

After the Gold Rush: Data Mining in the New Economy

David Stodder
Editorial Director
Intelligent Enterprise
www.intelligententerprise.com

Intelligent
ENTERPRISE

About the Speaker

- Founding editor of *Intelligent Enterprise* (Sept. 1998), published by CMP Media Inc.
- Chief editor of *Database Programming & Design* (Nov. 1987 - Sept. 1998)
- Editorial director, Database Summit Series (including the Data Mining Summit, 1995-1998)

From Database to “Intelligence”...

- OLTP: The undisputed system of record (albeit relational, hierarchical, etc.)
- SQL Commoditization: Oracle, Microsoft, DB2 UDB
- Does it work with SAP? (or Peoplesoft, Baan, other packaged apps)
- Object/Relational or OODB? Who cares: Just make me E!!

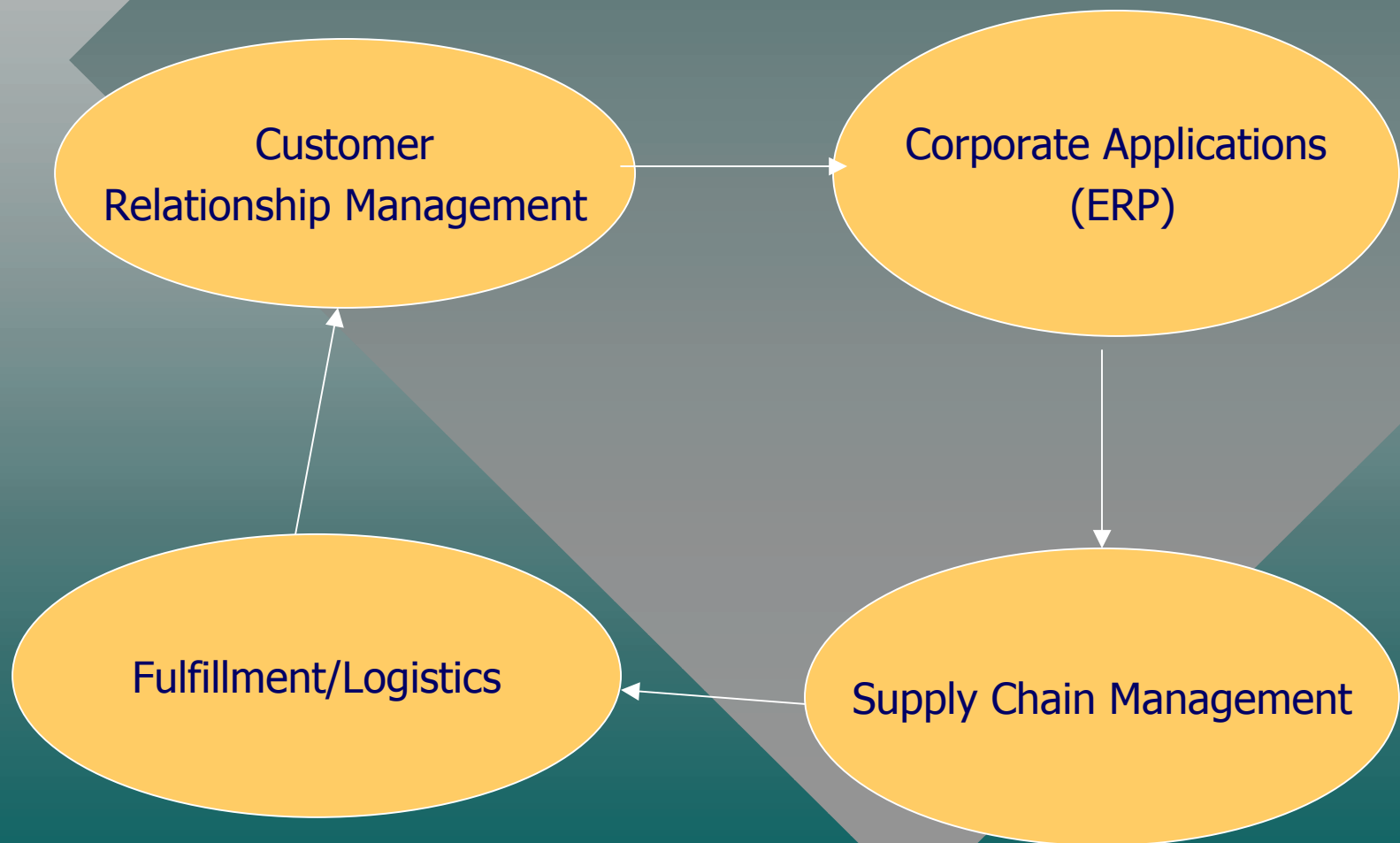
Intelligence, According to *Webster's*

- Ability to learn, to deal with new and trying situations
- Skilled use of reason; the ability to apply knowledge to manipulate one's environment
- Information concerning an enemy or possible enemy
- To be clever, alert, quick-witted, knowing

Elements of the Intelligent Enterprise

- Intelligence: Analysis, prediction, modeling, understanding
- Integration: Adaptable applications for intra- and inter-enterprise sharing
- Infrastructure: Platforms for rapid, efficient, affordable growth
- Collaboration: Partnerships to create and win new markets

Intelligence: In Demand Everywhere



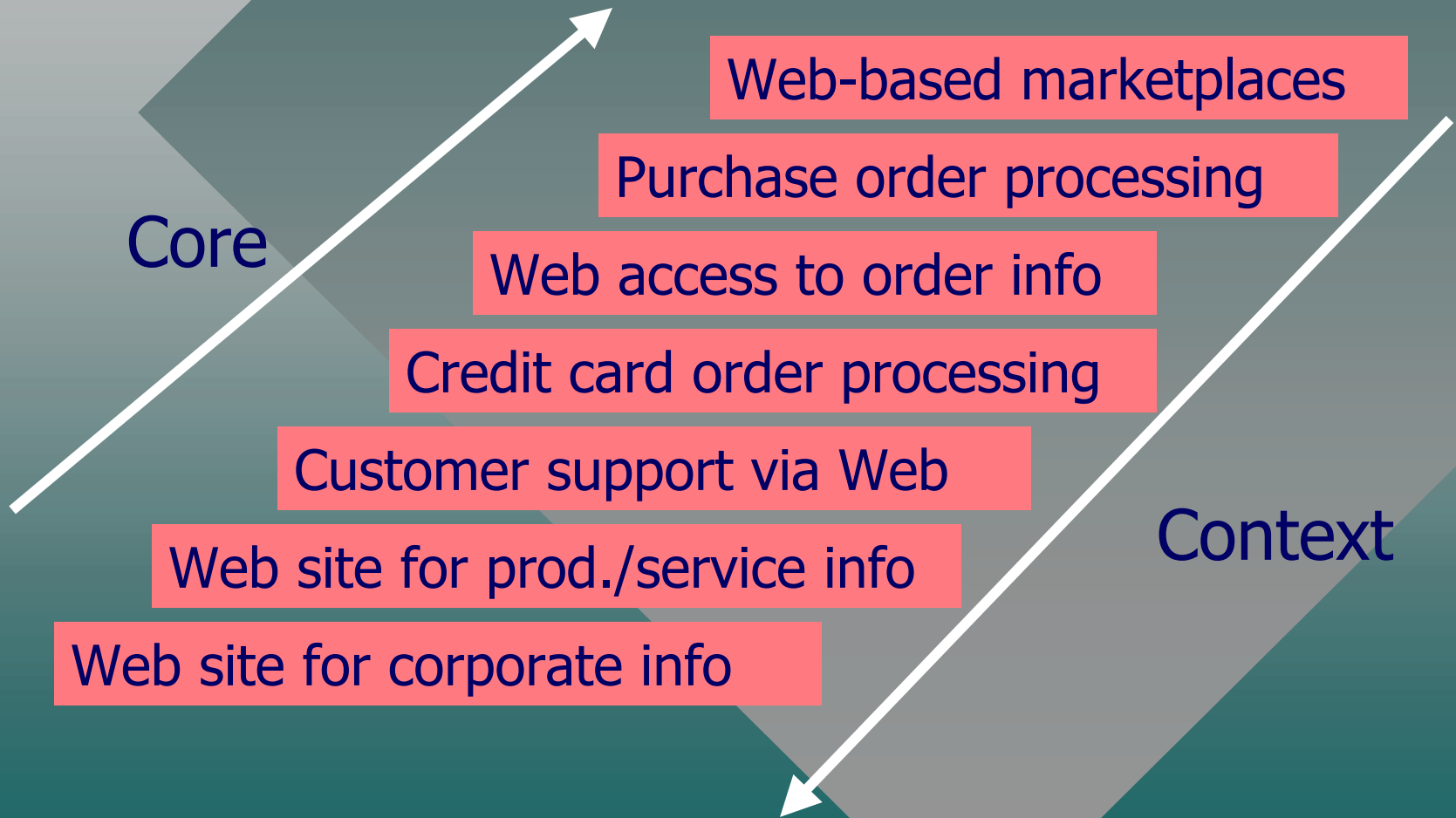
E-xplosion! What *was* that, anyway?

