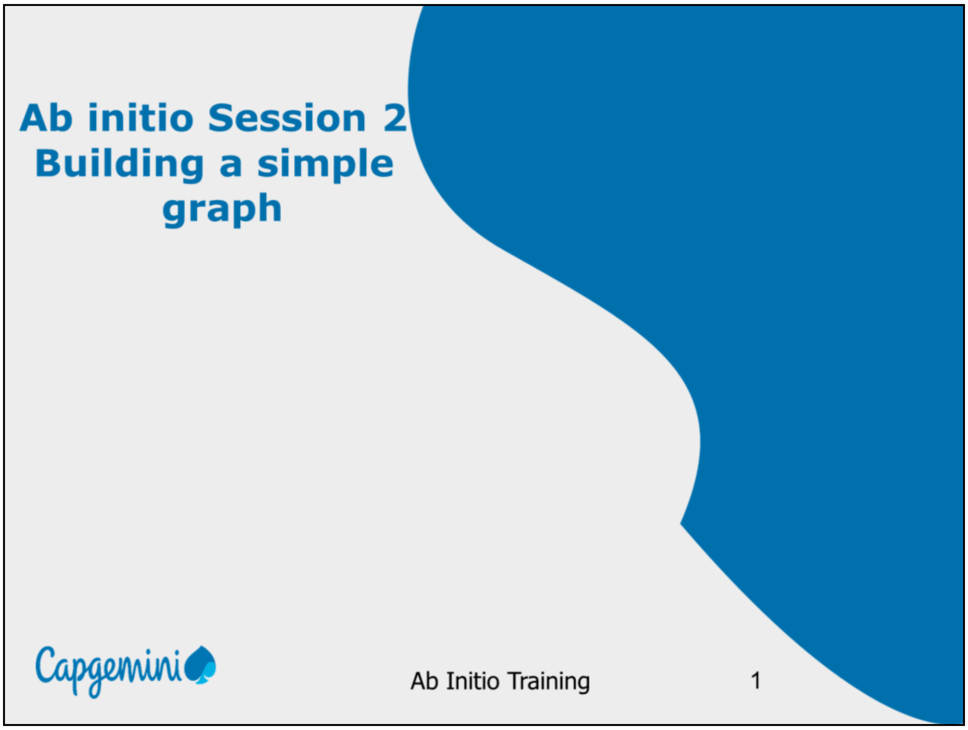
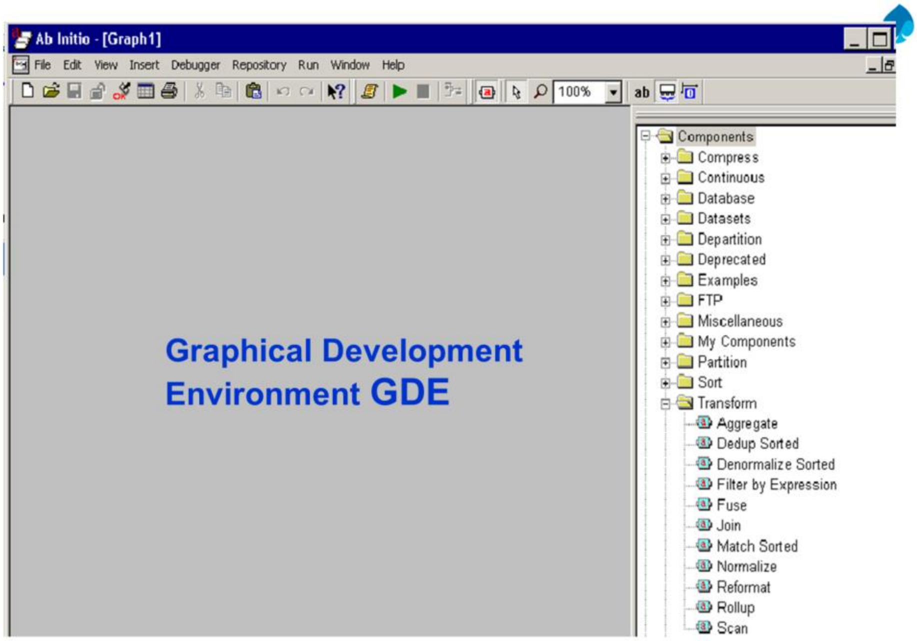


