BLUE MARTINI SOFTWARE

Information Organization and Retrieval Class, Berkeley
Oct 19, 2000

An Ideal E-Commerce Architecture for Building Web Sites Supporting Analysis and Personalization

Ronny Kohavi, Ph.D.
Director of Data Mining
Blue Martini Software
ronnyk@bluemartini.com
http://www.bluemartini.com
http://www.Kohavi.com

Overview

- **Warning: Your mileage may vary**
- **Introduction - the vision**
  - Webstore (interact with customers)
  - Analysis (understand)
  - Action (target)
- **Architecture**
  - Requirements
  - The unfair advantage
- **Summary**

Warnings

- **Ronny Kohavi's biased view**

  Your mileage may vary (standard disclaimers)

- **Real-life problems**
  - **Need effective solutions, not clean/beautiful solutions. Examples:**
    - Engine noise in planes 
    - Nomad robots at Stanford hospital
    - Structured data, not information extraction (when we can)
  - **Efficiency is paramount - software must be designed to run fast and scale well**
    - Start quickly with small/inefficient solutions, make sure it can grow. Measure with a micrometer; Mark with chalk; Cut with an axe. Design ideal architecture; Implement pieces; Code and ship, improve.
    - Use efficient algorithms (low complexity: O(n log n) for n records).

Introduction - The Vision

BLUE MARTINI SOFTWARE

International Data Corporation (IDC) reported in 1999 that (large) web sites costs $5.9M to assemble and $4.3M annually to maintain.

Vision: enterprise software application that allows companies to interact with, understand, and target customers

- Enterprise - allows integration (expensive)
- Interact - on the web and possibly through other “touch points” (e.g., phones)
- Understand - Analyze data (e.g., data mining)
- Target - personalize (web, e-mail)

Understand Customer Behavior

BLUE MARTINI SOFTWARE

- **Motivation: Improve the site over time**
  - **How many visitors?**
  - **Conversion rates (buyers to visitors) for products?**
  - **How are they traversing the site?**
    Killer pages
  - **Where are they coming from?**
    Which ads are effective?
  - **Failed searches?**
- **Solutions:**
  - **Reports**
  - **Data Mining and visualizations**

Conversion Rates

BLUE MARTINI SOFTWARE

Using hits and page views to judge site success is like evaluating a musical performance by its volume
-- Forrester Report, 1999

- **A key metric in e-commerce sites is the conversion rate (buyers to browsers)**
- **Especially useful by referrer (e.g., ad)**
- **What is a typical conversion rate (e.g., dell.com)**

A Real-World Referrer Example

BLUE MARTINI SOFTWARE

- On one of our sites, we saw the following in their initial rampup period

Referrer	# Sessions	% of traffic	# Sales	Conv rate
ShopNow	16,178	6.9%	6	0.04%
FashionMall	19,685	8.4%	17	0.09%
MyCoupons	2,052	0.9%	170	8.28%

Conversion rates differ by a factor of over 200!

What is Data Mining

- **More Data Mining and Viz at 11 today**
- **Elevator description (purple)**

For future predictions

Actionable

The non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data.

-- Fayyad, Piatetsky-Shapiro, Smyth [1996]

Examples of Patterns (real)

BLUE MARTINI SOFTWARE

- **Data from a legcare/legware e-retailer.**
- **Patterns for heavy purchasers:**
  - Not an AOL user (defined by browser)
  - Came to site from print-ad or news, not friends & family (reg form)
  - Very high and very low income
  - High home market value, owners of luxury vehicles
  - Repeat visitors (four or more times)
  - Visits to specific areas of site
- **Patterns for shoppers**
  - Those that came with a discount coupon (code)
  - Those that did not come from winnie-cooper.com
    Who is Winnie Cooper? 

Who is Winnie Cooper?

- **Winnie-cooper is a 31 year old guy who wears pantyhose**
- **He has a pantyhose site**
- **8,700 visitors came from his site to our legware/legcare site in three days (half the traffic at the time)**

Target / Personalize

- **Analysis through data mining and visualization yields insight**
- **Insight leads to action**
- **Examples:**
  - **Targeted campaigns - offer people what they are likely to want/buy**
  - **Personalize site (fewer images for modem users)**
  - **Different merchandise for different users**
  - **Jumbo pantyhose for visitors that come from Winnie-Cooper.com**

