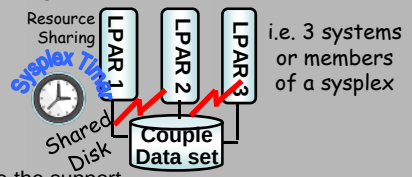
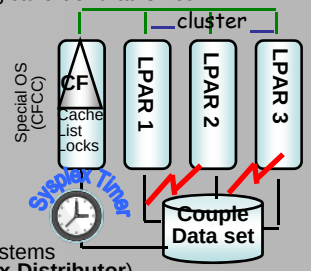
CheatSheet #38 zTidBits zOS' Sysplex & Parallel Sysplex

The quorum disk is to store the cluster log file and be a shared storage device separate from the rest of the shared storage.

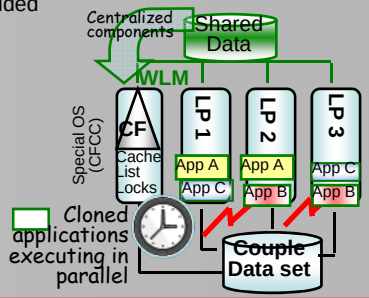
- In September 1990, IBM introduced the SYStems comPLEX, or sysplex, to help solve the complexity of managing multiple z/OS systems.
  - A **Sysplex** is a collection of z/OS systems that cooperate using hardware and software functions to process work, providing improved growth potential and higher level of availability.
  - The Sysplex architecture established the groundwork for simplified multisystem management through the **Cross-System Coupling Facility (XCF)** component of z/OS.
    > XCF services allow authorized applications on one system to communicate with applications on the same system or on other systems.
  - In a base sysplex, connectivity and communication between images is provided by high-speed interconnects called **channel-to-channel (CTC) links**.
  - The disk couple data set, which is shared between all of the images, holds control information and provides a mechanism for monitoring the status of the images.
  - When more than one image or CPC is involved, a **Sysplex Timer** is required to synchronize the time on all systems.
  - From a functionality perspective, the base sysplex did not change the clustering storage model.
    > z/OS provides a shared-disk model without shared cache.
  - What sysplex added was the following:
    > A standard way to communicate between systems
    > The support for quorum disks and member status
    > A common time source in the cluster



- The basic sysplex laid a foundations for communications between components and subsystems on the participating z/OS member images, but the basic sysplex did not provide the support for cache buffer coherency required for high-performance database subsystems.
- In addition, the existing Global Resource Serialization (locking) implementation could not provide the required performance for sharing high-volume transactional systems across a large number of systems.
- With the introduction of Parallel Sysplex technology in 1994 , high-performance data sharing through a new coupling technology (Coupling Facility) provided high-performance multisystem data sharing capability to authorized applications, such as system components and subsystems.
- The **Coupling Facility (CF)** is used by subsystems, such as IMS/DB DB2 and VSAM/RLS, to ensure the integrity and consistency of data throughout the entire sysplex.
  - One means of implementing a CF is in a Logical Partition (LP) just like a z/OS image but it runs its own highly specialized operating system, **Coupling Facility Control Code (CFCC)**.
  NOTE: Here is where customers can elect to purchase the Integrated Coupling Facility (ICF) engine.
  - It connects to all the z/OS images in a Sysplex by high-speed fibre or copper links.
  - The design of the CF ensures that requests are queued in a logical order that is not dependent on the speed of the connected CPCs.
    > This removes the unpredictable response time that is inherent in conventional devices such as DASD.
- This capability of linking together many systems and provides multisystem data sharing making the sysplex platform ideal for parallel processing, particularly for online transaction processing and decision support.
- The **Parallel Sysplex** technology builds on the base sysplex capability and allows an increased number of operating system images that can directly share work (up to 32).
  - The coupling facility enables high performance, multisystem data sharing across all the systems.
  - In addition, workloads can be dynamically balanced across systems with the help of new workload management functions (**Sysplex Distributor**).



- Parallel Sysplex enhanced the clustering storage model provided by System z to a *shared cache* model.
  - The following is what parallel sysplex added:
    > The high-performance shared memory external to the operating system images in the cluster.
    > The enhanced z/OS components exploiting the CF. For example, GRS (locks) could now move to a star topology using the CF to store global lock information, avoiding a continuous communication activity to keep all the systems synchronized improving performance when many systems were active in the sysplex.
    > The z/OS support for subsystems (DB2, VSAM, and IMS/DB) to exploit the shared cache and lock facilities in a standardized manner.