Ab initio Session 2 Building a simple graph

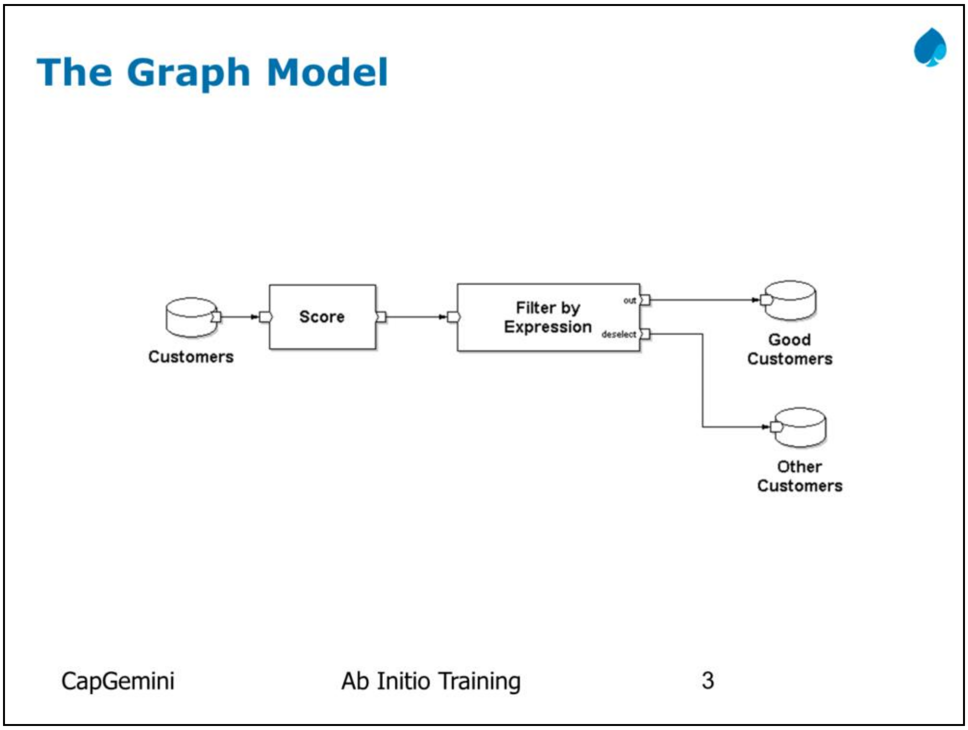
Capgemini

Ab Initio Training

1

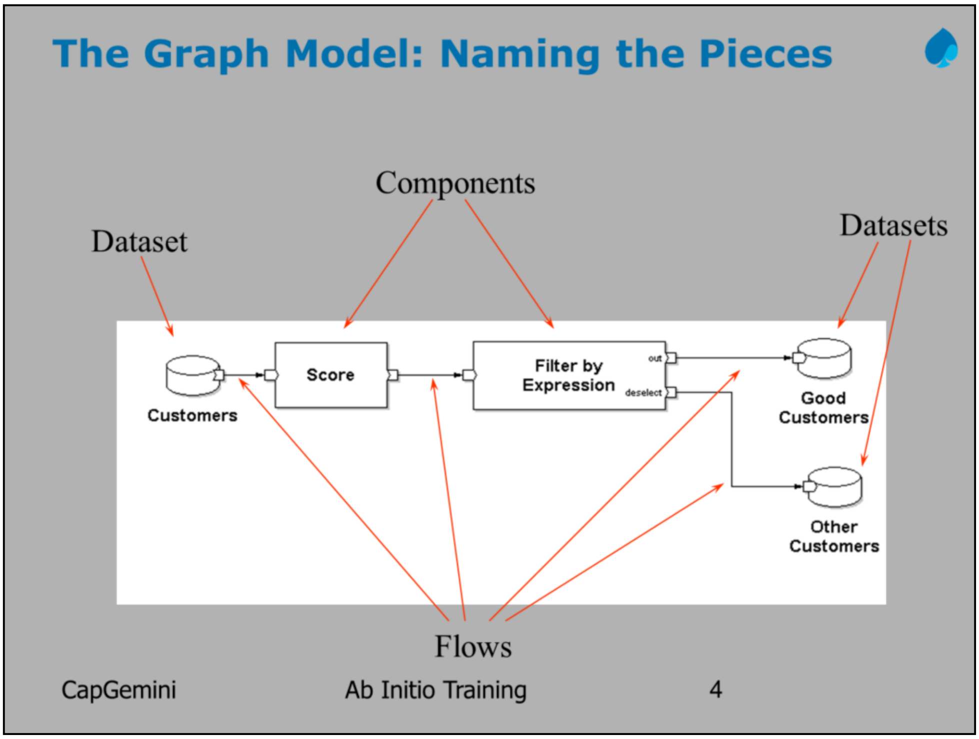


Graphical Development Environment GDE



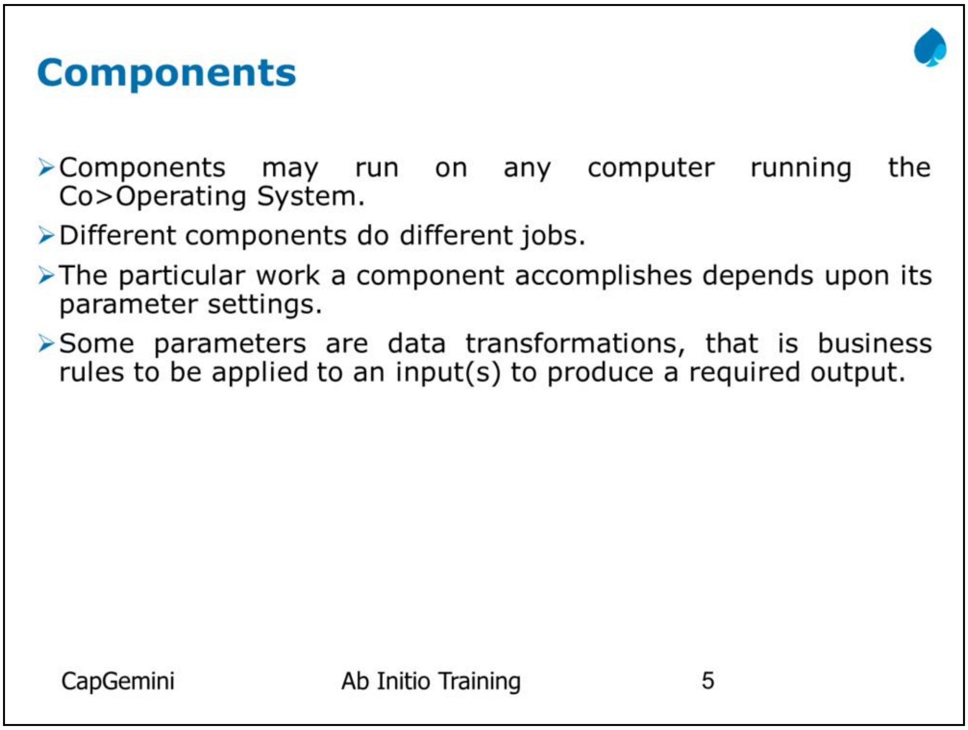
The Graph Model

CapGemini Ab Initio Training 3



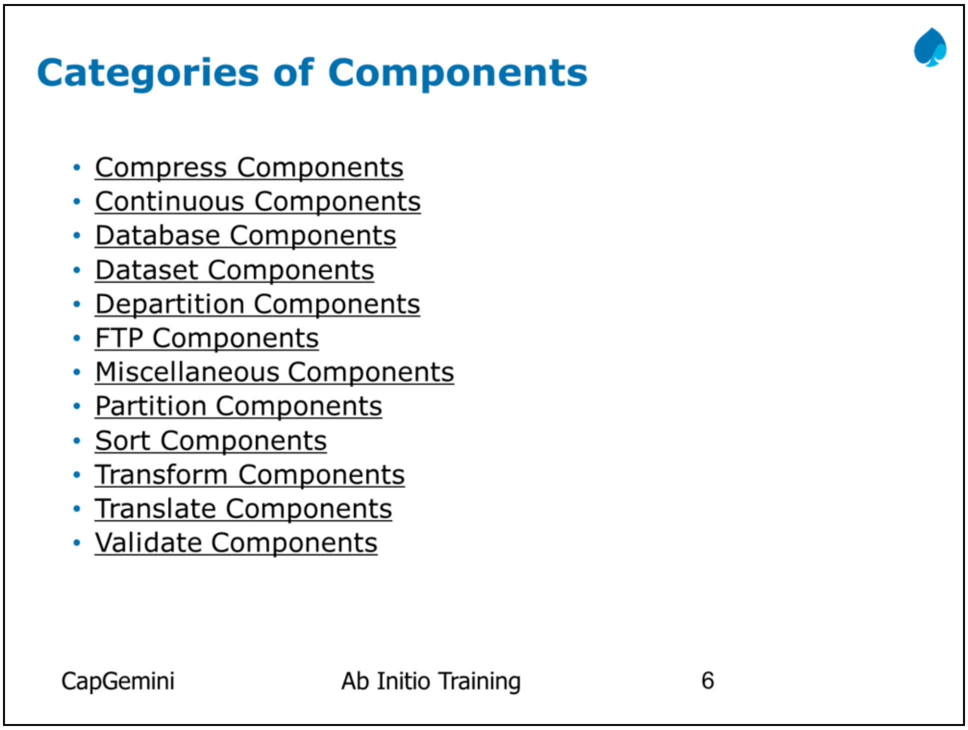
The Graph Model: Naming the Pieces

CapGemini Ab Initio Training 4



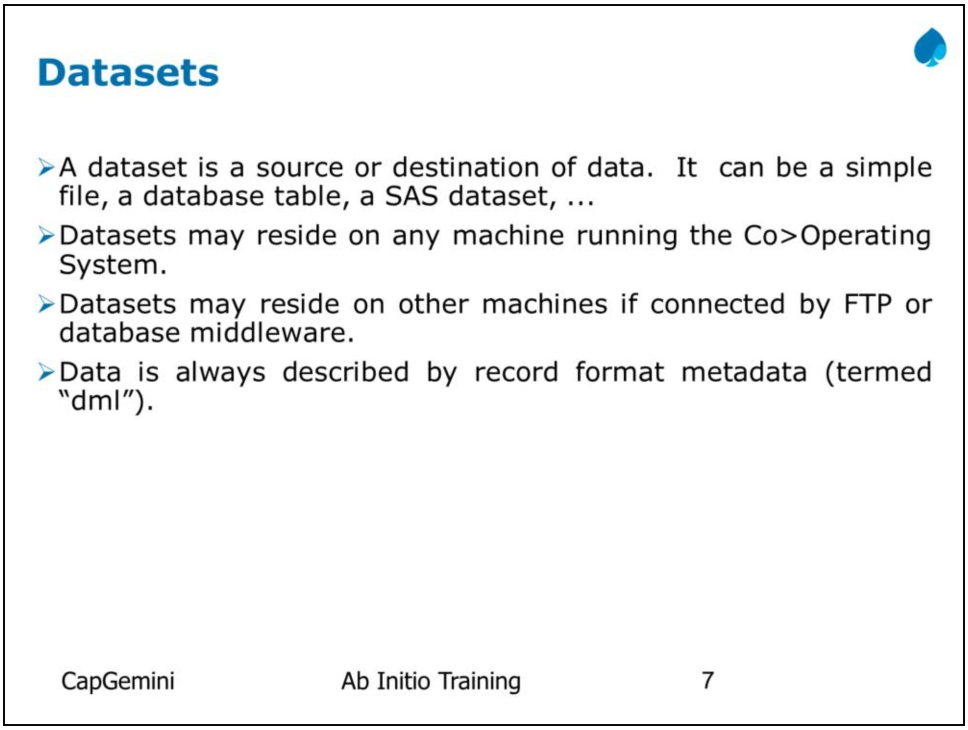
Components

- Components may run on any computer running the Co>Operating System.
- Different components do different jobs.
- The particular work a component accomplishes depends upon its parameter settings.
- Some parameters are data transformations, that is business rules to be applied to an input(s) to produce a required output.



