innovate achieve lead

SS ZG515 - Data Warehousing

BITS Pilani
Pilani Campus

Dr. Yashvardhan Sharma
CSIS Dept., BITS-Pilani

Contact Info

- **Email:** yash@pilani.bits-pilani.ac.in
- **Course Web site:**

http://taxila-aws.bits-pilani.ac.in/

Course Information

- Two Lectures / Week
  - Saturday –4:00 – 5:00 PM
  - Sunday – 4:00 -5:00 PM
- Total Lectures: 24 (Planned)
- Text Books:
  - Ponniah P, ***"Data Warehousing Fundamentals"***, John Wiley, 2003.
  - Kimball R, ***"The Data Warehouse Toolkit",*** 2e, John Wiley, 2002.

Main Topics

- Introduction & Background-Evolution of Data Processing
- Motivation for Data Warehousing
- Architecture
- Data Modeling-De-normalization
- Dimension Modeling
- Extract – Transform – Load (ETL)
- Data Quality Management (DQM)
- On Line Analytical Processing (OLAP), Data Mining
- Query Performance Enhancing Techniques-Need for speed (Parallelism, Join and Indexing techniques)
- DWH Implementation steps
- Complete implementation Case Studies
- Research Issues

Why this course?

- The world is changing (actually changed), either change or be left behind.
- Missing the opportunities or going in the wrong direction has prevented us from growing.
- What is the right direction?
- Harnessing the data, in a knowledge driven economy.

