an Oracle Instance:

- Establishing a user connection
- Creating a session



System users can connect to an Oracle database through SQLPlus or through an application program. This connection enables users to execute SQL statements.

The act of connecting creates a communication pathway between a user process and an Oracle Server. As is shown in the figure above, the User Process communicates with the Oracle Server through a Server Process. The User Process executes on the client computer. The Server Process executes on the server computer, and actually executes SQL statements submitted by the system user.

The figure shows a one-to-one correspondence between the User and Server Processes. This is called a **Dedicated Server** connection. An alternative configuration is to use a **Shared Server** where more than one User Process shares a Server Process.

Sessions: When a user connects to an Oracle server, this is termed a session. The **User Global Area** is session memory and these memory structures are described later in this document. The session starts when the Oracle server validates the user for connection. The session ends when the user logs out (disconnects) or if the connection terminates abnormally (network failure or client computer failure).

A user can typically have more than one concurrent session, e.g., the user may connect using SQLPlus and also connect using Internet Developer Suite tools at the same time. The limit of concurrent session connections is controlled by the DBA.

If a system user attempts to connect and the Oracle Server is not running, the system user receives the **Oracle Not Available** error message.

Physical Structure – Database Files

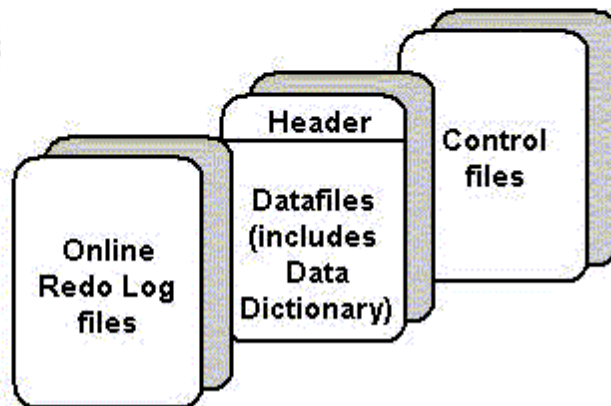
As was noted above, an Oracle database consists of physical files. The database itself has:

- **Datafiles** – these contain the organization's actual data.
- **Redo log files** – these contain a chronological record of changes made to the database, and enable recovery when failures occur.
- **Control files** – these are used to synchronize all database activities and are covered in more detail in a later module.

Physical Structure

The physical structure includes three types of files:

- **Control files**
- **Datafiles**
- **Redo log files**



Memory Management and Memory Structures

Oracle Database Memory Management

Memory management - focus is to maintain optimal sizes for memory structures.

- Memory is managed based on memory-related [initialization parameters](#).
- These values are stored in the init.ora file for each database.

Three basic options for memory management are as follows:

- **Automatic memory management:**
 - DBA specifies the target size for instance memory.
 - The database instance automatically tunes to the target memory size.
 - Database redistributes memory as needed between the SGA and the instance PGA.
- **Automatic shared memory management:**
 - This management mode is partially automated.
 - DBA specifies the target size for the SGA.
 - DBA can optionally set an aggregate target size for the PGA or managing PGA work areas individually.
- **Manual memory management:**
 - Instead of setting the total memory size, the DBA sets many initialization parameters to manage components of the SGA and instance PGA individually.

The memory structures include three areas of memory:

- System Global Area (SGA) – this is allocated when an Oracle Instance starts up.
- Program Global Area (PGA) – this is allocated when a Server Process starts up.
- User Global Area (UGA) – this is allocated when a user connects to create a session.

System Global Area

The **SGA** is a read/write memory area that stores information shared by all database processes and by all users of the database (sometimes it is called the **Shared Global Area**).

- This information includes both organizational data and control information used by the Oracle Server.

- The SGA is allocated in memory and virtual memory.
- The size of the SGA can be established by a DBA by assigning a value to the parameter **SGA_MAX_SIZE** in the parameter file—this is an optional parameter.

