Oracle RAC Architecture

By Ahmed Baraka

Objectives

In this lecture, you should learn how to describe the following:

- Oracle RAC hardware requirements
- Shared storage configuration options
- Oracle Grid Infrastructure components
- Oracle Clusterware network configuration
- Oracle RAC connectivity cycle
- Oracle RAC database files locations

Oracle RAC History

- Oracle Parallel Server (OPS)
  - Oracle 6.2 on VAX/VMS.
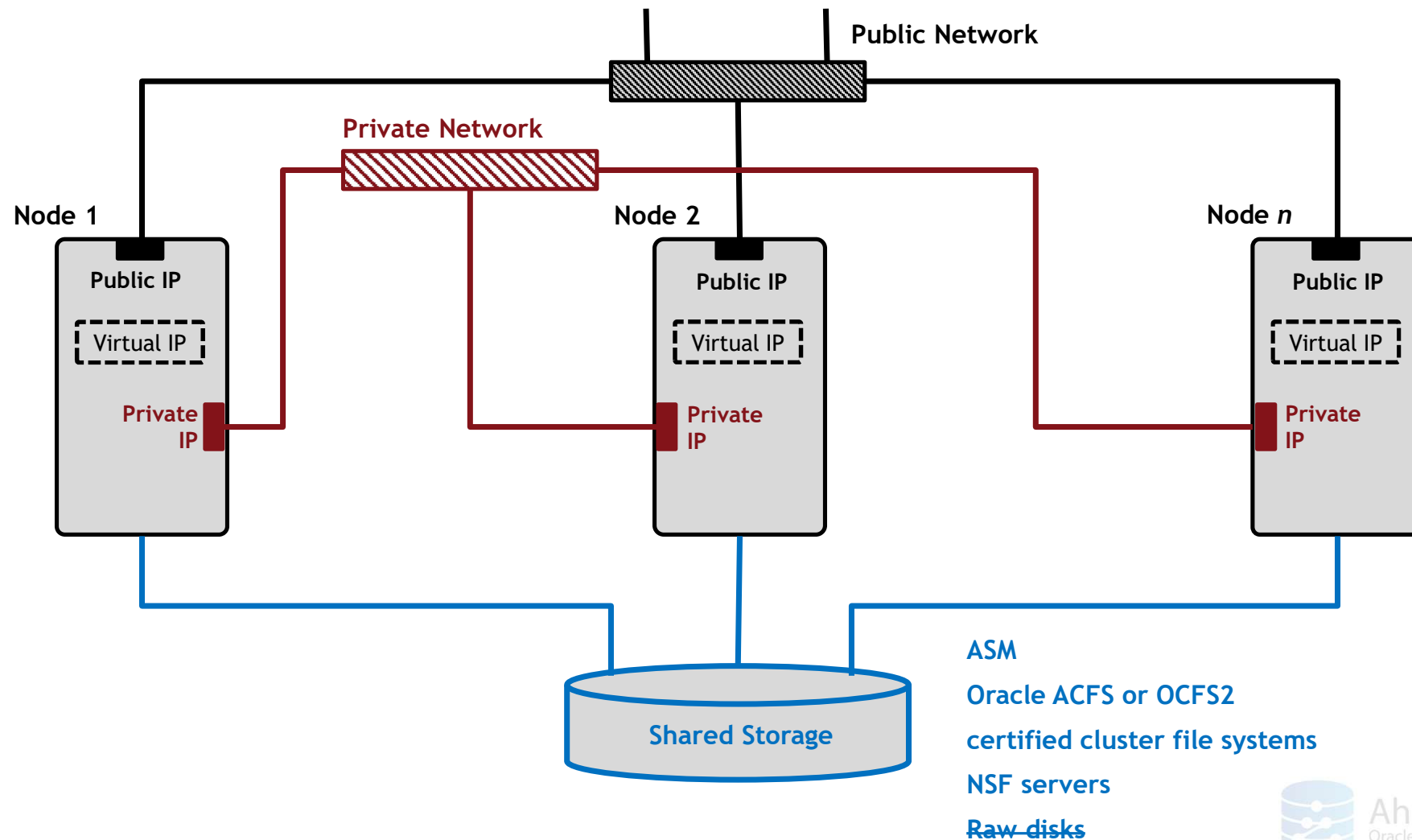- Oracle Real Application Cluster (RAC)
  - Oracle 9i

Oracle RAC: Software Stack

- Oracle Grid Infrastructure
  - Clusterware
    - Manage the whole cluster
    - Monitor the availability of the resources
    - Controls the resources startup order
    - Can be programmed for user applications
  - ASM
- Oracle Database

Oracle RAC Basic Architecture

Public Network

Private Network

Node 1

Public IP

Virtual IP

Private IP

Node 2

Public IP

Virtual IP

Private IP

Node _n_

Public IP

Virtual IP

Private IP

Shared Storage

ASM

Oracle ACFS or OCFS2

certified cluster file systems

NSF servers

~~Raw disks~~

Oracle RAC Architecture Overview

- Every server that is part of the RAC is called node.
- Clients connect to the database through public network.
- Nodes connect to each other via private network (called interconnect)
- Database datafiles are stored in shared storage.
- A RAC database has multiple instances. Every instance is running in a single node.
- For a database to be available to the clients, at least one instance should be up and running.