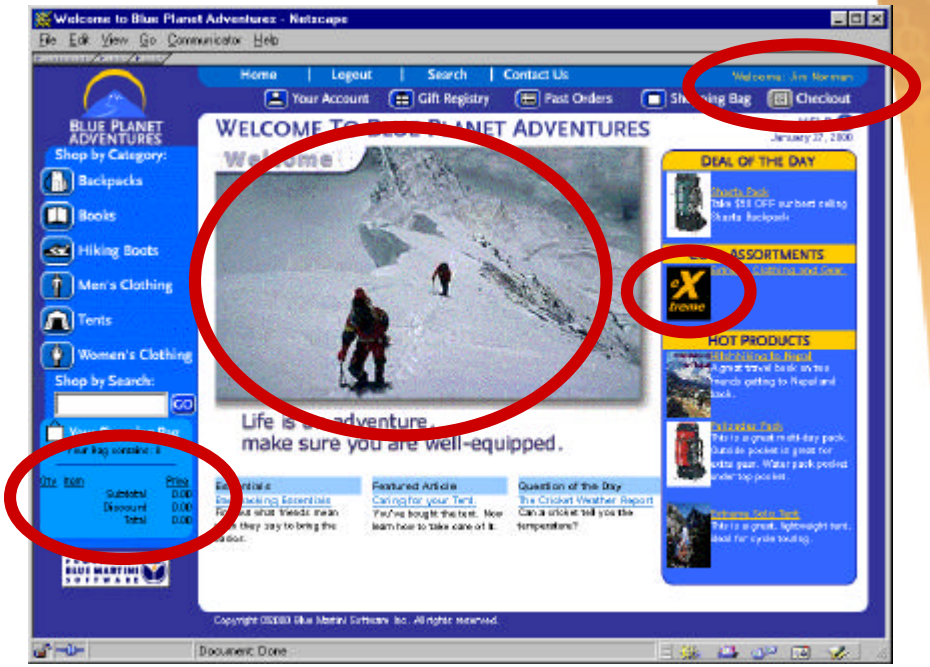
Personalization Benefits

BLUE MARTINI SOFTWARE

- Increase Conversion
- Increase Basket Size
- Increase Customer Retention

Name : Anonymous

Name : Joe Smith
Income : $50,000-$75,000
Attitude : Extreme
Gender : Male



Interact - Touch Points

- **Interact with customers across touch point**
  - **Webstore**
  - **Phone**
  - **Wireless**
  - **Bricks-and-mortar**
- **We want consistent messages across**

  Example: same promotions and cross-sells on webstore, wireless PDA, and phone call to purchase

Overview

- **Warning: Your mileage may vary**
- **Introduction - the vision**
  - Webstore (interact with customers)
  - Analysis (understand)
  - Action (target)
- **Architecture** Ü
  - Requirements
  - The unfair advantage
- **Summary**

Requirements - Biz Logic

- **Business logic must be shared across channels (webstore, customer service, etc)**
  - **Everything must be stored in a database**
  - **Web pages use API calls to access database.**
  - **Rules store logic for recommendations, promotions.**
  - **On the web, use Java Server pages (JSP), which consits of HTML with embedded Java code.**
    For example, the following displays the home page image, which may be different for each user:

```
<%homeImage=webstore.getCollectionRecommendation("Images")%>
<img src="<%= homeImageObject.getPath() %>" align="center">
```

Requirements - Attributes

BLUE MARTINI SOFTWARE

- **Attributes everywhere**

  Every object should support description through any number of attributes

  - Examples of attributable object:
    - Customers (name, address, age, gender, income)
    - Product (short & long description, waterproof,….)
    - Order header (customer, ship address, price, coupon)
    - Order line (product, quantity, color, size)
    - Web page template (site area, designer)
    - Image (size, image/drawing, caption)
  - Why?
    - Multiple attributes for different touch points (e.g., long description for web, short for wireless PDA)
    - Structured data - makes data mining easier

Requirement - Hierarchies

BLUE MARTINI SOFTWARE

- **Hierarchies everywhere (trees)**
  - **Examples:**
    - Product arranged in hierarchy
    - Assortments (collection of products) are hierarchical
    - Promotions
    - Analyses
  - **Why?**
    - Manageability - humans can't deal with lists
      Much better at traversing trees/hierarchies
    - Inheritance - Children inherit properties from parents.
      For example, all children of "Jeans" automatically inherit properties from parent
    - Abstraction levels for data mining patterns

      Diapers and Beer sell together, but there is no specific diaper that sells with a specific beer

Site Versioning

- Site must be up while the next site is being designed
- Switch from old to new site must be smooth
- Architecture must support multiple versions
  - Deployment of new site
  - Users who are in mid session continue to see “old site.”
  - New users see new site

Clickstream Collection

BLUE MARTINI SOFTWARE

- **Track user actions on site for analysis**
- **Web logs insufficient**
  - **Don't know what they typed during search**
  - **HTTP is stateless - need to sessionize visits**
  - **URLs are meaningless in changing sites**
  - **Dynamic sites / personalized sites show different content for same URL**
- **Solution:**
  - **Create our own clickstream log**
  - **Very rich, including meta data (e.g., what was on the page).**