The SGA is allocated when an Oracle instance (database) is started up based on values specified in the initialization parameter file (either PFILE or SPFILE).

The SGA has the following mandatory memory structures:

- Database Buffer Cache
- Redo Log Buffer
- Java Pool
- Streams Pool
- Shared Pool – includes two components:
 - Library Cache
 - Data Dictionary Cache
- Other structures (for example, lock and latch management, statistical data)

Additional optional memory structures in the SGA include:

- Large Pool

The **SHOW SGA** SQL command will show you the SGA memory allocations.

```
SQL> connect system/ system as sysdba
Connected.
```

```
SQL> show sga
```

```
Total System Global Area 1610612736 bytes
Fixed Size                  2084296 bytes
Variable Size               1006633528 bytes
Database Buffers           587202560 bytes
Redo Buffers                 14692352 bytes
```

Early versions of Oracle used a **Static SGA**. This meant that if modifications to memory management were required, the database had to be shutdown, modifications were made to the **init.ora** parameter file, and then the database had to be restarted.

Oracle 11g uses a **Dynamic SGA**. Memory configurations for the system global area can be made without shutting down the database instance. The DBA can resize the Database Buffer Cache and Shared Pool dynamically.

The size of the SGA cannot exceed the parameter **SGA_MAX_SIZE** minus the combination of the size of the additional parameters, **DB_CACHE_SIZE**, **LOG_BUFFER**, **SHARED_POOL_SIZE**, **LARGE_POOL_SIZE**, and **JAVA_POOL_SIZE**.

Memory is allocated to the SGA as contiguous virtual memory in units termed granules. Granule size depends on the estimated total size of the SGA, which as was noted above, depends on the SGA_MAX_SIZE parameter. Granules are sized as follows:

- If the SGA is less than **1 GB** in total, each granule is **4 MB**.
- If the SGA is greater than **1 GB** in total, each granule is **16 MB**.

Granules are assigned to the Database Buffer Cache, Shared Pool, Java Pool, and other memory structures, and these memory components can dynamically grow and shrink. Using contiguous memory improves system performance. The actual number of granules assigned to one of these memory components can be determined by querying the database view named **V\$BUFFER_POOL**.

Granules are allocated when the Oracle server starts a database instance in order to provide memory addressing space to meet the SGA_MAX_SIZE parameter. The minimum is 3 granules: one each for the fixed SGA, Database Buffer Cache, and Shared Pool. In practice, you'll find the SGA is allocated much more memory than this. The SELECT statement shown below shows details of the granules allocated to BUFFER_POOL.

```
SELECT name, block_size, current_size, prev_size, prev_buffers
FROM v$buffer_pool;
```