Oracle RAC: Hardware Requirements

- Two or more servers
  - Similar hardware architecture and same OS

- Two Network cards installed into every node:
  - Public network: connected to the clients
  - Private network: interconnect network among the nodes
  - Redundant private interconnect configuration is highly recommended

- Shared storage
  - High speed storage network required, such as 16GbE FC
  - Redundant connections to the storage with multipathing configuration

Oracle RAC: Memory Requirements

- If the same workload is kept on an instance, when migrating a single-instance database to RAC, more memory is need:
  - 10% more buffer cache
  - 15% more shared pool
  - Memory used by ASM and Clusterware processes
- Memory consumed by client sessions will be distributed among the instances

Shared Storage Options

- ASM (the recommended option)
  - Volume manager and file system for Oracle Databases
  - More efficient for Oracle Databases
  - The only supported option with Oracle Standard Edition
- Oracle ACFS
  - Extends ASM functionality to support all files.
  - Provides: dynamic FS resizing, direct access to ASM DG, I/O parallelism
- OCFS2 (Oracle Cluster File System 2)
  - Open source general-purpose file system developed by Oracle
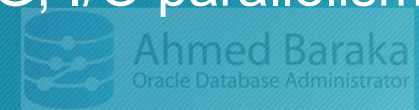
Shared Storage Options (cont)

- Third-party Certified cluster file systems
  - Examples: Veritas Storage Foundation Cluster File System (SFCFS) for Oracle RAC

- NFS Server
  - Only NFS protocol version 3 (NFSv3) is currently supported

Oracle RAC: Engineered Systems

- Oracle Database Appliance (ODA)
  - Pros: tested for maximum performance, easy management
  - Cons: only two nodes
- Oracle Exadata Database Machine
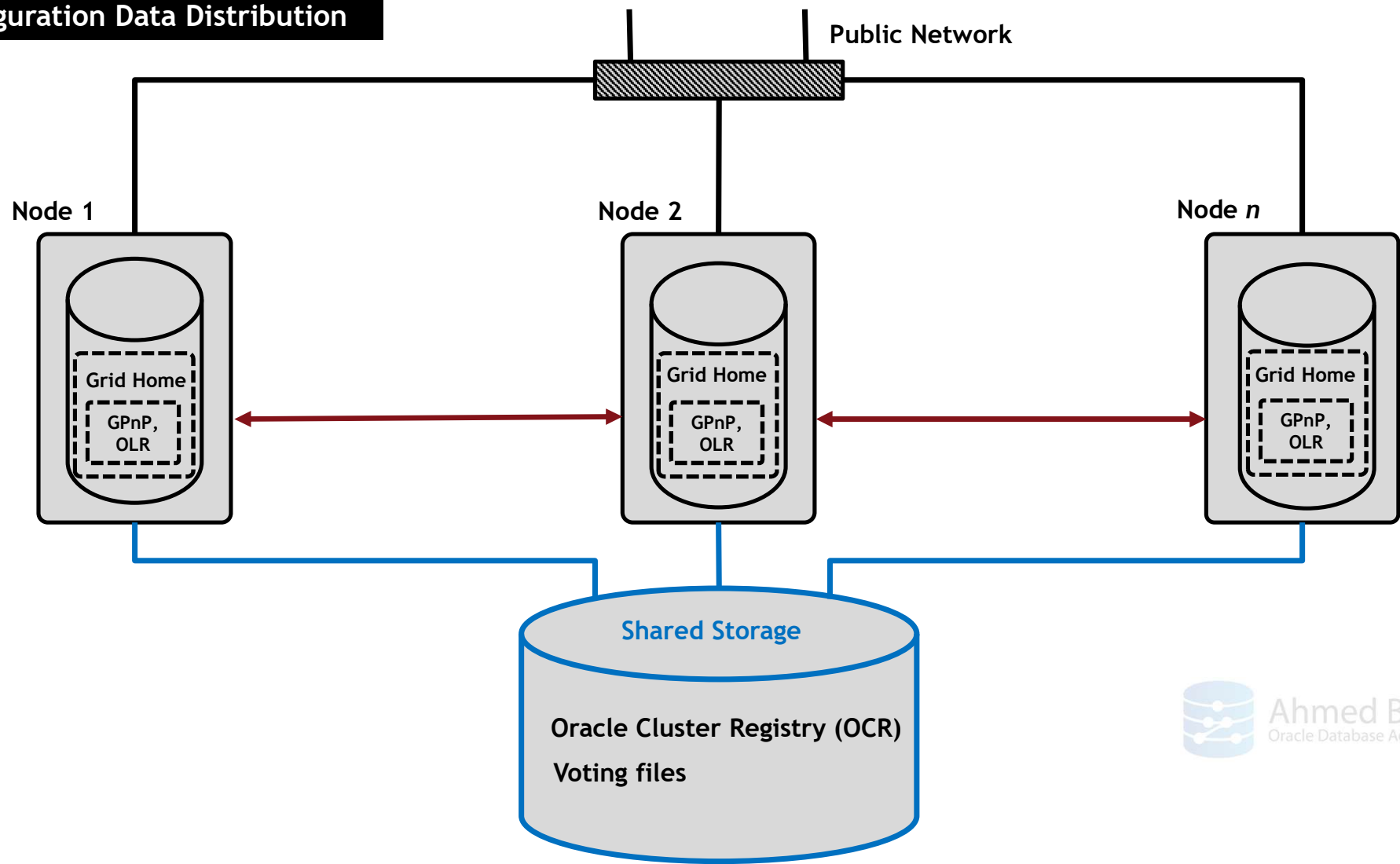
About Oracle Grid Infrastructure