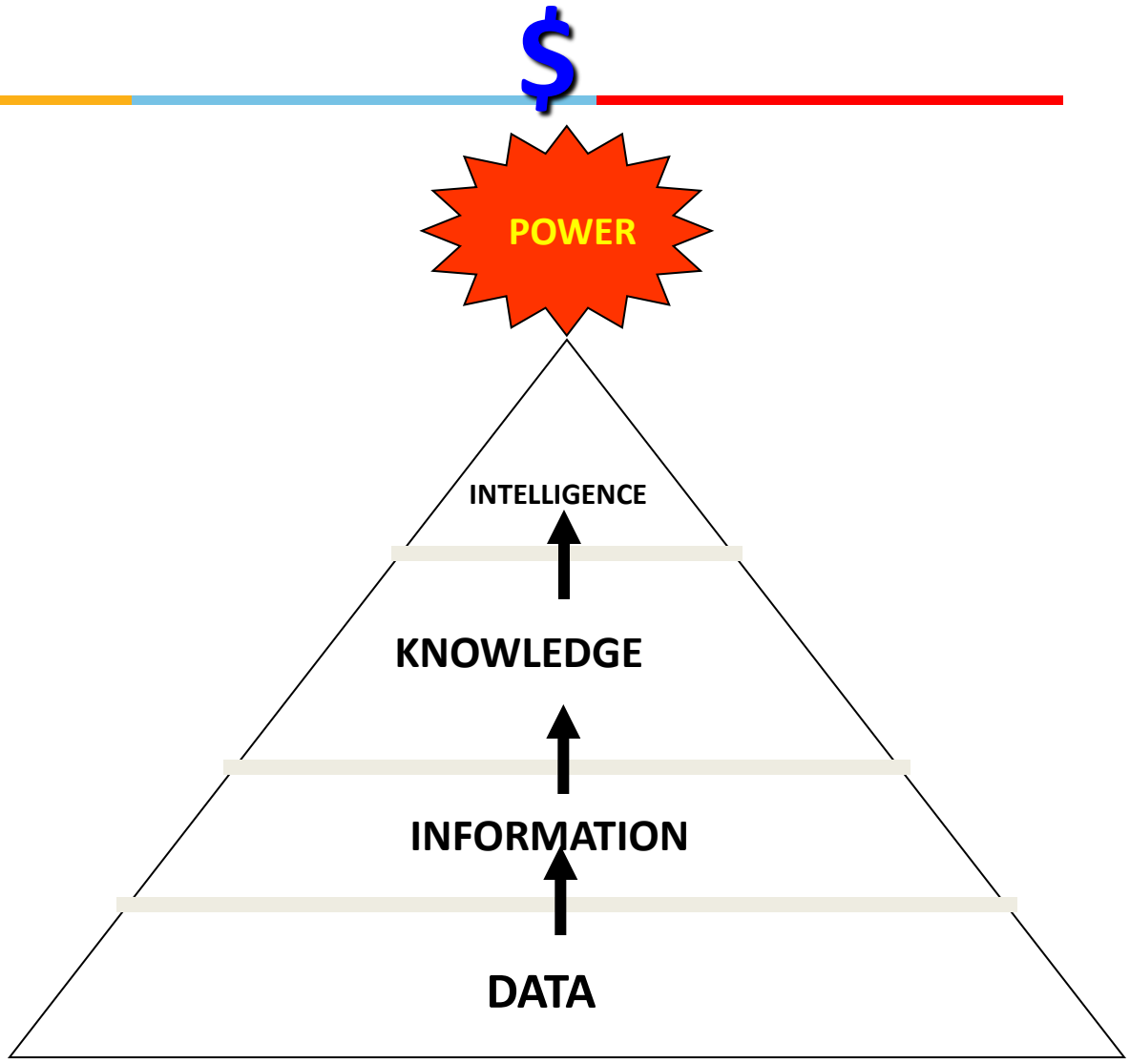
The need

"Drowning in data and starving for information"

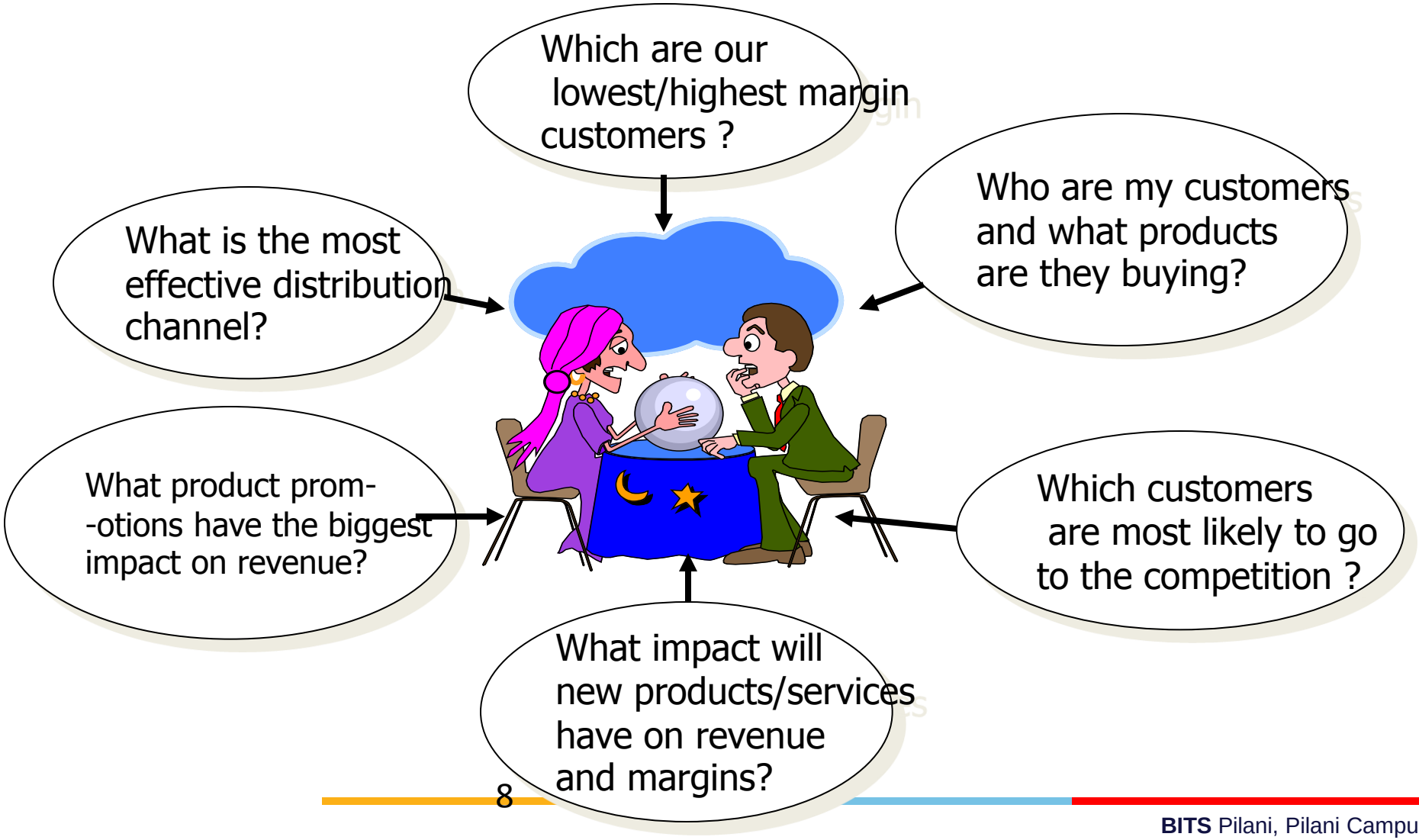
-John Naisbett

Knowledge is power, Intelligence is absolute power!