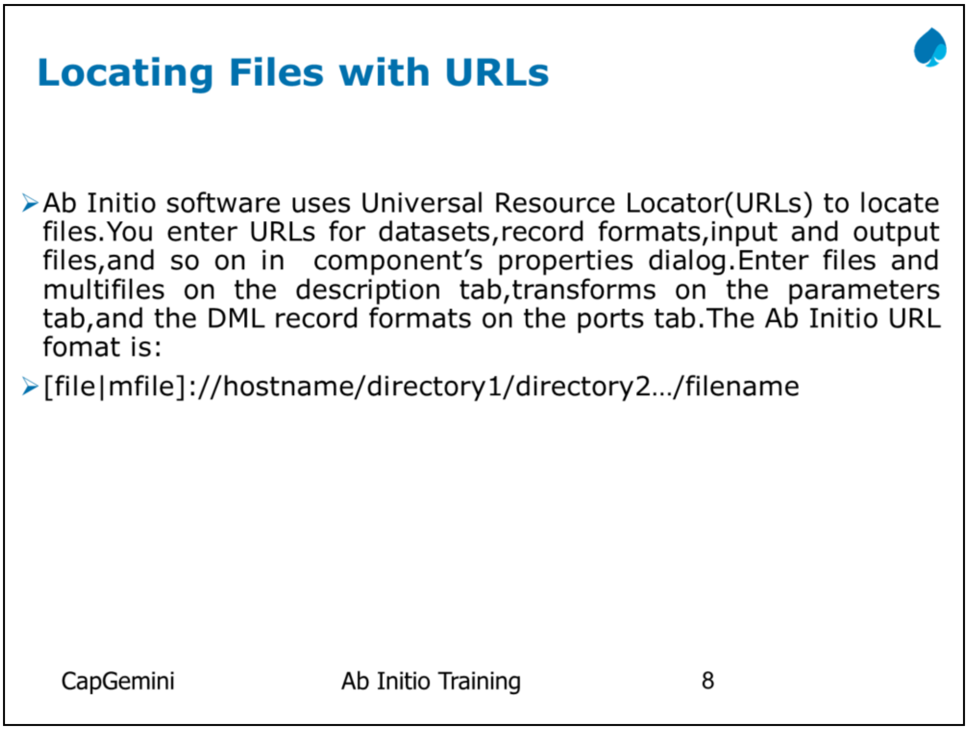
Categories of Components

- Compress Components
- Continuous Components
- Database Components
- Dataset Components
- Departition Components
- FTP Components
- Miscellaneous Components
- Partition Components
- Sort Components
- Transform Components
- Translate Components
- Validate Components



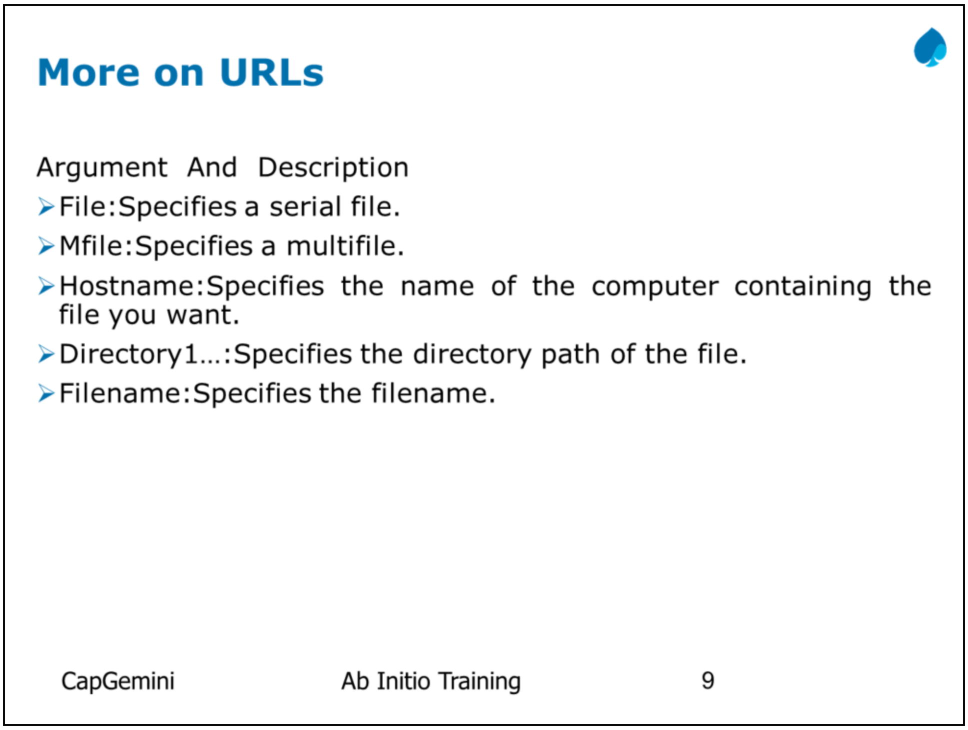
Datasets

- A dataset is a source or destination of data. It can be a simple file, a database table, a SAS dataset, ...
- Datasets may reside on any machine running the Co>Operating System.
- Datasets may reside on other machines if connected by FTP or database middleware.
- Data is always described by record format metadata (termed “dml”).



Locating Files with URLs

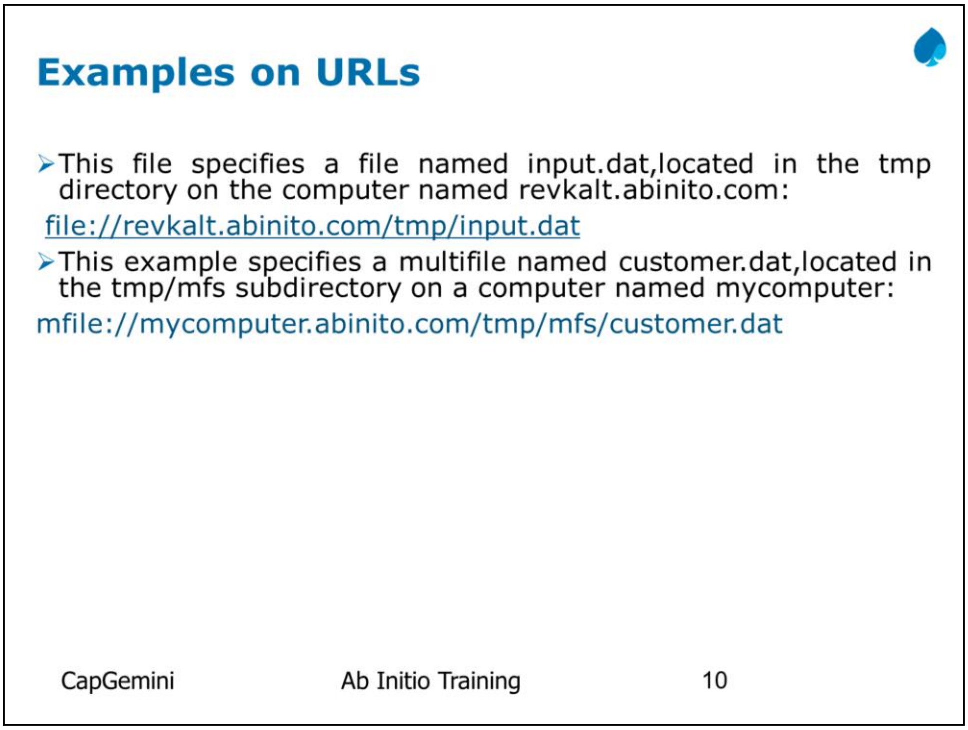
- Ab Initio software uses Universal Resource Locator(URLs) to locate files.You enter URLs for datasets,record formats,input and output files,and so on in component's properties dialog.Enter files and multifiles on the description tab,transforms on the parameters tab,and the DML record formats on the ports tab.The Ab Initio URL fomat is:
- [file|mfile]://hostname/directory1/directory2.../filename