- Provide clustering features (clusterware) plus ASM
- It must be installed in its own, local home
- Usually owned by a dedicated OS user "`grid`"
- Cluster Ready Service (CRS) resources:

  - ASM instances
  - ASM Diskgroup
  - Virtual IP (VIP) services
  - Single Client Access Name (SCAN)

  - SCAN listener
  - database services
  - Oracle Net listener
  - Oracle Notification Service (ONS)

About Oracle Clusterware

- Requires shared storage to store:
    - Voting files for node membership
    - Oracle Clusterware Registry (OCR) for cluster configuration information
    - They must be saved in a high-redundant storage

- Further configuration data stored in two local files:
    - Oracle Local Registry (OLR): metadata for the local node
    - Grid Plug and Play (GPnP) profile: network profile and VD

- It used private interconnect network to carry the heartbeat communication

Oracle Clusterware
Configuration Data Distribution

Public Network

Node 1

Node 2

Node *n*

Grid Home

GPnP,
OLR

Grid Home

GPnP,
OLR

Grid Home

GPnP,
OLR

Shared Storage

Oracle Cluster Registry (OCR)

Voting files

Clusterware: CRS Daemons and Services

Daemon/Service	Process Name	Description
CRS	`CRSD`	Start, stop, monitor, and failover operations on resources
CSS Service	`ocssd.bin`	Monitors the node membership in the cluster and updates the node status information in VD
CSS Agent	`cssdagent`	Monitors, starts, and stops the CSS.
CSS Monitor	`cssdmonitor`	It works with the CSS agent to provide data integrity in the clusterware. Base on the CPU startvation or OS status, it may reboot the node.
CTSS	`octssd.bin`	Maintains time synchronization among all the nodes in the cluster
EVM	`evmd.bin`	Publishes events to all the cluster nodes
ONS	`ons`	Manages the Fast Application Notification (FAN) events

Clusterware: CRS Daemons and Services (cont)

Daemon/Service	Process Name	Description
Oracle Agent	`oraagent`	Managing ohasd resources owned by `oracle`
Oracle Root Agent	`orarootagen`	Managing ohasd resources owned by `root`
Oracle ASM		Managing the shared storage and disk volumes

Note: this list is not all the Clusterware processes.

Oracle Clusterware Network Configuration

- Oracle Grid Naming Service (GNS)
  - A subdomain is delegated in the DNS server for the cluster
  - You only define the GNS server VIP in the DNS
  - DHCP is supported only in Linux systems

- Static Configuration:
  - No subdomain is delegated in the DNS server
  - All names and addresses are resolved by DNS:
    - A SCAN name to resolve to three static IP addresses for the cluster
    - A public and virtual IP names and addresses for each node
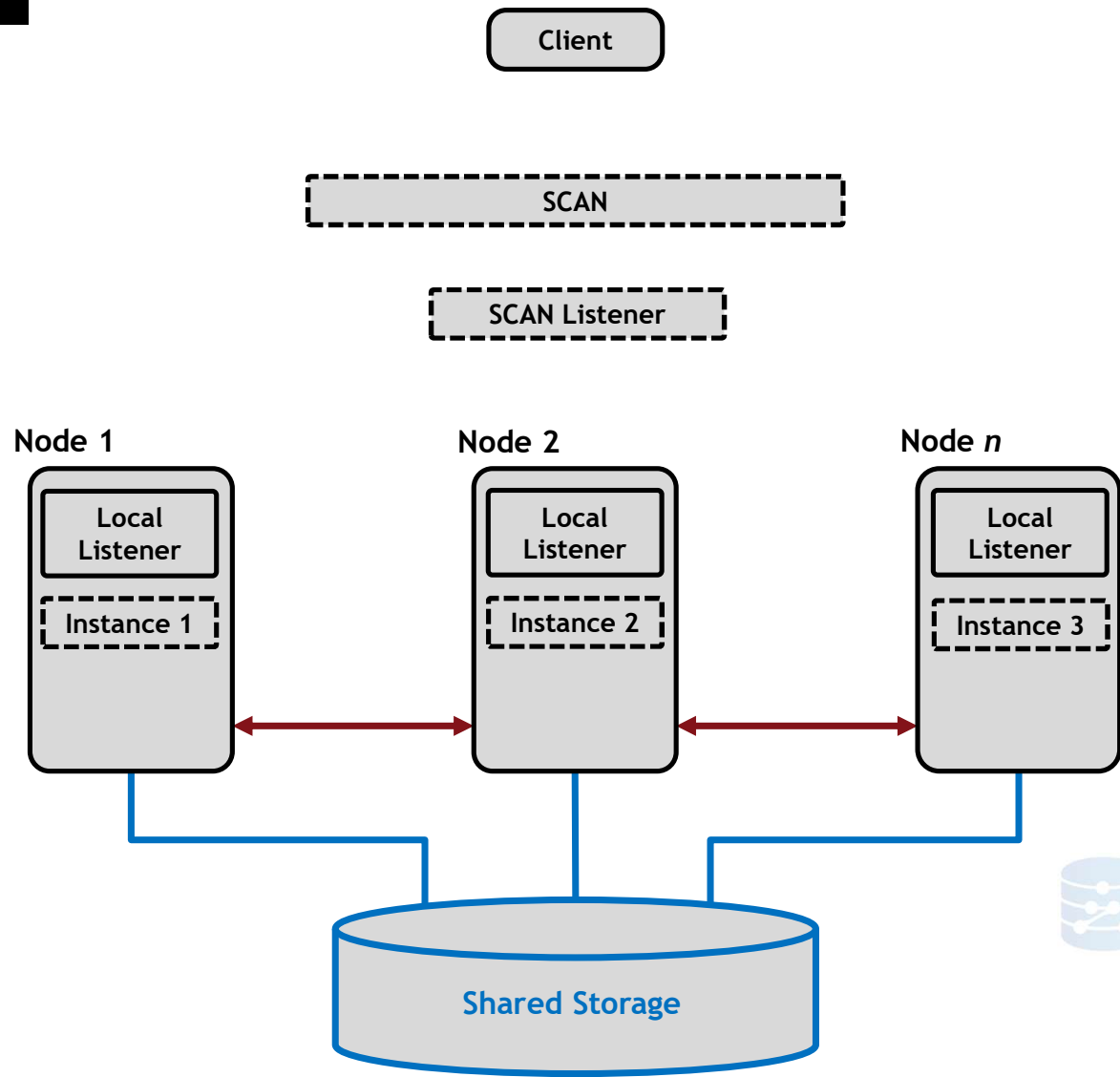   - A virtual IP name and address for each node

