Welcome!

Module 1

- Data Warehouse – Business Intelligence Concepts

"A collection of integrated, subject-oriented databases designed to support the DSS function where each unit of data is relevant to some moment in time..."

Inmon, Imhoff and Sousa, The Corporate Information Factory

"A copy of transaction data specifically structured for query and analysis."

Ralph Kimball, The Data Warehouse Toolkit

IBM Global Business Services

Data Warehouse – Concepts

Date:

Duration of course:

Introduction

About Me

- **Parwaz Dalvi**
  Sr. Architect / Consultant DW-BI & Database

  TOGAF 8 Certified (The Open Group Architecture Framework)

My Session For you

- **Data Warehouse Concepts**

Session‘s Objective

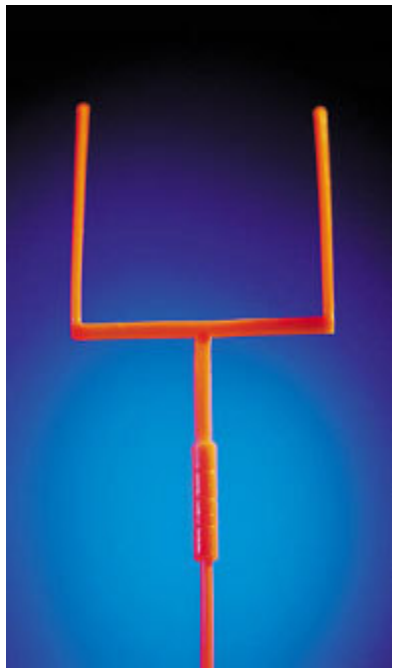
- Understand what Data Warehousing means
- Realize the Need, Advantages & Challenges in implementation of a DW Solution
- Understand Data Warehouse Architecture and its components
- Understand IBM Reference DW-BI Architecture
- Understand IBM’s IOD initiative and realize how DW-BI helps in achieving this objective
- Know the DW-BI Tools and Products, the trends in DW-BI
- Know your Growth Prospects in the DW-BI Arena within IBM

Course Content

Module	Content	Duration
1	***Data Warehouse Evolution***	
2	Data Warehouse Concepts	
3	Data Warehouse Architecture – Part 1 – GENERIC	
4	Data Modeling in DW-BI	
5	Data Warehouse Architecture – Part 2 – SPECIFIC	
6	DW-BI - IBM Reference Architecture & IOD	
7	DW-BI Tools and Products	
8	Trends in DW-BI	
9	Growth Path of DW-BI Professionals	

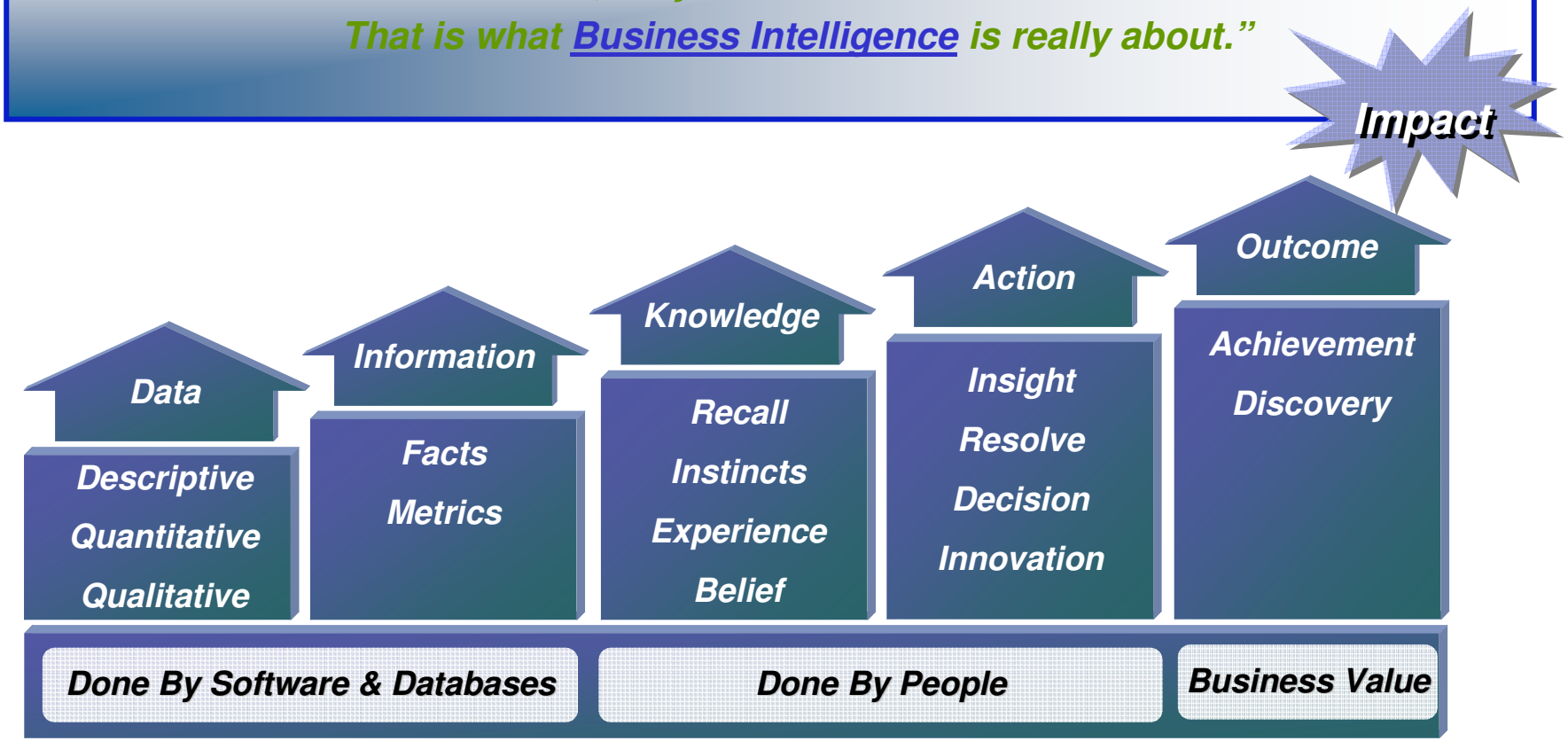
Course Objectives

- Upon completing this course, you should be able to:
  - Reason the Evolution of Data Warehouse & BI
  - Understand Data Warehouse Concepts
  - Differentiate between OLTP, OLAP, ODS, DW, BI, & DSS
  - Relate Data Warehouse Architecture & its Components
  - Relate Ralph Kimball and Bill Inmon Approaches to DW
  - Understand ETL, Metadata, MDM, DQ, Analytics, Mining
  - Understand IBM Reference Architecture
  - IBM IOD Initiatives and its relation to DW-BI
  - Know DW-BI Products & Tools
  - Trends In DW-BI & Growth path for all of us



Data, Information & Knowledge!

" If you give somebody a fish, they can eat for the rest of the week, but if you teach someone to fish, they can eat for the rest of their lives.
That is what Business Intelligence is really about."



Data, Information & Knowledge!

