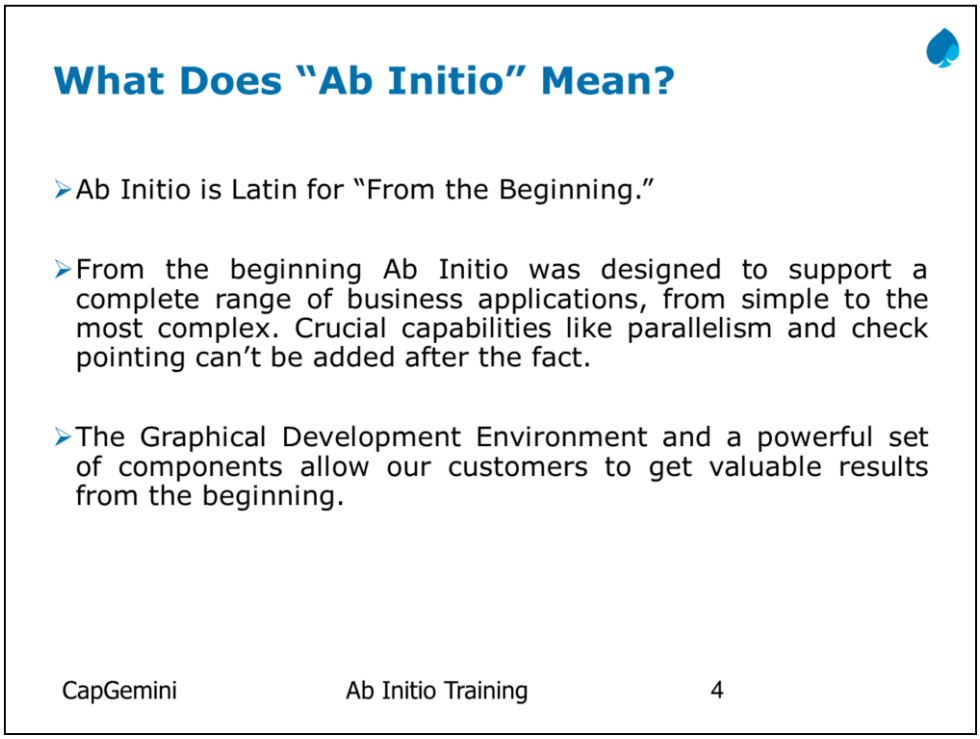

Ab initio Session 1
Introduction to
Ab Initio
Capgemini
Ab Initio Training
1

History of Ab Initio

- **Ab Initio Software Corporation** was founded in the mid **1990's** by Sheryl Handler, the former CEO at Thinking Machines Corporation, after TMC filed for bankruptcy. In addition to Handler, other former TMC people involved in the founding of Ab Initio included Cliff Lasser, Angela Lordi, and Craig Stanfill.
- Ab Initio is known for being very secretive in the way that they run their business, but their software is widely regarded as top notch.

History of Ab Initio

- The Ab Initio software is a fourth generation data analysis, batch processing, data manipulation graphical user interface (GUI)-based parallel processing tool that is used mainly to extract, transform and load data.
- The Ab Initio software is a suite of products that together provides platform for robust data processing applications. The Core Ab Initio Products are: The [Co>Operating System] The Component Library The Graphical Development Environment.



Ab Initio software helps you build large-scale data processing applications and run them in parallel environments.

Ab Initio software consists of two main programs:

Ab Initio's focus

- "Moving Data"
  - move small and large volumes of data in an efficient manner
  - deal with the complexity associated with business data
- High Performance
  - scalable solutions
- Better productivity

Ab Initio's Software

- Ab Initio software is a general-purpose data processing platform for mission-critical applications such as:
  - Data warehousing
  - Batch processing
  - Click-stream analysis
  - Data movement
  - Data transformation