Oracle Clusterware Network Configuration (cont)

- Single Client Access Name (SCAN)
  - A domain name registered to at least one and up to three IP addresses, either in DNS or GNS.
  - Must be used by the clients when connecting to the RAC database
- Node Virtual IP (VIP) is managed by the Clusterware:
  - Must be assigned to each node
  - Must be on the same subnet as the public IP address
  - Provides rapid failure notification and failover
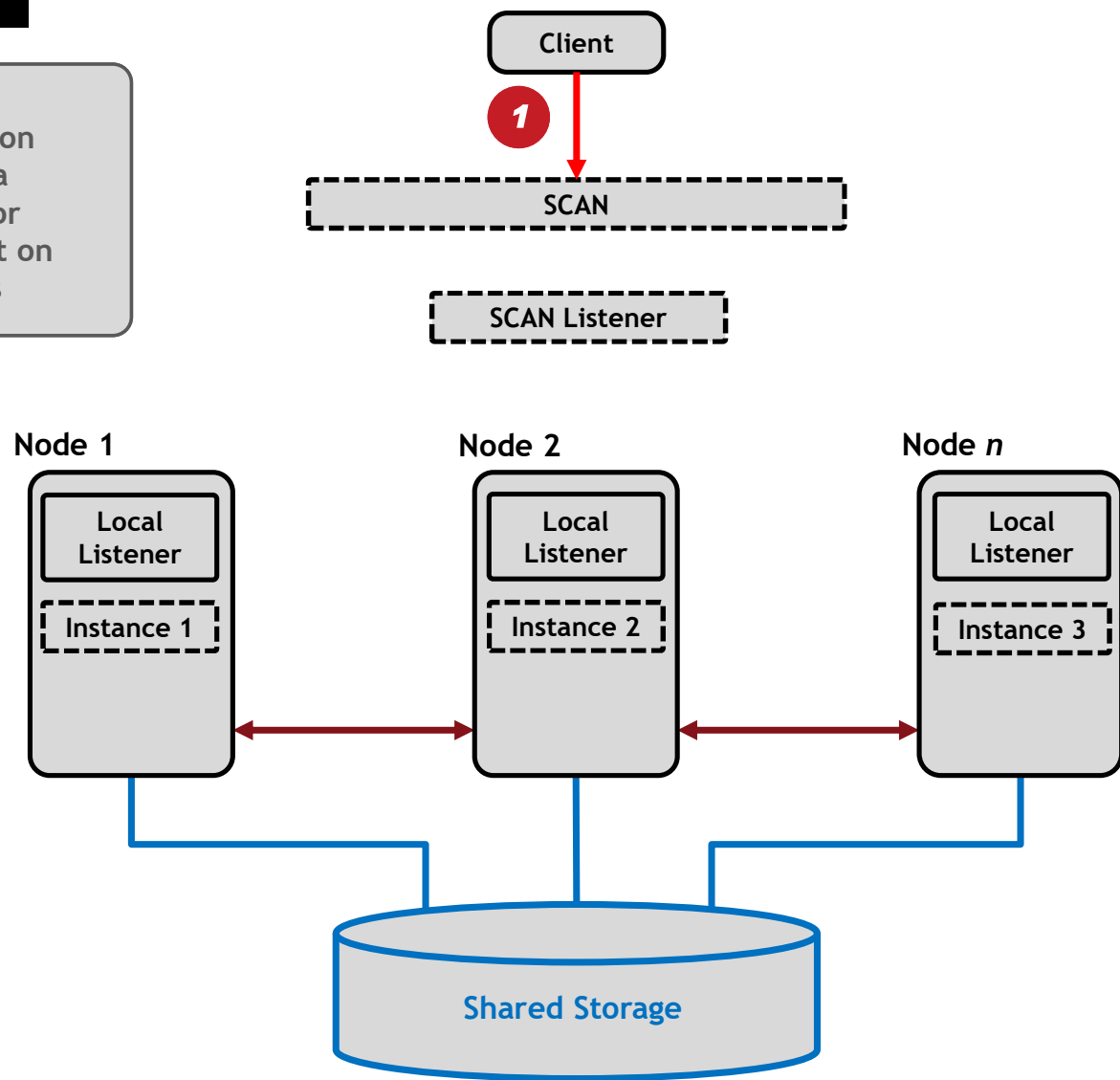
Oracle RAC Connectivity Cycle

