

Ab initio Session 15

Introduction to Ab Initio



Ab Initio Training

1



Remaining Concepts like

- **Replicate**
- **What is a Phase?**
- **What is a Checkpoint?**
- **What Is a Subgraph?**

Replicate



- Replicate arbitrarily combines all the data records it receives into a single flow and writes a copy of that flow to each of its output flows.





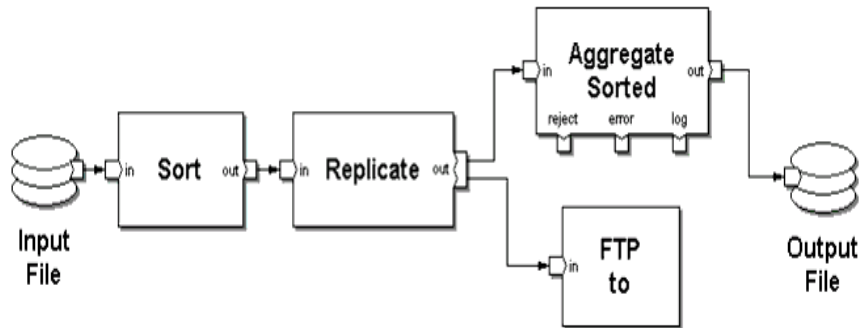
Runtime Behavior :Replicate

Replicate does not support using default record assignment. As a result, make the record format of the **in** and **out** ports identical. If you do not, execution of the graph stops when it reaches Replicate. The Replicate component:

- Arbitrarily combines the data records from all the flows on the **in** port into a single flow
- Copies that flow to all the flows connected to the **out** port Use Replicate to support component parallelism — for example, when you want to perform more than one operation on a flow of data records coming from an active component.



Suppose you want to aggregate a flow of records and also send them to another computer.
You can accomplish this by using Replicate as in the following graph:



CapGemini

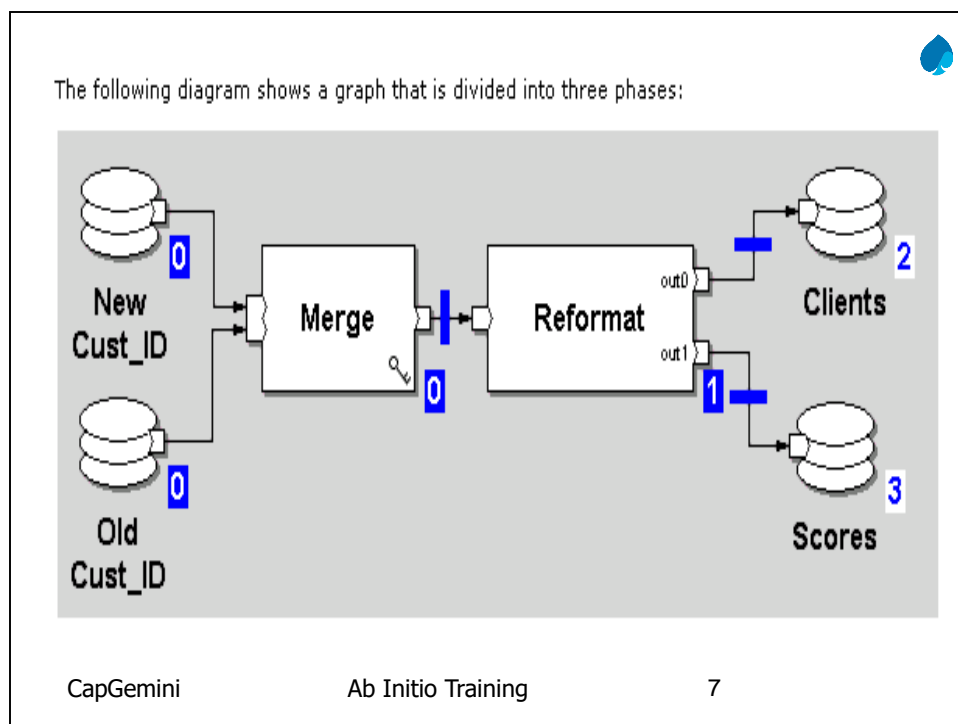
Ab Initio Training

5

What is a Phase?



A phase is a stage of a graph that runs to completion before the start of the next stage. By dividing a graph into phases, you can [save resources](#), avoid [deadlock](#), and safeguard [against failures](#). To protect a graph, all phases are [checkpoints](#) by default.



Phase-Description



- Phase numbers are assigned to each component.
- White phase numbers inside blue boxes denote phases that are checkpoints.

0

Phase-Description



- Blue phase numbers against a white background denote phases that are *not* checkpoints. Phases with checkpoints save temporary files until the next checkpoint completes; phases without checkpoints delete all temporary files at the end of the phase.

2

Phase-Description



- When there are phase breaks inside a subgraph, the phase number shows the range of phases that occur inside.
- When you insert an Intermediate File component into a graph, it automatically creates a phase break.

0-3

Phase-Description



- Blue bars across flows indicate phase breaks that are checkpoints, unfilled blue bars across flows indicate phase breaks that are not checkpoints.



What is a Checkpoint



- A checkpoint is a **phase** that acts as an intermediate stopping point in a graph and saves status information to allow you to **recover** from failures. By assigning phases with checkpoints to a graph, you can recover completed stages of the graph if failure occurs.

Setting Phases and Checkpoints



- In the graph, select all the components you want to place in phase **1**.
- Click the Increment Phase button.



Setting Phases and Checkpoints



- The phase number beside each selected component, and all downstream components, becomes **1**. Phase breaks appear on the flows between phases



Recovery



When you encounter a graph that fails (for example, one **phase** of a graph encounters an error or an operator aborts the program during execution), you do not need to rerun checkpointed phases that completed prior to failure. When you run the job again, the graph *recovers* from the last **checkpoint**. For the failed phases, all processes will be terminated, all temporary files will be deleted, and all nodes and their respective files will be rolled back to their initial state unless you select **Don't roll back file if job fails** on the File Properties: Access tab.

- If you find a hardware failure, (for example, one or more nodes loses power or any of the UNIX operating systems crash), rerun the job to *recover* it.

What Is a Subgraph?



- A subgraph is a **graph fragment**.
- Just like graphs, subgraphs contain components and flows.
- A **subgraph groups** together components that perform a subtask in a graph. The subgraph creates a reusable component that performs the subtask.



Thank You

End of Session 15

CapGemini

Ab Initio Training

18