- ***Data*** *- Is Composed of individual discreet facts that collect descriptive, quantitative, and qualitative value of business interests. Data Warehousing involves two types of data –*
  - ***Operational Data*** *: It describes the day to day events and transactions of the business*
  - ***Informational Data*** *: It is one that is reconciled, integrated, and cleaned to constitute the raw material from which information is constructed*

- ***Information*** *- Is an organized collection of data presented in a specific and meaningful way*

- ***Knowledge*** ***-*** *It encompasses the familiarity, awareness, understanding, and perceptions of a person about a given subject*

Data, Information & Knowledge!

- ***Action*** *- Is the process of doing something. Effective action is the process of doing the right thing.*

- ***Outcomes*** *- These are the result of actions.* ***Favorable*** *business actions are those that reduce cost, save time, optimize resources, increase revenue, satisfy clients, or otherwise fulfill the business mission & goals.*

- ***Impact & Value*** *– Value is realized at the bottom line of business – when outcome reduce costs or increase revenue, either directly or indirectly.*
  - *The Value of an Action is determined by the* ***Outcomes Produced***
  - *The Value of Information is derived through contribution to valued action –* ***Providing support for insight, Resolve, Decision, & Innovation***
  - *The Value of a* ***Data Warehouse*** *depends entirely on the* ***value of Information Services that it delivers***

IBM Global Business Services

Course Title:

Module 1 : Data Warehouse Evolution

Module Objectives

- At the completion of this chapter you should be able to understand:
  - Data Warehouse Definitions
  - Data Warehouse Evolution
  - Why build a Data Warehouse
  - Advantages of Data Warehouse

Module 1: Evolution of Data Warehouse: Agenda

- Topic 1. Data Warehouse Definitions
- Topic 2. Data Warehouse Evolution
- Topic 3. Need for Data Warehouse & its Advantages

Module 1: > Topic 1: Data Warehouse Definitions

- Definitions of Data Warehouse
- Definition of Business Intelligence
- Relationship between Data Warehouse and Business Intelligence

Module 1: > Topic 1: Data Warehouse Definitions

Definitions of a Data Warehouse

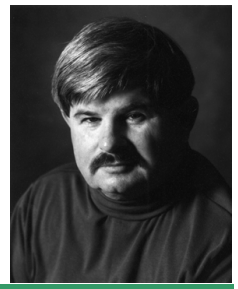
A Data Warehouse is a subject-oriented, integrated, nonvolatile, time-variant collection of data in support of management's decisions

- WH Inmon Newsletter July/August 1992

I want to see the Credit Report for Accounts, Sales & Marketing for this customer

Data From Various Scources is Integrated for a subject or business process

Data Store for historical period & is populated on a daily/weekly basis



WH Inmon - Father Of Data Warehousing

Module 1: > Topic 1: Data Warehouse Definitions

Definitions of a Data Warehouse

A Data Warehouse is a nothing more than the union of all the constituent Data Marts

- Ralph Kimball, The Data Warehouse Lifecycle Toolkit, 1998

A collection of integrated, subject oriented databases designed to support the DSS function, where each unit of data is relevant to some moment of time

- Imhoff

Data Warehouse is a repository of data summarized or aggregated in simplified form from operational systems. End user orientated data access and reporting tools let user get at the data for decision support

- Babcock

Module 1: > Topic 1: Data Warehouse Definitions

Definition of Business Intelligence

Business intelligence (BI) - Refers to technologies, applications and practices for the collection, integration, analysis, and presentation of business information and sometimes to the information itself.

The purpose of business intelligence is to support better business decision making

Business intelligence applications can be:

Mission-critical and integral to an enterprise's operations or occasional to meet a special requirement

Enterprise-wide or local to one division, department, or project

Centrally initiated or driven by user demand

Module 1: > Topic 1: Data Warehouse Definitions

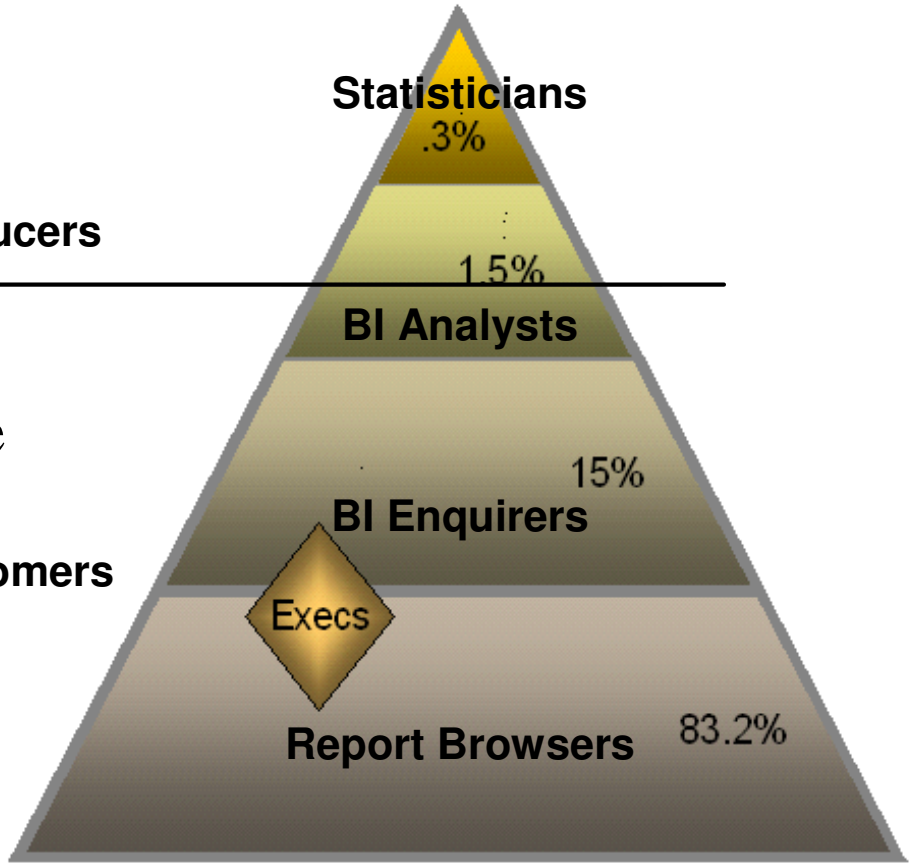
Types of BI ***Information Pyramid – Analyst Role Based***

Inferential

Information Producers

Descriptive

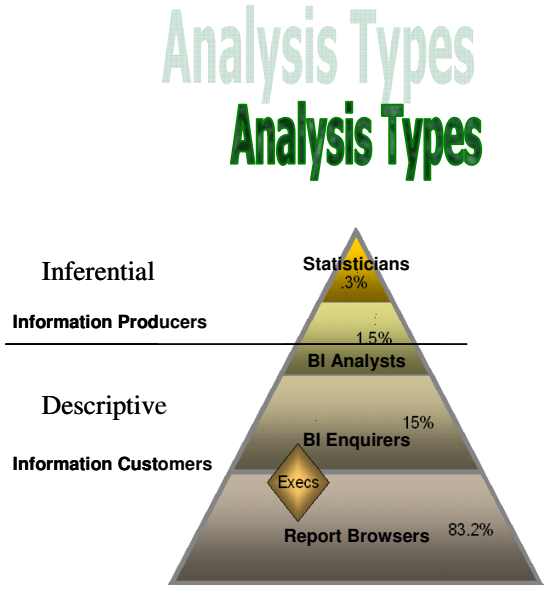
Information Customers



Module 1: > Topic 1: Data Warehouse Definitions

Types of BI

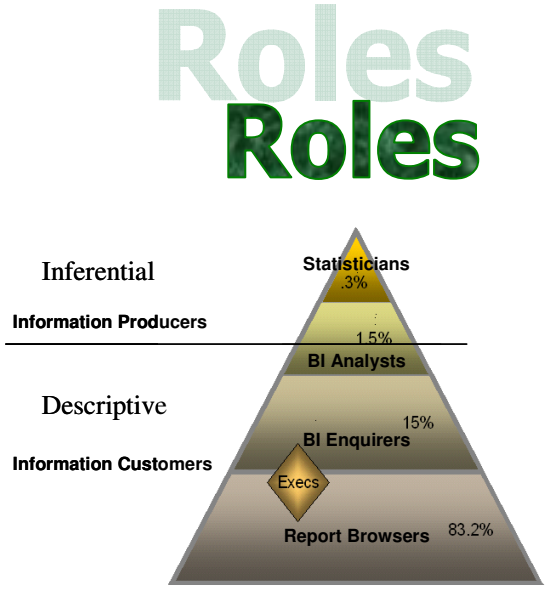
- ***Inferential Analysis*** *leverages statistical techniques to develop inferences to forecast results or to validate that the results of an experimental model provide support for a particular theory or practical innovation. If successful, the model can be generalised to the target population that it represents. An effective model creates incremental revenue by driving revenue-positive decisions consistently across the entire corporation*
- ***Descriptive Analysis*** *leverages more basic metrics such as mean, mode, standard deviation, range and data distribution to provide insight into historical business phenomena. The revenue impact of this information tends to be localised to the individual who is using it to make decisions, and the ability of that individual to leverage information to improve results within the individual's span of control*



Module 1: > Topic 1: Data Warehouse Definitions

Types of BI

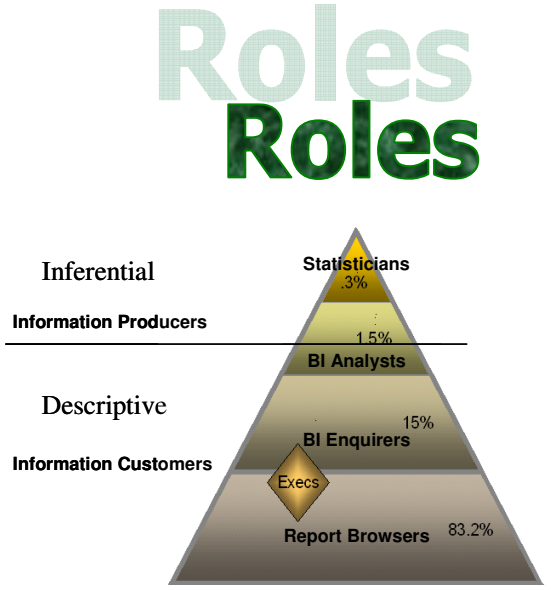
- ***Information producers*** *are those individuals who develop formal reports, build dimensional databases, implement analytic environments and create analytic metrics and models. They require BI tools that offer complex functionality or advanced analytics*
- ***Information consumers*** *are decision-makers, such as business managers, executives and front line workers. They need information to be presented in a familiar, repeatable, easy-to-access format. Above all, they need information that is actionable*
- *Some analysts fit both categories, functioning as information producers who develop information for their own consumption*
- ***Statisticians*** *employ complex statistical and mathematical tools to analyse relatively raw data. They are usually organised to support significant segments of the business through project-based research*



Module 1: > Topic 1: Data Warehouse Definitions

Types of BI

- ***BI Analysts*** *are sometimes described as knowledge workers, performing ad hoc analysis to explore more immediate opportunities and risks from raw or prepared data. They are usually organised to support particular business units or a specific function that crosses business units*
- ***BI Enquirers*** *are business users who have generally well-defined requirements with a need for adaptable, on-demand reporting with some drill-down facilities*
- ***Report Browsers*** *are business users with very specific, seldom-changing information needs. They require highly structured BI access, including prepared reports or targeted queries that provide, for example, results of work carried out,* actionable data, or high-level business performance

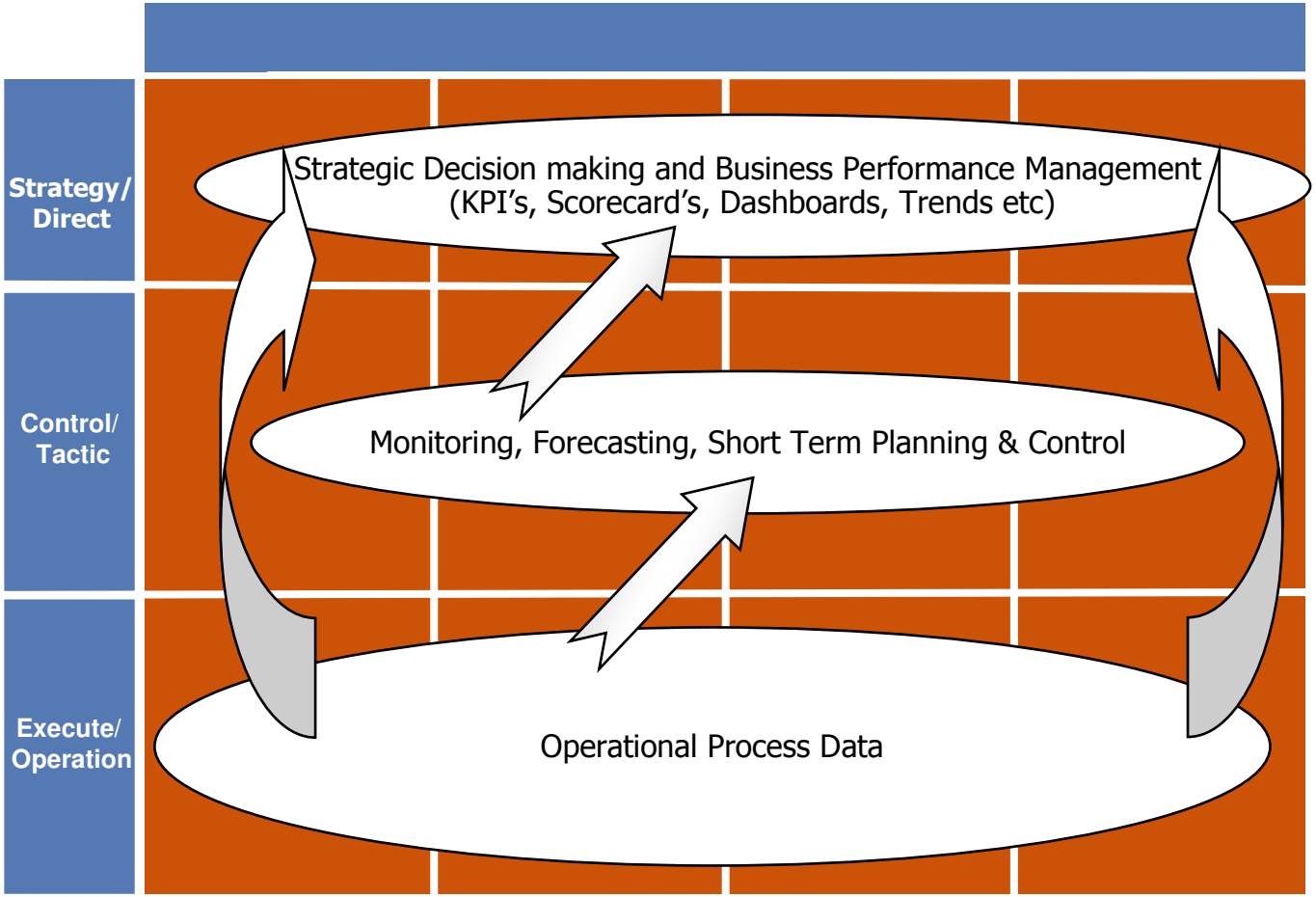


Module 1: > Topic 1: Data Warehouse Definitions

Types of BI

Analysis Types

Analysis Types



Module 1: > Topic 1: Data Warehouse Definitions

Relationship between DW and BI

- *Data Warehousing*

  Is a way of storing data and creating information through leveraging data marts. DM's are segments or categories of information and/or data that are grouped together to provide 'information' into that segment or category. DW does not require BI to work. Reporting tools can generate reports from the DW