Client

SCAN

SCAN Listener

Node 1

Local Listener

Instance 1

Node 2

Local Listener

Instance 2

Node *n*

Local Listener

Instance 3

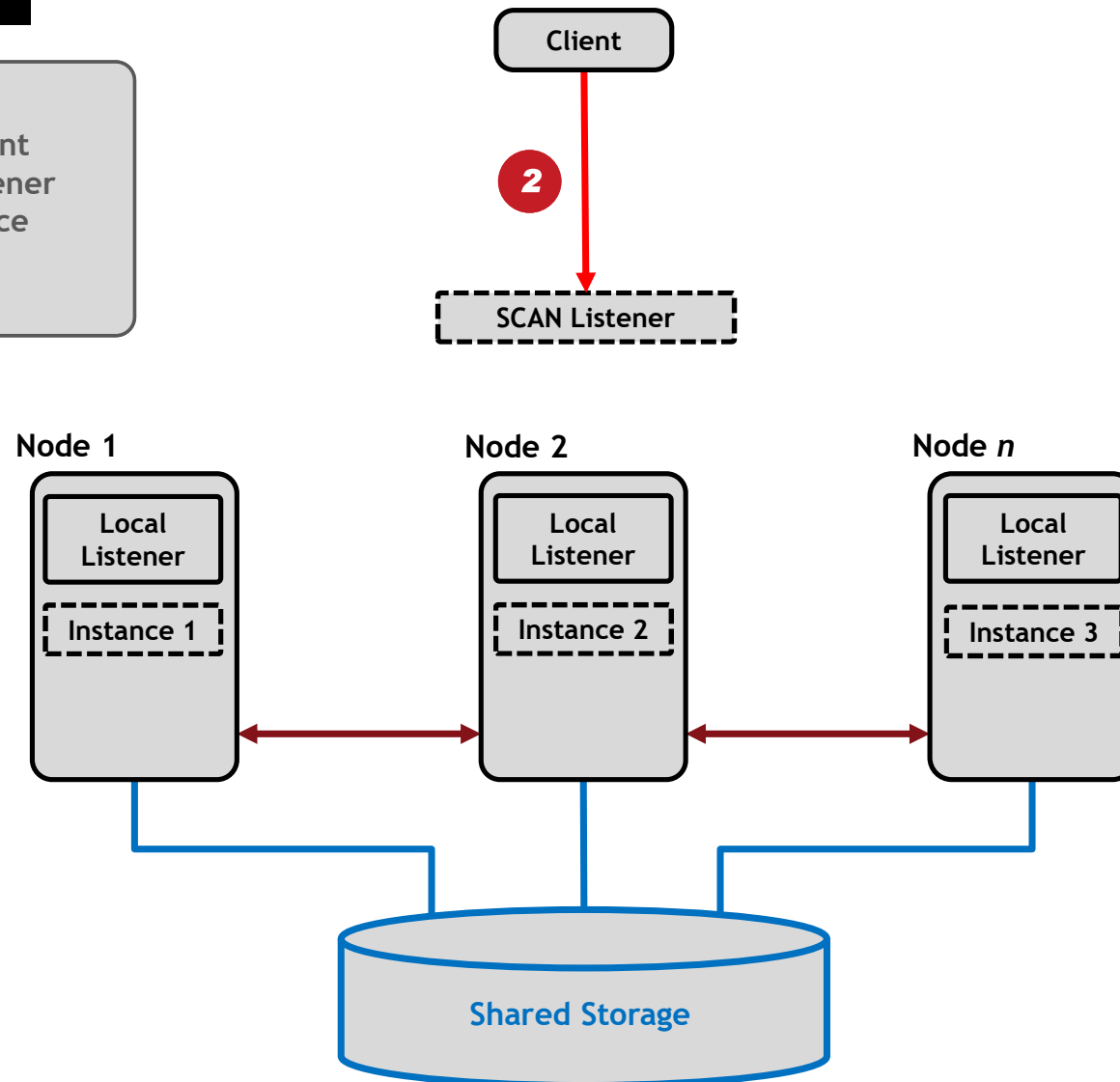Shared Storage

Oracle RAC Connectivity Cycle

Steps:

1 Client requests connection using a **SCAN** name and a **service name.** The DNS or GNS returns to the client on of the SCNA IP addresses

Client

1

SCAN

SCAN Listener

Node 1

Local Listener

Instance 1

Node 2

Local Listener

Instance 2

Node *n*

Local Listener

Instance 3

Shared Storage

Oracle RAC Connectivity Cycle

Steps:

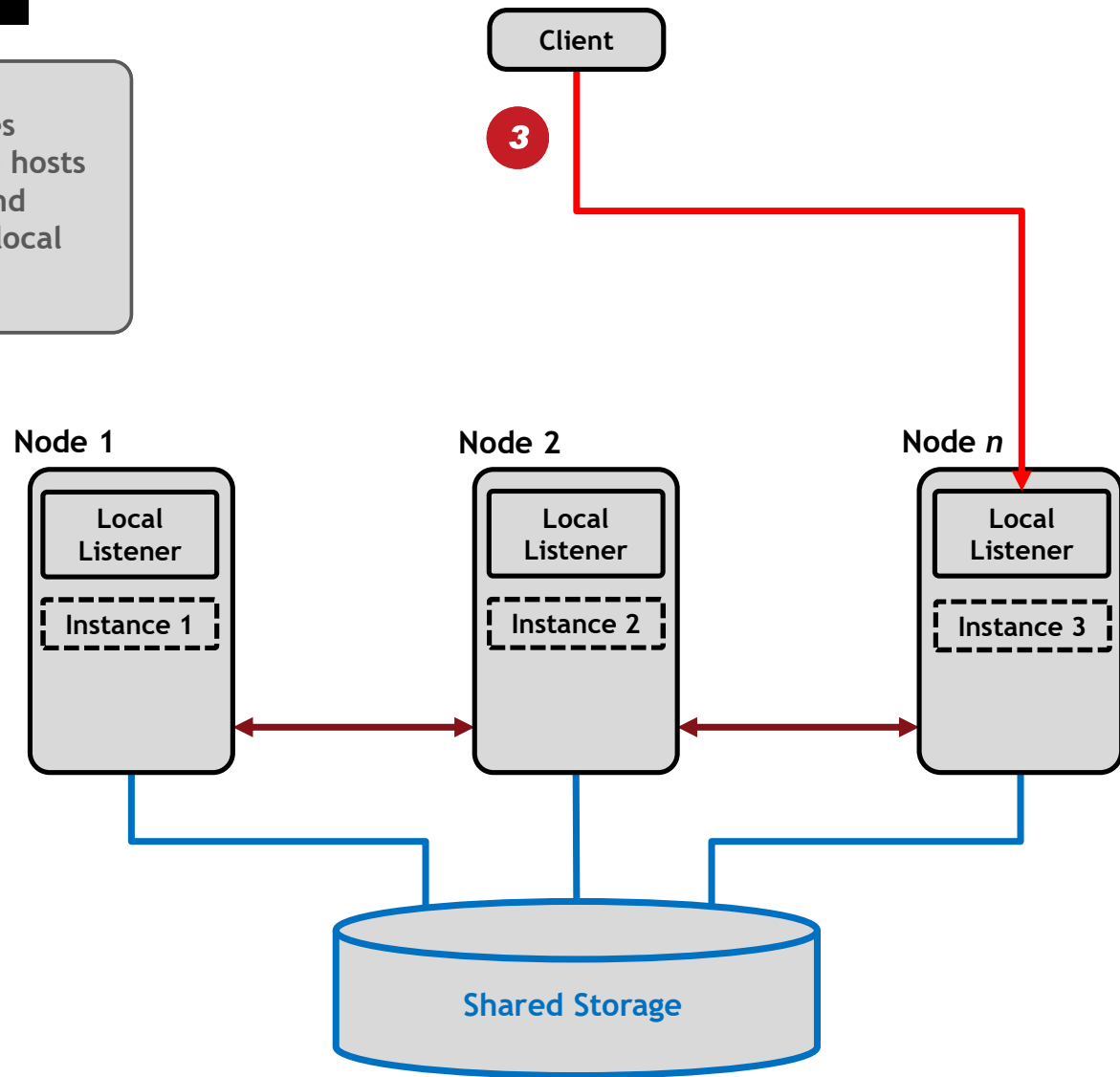**2** Using a SCAN IP, the client connects to a SCAN Listener providing it with a service name.