More on URLs

Argument And Description

- File:Specifies a serial file.
- Mfile:Specifies a multifile.
- Hostname:Specifies the name of the computer containing the file you want.
- Directory1…:Specifies the directory path of the file.
- Filename:Specifies the filename.



Examples on URLs

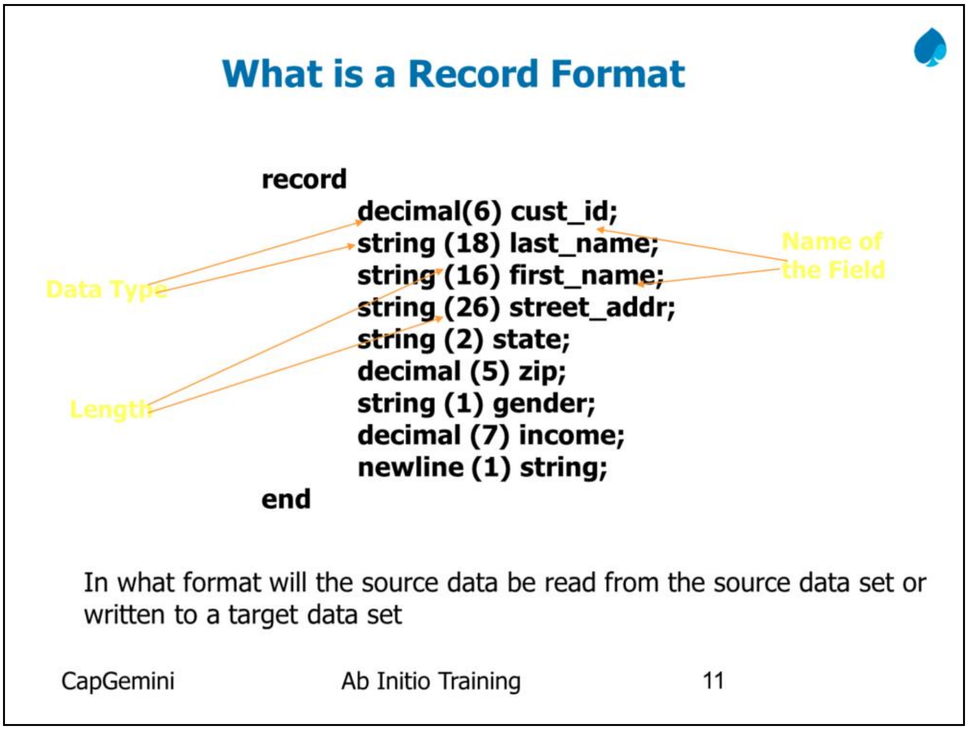
- This file specifies a file named input.dat,located in the tmp directory on the computer named revkalt.abinito.com:

file://revkalt.abinito.com/tmp/input.dat

- This example specifies a multifile named customer.dat,located in the tmp/mfs subdirectory on a computer named mycomputer:

mfile://mycomputer.abinito.com/tmp/mfs/customer.dat

CapGemini Ab Initio Training 10



What is a Record Format

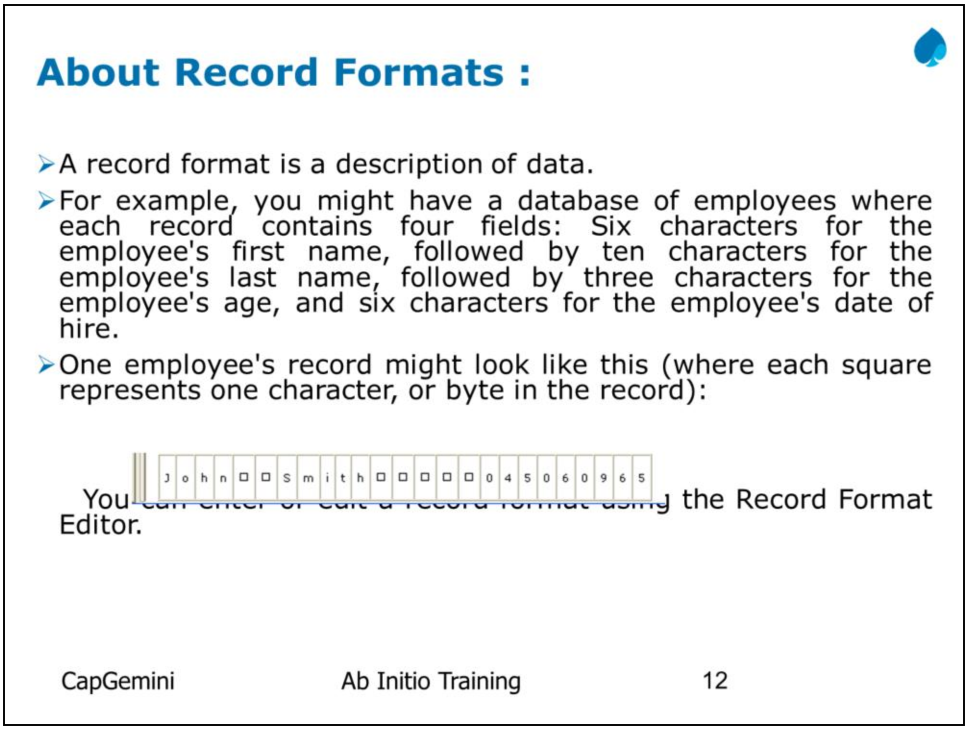
```
record
        decimal(6) cust_id;
        string (18) last_name;
        string (16) first_name;
        string (26) street_addr;
        string (2) state;
        decimal (5) zip;
        string (1) gender;
        decimal (7) income;
        newline (1) string;
end
```

Data Type
Length
Name of the Field

In what format will the source data be read from the source data set or written to a target data set