- *Business Intelligence*

  Is the leveraging of DW to help make business decisions and recommendations. Information and data rules engines are leveraged here to help make these decisions along with statistical analysis tools and data mining tools

Module 1: > Topic 1: Data Warehouse Definitions Summary

- Having completed this topic, you should be able to:
  - Know Definitions of Data Warehouse
  - Know definitions of Business Intelligence
  - Relate Difference between Data Warehouse and Business Intelligence at a broader level

Module 1: > Topic 1: Data Warehouse Definitions Review

Module 1: > Topic 2: Data Warehouse Evolution

- Evolution of DW-BI
- 1960 -1970 - MIS Era
- 1970 – 1980 - Querying Era
- 1980 – 1990 - Naturally Evolving Architecture Era
- 1990-200x - Analysis Era

Module 1: > Topic 2: Data Warehouse Evolution

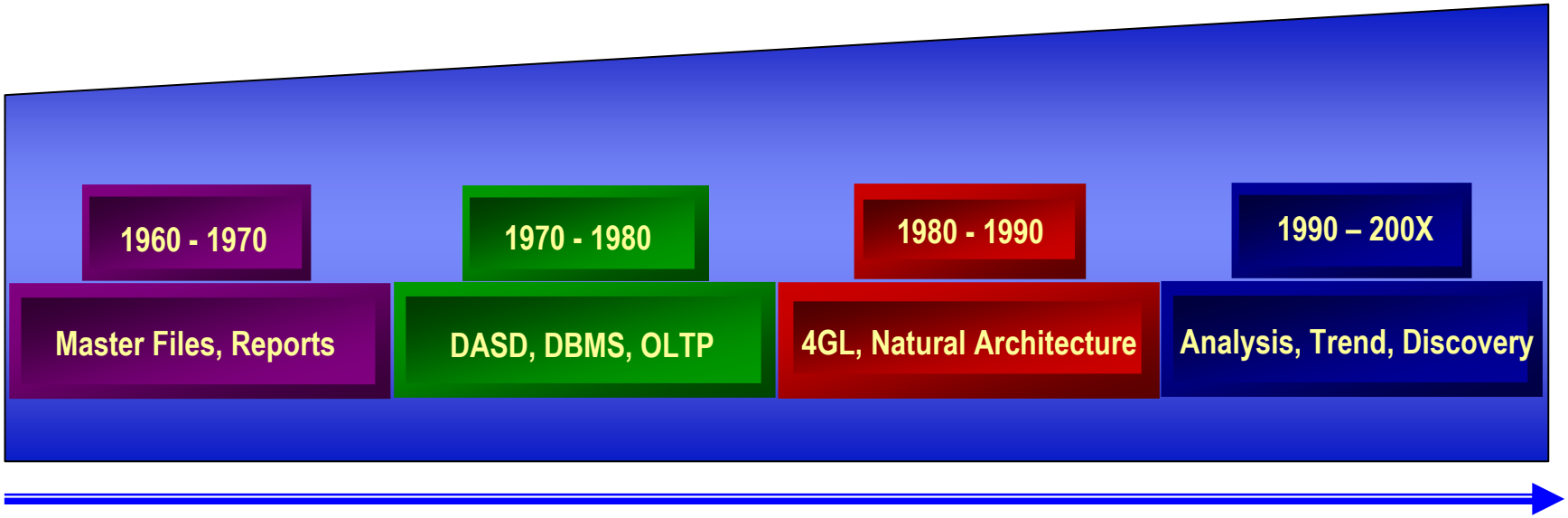
Evolution of DW-BI – Some Facts – How it all began

- Examination of bones found in archeological excavations in Chile shows that medicine — in, at least, a rudimentary form — was practiced as far back as 10,000 years ago
- Some of the streets in Rome were laid out by civil engineers more than 2,000 years ago
- Other professions have roots that can be traced to antiquity
- The profession and practice of information systems and processing are certainly immature, because they have existed only since the early 1960s. Data Warehousing & BI has been evolving simultaneously with the evolution of Information Technology
- In early days the notion of DW-BI was that if we get the details right the end result will somehow take care of itself

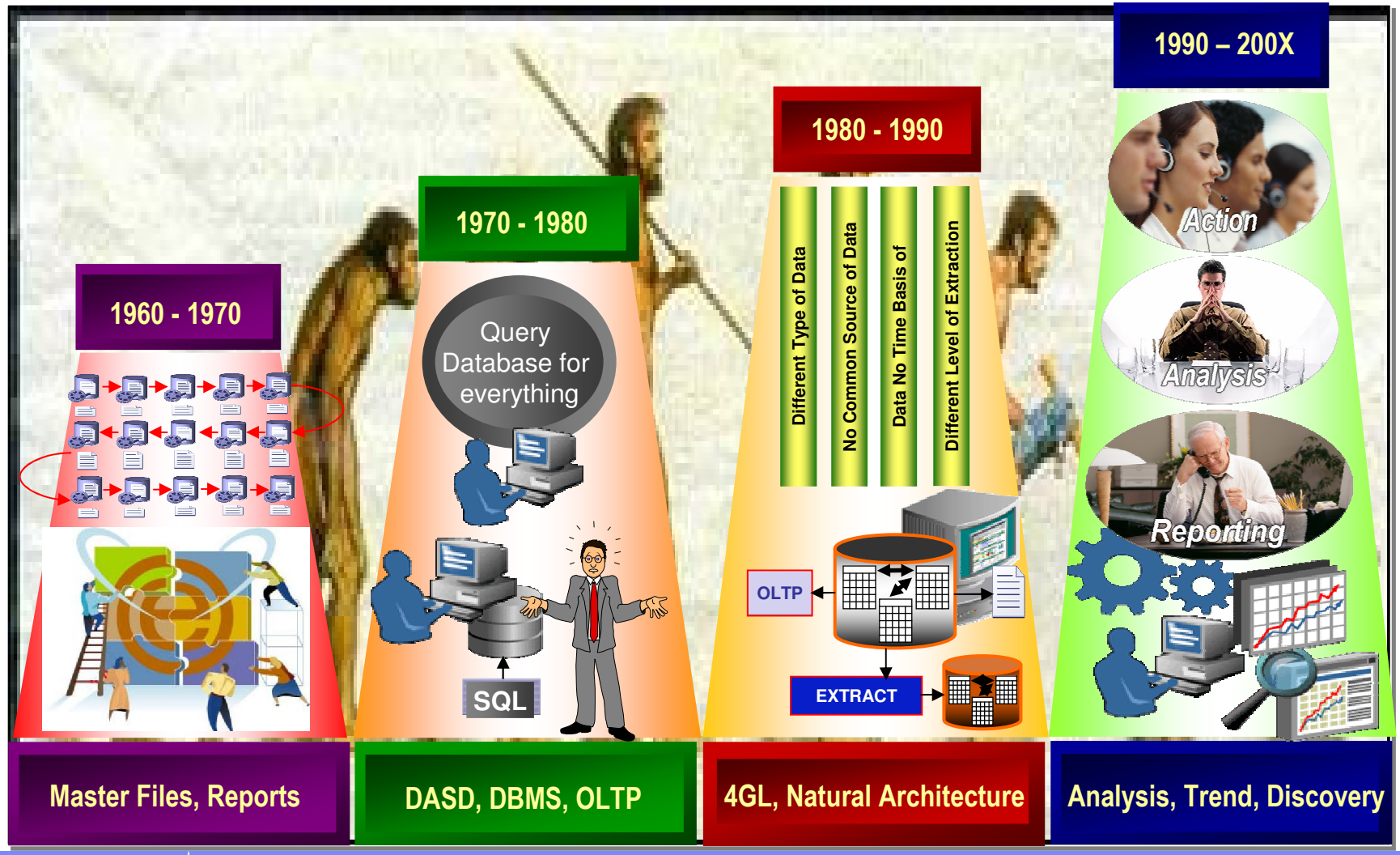
Module 1: > Topic 2: Data Warehouse Evolution