Client

2

SCAN Listener

Node 1

Local Listener

Instance 1

Node 2

Local Listener

Instance 2

Node *n*

Local Listener

Instance 3
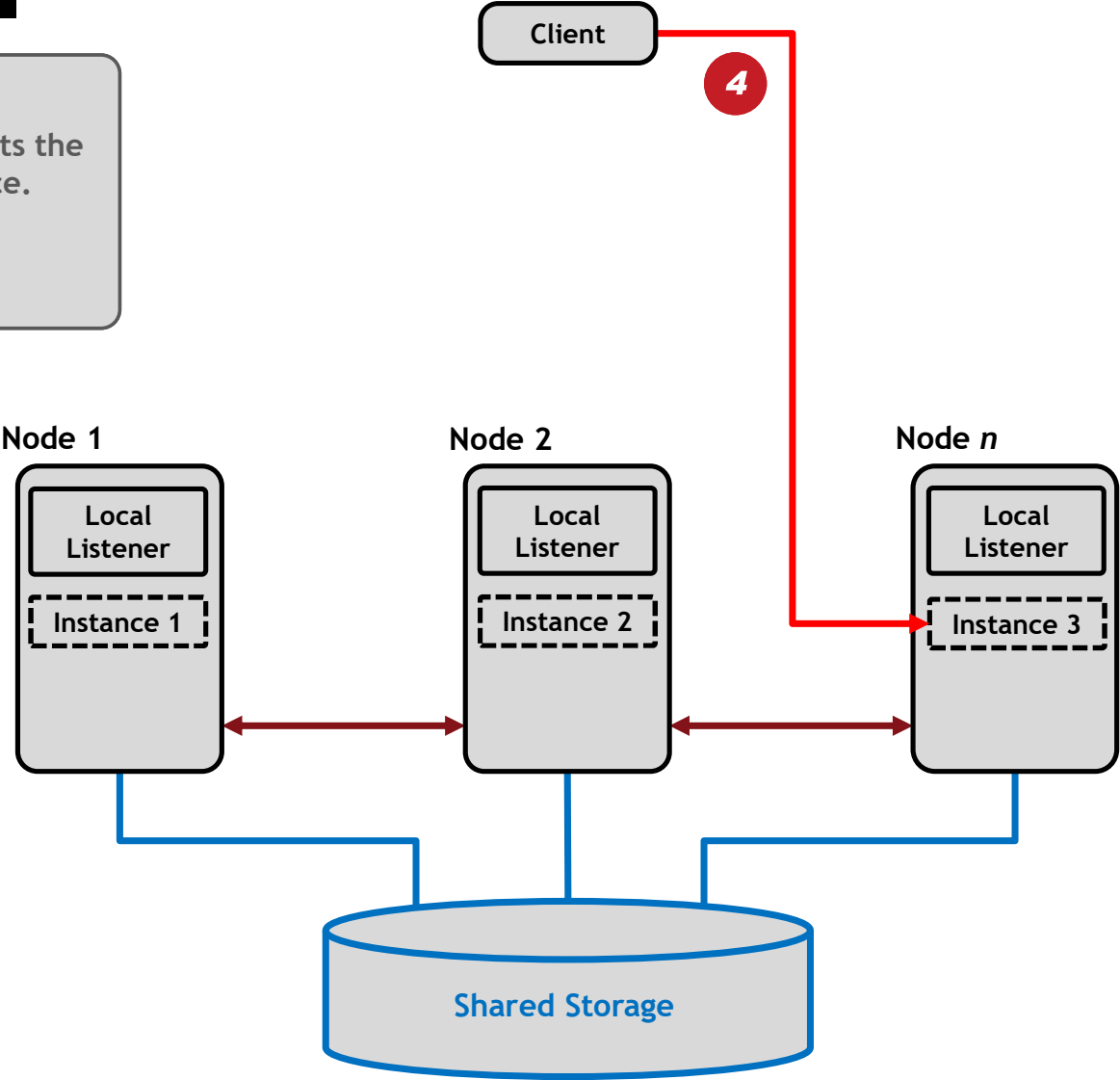
Shared Storage

Oracle RAC Database File Locations

Public Network

Node 1
Instance 1
DB Home

Node 2
Instance 2
DB Home

Node *n*
Instance 3
DB Home

Shared Storage

System and user datafiles

Undo tablespace *per inst.*

2 Redo G. *per instance*

Control F., SPFILE, psswd f.

DB Home (not recommended)
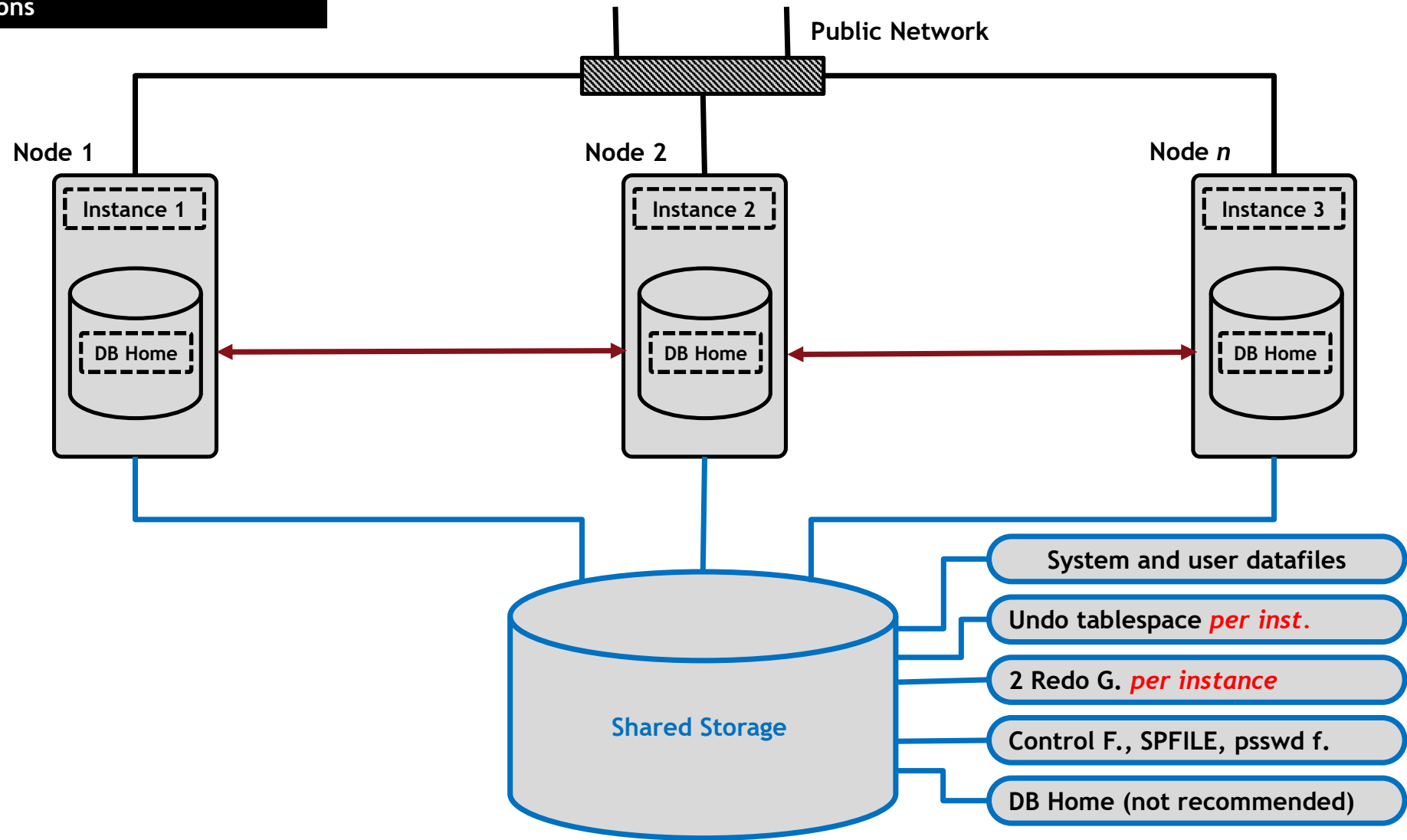
Oracle RAC Clusterware and Database Stack Components