Applications of Ab Initio Software

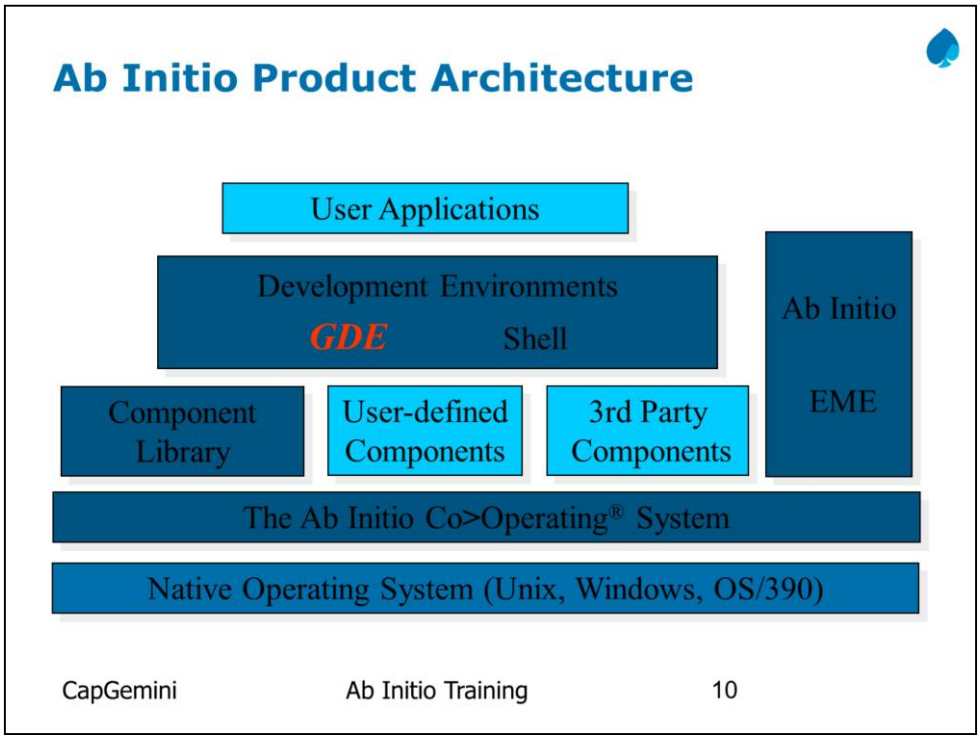
- Processing just about any form and volume of data.
- Parallel sort/merge processing.
- Data transformation.
- Rehosting of corporate data.
- Parallel execution of existing applications.

Ab Initio Provides For:

- Distribution - a platform for applications to execute across a collection of processors within the confines of a single machine or across multiple machines.
- Reduced Run Time Complexity - the ability for applications to run in parallel on any combination of computers where the Ab Initio Co>Operating System is installed from a single point of control.

Applications of Ab Initio Software in terms of Data Warehouse

- Front end of Data Warehouse:
  - Transformation of disparate sources
  - Aggregation and other preprocessing
  - Referential integrity checking
  - Database loading
- Back end of Data Warehouse:
  - Extraction for external processing
  - Aggregation and loading of Data Marts


Ab Initio Product Architecture
User Applications
Development Environments
GDE
Shell
Ab Initio
EME
Component Library
User-defined Components
3rd Party Components
The Ab Initio Co>Operating® System
Native Operating System (Unix, Windows, OS/390)
CapGemini
Ab Initio Training
10

GDE :

Ab Initio Architecture-Explanation

- The Ab Initio Co operating system unites the network of computing resources-CPUs, storage disks , programs , datasets into a production quality data processing system with scalable performance and mainframe class reliability.
- The Cooperating system is layered on the top of the native operating systems of the collection of servers .It provides a distributed model for process execution, file management ,debugging, process monitoring , checkpointing .A user may perform all these functions from a single point of control.

Co>Operating System Services

- Parallel and distributed application execution
  - Control
  - Data Transport
- Transactional semantics at the application level.
- Checkpointing.
- Monitoring and debugging.
- Parallel file management.
- Metadata-driven components.

Ab Initio: What We Do

- Ab Initio software helps you build large-scale data processing applications and run them in parallel environments. Ab Initio software consists of two main programs:
- **Co>Operating System:**

  which your system administrator installs on a *host* Unix or Windows NT server, as well as on processing computers.
- **The Graphical Development Environment (GDE):**

  which you install on your PC (*GDE Computer*) and configure to communicate with the host.