Evolution of DW-BI – Some Facts – How it all began

- The origins of data warehousing and *decision support systems (DSS)* processing hark back to the very early days of computers and information systems. It is interesting that DSS processing developed out of a long and complex evolution of information technology. Its evolution continues even today …
- Broad Classification of Evolution of DW-BI



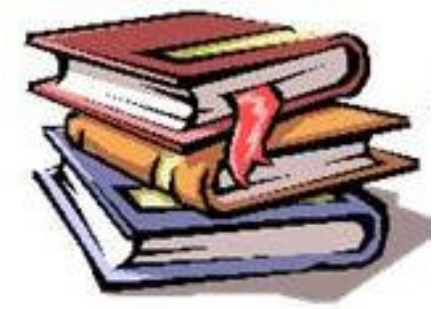
Module 1: > Topic 2: Data Warehouse Evolution

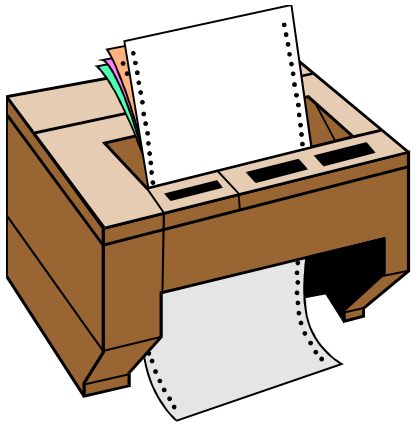


Module 1: > Topic 2: Data Warehouse Evolution

1960 - 1970 : MIS Era – Master Files & Reports

- Unfriendly
- Slow
- Dependent on IS programmers
- Inflexible
- Analysis limited to defined reports



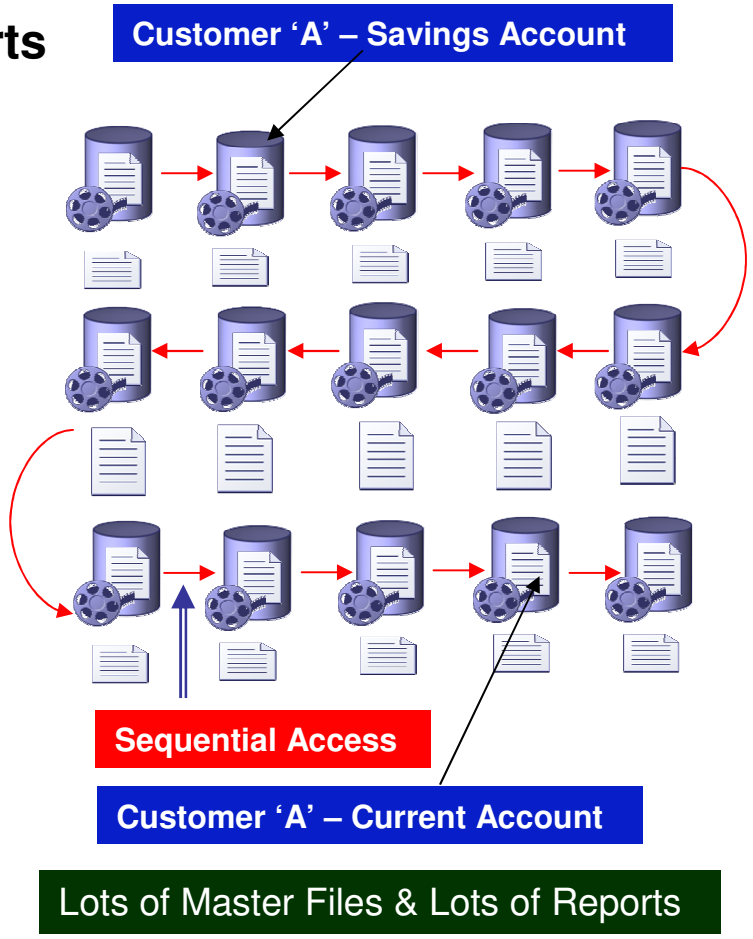


Focus on Reporting Only – Sequential access to data - Manual comparision of data

Module 1: > Topic 2: Data Warehouse Evolution

1960 - 1970 : MIS Era – Master Files & Reports

- Individual applications run using Master Files
- Programs built in early language – FORTRAN, COBOL
- Master Files were housed on Magnetic Tapes
- Magnetic Tapes could store good amount of data but had to be accessed sequentially
- Accessing an entire tape would take as long as 20-30 minutes
- Growth of Master Files & Magnetic Tapes lead to huge amount of redundant data



Module 1: > Topic 2: Data Warehouse Evolution

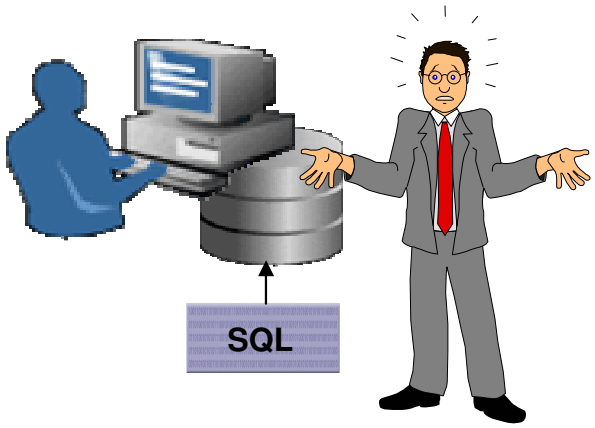
1960 - 1970 : MIS Era – Master Files & Reports

- *Drawbacks -*

– *Complexity of Maintenance & Development*
– *The Complexity of Maintaining Programs*
– *Data Access only to Coders & Programmers*
– *Synchronization of data*
– *Sequential Access to data leading to enormous delay in access of related data*
– *Cost of Hardware to support all Master Files was considerable*
– *Manual Comparison of data*
– *The problems of master files — problems inherent to the medium itself was stifling*