Efficiency/Scalability

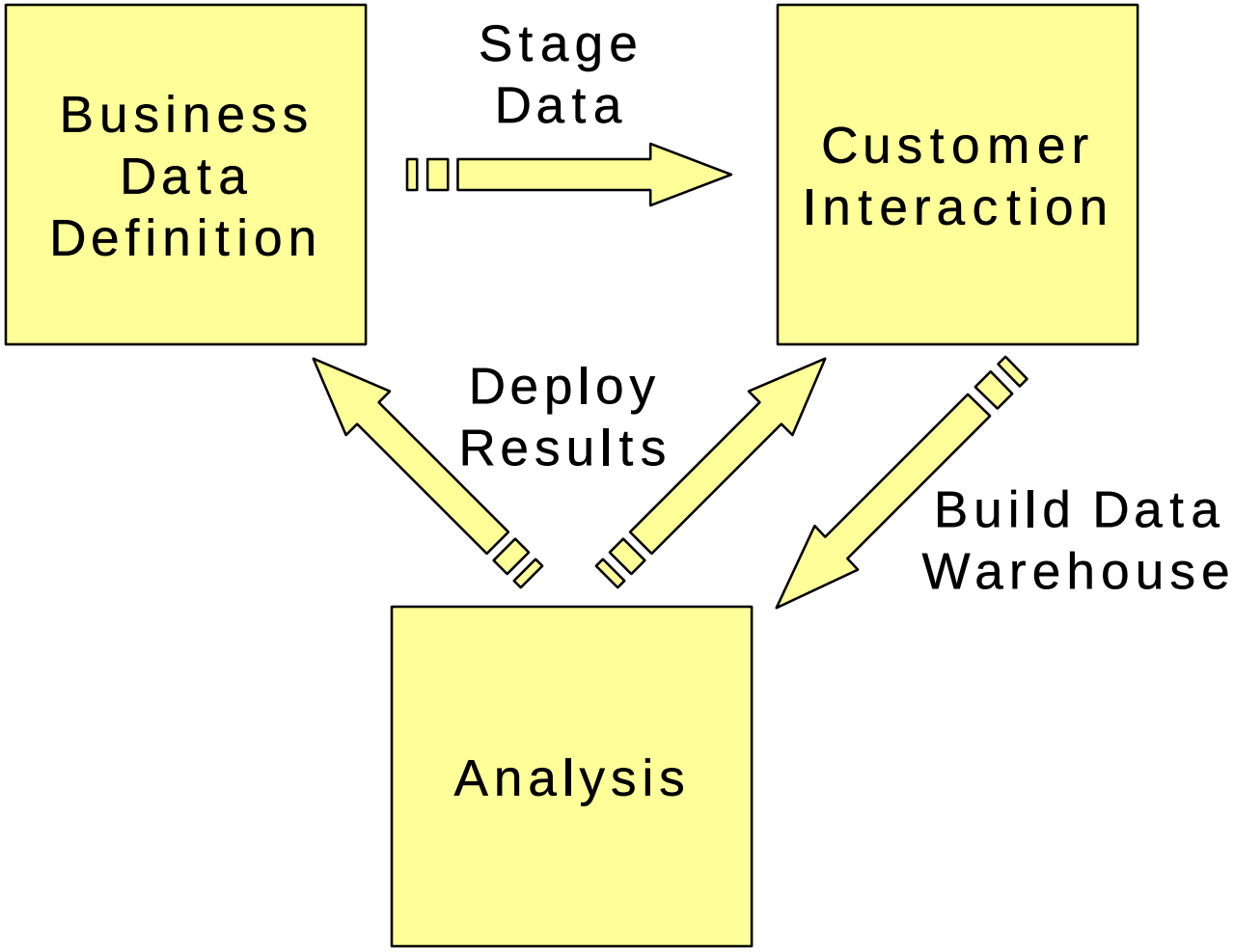
BLUE MARTINI SOFTWARE

- **Site must be efficient/distributed**
  - **Multiple web servers and application servers (application servers control the logic and generate the HTML pages; webservers just serve them)**
  - **Requires data replication**
- **Solution:**
  - **Site definition and design done against an inefficient database schema that is easy to work with**
  - **"Staging" process transforms data to a very efficient (time-wise) format for deployment**
    Deployment format can change over time as we find more tricks to improve efficiency