- This infrastructure implementation aligns with the mainframe feature of "*Share-Everything*"
- The CFCC interacts with the Cross systems Extended Services (XES) component on z/OS to create and maintain data in the CF memory data.
  - In the CF, data is organized into distinct objects called **structures.**
    > CF structures are used by authorized programs to implement data sharing and high-speed serialization.
  - There are a number of different types of CF structures that can be created: **cache, list, and lock**, each providing a specific function to the application.
    NOTE: Some storage in the coupling facility can also be allocated as a dedicated dump space for capturing structured information for diagnostic purposes.
    > **Cache structure** supplies a mechanism called buffer invalidation to ensure consistency of cached data. The cache structure can be used as a high-speed buffer for storing shared data with common read/write access. It is used by subsystems like database managers to provide shared buffers and maintain buffer coherency across multiple systems.
    > **Lock structures** supply shared and exclusive locking capability for serialization of shared resources down to a very small unit of data. The main users of lock structures are GRS (for sharing any type of resource) and the data sharing lock managers (IRLM and VSAM/Record Level Sharing: RLS). Associated with the lock structures are XES lock services, which come into play in case of lock contention. The user of the request can specify shared or exclusive access. Normally, shared access is associated with reads and exclusive access with a write, but it depends on the program issuing the request. As long as no one has exclusive access to the resource, shared access requests are always granted by the Coupling Facility.
    NOTE: If there is anyone with exclusive access, the allow/disallow decision must be made by the resource serialization manager for that resource.
    > **List structures** are the most general purpose type of structure and as such are not easy to define in a concise manner. List structures enable authorized applications to share data that is organized in a set of lists for implementing functions such as shared work queues and shared status information. List entries can be processed as FIFO or LIFO, and access to list entries can be serialized or unserialized. List structures provide the most flexibility and functionality of all the CF structuretypes.
- **Coupling Facility Links (CF links)** are high speed fibre or copper interconnects used to communicate between CF LPARs and z/OS LPARs. In the case of System Managed structure duplexing, CF links are also used for CF LPAR to CF LPAR communication. There are a number of types of CF links.
  - ISC3: Fibre, 200MB/sec, up to 100km - Standard fibre links
  - ICB4: Copper, 2000MB/sec, 10m, z990 technology - High speed over low distance
  - ICP: Microcode, 3500MB/sec - Very high speed within one CPC or System
  - PSIFB: Parallel Sysplex coupling over InfiniBand (PSIFB). This link has a bandwidth of up to 6GBps between two z10 EC servers and of up to 3GBps between z10 EC and System z9 servers.
- **Timers:** The presence of the timers is due to a long standing requirement for accurate time and date information in data processing.
  - As single operating systems have been replaced by multiple, coupled operating systems on multiple servers, this need has evolved into a requirement for both accurate and consistent clocks among these systems.
  - This capability has been provided by the Sysplex Timer device (IBM 9037) and by its follow-on and preferred the **Server Time Protocol (STP)**.
    > STP is designed for servers that are configured in a Parallel Sysplex or a standard sysplex without a Coupling Facility, as well as servers that are not in a sysplex but need to be time-synchronized. STP supports a multi-site timing network of up to 100 Km (62 miles) over fiber optic cabling. STP is a message-based protocol in which time keeping information is passed over data links between servers.
  - All systems in a sysplex must operate from the same time source.
  - Time synchronization between all systems in a sysplex is required because there are many instances where events that happen on different systems need to be aggregated for processing.
    > An example is a database restart that needs to apply log records from several systems in the correct time sequence.
- **XCF** provides the signaling services in a Parallel Sysplex.
  - Authorized applications can call XCF services to pass messages to other systems in the sysplex.
  - XCF uses signaling paths over high-speed fibre interconnects between systems or through CF list structures and for maintaining the status of members of the sysplex.
- **XES** is an extension to the XCF componen and provides the sysplex services that applications and subsystems use to share data held in the coupling facility.
  - These services support the sharing of cached data and common queues while maintaining data integrity and consistency.
  - Any subsystem, i.e. DB2, that needs to perform activity against the CF structures uses XES services.