**Informatica
PowerCenter**

Lesson 16: Partitioning

Lesson Objectives

- In this Lesson you will learn about:
  - Introduction to Informatica Partitioning
  - Different Types Of Partitioning
    - Database partitioning
    - Hash auto-keys
    - Hash user keys
    - Key range
    - Pass-through
    - Round-robin

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

16.1. Introduction To Partitioning

- The partition type controls how the Integration Service distributes data among partitions at partition points.
  - The Integration Service creates a default partition type at each partition point.
  - After tuning all the performance bottlenecks we can further improve the performance by addition partitions.
- Partitions are widely divided into –
  - Dynamic Partitioning (No. of partitions passed as parameters)
  - Non-Dynamic Partitioning (No. of partitions fixed while coding)
  - The partition attributes include setting the partition point, the number of partitions, and the partition types.
  - Informatica partition are useful in situations where
    - we need to load huge volume of data or
    - when using Informatica source which already has partitions defined, and using those partitions will allow to improve the session performance.

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

16.1. Introduction To Partitioning

- Partition Point: There can be one or more pipelines inside a mapping.
  - Adding a partition point will divide this pipeline into many pipeline stages.
  - Informatica will create one partition by default for every pipeline stage.
  - As we increase the partition points it increases the number of threads.
    - Informatica has mainly three types of threads –Reader, Writer and Transformation Thread.
  - Maximum 64 partition point can be define.
    - However, creating large number of partitions or partition points in a session that processes large amounts of data, can overload the system
  - Partition points cannot be created for the following transformations:
    - Source definition
    - Sequence Generator
    - XML Parser
    - XML target
    - Unconnected transformations

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

16.2. Types Of Partition

- Type of partitions are
  - Database partitioning
  - Pass-through
  - Hash auto-keys
  - Hash user keys
  - Key range
  - Round-robin
- Database Partitioning:
  - The PowerCenter Integration Service queries the IBM DB2 or Oracle system for table partition information.
  - It reads partitioned data from the corresponding nodes in the database.
  - Use database partitioning with Oracle or IBM DB2 source instances on a multi-node table space.
  - Use database partitioning with DB2 targets

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

16.2. Types Of Partition

- **Pass Through Partitioning**
  - This is default configuration for session.
  - The PowerCenter Integration Service passes all rows at one partition point to the next partition point without redistributing them
  - Choose pass-through partitioning where you want to create an additional pipeline stage to improve performance, but do not want to change the distribution of data across partitions

- **Hash auto-keys Partitioning**
  - The PowerCenter Integration Service uses all grouped or sorted ports as a compound partition key
  - You may need to use hash autokeys partitioning at Rank, Sorter, and unsorted Aggregator transformations.

- **Hash user keys.**
  - The PowerCenter Integration Service uses a hash function to group rows of data among partitions.
  - You define the number of ports to generate the partition key

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING
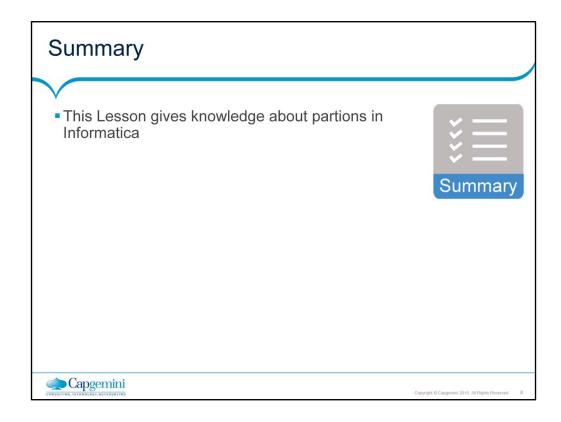
16.2. Types Of Partition

- Key Range
  - You specify one or more ports to form a compound partition key.
  - The PowerCenter Integration Service passes data to each partition depending on the ranges you specify for each port.
  - Use key range partitioning where the sources or targets in the pipeline are partitioned by key range.
- Round-robin
  - The PowerCenter Integration Service distributes blocks of data to one or more partitions.
  - Use round-robin partitioning so that each partition processes rows based on the number and size of the blocks

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

Summary

- This Lesson gives knowledge about partions in Informatica

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

Review Question

- Question 1: Partitions are used for
  - Option 1: Performance Improvement
  - Option 2: Remove performance bottleneck
  - Option 3: None of above

- Question 2: Is default partition
  - Option 1: Round Robin
  - Option 2: Pass-trough
  - Option 3: Hash key

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING