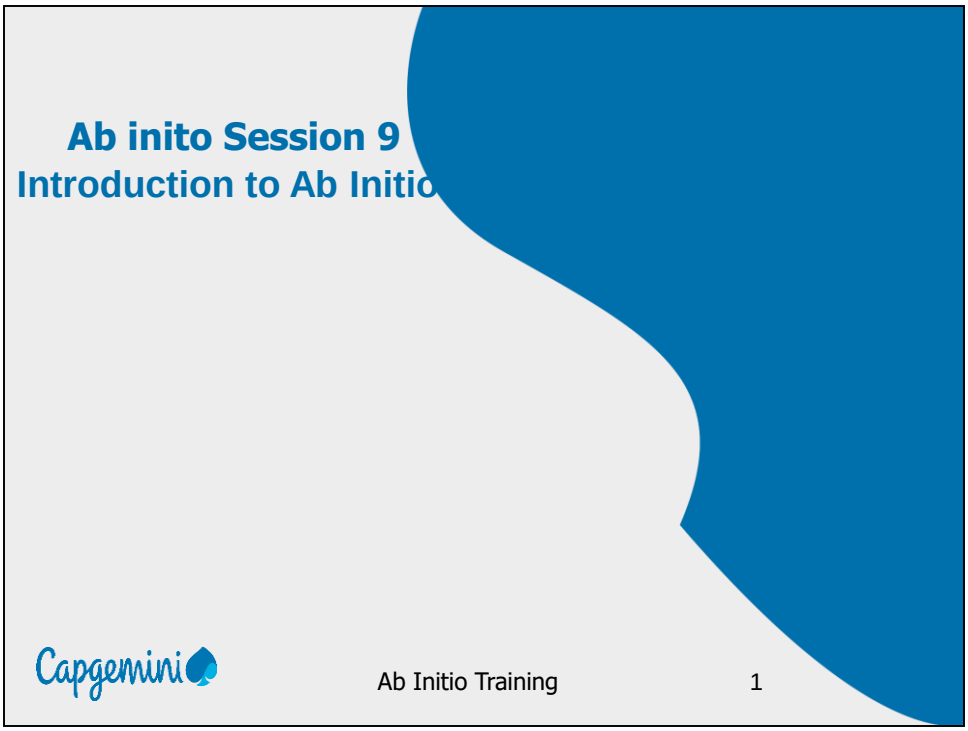
Ab inito Session 9
Introduction to Ab Initio
Capgemini
Ab Initio Training
1

Partitioning Components

- Partition by Key
- Partition by Round-robin
- Partition by Expression
- Partition by Percentage
- Partition by Range
- Broadcast
- Partition with Load Balance

Partition by Key

Partition by Key reads records from the in port and distributes data records to its output flow partitions according to key values.

A ***partition by key*** component is generally followed by a ***sort*** component

The Partition by Key component:

- Reads records in arbitrary order from the in port
- Distributes them to the flows connected to the out port, according to the key parameter, writing records with the same key value to the same output flow.

Parameter:-Partition by Key

- key

  Names(s) of the key field(s) you want Partition by Key to use when it distributes data records among flow partitions.

Partition by Round Robin

- Partition by Round-robin distributes blocks of data records evenly to each output flow in round-robin fashion.
- The difference between Partition by Key and Partition by Round Robin is the 1st one may not distribute data uniformly across the all partition in a multi file system but the latter does.

Parameters:- Partition by Round Robin

- Blocksize

  Number of records distributed to one flow before distributing the same number to the next flow.

  Default is 1.

The Partition by Round-robin component:

- Reads records from the in port.
- Distributes them in blocksize chunks to its output flows according to the order in which the flows are connected .
- The effect is like dealing a deck of cards.

Partition by Expression

- Partition by Expression distributes data records to its output flow partitions according to a specified DML expression.

Partition by Expression-cont.

The Partition by Expression component:

- Reads records in arbitrary order from the flows connected to the in port.
- Distributes the records to the flows connected to the out port, according to the expression in the function parameter.

Parameter:- Partition by Expression

- Function

DML expression using a field or fields from the input record format:

- **The expression must evaluate to a number between 0 and the number of flows connected to the out port minus 1.**
- **Partition by Expression routes the record to the flow number returned by this expression.**
- **Flow numbers start at 0.**

Partition by Percentage

- Partition by Percentage distributes a specified percentage of the total number of input data records to each output flow .

The Partition by Percentage component

- Reads records from the in port
- Writes a specified percentage of the input records to each flow on the out port
- You can supply the percentages that Partition by Percentage uses to partition data records in either of two ways:
  - By specifying the percentages in the percentages parameter.
  - By connecting the output of any component that produces a list of percentages to the pct port of Partition by Percentage.

Parameter:- Partition by Percentage

- List of percentages expressed as integers from 1 to 100, separated by spaces.

Partition by Range

- Partition by Range distributes data records to its output flow partitions according to the ranges of key values specified for each partition.

Parameters

- key
- Name(s) of the field(s) containing the key values you want Partition by Range to use when it distributes data records among output partitions.
- The field(s) specified must exist in the record formats for both the in and split ports, and must be of the same type in both record formats.

The Partition by Range component:

- Reads splitter records from the split port, and assumes that these records are sorted according to the key parameter.
- Determines whether the number of flows connected to the out port is equal to n (where n-1 represents the number of splitter records).
- If not, Partition by Range writes an error message and stops the execution of the graph.

The Partition by Range component:

- Reads data records from the flows connected to the in port in arbitrary order.
- Distributes the data records to the flows connected to the out port according to the values of the key field(s), as follows:
- Assigns records with key values less than or equal to the first splitter record to the first output flow
- Assigns records with key values greater than the first splitter record, but less than or equal to the second splitter record to the second output flow, and so on.

BROADCAST

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

The Broadcast component:

- Reads records from all flows on the in port
- Combines the records arbitrarily into a single flow
- Copies all the records to all the flow partitions connected to the out port
- Use Broadcast to increase data parallelism when you have connected a single fan-out flow to the out port or to increase component parallelism when you have connected multiple straight flows to the out port.

Thank You

End of Session 11