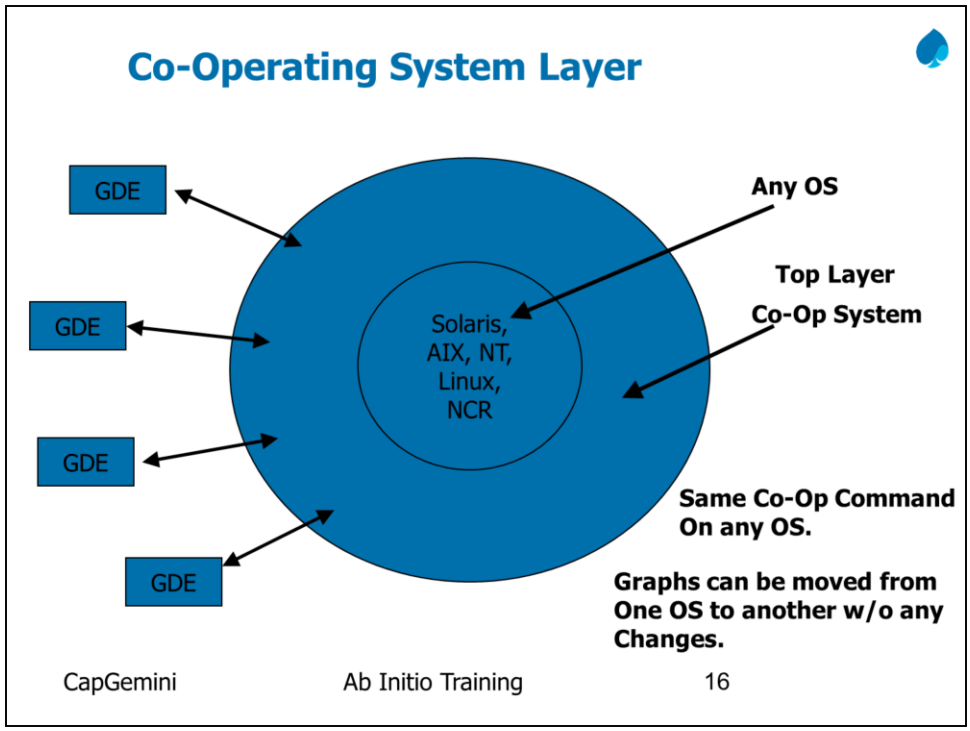
CapGemini Ab Initio Training 13

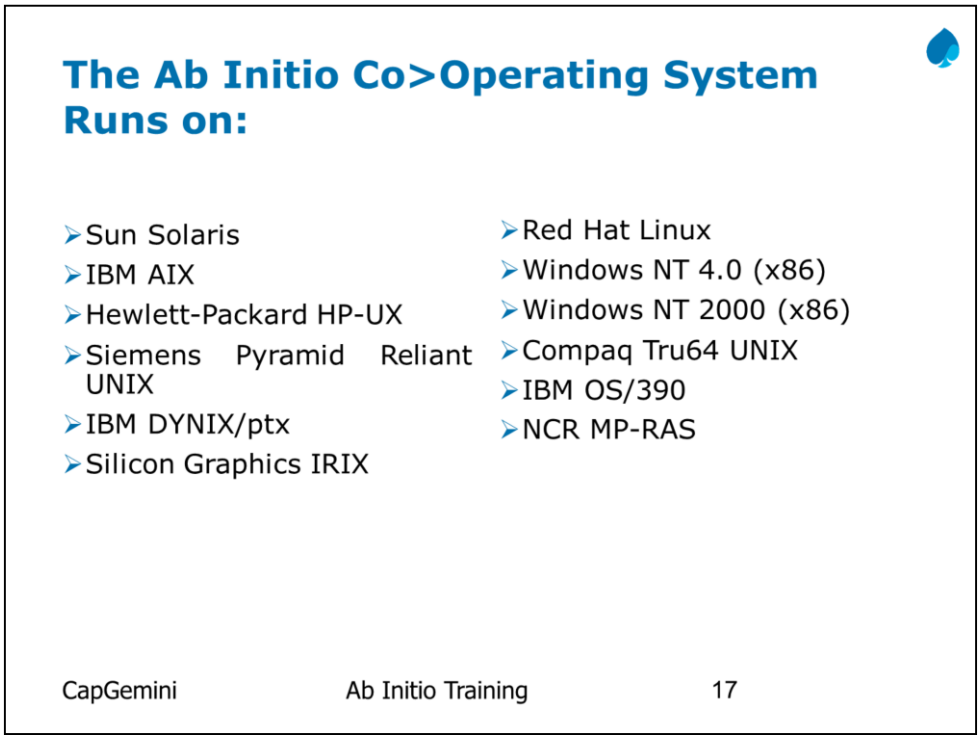
The Ab Initio Co>Operating® System

- **The Co>Operating System** Runs across a variety of Operating Systems and Hardware Platforms including OS/390 on Mainframe, Unix, and Windows. Supports distributed and parallel execution. Can provide scalability proportional to the hardware resources provided. Supports platform independent data transport.