CapGemini Ab Initio Training 11

Notes

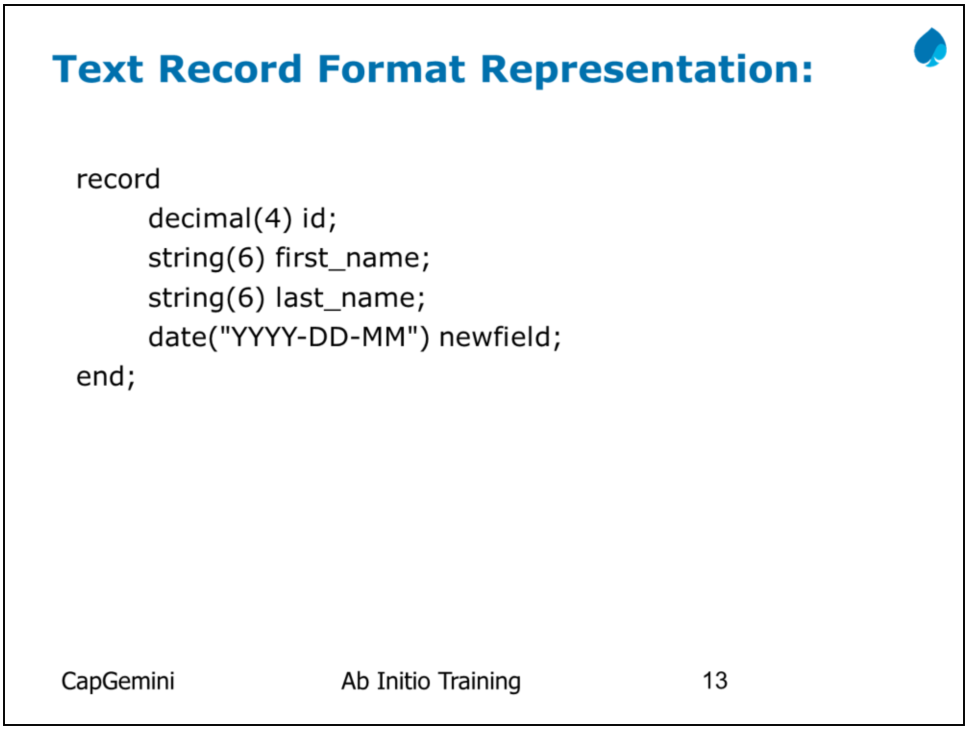


About Record Formats :

- A record format is a description of data.
- For example, you might have a database of employees where each record contains four fields: Six characters for the employee's first name, followed by ten characters for the employee's last name, followed by three characters for the employee's age, and six characters for the employee's date of hire.
- One employee's record might look like this (where each square represents one character, or byte in the record):

J	o	h	n	□	□	S	m	i	t	h	□	□	□	□	□	0	4	5	0	6	0	9	6	5

You can enter or edit a record format using the Record Format Editor.

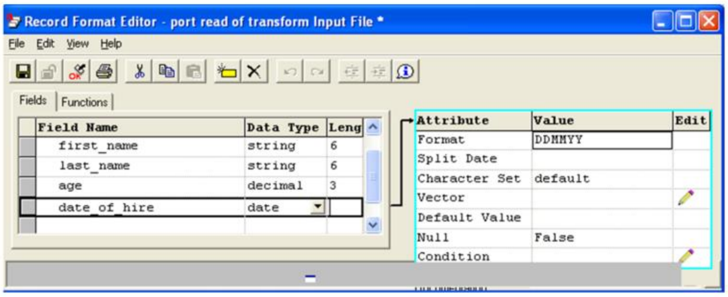


Text Record Format Representation:

```
record
        decimal(4) id;
        string(6) first_name;
        string(6) last_name;
        date("YYYY-DD-MM") newfield;
end;
```

Specifying the Record Format of a Port

Record Format Editor

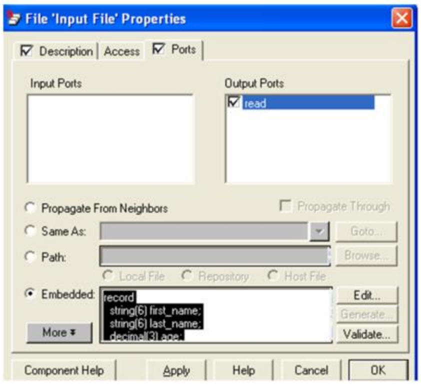


Field Name	Data Type	Leng
first_name	string	6
last_name	string	6
age	decimal	3
date_of_hire	date	

Attribute	Value	Edit
Format	DDMMYY	
Split Date		
Character Set	default	
Vector		✎
Default Value		
Null	False	
Condition		✎

Specifying the Record Format of a Port

- You can assign a record format to a dataset component or program component by viewing the component's properties dialog, and specifying the record format on the Ports tab.



File 'Input File' Properties

☑ Description | Access | ☑ Ports

Input Ports

Output Ports: ☑ read

○ Propagate From Neighbors

□ Propagate Through

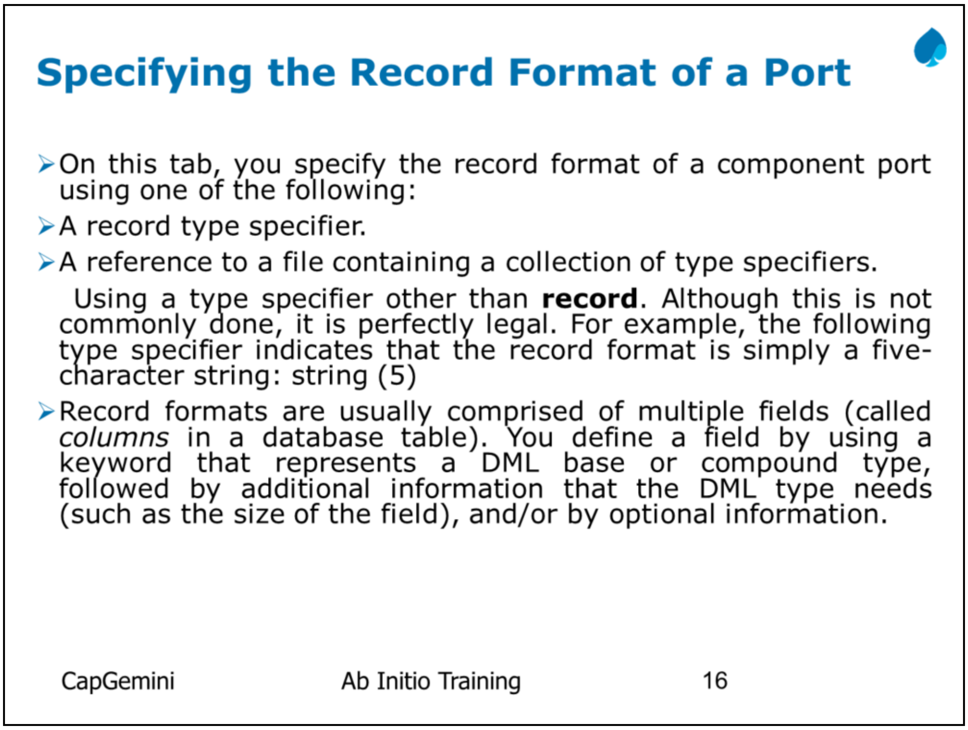
○ Same As:

○ Path:

○ Local File ○ Repository ○ Host File

● Embedded:

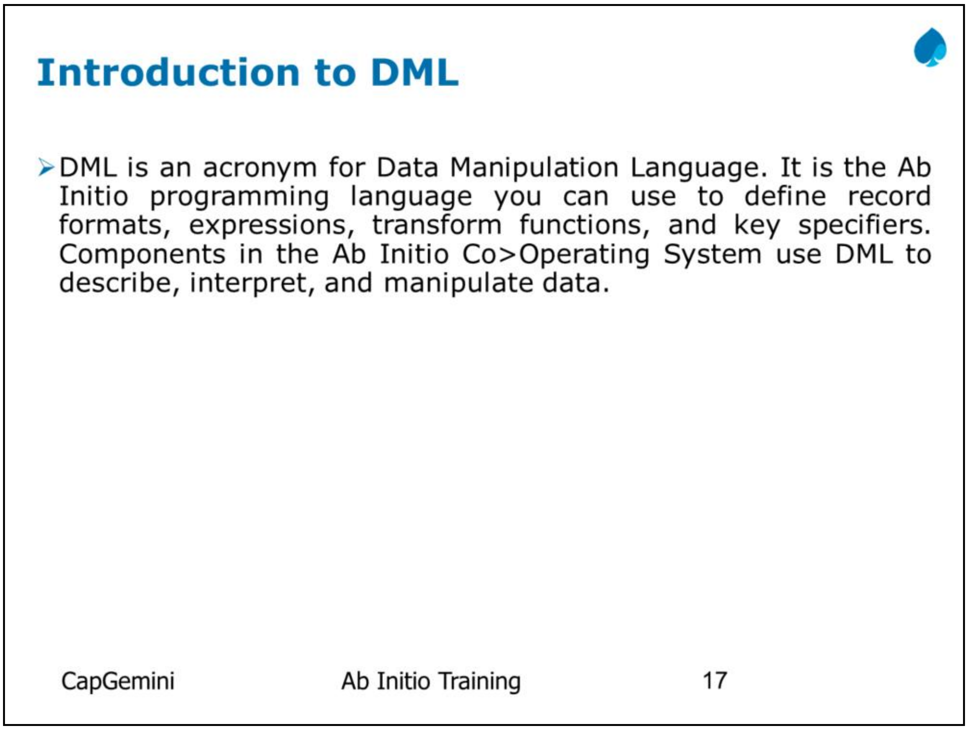
```
record
  string(6) first_name;
  string(6) last_name;
  decimal(3) age;
```



Specifying the Record Format of a Port

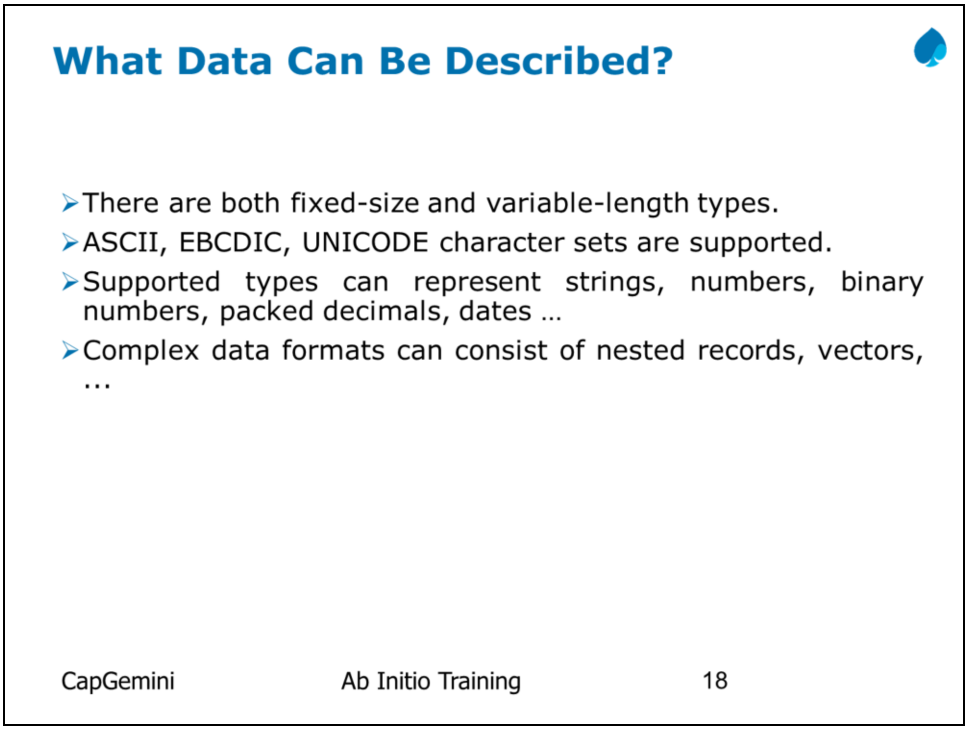
- On this tab, you specify the record format of a component port using one of the following:
- A record type specifier.
- A reference to a file containing a collection of type specifiers.

  Using a type specifier other than **record**. Although this is not commonly done, it is perfectly legal. For example, the following type specifier indicates that the record format is simply a five-character string: string (5)
- Record formats are usually comprised of multiple fields (called *columns* in a database table). You define a field by using a keyword that represents a DML base or compound type, followed by additional information that the DML type needs (such as the size of the field), and/or by optional information.



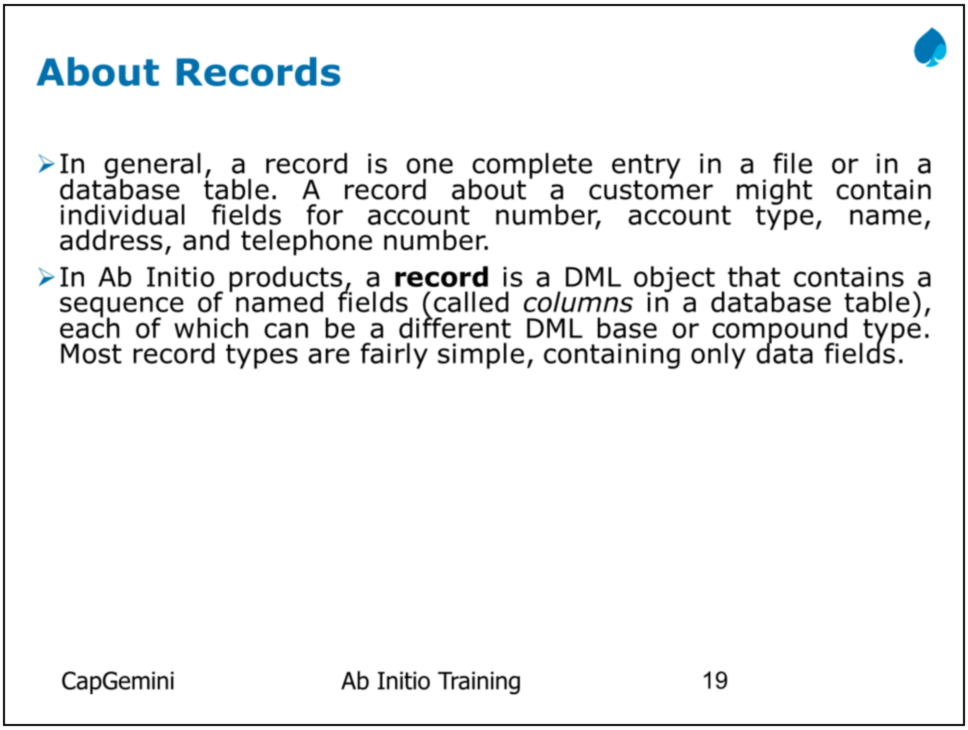
Introduction to DML

- DML is an acronym for Data Manipulation Language. It is the Ab Initio programming language you can use to define record formats, expressions, transform functions, and key specifiers. Components in the Ab Initio Co>Operating System use DML to describe, interpret, and manipulate data.