Module 1: > Topic 2: Data Warehouse Evolution

1970 - 1980 : Querying Era – DASD, DBMS, OLTP



- Database – A single source of data for all kind of processing
- Ad hoc, unstructured access to corporate data
- SQL as interface not scalable
- Cannot handle complex analysis

Focus on Querying Only

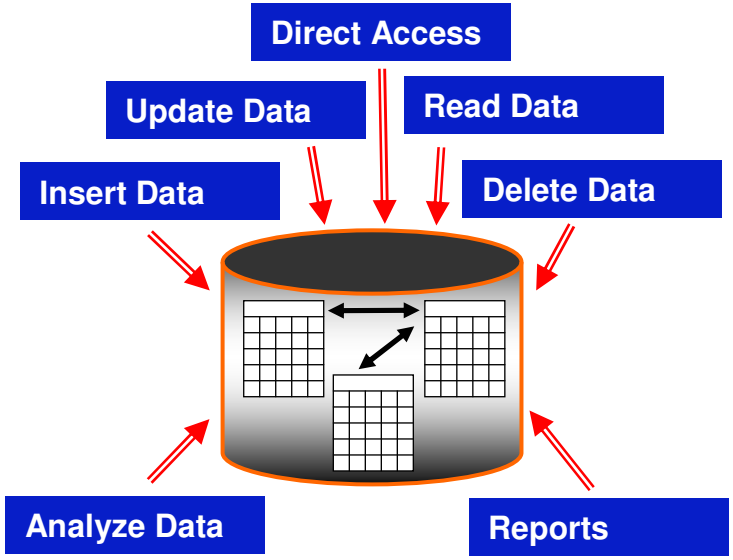
Module 1: > Topic 2: Data Warehouse Evolution

1970 - 1980 : Querying Era – DASD, DBMS, OLTP

- Advent of disk storage or the **direct access storage device (DASD)**. Data could be accessed directly on DASD, no sequential data read, time required to locate a record on the DASD was significantly faster (milliseconds) than scanning a magnetic tape to locate the record

- With the advent of DASD came a new type of system software namely, **database management system (DBMS)**, The purpose of the DBMS was to make it easy for the programmer to store and access data on a DASD

- With DBMS came the notion of Database,
  A Database was defined as a single source of data for all processing.
  By mid 1970's **online transaction processing (OLTP)** made faster access to data possible



Database –

Single source of data for all kind of processing

Module 1: > Topic 2: Data Warehouse Evolution

1970 - 1980 : Querying Era – DASD, DBMS, OLTP

- *Drawbacks -*

– *Corporate data still resided on Silos databases, & comparison of data meant creation of new interfaces and schemas, this was time consuming and involved cost*

– *Emphasis was more on querying data, representation of the data to put in business use was at a very premature level*

– *Operational & Analytical processing happening on the same database at the same time, affecting database performance & in turn affecting the applications running on these databases*

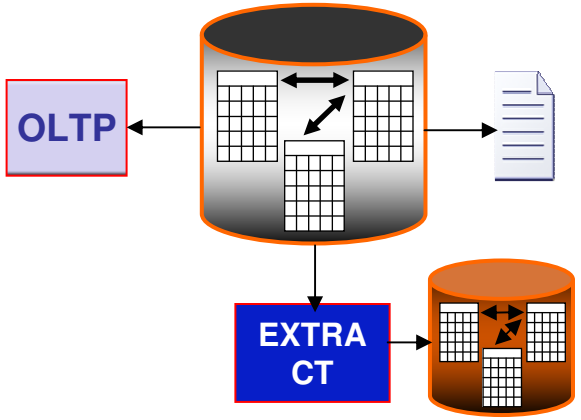
– *SQL provided access to corporate data, but the modeling of data in TNF and limitations within SQL Language itself, meant that there was a limitation to which data residing in these databases could be exploited*

– *Technologies relating to DW & BI were themselves very not very mature hence too much reliability and dependency on SQL to access & manipulate data*

Module 1: > Topic 2: Data Warehouse Evolution

1980 – 1990 : Naturally Evolving Architecture Era

- PC's & Fourth generation languages (4GL's)
- Segregation of data for Analytical processing
- End user directly controlling the data & systems
- Advent of "EXTRACT" processing
- No common source of data to begin with
- No Time basis for data comparison