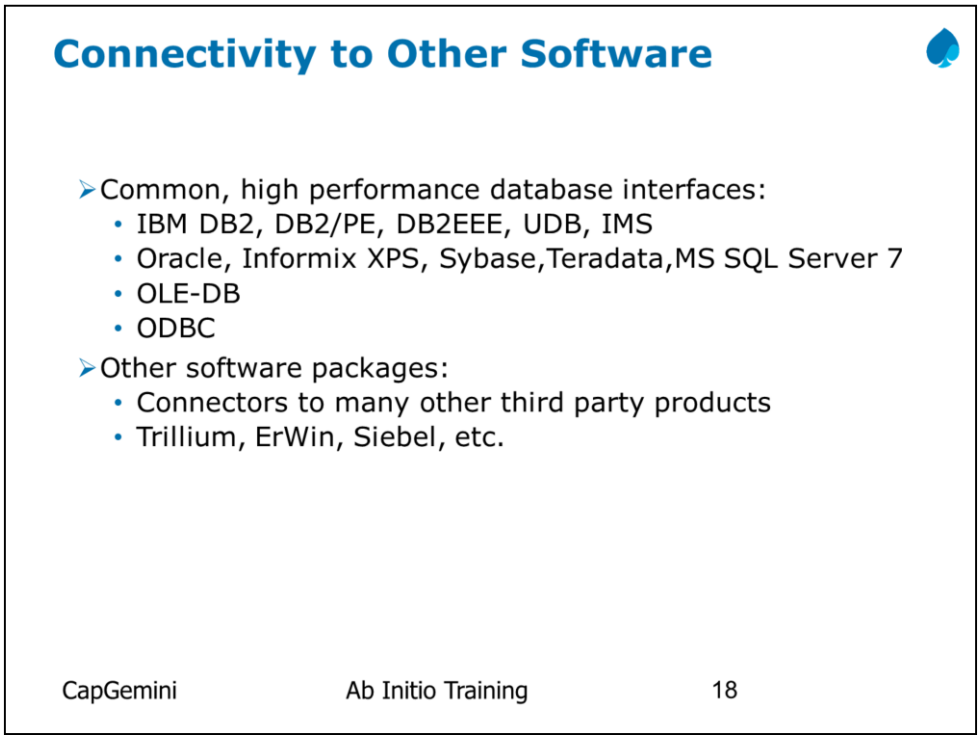
The Ab Initio Co>Operating® System-Continued

The Ab Initio Co>Operating System depends on parallelism to connect (i.e., cooperate with) diverse databases. It extracts, transforms and loads data to and from Teradata and other data sources.


Co-Operating System Layer
GDE
GDE
GDE
GDE
Solaris,
AIX, NT,
Linux,
NCR
Any OS
Top Layer
Co-Op System
Same Co-Op Command
On any OS.
Graphs can be moved from
One OS to another w/o any
Changes.
CapGemini
Ab Initio Training
16

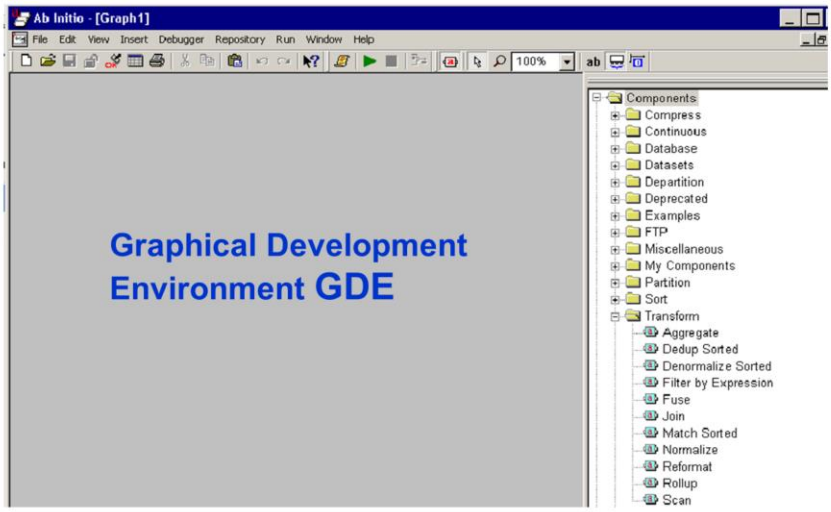


Who is this aimed at? For a course held here, this might be useful. For a course held at a client site, they probably have a specific environment. Do we have a 1 page brochure that says all this?



See comments on last notes page.