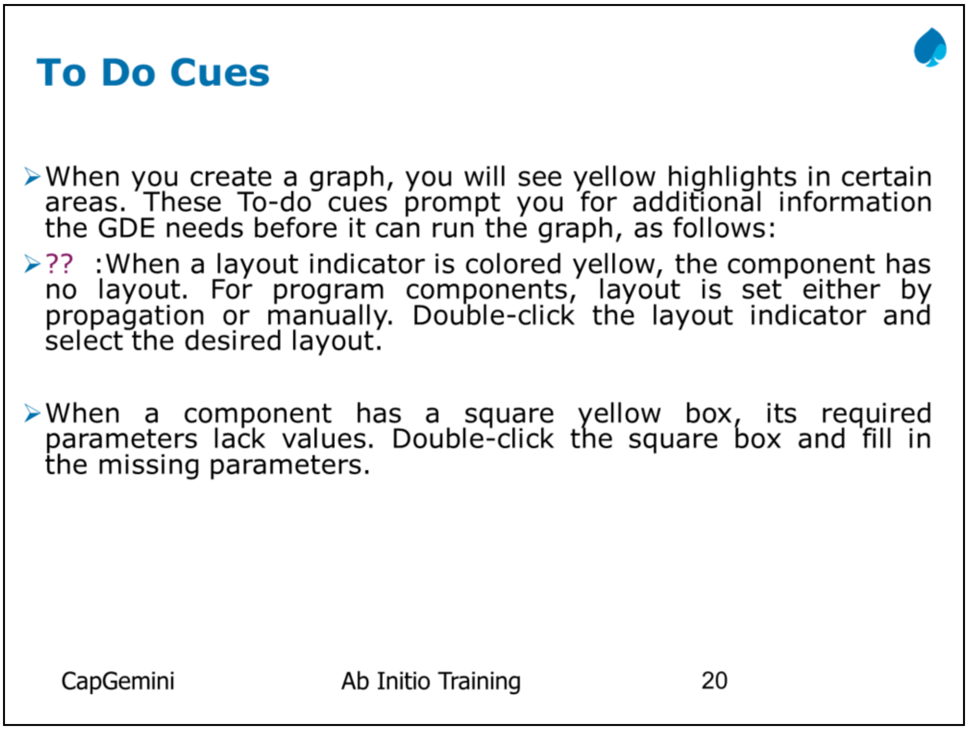
What Data Can Be Described?

- There are both fixed-size and variable-length types.
- ASCII, EBCDIC, UNICODE character sets are supported.
- Supported types can represent strings, numbers, binary numbers, packed decimals, dates …
- Complex data formats can consist of nested records, vectors, …



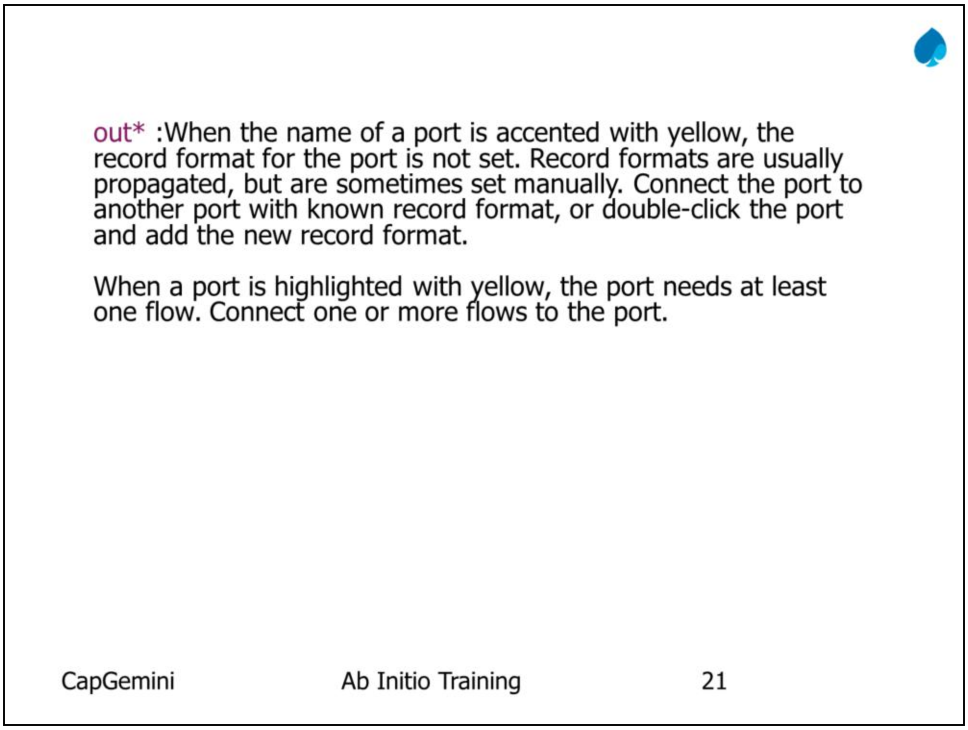
About Records

- In general, a record is one complete entry in a file or in a database table. A record about a customer might contain individual fields for account number, account type, name, address, and telephone number.
- In Ab Initio products, a **record** is a DML object that contains a sequence of named fields (called *columns* in a database table), each of which can be a different DML base or compound type. Most record types are fairly simple, containing only data fields.



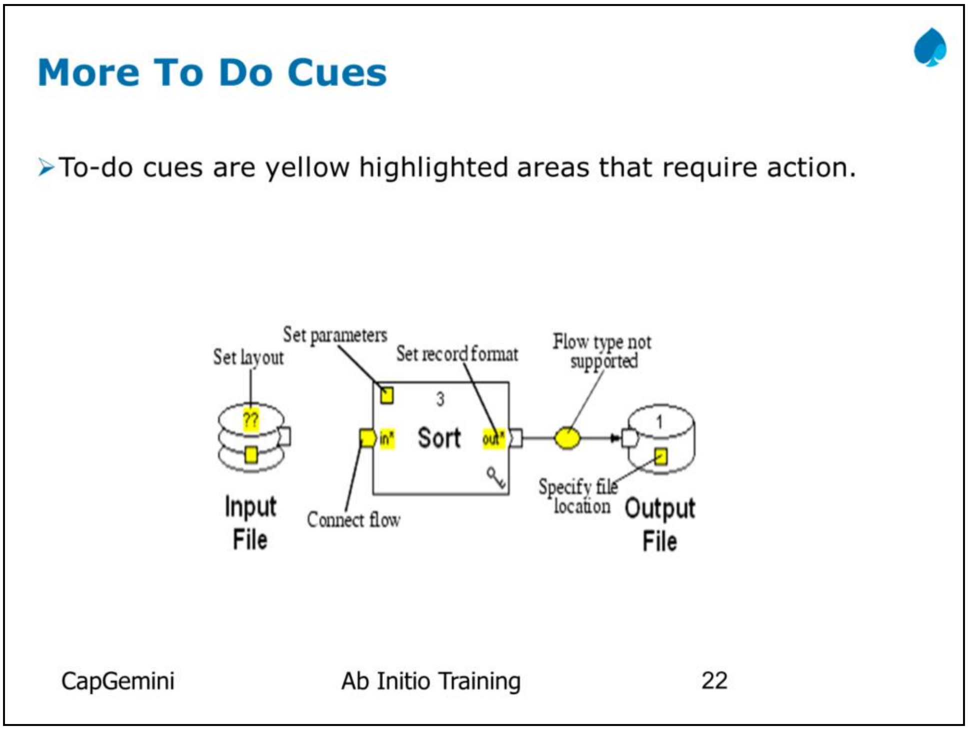
To Do Cues

- When you create a graph, you will see yellow highlights in certain areas. These To-do cues prompt you for additional information the GDE needs before it can run the graph, as follows:
- ?? :When a layout indicator is colored yellow, the component has no layout. For program components, layout is set either by propagation or manually. Double-click the layout indicator and select the desired layout.
- When a component has a square yellow box, its required parameters lack values. Double-click the square box and fill in the missing parameters.