The need



A producer wants to know....



Data, Data everywhere yet …

- I can't find the data I need
  - data is scattered over the network
  - many versions, subtle differences

⌘ I can't get the data I need
  - need an expert to get the data

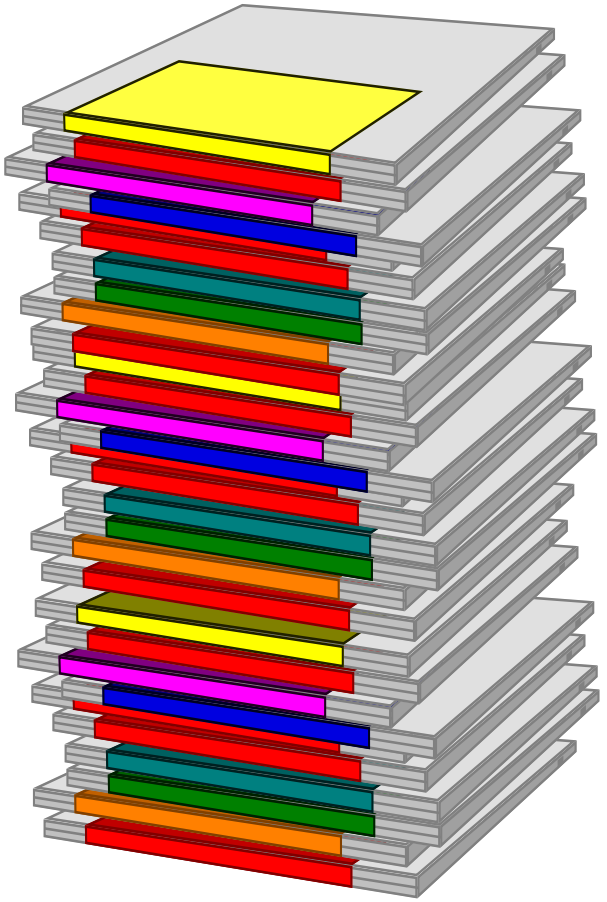
⌘ I can't understand the data I found
  - available data poorly documented

⌘ I can't use the data I found
  - results are unexpected
  - data needs to be transformed from one form to other

What is a Data Warehouse?

A single, complete and consistent store of data obtained from a variety of different sources made available to end users in a what they can understand and use in a business context.

[Barry Devlin]



What are the users saying...

- Data should be integrated across the enterprise
- Summary data has a real value to the organization
- Historical data holds the key to understanding data over time
- What-if capabilities are required

What is Data Warehousing?

A process of transforming data into information and making it available to users in a timely enough manner to make a difference

[Forrester Research, April 1996]

Background

- 1980's to early 1990's
  - Focus on computerizing business processes
  - To gain competitive advantage
- By early 1990's
  - All companies had operational systems
  - It no longer offered any advantage
- How to get competitive advantage??

OLTP Systems: Primary Purpose

Run the operations of the business

- For example: Banks, Railway reservation etc.
- Based on ER Data Modeling
- Transaction based system
- Data is always current valued
- Little history is available
- Data is highly volatile
- Has “Intelligent keys”

OLTP Systems

- Has relational normalized design
- Redundant data is undesirable
- Consists of many tables
- High volume retrieval is inefficient
- Optimized for repetitive “narrow” queries
- Common data in many applications

Need for Data Warehousing

- Companies, over the years, gathered huge volumes of data
- “Hidden Treasure”
- Can this data be used in any way?
- Can we analyze this data to get any competitive advantage?
- If yes, what kind of advantage?

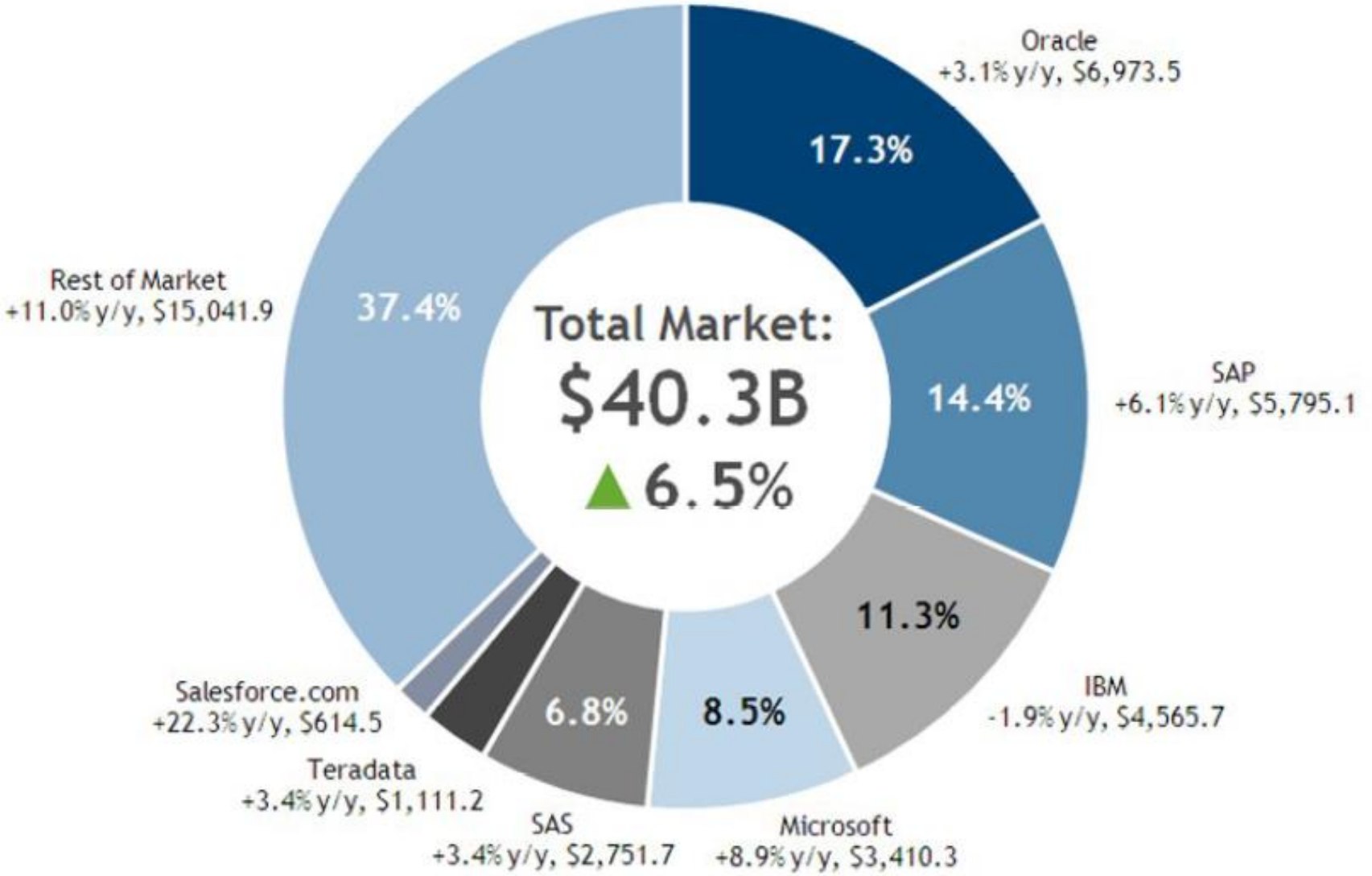
Benefits of Data Warehousing

- Allows “efficient” analysis of data
- Competitive Advantage
- Analysis aids strategic decision making
- Increased productivity of decision makers
- Potential high ROI
- Classic example: Diaper and Beer