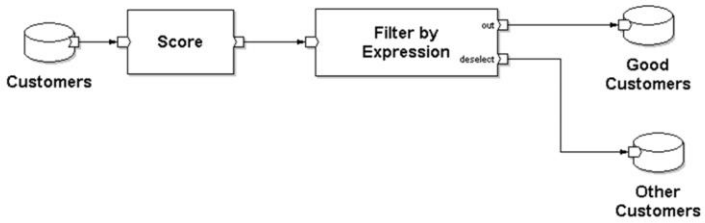
The GDE


Ab Initio - [Graph1]
File Edit View Insert Debugger Repository Run Window Help
100%
ab
Graphical Development Environment GDE
Components
Compress
Continuous
Database
Datasets
Departition
Deprecated
Examples
FTP
Miscellaneous
My Components
Partition
Sort
Transform
Aggregate
Dedup Sorted
Denormalize Sorted
Filter by Expression
Fuse
Join
Match Sorted
Normalize
Reformat
Rollup
Scan

The GDE

The Graphical Development Environment (GDE) provides a graphical user interface into the services of the Co>Operating System. **The Graphical Development Environment** Enables you to create applications by dragging and dropping Components. Allows you to point and click operations on executable flow charts. The Co>Operating System can execute these flowcharts directly. Graphical monitoring of running applications allows you to quantify data volumes and execution times, helping spot opportunities for improving performance.

The Graph Model



The Component Library:

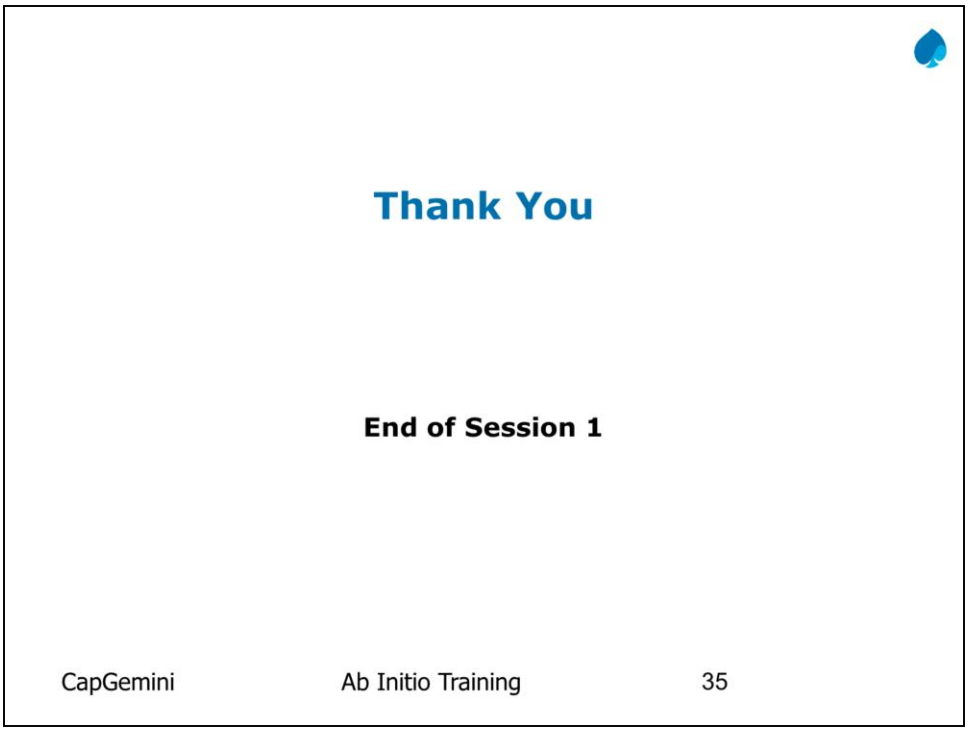
- **The Component Library:** Reusable software Modules for Sorting, Data Transformation, database Loading Etc. The components adapt at runtime to the record formats and business rules controlling their behavior.
- Ab Initio products have helped reduce a project's development and research time significantly.

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.

EME

- The Enterprise Meta>Environment (EME) is a high-performance object-oriented storage system that inventories and manages various kinds of information associated with Ab Initio applications. It provides storage for all aspects of your data processing system, from design information to operations data.
- The EME also provides rich store for the applications themselves, including data formats and business rules. It acts as hub for data and definitions . Integrated metadata management provides the global and consolidated view of the structure and meaning of applications and data- information that is usually scattered throughout you business .


Thank You
End of Session 1
CapGemini
Ab Initio Training
35