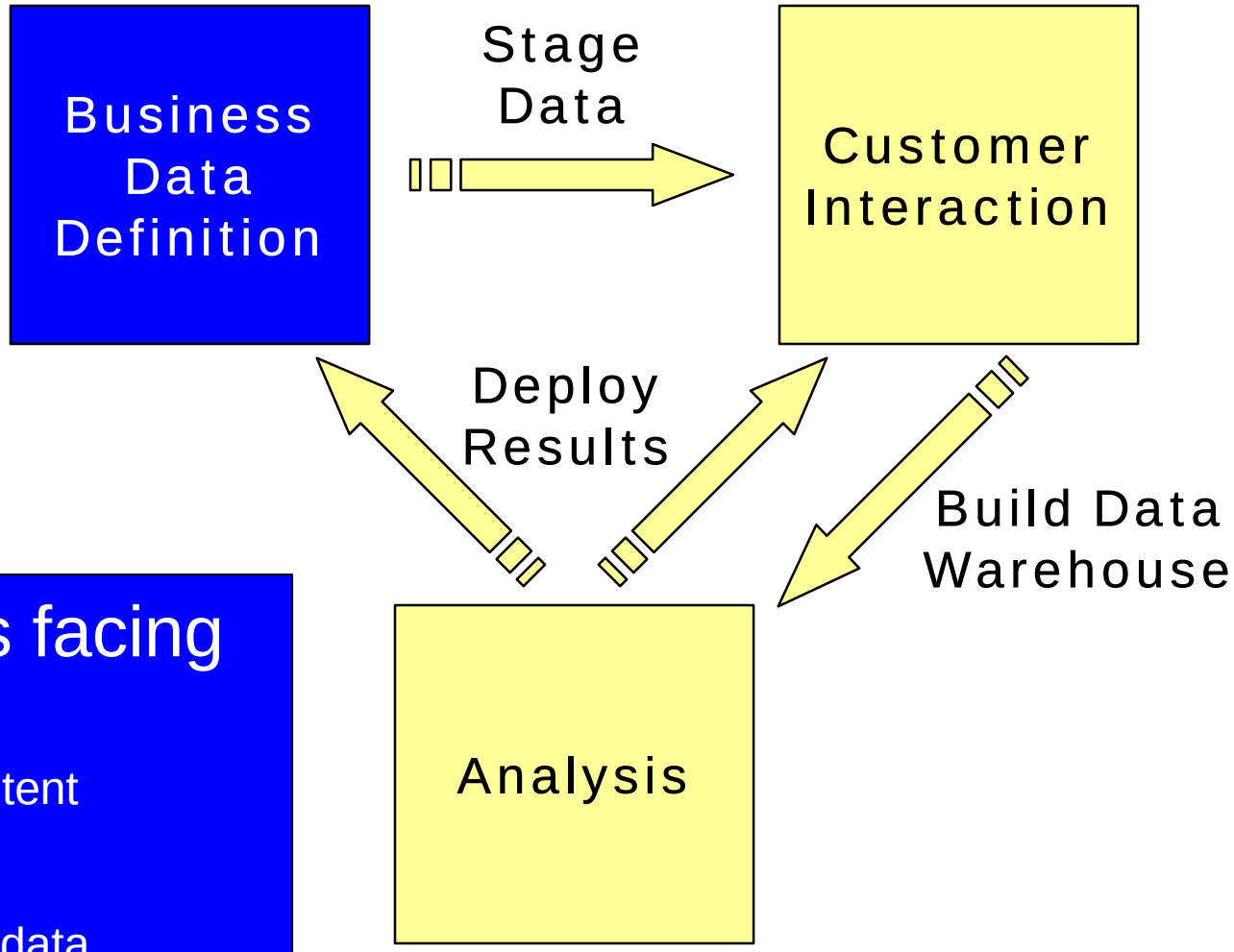
Data Warehouse

- Analysis must never be done at the webstore, which is an OLTP system (On-Line Transaction Processing)
- Data must be copied, joined with external data, transformed, cleaned:
  a Data Warehouse
- Reporting, data mining, and visualizations, are all done against data warehouse

