

# Data Warehouse Architecture and Project Management

For  
Senior Database Programmers,  
Data Architects,  
and Project Managers

Copyright © 2000 Computer Adept, Inc.

**COMPUTER *adept***

# Data Warehouse Architecture and Project Management

CLICKSTREAM

DATA WAREHOUSING

# Clickstream Data Warehousing

→ Customer Relationship Management and the Internet

× Amazon

× CNN

→ The "Clickstream"

# Clickstream Data Warehousing

## ← Clickstream Log Data Captured

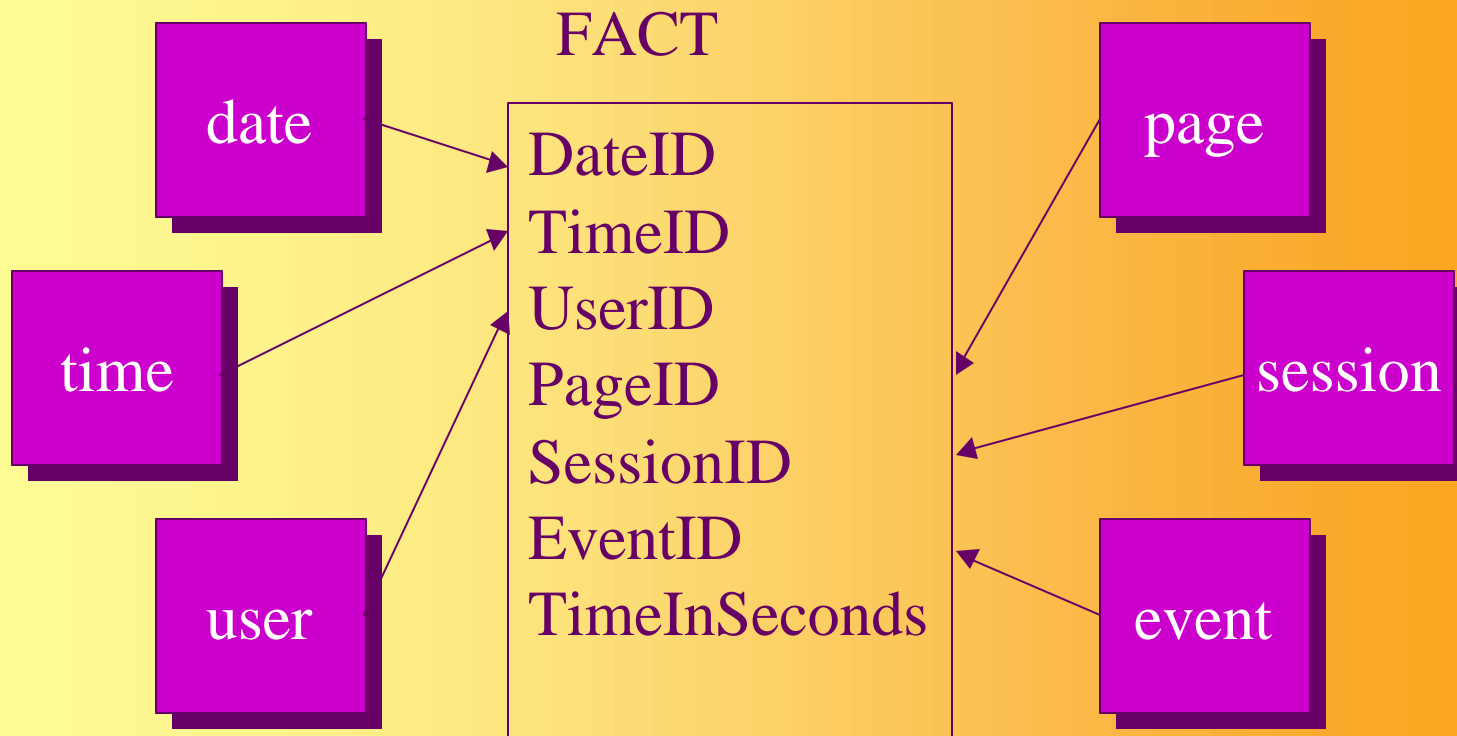
- ※ IP Address
- ※ ISP
- ※ Cookie ID
- ※ Referrer
- ※ Browser version
- ※ Date/Time
- ※ Object

# Clickstream Data Warehousing

- What do we want to know?
- What can we always track?
- How can we get more information?
- What are the privacy issues?

# Clickstream Data Warehousing

## Dimensions



# Clickstream Data Warehousing

← What can we always track?

※ When?

- Date of the page request
  - Weekend?
  - Workday?
- Time of the page request
  - Load balancing
  - 9-5 or evenings?
  - Time zones? (Combine with ISP info)
- Elapsed Time (inferred)

# Clickstream Data Warehousing

← What can we always track?

✧ Visitor:

- IP address of the visitor (usually ISP)
- Cookie ID of the visitor (if they accept cookies)
  - Identifying a return visitor
- What can we learn from this basic information?
- How can we get more information?



# Clickstream Data Warehousing

← What can we always track?

## ※ Page

- Referrer
  - Last page
  - Do we want to advertise with common referrers?
- Requested
  - URL
  - Object

# Clickstream Data Warehousing

← What can we always track?

※ Event Type

- Submit
- Link
- Search

※ Browser version making the request

- If a particular browser version results in different behavior - clue to bugs and/or performance.

# Clickstream Data Warehousing

## → Questions we can answer

- ※ What parts of our Web site get the most visitors?
- ※ What parts of the Web site do we associate most frequently with actual sales?
- ※ What parts of the Web site are superfluous or visited infrequently?

# Clickstream Data Warehousing

## → Questions we can answer

### ※ What are some page profiles?

- "session killers"
- links
- long views
- lead to sales

### ※ What are some customer profiles?

- new
- returning
- unhappy

# Clickstream Data Warehousing

← What do we want to know?

- ※ How can we induce the customer to register with our site so we learn some useful information about that customer?
- ※ How many visits do unregistered customers typically make with us before they are willing to register? Before they buy a product or service?

# Clickstream Data Warehousing

← What do we want to know?

- ※ Demographics

- ※ Marketing Campaign Effectiveness

- ※ Click/Buy Ratios

- ※ Cross-selling Opportunities

# Clickstream Data Warehousing

→ How can we get more information?

- ✧ Registration

- ✧ Purchases

- ✧ Ethical Issues

# Clickstream Data Warehousing

## → Privacy Principles

- ※ Choice/Consent - get permission to collect data
- ※ Notice - give notice to user of what data is collected
- ※ Purpose Specifications/Limitation - give notice to user of how data will be used



# Clickstream Data Warehousing

## → Privacy Principles

- ※ Use Limitation - ensure that the data is used for reasons described in notice above
- ※ Collection Limitation - only collect relevant information
- ※ Data Quality/Accuracy - ensure the quality and accuracy of data that is collected
- ※ Data Security/Confidentiality - maintain security for collected data

# Clickstream Data Warehousing

## → Privacy Principles

- ※ Access/Correction - give users the ability to access and correct personal information
- ※ Onward Transfer - ensure data transfers only occur to third parties with the same level of privacy protection
- ※ Enforcement/Accountability/Redress - industry responsibility to enforce principles through dispute resolution process