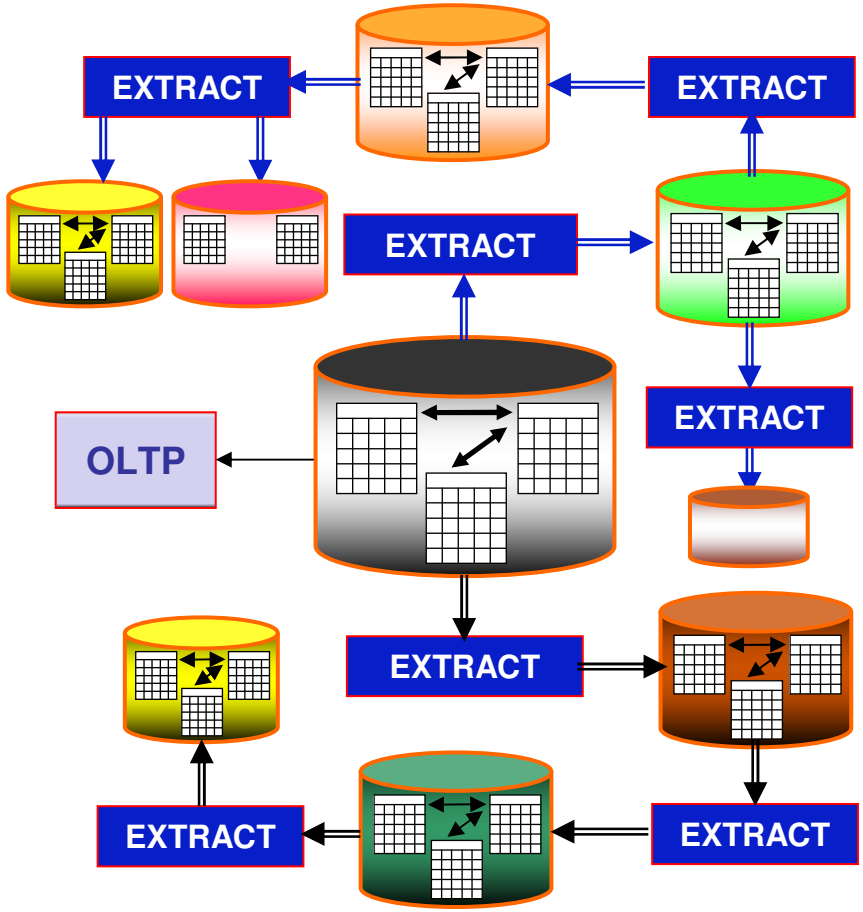


Focus on Extracts, Extracts of Extracts, Extracts of Extracts of Extracts

Module 1: > Topic 2: Data Warehouse Evolution

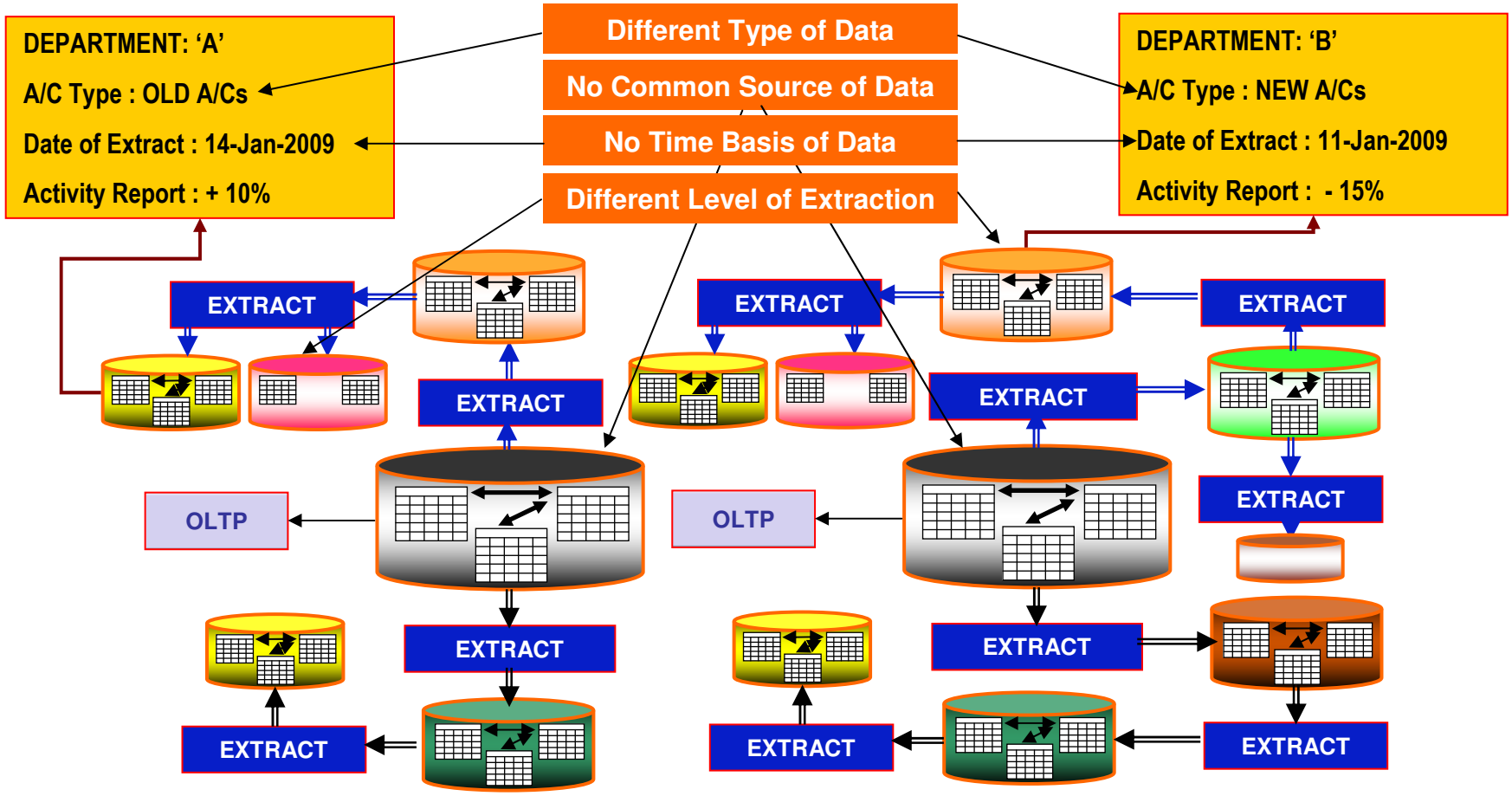
1980 – 1990 : Naturally Evolving Architecture Era

- The notion was “Start with some parameters, search a file/database based on satisfaction of these parameters, then pull the data elsewhere
- Extract processing could move the data out of the high- performance online processing (OLTP), there is no conflict in terms of performance when data needs to be analyzed en masse
- Data moved out of the operational, transaction processing with an **extract** program facilitates a shift in the data control. The **End User** can now own the data & take control of it



Module 1: > Topic 2: Data Warehouse Evolution

1980 – 1990 : Naturally Evolving Architecture Era



Module 1: > Topic 2: Data Warehouse Evolution

1980 – 1990 : Naturally Evolving Architecture Era

- *Drawbacks -*

– ***Data Credibility*** *-*
  - *No common source of data from the beginning*
  - *No Time basis of data*
  - *Algorithmic differential of data*
  - *Different levels of extraction of data for comparison*
  - *Synchronization of data is difficult & comparison of un-synchronized data would yield incorrect results*

– ***Productivity*** *-*
  - *In order to locate data for analysis, many files & layouts of data must be analyzed*
  - *Different skill sets would be required in order to access disparate data across the enterprise*
  - *Rationalization of data for standards & nomenclature across an enterprise is a tedious process*
  - *In order to produce a report, data would need to be compiled from various sources, This would involve writing lots of programs & customized codes*

Module 1: > Topic 2: Data Warehouse Evolution

1980 – 1990 : Naturally Evolving Architecture Era

- *Drawbacks -*

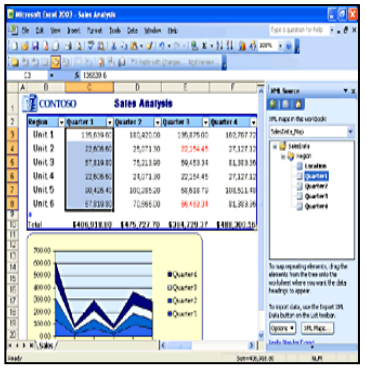
– ***Productivity*** *-*
  - *The time required to locate the data, get the data in a specific format and then code to get the required report would consumes months & years*
  - *When any report is being developed, the requirements for future reports are not known, thus each new corporate report would require the same amount of overhead as the previous one*

– ***Inability to convert Data into Information*** *-*
  - *Data integration from disparate legacy applications would be a major constraint in trying to satisfy an information request*
  - *There is not enough historical data stored in the applications to meet the needs of a DSS request*
  - *The systems found in the naturally evolving architecture are simply inadequate for supporting information needs. They lack integration and there is a discrepancy between the time horizon (or parameter of time) needed for analytical processing and the available time horizon that exists in the applications*
  - *Information is expensive to access & takes a long time to create*

Module 1: > Topic 2: Data Warehouse Evolution

1990 – 200x : Analysis Era – From Data to Information

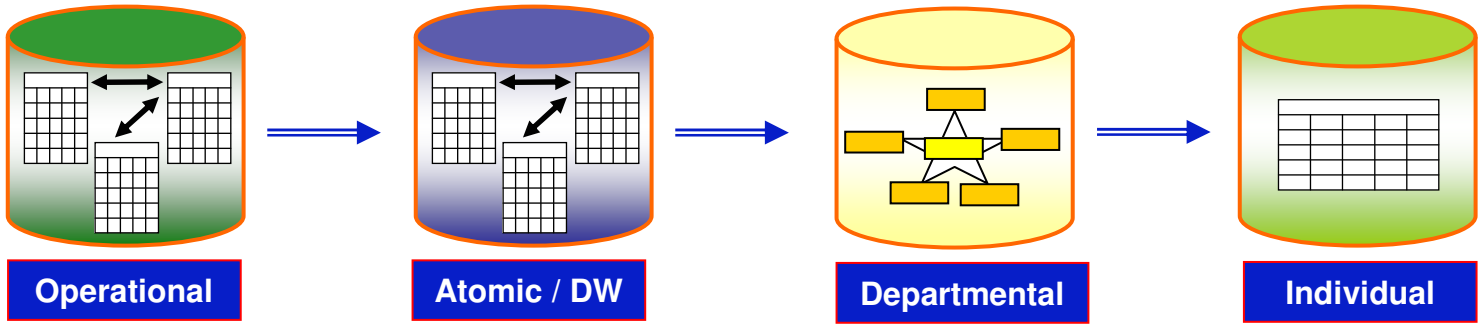
- Trend Analysis
- What If ?
- Moving Averages
- Cross Dimensional Comparisons
- Statistical profiles
- Automated pattern and rule discovery