Integrated Architecture



Integrated Architecture



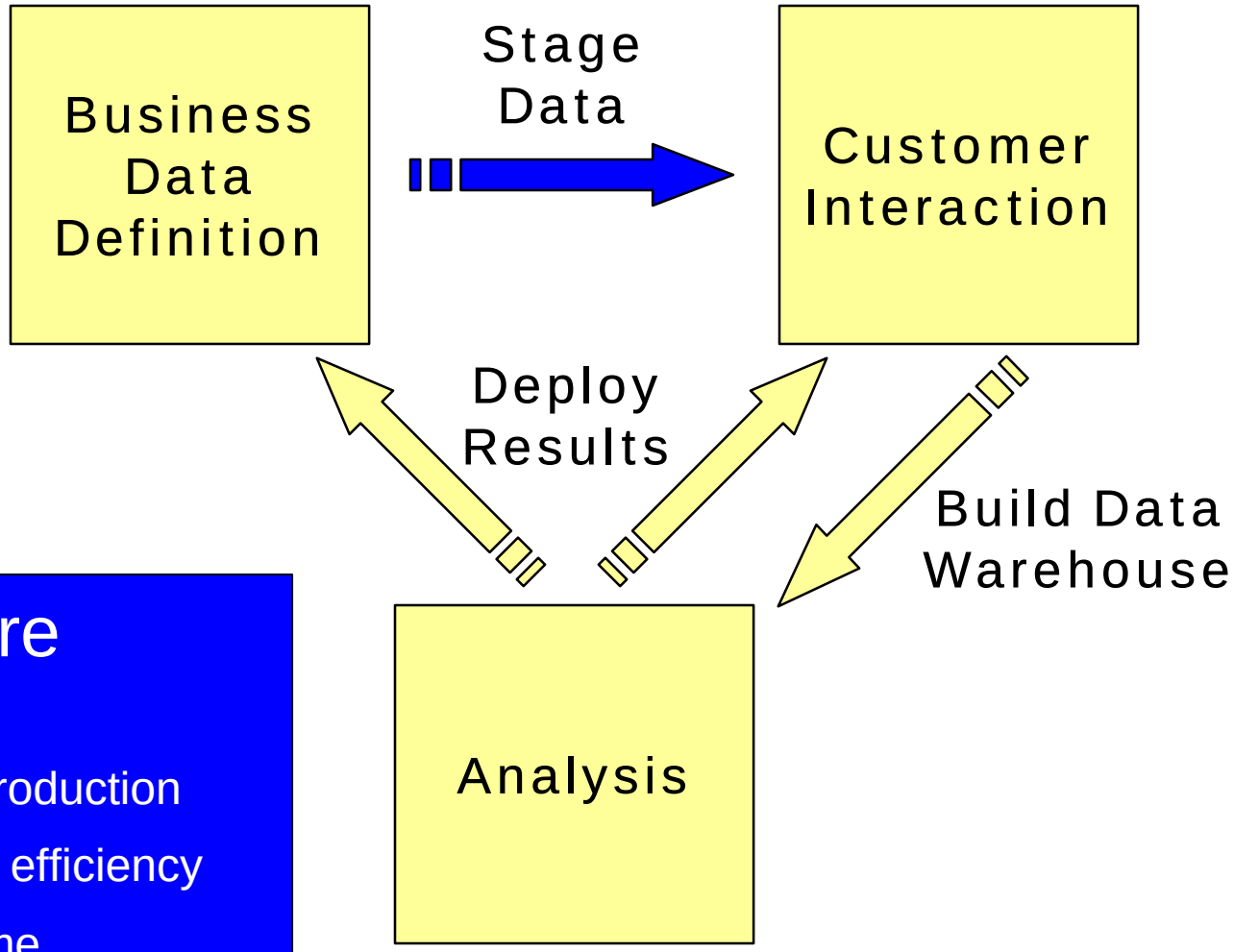
Business facing

Products, content

Attributes

Shared meta-data

Integrated Architecture



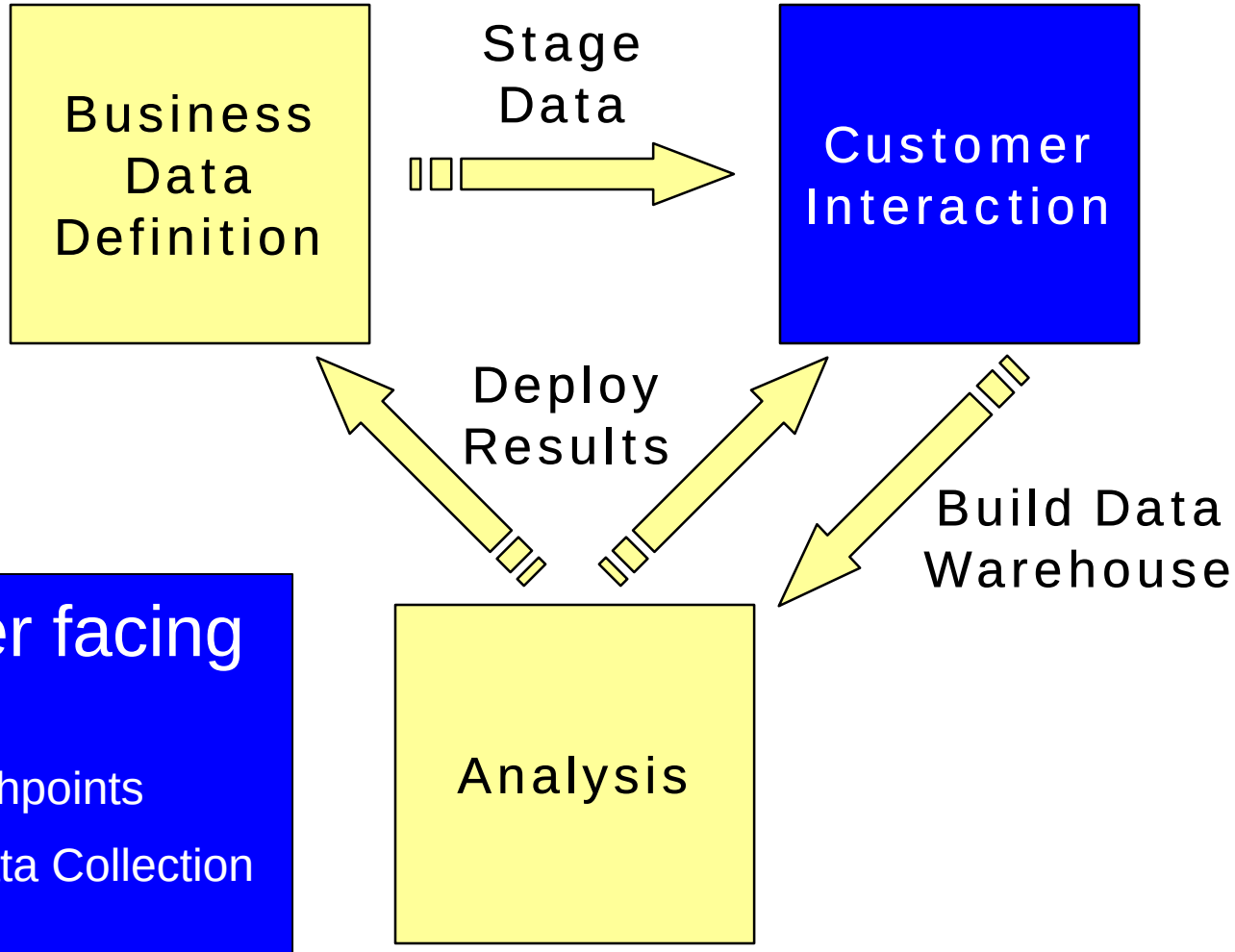
Build store

Test before production

Transform for efficiency

Zero down-time

Integrated Architecture

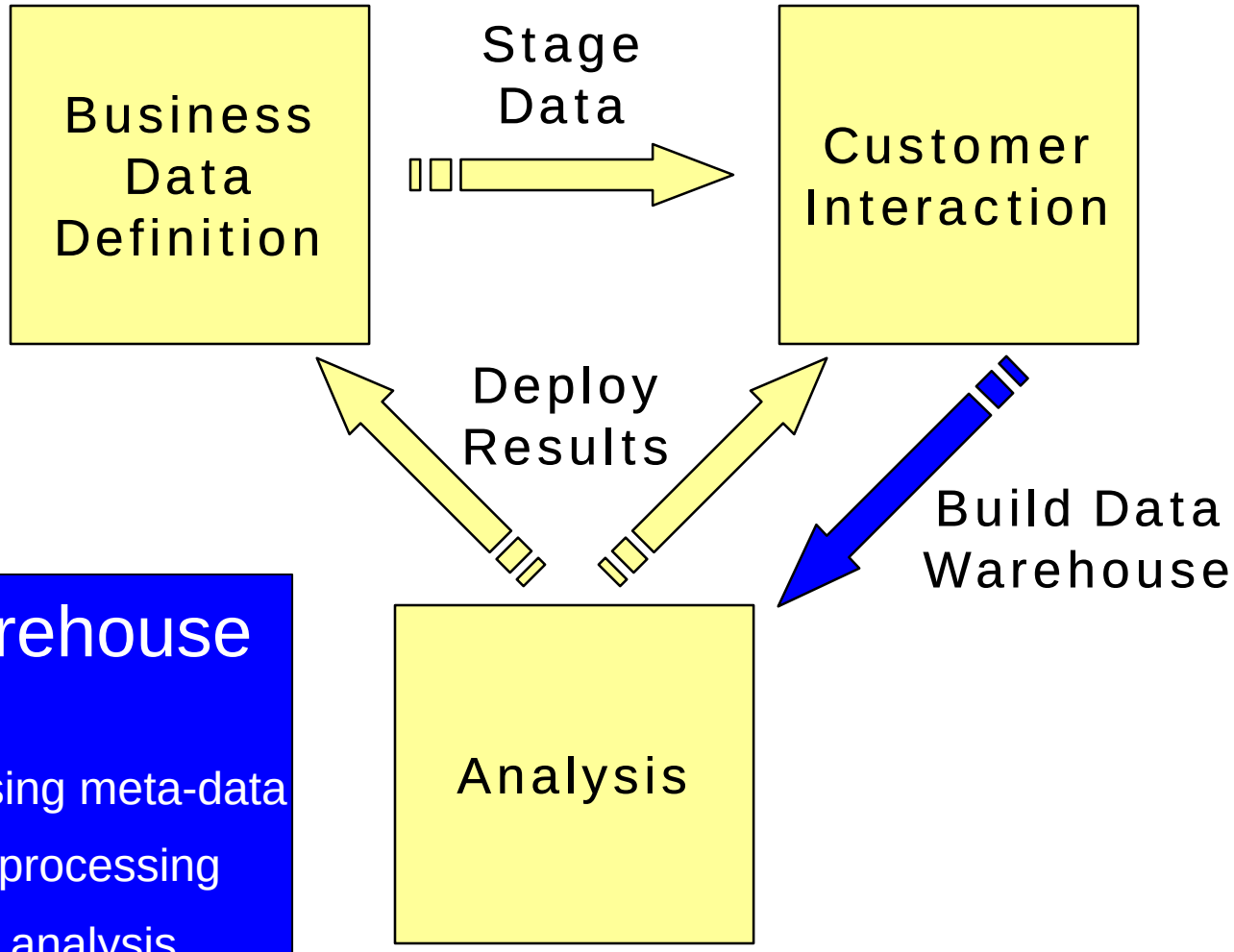


Customer facing

Multiple Touchpoints

Integrated Data Collection