Decision Support Systems, DW, & OLAP

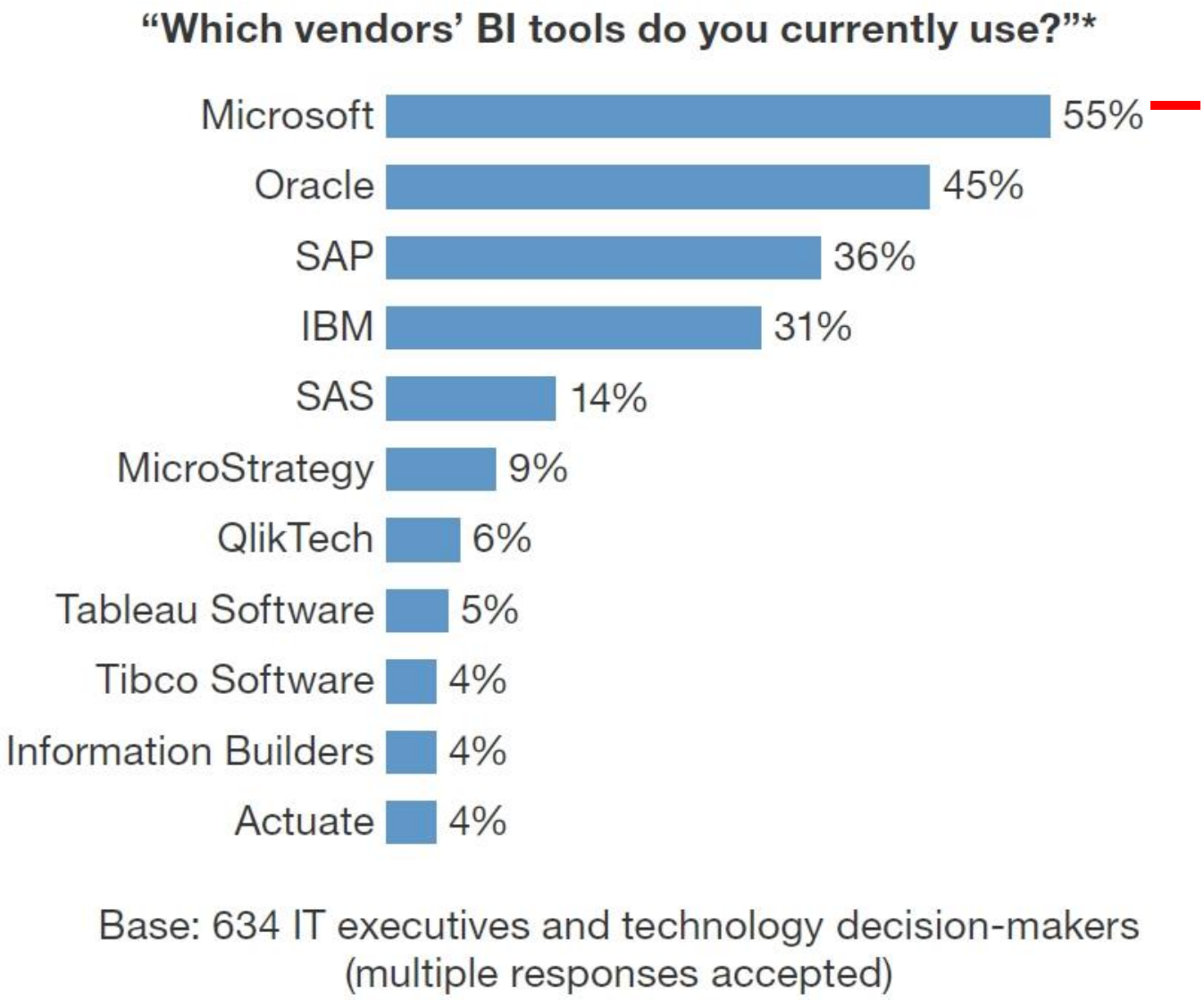
- Information technology to help the knowledge worker (executive, manager, analyst) make faster and better decisions.
- Data Warehouse is a DSS
- A data warehouse is an architectural construct of an information system that provides users with current and historical decision support information that is hard to access or present in traditional operational systems.
- Data Warehouse is not an Intelligent system
- On-Line Analytical Processing (OLAP) is an element of DSS

Worldwide Business Analytics Software 2014 Share Snapshot



Note: 2014 Share (%), Growth (%), and Revenue ($M)

Source: IDC, 2015


"Which vendors' BI tools do you currently use?"*
Microsoft 55%
Oracle 45%
SAP 36%
IBM 31%
SAS 14%
MicroStrategy 9%
QlikTech 6%
Tableau Software 5%
Tibco Software 4%
Information Builders 4%
Actuate 4%
Base: 634 IT executives and technology decision-makers
(multiple responses accepted)