SCAN

Node *x*

Grid Infrastructure

VIP

ASM

Dynamic Service

SCAN Listener

Local Listener

Database Home

Database Instance

Node *y*

Grid Infrastructure

VIP

ASM

Dynamic Service

SCAN Listener

Local Listener

Database Home

Database Instance

Interconnect

Shared Storage:
(OCR – Vote files) – (DB files)
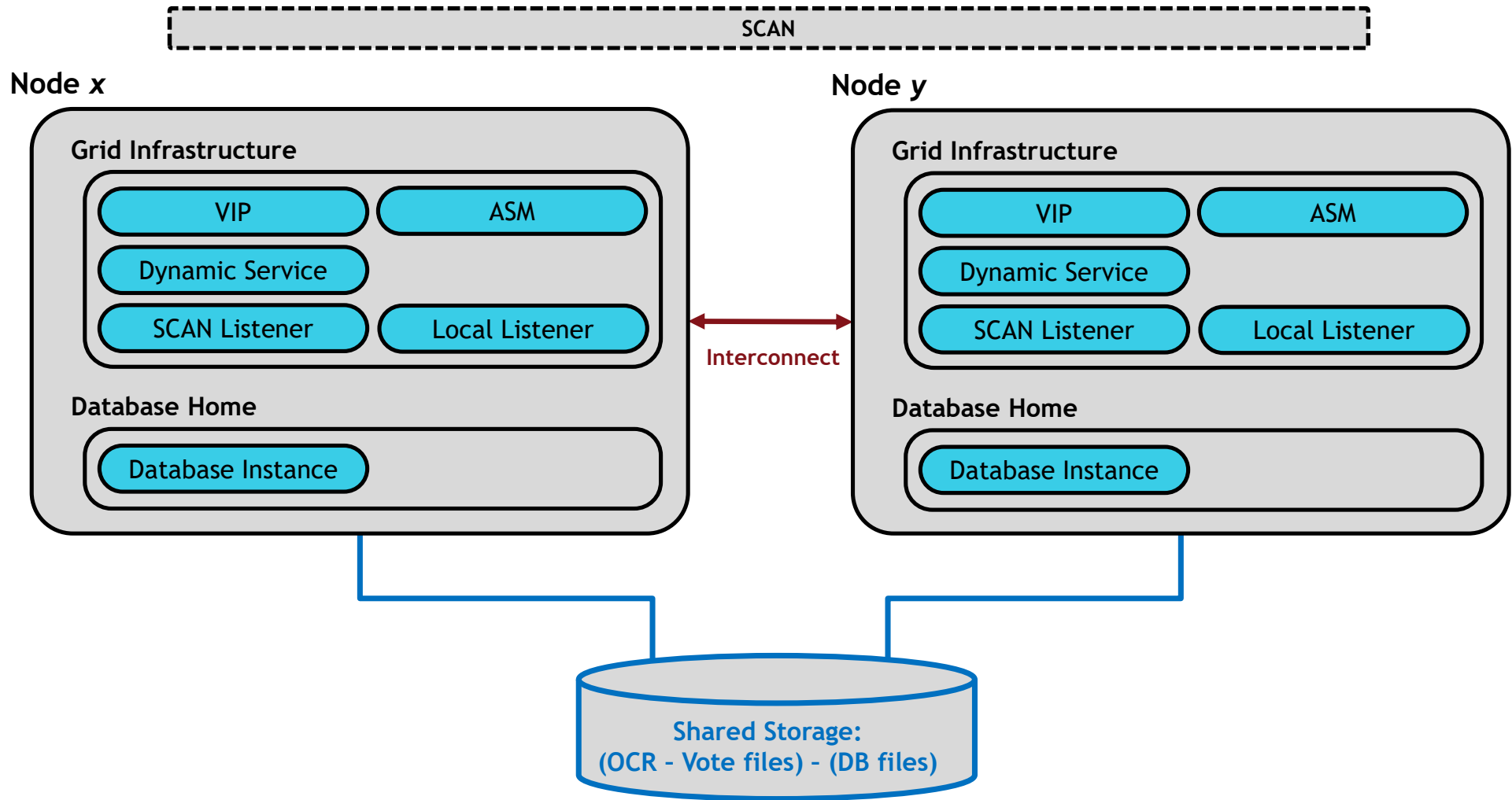
Summary

By the end of this lecture, you should have learnt how to describe the following:

- Oracle RAC hardware requirements
- Shared storage configuration options
- Oracle Grid Infrastructure components
- Oracle Clusterware network configuration
- Oracle RAC connectivity cycle
- Oracle RAC database files locations