Integrated Architecture



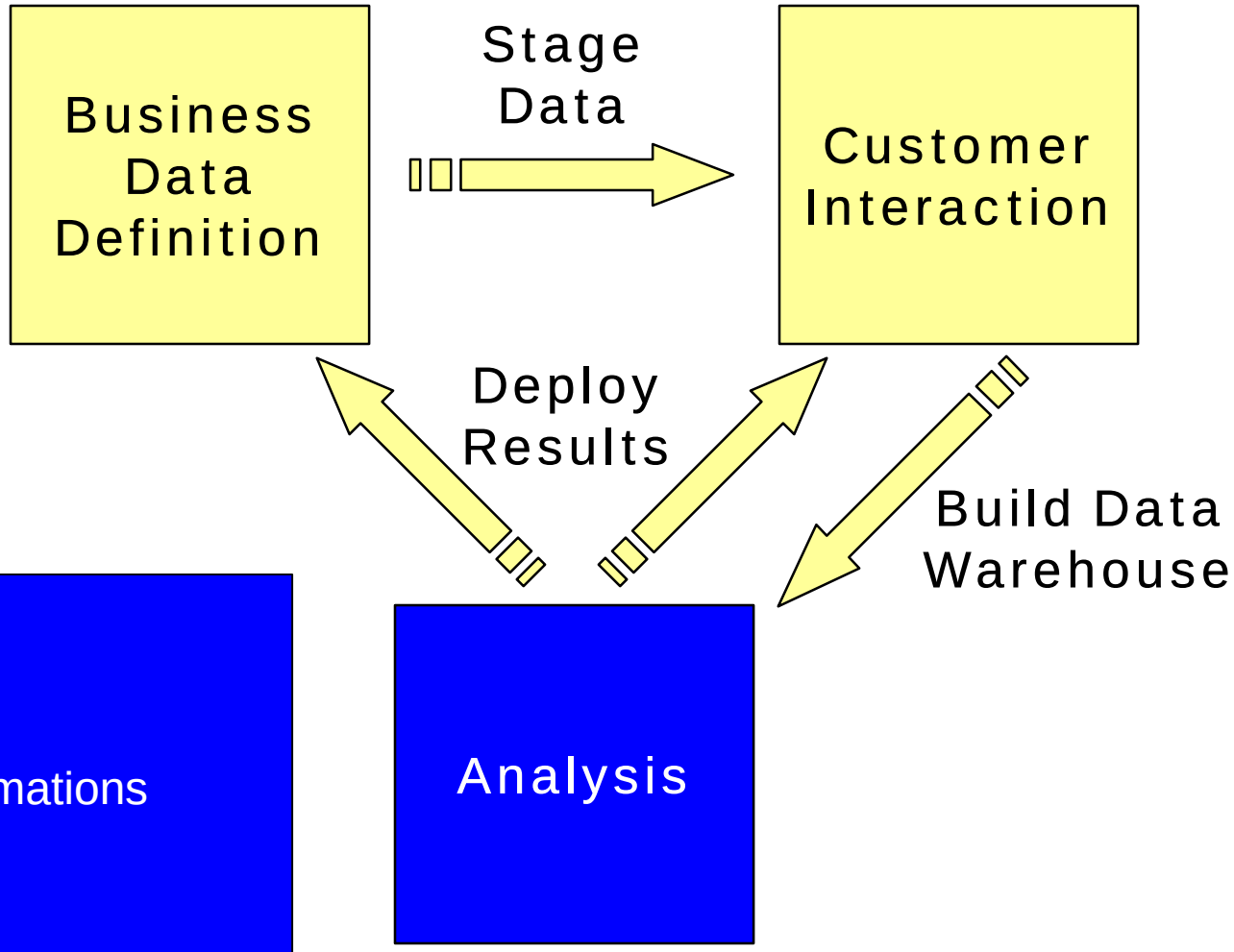
Build warehouse

Automated using meta-data

Reduces pre-processing

Transform for analysis

Integrated Architecture



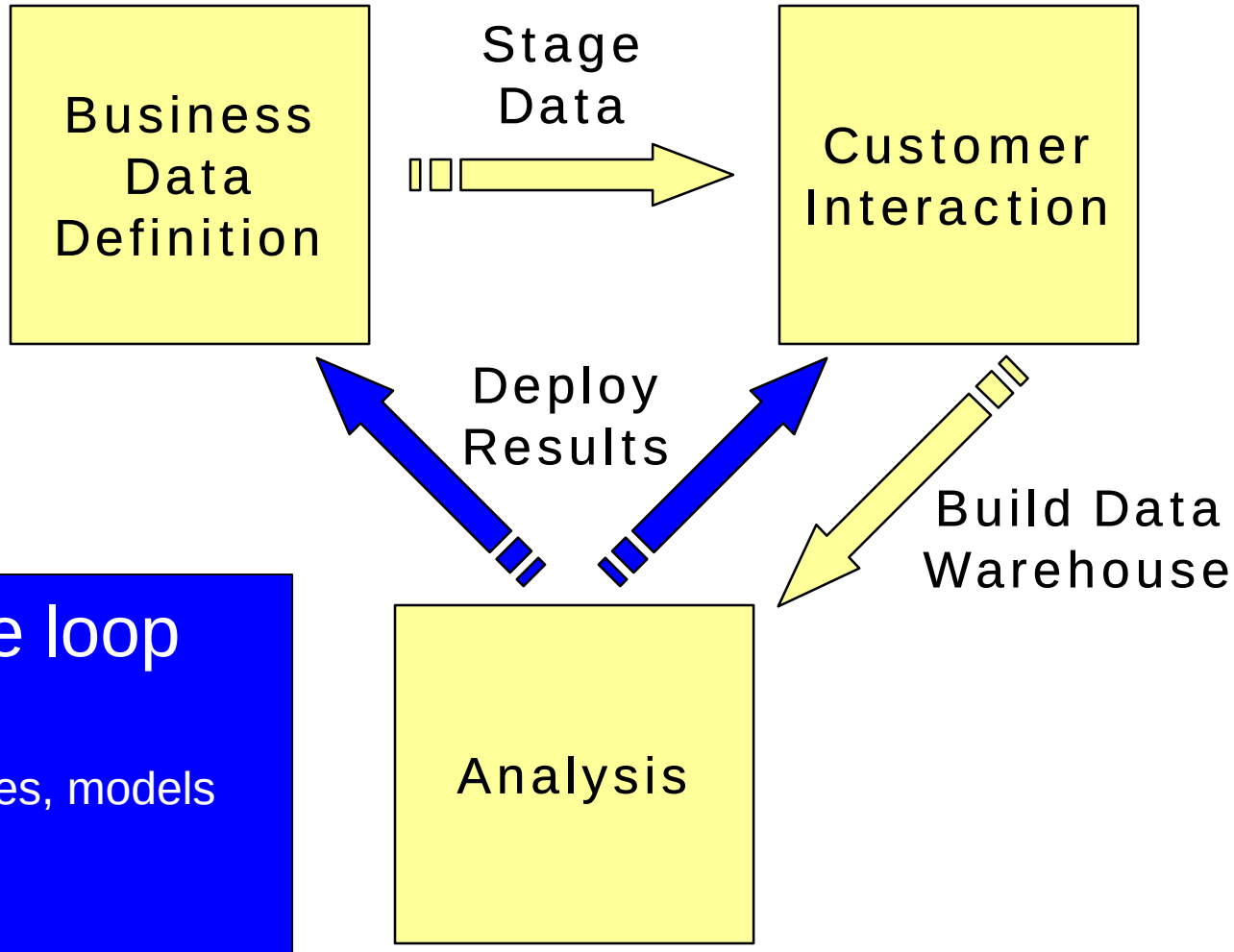
Analysis

Data transformations

Exploration

Modeling

Integrated Architecture



Close the loop

Transfer scores, models

Personalize

Overview

- **Warning: Your mileage may vary**
- **Introduction - the vision**
  - Webstore (interact with customers)
  - Analysis (understand)
  - Action (target)
- **Architecture**
  - Requirements
  - **The unfair advantage** Ü
    The integrated system provides much more than each component alone
- **Summary**

Trackers in the Real World

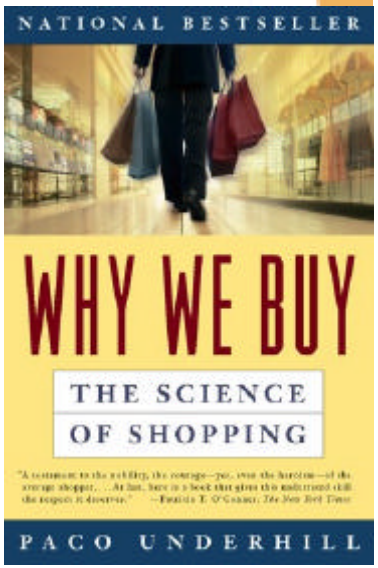
BLUE MARTINI SOFTWARE

Experiments in bricks-and-mortar stores are hard. Here is an excerpt from *Why We Buy: the Science of Shopping* describing a "log":