Focus on Reporting, Querying and Online Analysis

Module 1: > Topic 2: Data Warehouse Evolution

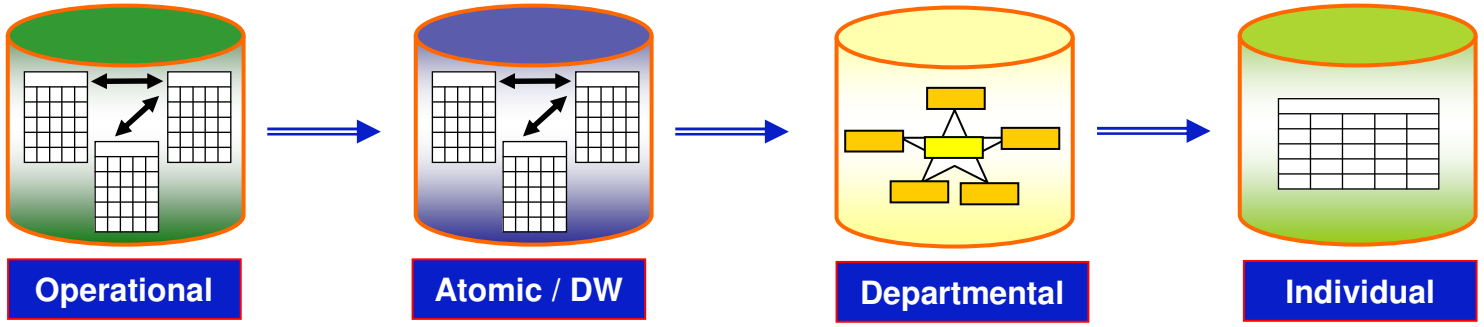
1990 – 200x : Analysis Era – Architected Environment



- Architected Data Warehouse Approach
- Data segregated as “Primitive” or “RAW” data & “Derived” or “Processed” data
- Primitive data – Detailed data required to run the day to day operations
- Derived data – Data that has been processed, integrated, cleansed and summarized
- Primitive data is current valued data & derived data is historical data
- Primitive data can be updated, Derived data can be recalculated but not directly updated
- **Operational** data is **Primitive , DSS** data is **Derived**
- **Primitive** data would support the **Clerical** function, **Derived** data would support the **Managerial** function

Module 1: > Topic 2: Data Warehouse Evolution

1990 – 200x : Analysis Era – Architected Environment



- Detailed
- Day to Day
- Current Valued
- Application Oriented

- Most Granular
- Time variant
- Integrated
- Subject Oriented
- Little Summary

- Derived data
- Dimensional Model
- Business process specific

- Temporary
- Ad Hoc
- Non Repetitive
- PC, Workstation based

These different levels of data are the basis of a larger architecture called the **Corporate Information Factory**

Module 1: > Topic 2: Data Warehouse Evolution

1990 – 200x : Analysis Era – Architected Environment – Example - Customer

Operational	Atomic / DW	Dept./ MART	Individual
		Customers by Month	Temporary
Name : John Miller Date: 14-Jan-2009 Address : 123, Main Street Credit : AA	Name : John Miller Date : 01-Jan-2000 Address : 345, High Street Credit : B	Jan 2005 : 4001 Feb 2005 : 3567 Mar 2005 : 3878 Apr 2005 : 3999	Customers since year 2000 with account balances > 20K$ & with credit rating A or higher
	Name : John Miller Date : 10-Aug-2005 Address : 678, C Street Credit : A		
	Name : John Miller Date: 14-Jan-2009 Address : 123, Main Street Credit : AA		
What is Jones Credit rating right now	What is Jones Credit history over the years	Are we getting more or fewer customers	What are the trends of Customers we are analyzing

Module 1: > Topic 2: Data Warehouse Evolution

1990 – 200x : Analysis Era

- *Benefits -*

– *Single source of origination of data, leading to better faith in data by the business*

– *Operational & Analytical data segregated, thus maintaining high performance on both ends*

– *Data Access to End Users*

– *Analytical Data Model is business friendly, subject oriented, represents data values over time & summarized*

– *Analysis of data over a period of time leads to knowledge, insight & discovery*

– *Integration of data & consolidation is better governed, this in turn reduces the data quality issues*

– *Comparison of data is automated using latest BI & Analytical tools, helps in Reporting , Analysis of data, Monitoring of data & Prediction*

– *Serves the Managerial Community in taken better business decisions*

Module 1: > Topic 2: Data Warehouse Evolution Summary

- Having completed this topic, you should be able to:
  - Know the Evolution of the Data Warehouse over the years

Module 1: > Topic 2: Data Warehouse Evolution Review

Module 1: > Topic 3: Why build a Data Warehouse

- Need & Advantages of Data Warehousing
- Challenges in Implementation of a Data Warehouse

Module 1: > Topic 3: Why build a Data Warehouse

Need & Advantages for Data Warehousing

1. **Operational Efficiency**
   - Make the right information available at the right time
   - Manage data volumes and business complexities
2. **Compliance and Transparency**
   - Leverage value of data across the enterprise
   - Adherence to federal rules and regulations
3. **Information Integration**
   - Manage information complexity with data integration
   - Manage IT costs
4. **Competitive Differentiation**
   - Outperform the competition rather than just stay in business
   - Use Analytics to get insights into information within data
5. **Data Governance**
   - Manage data as an asset
   - Make data secure and reliable

Module 1: > Topic 3: Why build a Data Warehouse

Challenges in Implementing a Data Warehouse System

1. **Cost in Building the Data Warehouse**
   - Operational Cost for DW
   - Maintenance Cost for DW
2. **Big Bang Approach**
   - Justifying ROI on DW development
   - Time required to build & get results in DW
3. **Architectural Challenges**
   - Manage information consistency with architectural changes
   - Manage information complexity with data consolidation
4. **Data Governance Issues**
   - Data Ownership issues
   - Data Integration, Consistency & Quality issues
5. **Business Complexities**
   - Adaptability of DW ever changing business scenarios
   - Business complexities due to mergers & acquisitions

Module 1: > Topic 3: Why Data Warehouse Summary

- Having completed this topic, you should be able to:
  - Know the Need and Advantages of having a Data Warehouse System
  - Know the Challenges faced in implementation of a Data Warehouse System

Module 1: > Topic 2: Why Data Warehouse Review

References

- TDWI – The Data Warehousing Institute … www.tdwi.org
- Data Warehouse Life Cycle Toolkit – Ralph Kimball
- Newsletter July/August 1992 - WH Inmon
- Building the Data Warehouse – WH Inmon
- Forrester Research … www.forrester.com
- DM Review … www.dmreview.com