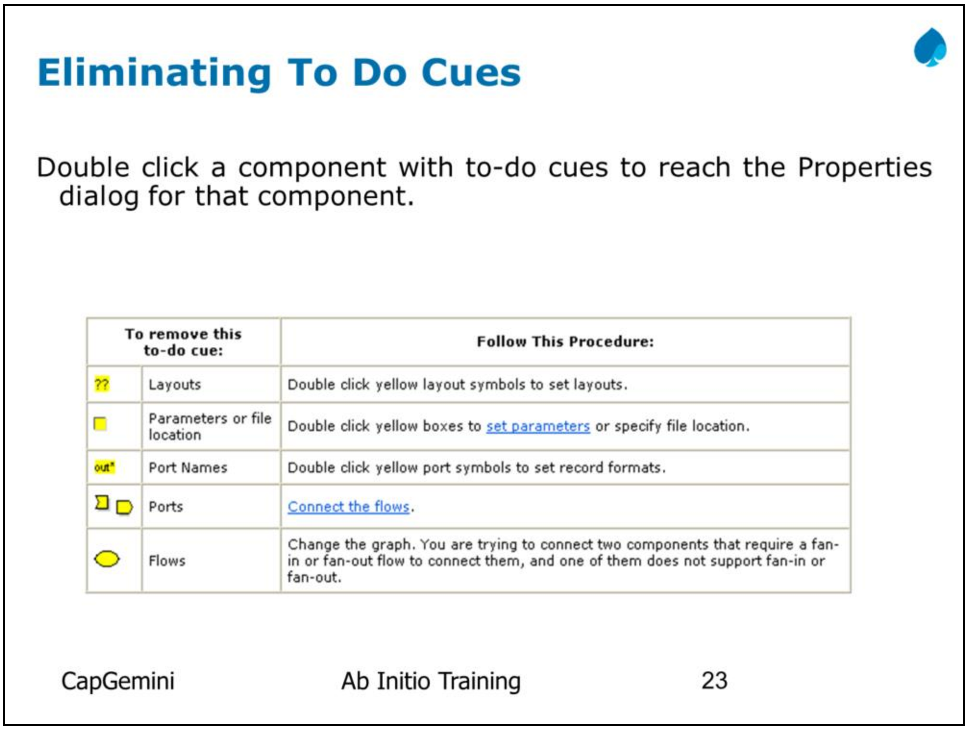
out* :When the name of a port is accented with yellow, the record format for the port is not set. Record formats are usually propagated, but are sometimes set manually. Connect the port to another port with known record format, or double-click the port and add the new record format.

When a port is highlighted with yellow, the port needs at least one flow. Connect one or more flows to the port.



More To Do Cues

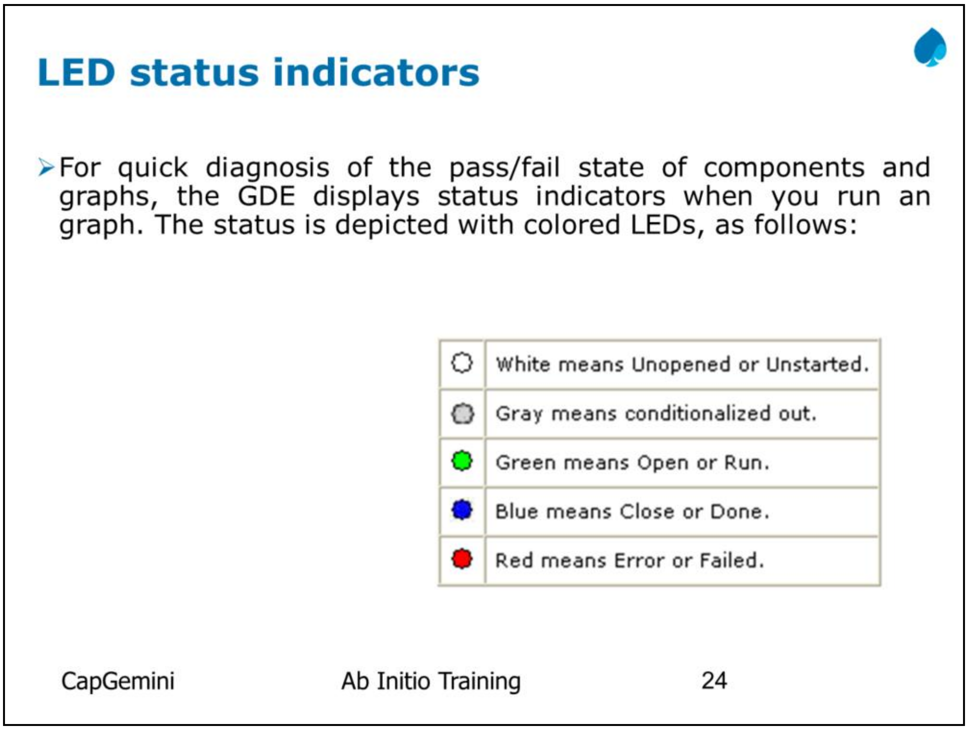
- To-do cues are yellow highlighted areas that require action.



Eliminating To Do Cues

Double click a component with to-do cues to reach the Properties dialog for that component.

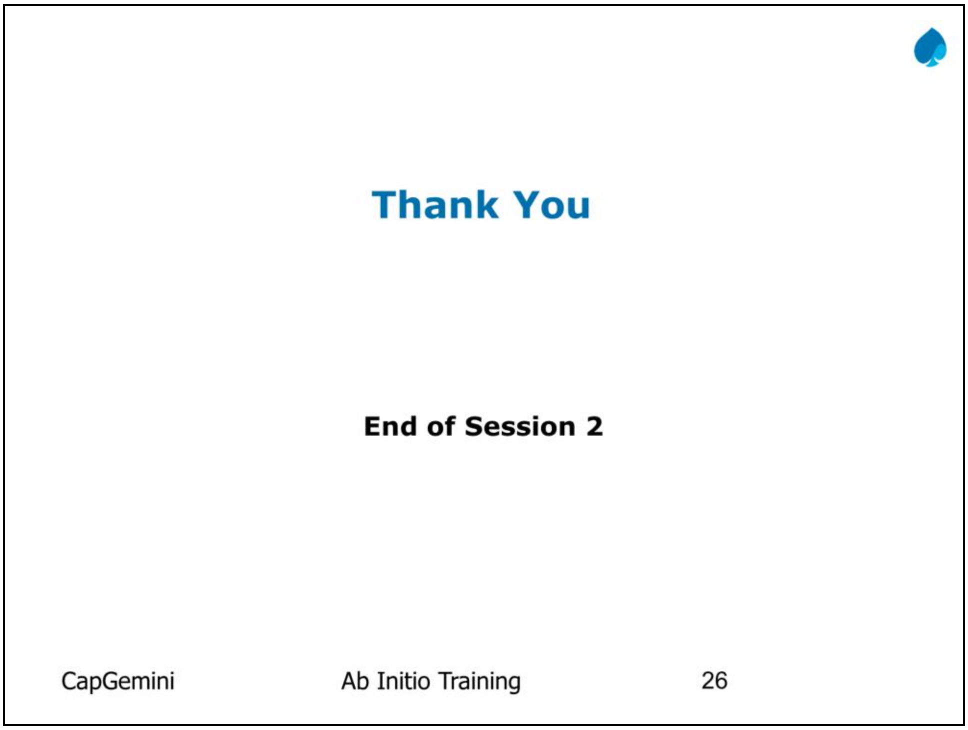
To remove this to-do cue:		Follow This Procedure:
??	Layouts	Double click yellow layout symbols to set layouts.
🟨	Parameters or file location	Double click yellow boxes to set parameters or specify file location.
out*	Port Names	Double click yellow port symbols to set record formats.
□ □	Ports	Connect the flows.
⬭	Flows	Change the graph. You are trying to connect two components that require a fan-in or fan-out flow to connect them, and one of them does not support fan-in or fan-out.



LED status indicators

- For quick diagnosis of the pass/fail state of components and graphs, the GDE displays status indicators when you run an graph. The status is depicted with colored LEDs, as follows:

○	White means Unopened or Unstarted.
◍	Gray means conditionalized out.
🟢	Green means Open or Run.
🔵	Blue means Close or Done.
🔴	Red means Error or Failed.



Thank You

End of Session 2