> **She's in the bath section. She's touching towels. Mark this down -- she's petted one, two, three, four of them so far. She just checked the price tag on one. Mark that down, too. Careful, her head's coming up -- blend into the aisle. She's picking up two towels from the tabletop display and is leaving the section with them. Get the time. Now, tail her into the aisle and on to her next stop.**

EnviroSell Inc. goes through 14,000 hours of store videotapes a year to do behavioral research

The web changes everything: clickstreams



The Web Advantage

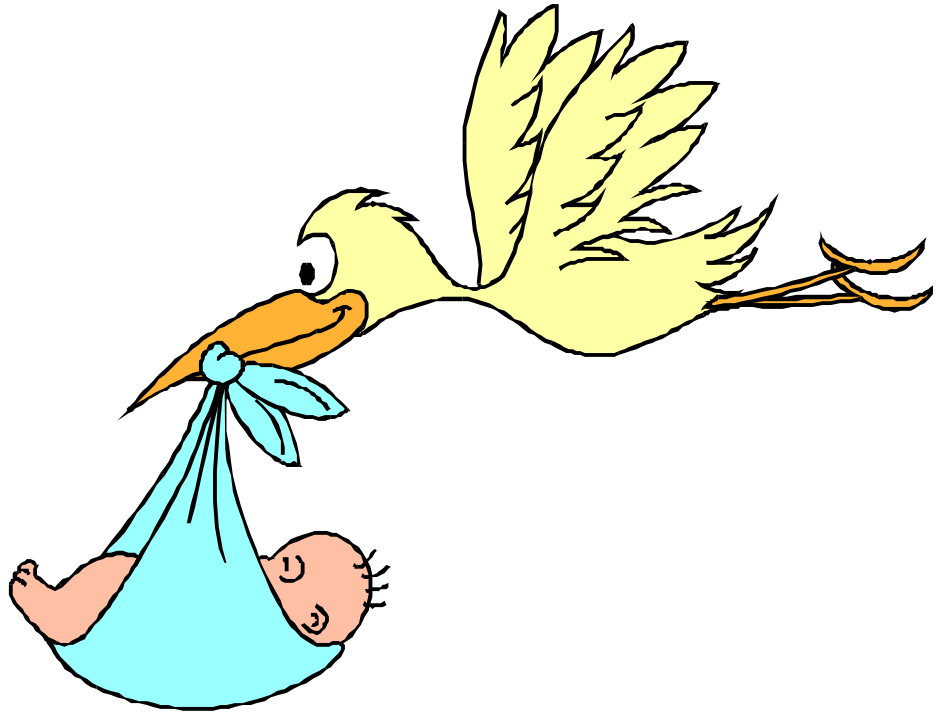
- **In e-commerce it is easy to change a site and measure the *effect* of changes**
  - One can easily set control groups on a web site
  - Easy to offer cross-sells or up-sells
  - Contrast with changing actual store layouts
- **Response to e-mails and surveys is days, not weeks and months**
- **Data is clean (unlike legacy data)**

Clean Data - Birth Dates

A bank discovered that almost 5% of their customers were born on the exact same date

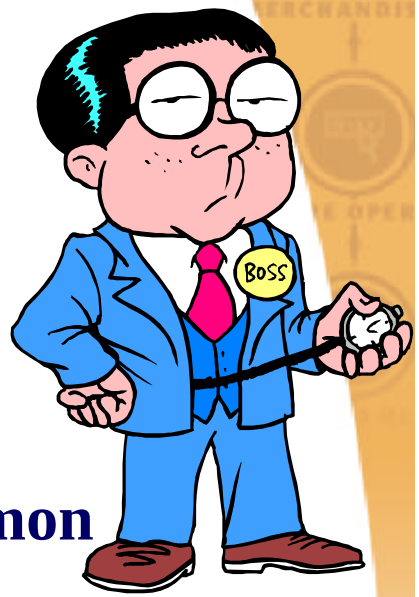
Why?

Hint: 11 Nov 1911



Data Transformations

- **80% of the time spent in data analysis is typically spent transforming data**
- **An integrated architecture can**
  - **Automate transfer of data from webstore environment to data warehouse**
  - **Provide data transformation UI**
  - **Provided "canned" transformations for common business problems**



Data Mining Advantage

BLUE MARTINI SOFTWARE

- **Humans have terrible intuition when there is a lot of data**
- **Example:**
  - **400,000 Americans/year die from cigarette smoking**
  - **Quick, how many fully-loaded Jumbo 747 planes crashes is this equivalent do?**

3 crashes every day, 365 days a year

Averages, Medians, Modes

- A person invests $100,000 in a volatile stock
- Each year it either rises by 60% or falls by 40%
- After 100 years, what is the
  - Expected value
  - Mode (most likely value)
  - Median (half the people will earn less than this, half more than this)

$1,378,061,234 (over $1B) =

100K x 1.1^100

$13,000=

100K x (1.6)^50 x (.6)^50

$13,000

(same as mode)

Summary

- Many sites spend millions of dollars in maintenance because they lack a good architecture
- Architect your solution early
- Think of scalability and efficiency
- Think ahead:
  - Many sites are beautiful but it's all CGI, which doesn't scale
  - Analysis is key - what are customers doing? Failed searches. Killer pages. Referrer pages/ads
  - Close the loop. Analysis without action has no ROI

Web Data and Contact Info

- **Clickstream data available for research/educational purposes at http://www.ecn.purdue.edu/KDDCUP/**
- **More questions? Ronnyk@bluemartini.com**