Data Warehouse: Major Players

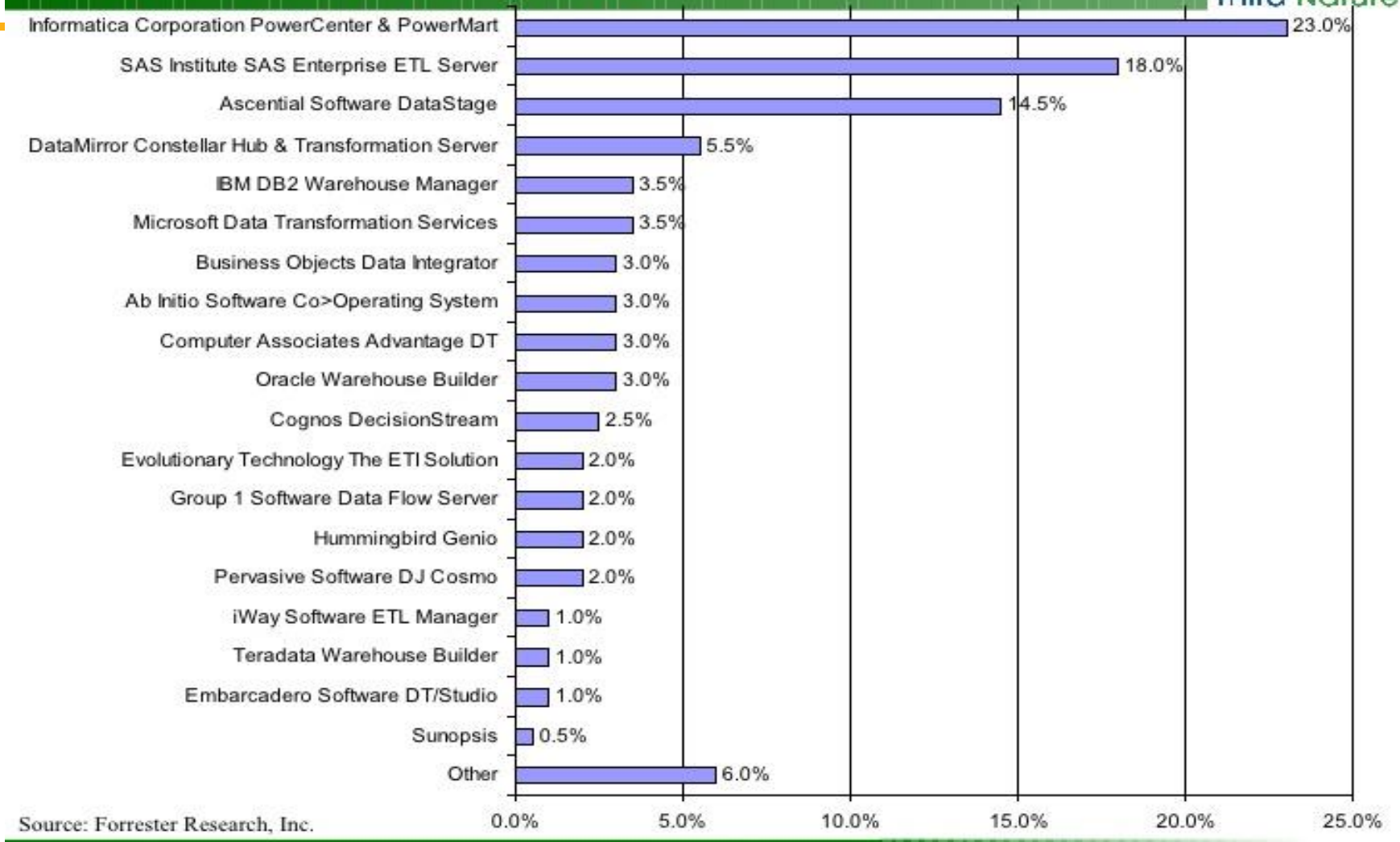
innovate achieve lead

BI Vendor Products OLAP (2011)

Vendor	Product(s)
SAP Business Objects	SAP NetWeaver BW (InfoCubes)
Oracle	Hyperion Essbase
IBM Cognos	PowerPlay TM1
MicroStrategy	Intelligence Server
Microsoft	Analysis Services
SAS	OLAP Server
Pentaho	Mondrian
JasperSoft	Jasper Analysis

Vendor Market Share

Third Nature



Source: Forrester Research, Inc.

Data Warehouse: Characteristics

- Analysis driven
- Ad-hoc queries
- Complex queries
- Used by top managers
- Based on Dimensional Modeling
- Denormalized structures

Data Warehouse

- A decision support database that is maintained separately from the organization's operational databases.
- A data warehouse is a
  - subject-oriented,
  - integrated,
  - time-varying,
  - non-volatile

  collection of data that is used primarily in organizational decision making