NAME	BLOCK_SIZE	CURRENT_SIZE	PREV_SIZE	PREV_BUFFERS
-----	-----	-----	-----	-----

DEFAULT	8192	560	576	7
1244				

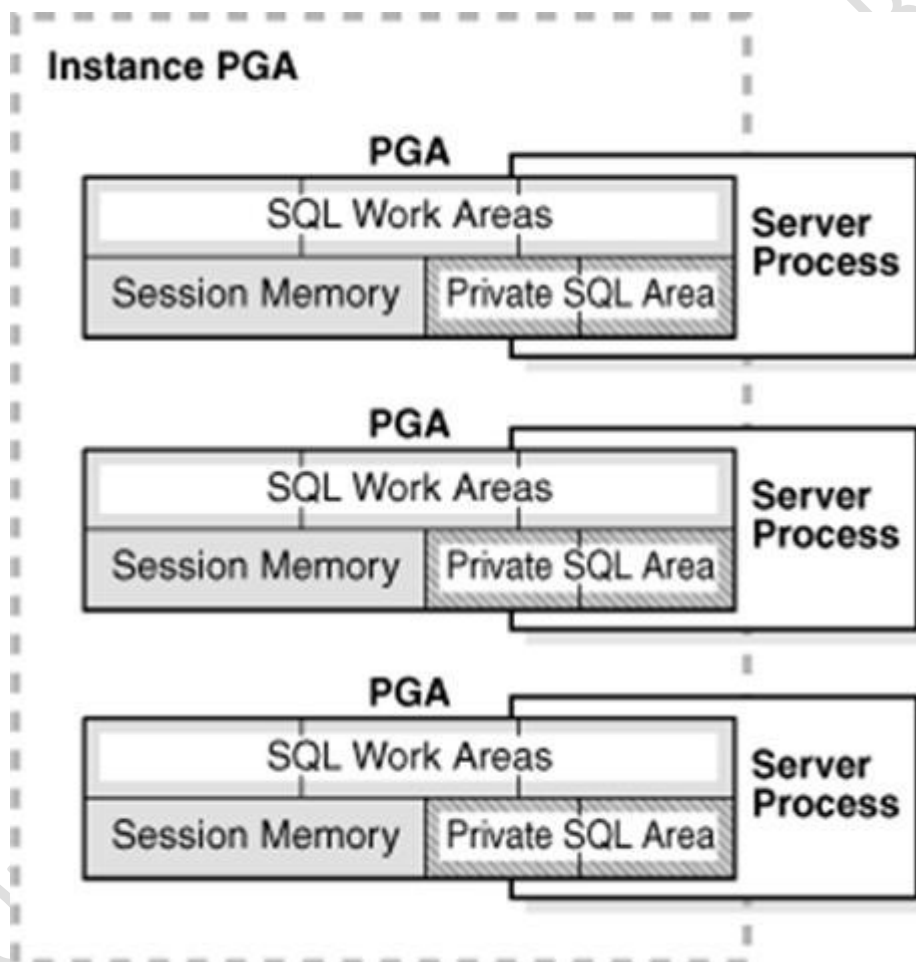
Program Global Area (PGA)

A **PGA** is:

- a **nonshared** memory region that contains data and control information exclusively for use by an Oracle process.

- A PGA is created by Oracle Database when an Oracle process is started.
- One PGA exists for each **Server Process** and each **Background Process**. It stores data and control information for a single **Server Process** or a single **Background Process**.
- It is allocated when a process is created and the memory is scavenged by the operating system when the process terminates. This is **NOT** a shared part of memory – one PGA to each process only.
- The collection of individual PGAs is the **total instance PGA**, or **instance PGA**.
- Database initialization parameters set the size of the instance PGA, not individual PGAs.

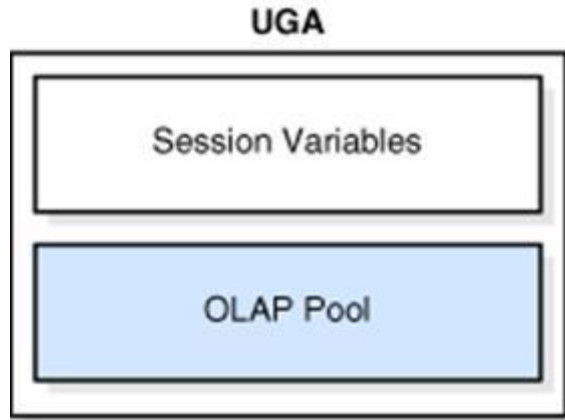
The **Program Global Area** is also termed the **Process Global Area (PGA)** and is a part of memory allocated that is outside of the **Oracle Instance**.



The content of the PGA varies, but as shown in the figure above, generally includes the following:

User Global Area

The **User Global Area** is session memory.



A session that loads a PL/SQL package into memory has the **package state** stored to the UGA. The **package state** is the set of values stored in all the package variables at a specific time. The state changes as program code the variables. By default, package variables are unique to and persist for the life of the session.

The **OLAP page pool** is also stored in the UGA. This pool manages OLAP data pages, which are equivalent to data blocks. The page pool is allocated at the start of an OLAP session and released at the end of the session. An OLAP session opens automatically whenever a user queries a dimensional object such as a cube.