- Lots of credit: Market capitalization becomes the measure of success; capital flows to find highest risk-adjusted return
- Clicks and cookies: E-OLTP is a data Vesuvius
- We share: Networking becomes mainstream
- Darwin on steroids: Rapid differentiation becomes key to competitive advantage

Differentiation Based on Core Competence

- Core versus Context (Geoffrey Moore, "Living on the Faultline")
- Core: "When the outcome directly affects the competitive advantage of the company in its targeted markets."
- Context is like hygiene: What "the marketplace expects you to do well but gives you no credit for doing *exceptionally* well."

Moore's E-Commerce Escalator



Synchronicity: Timing is Everything

- Like a living organism, a business enterprise must develop within each window of opportunity
- A business enterprise must expect and anticipate change
- Short-term profits may be insignificant compared to fulfilling opportunities for market-leading growth

“Get Big Fast”: Still the Dot-Com Goal?

- Metcalfe’s “Network externalities”: The bigger you are, the more valuable your product
- Use economies of scale, but the future must now look profitable
- Positive feedback loop: Use every means to reach critical mass
- Advertise, advertise: The last Dot-com Super Bowl?

B2C: E-Tailing Was Just the Beginning

- Many information-intensive industries just getting started
- Online healthcare to reach \$370 billion by 2004: Forrester Research
- Insurance: eCoverage, Esurance, etc. attacking paper-heavy industry
- Publishing, entertainment: Industries under siege

IT's Challenge: Not Assets, but Information *About* Assets

- IT must become the “treasurer” of the information currency
- Riding networks, information and knowledge bind customers, partners, suppliers, employees
- Customer service becomes interaction
- Profit established through retention

Database: Defender of the Status Quo?

- Critical platform for the intelligent enterprise, but:
- Are designs stuck in the classic P&L view?
- Can the database architecture handle the data windfall?
- Sharing confronts centralized optimization, administration
- Databases normalize...

Data Mining: Instrument of Change

- Serving the organization's need to find value for competitive advantage
- Empowering domain experts, rather than financials/operations managers
- E-Business: Data comes from interaction
- Action is implicit: Without action, effort is too costly--at least right now

1995-1997: The Data Mining Bubble

- Refuge from the collapse of AI
- The nexus of machine learning/AI, predictive modeling/statistics, pattern recognition, visualization, decision support
- Market unfettered by 800-pound Gorilla
- Lots of hammers looking for nails: Over 100 tools for classification (decision tree, neural nets, etc.)

Knowledge Discovery Meets the Database

- Big-ticket application for massively parallel platforms and high-end hardware
- Tool for analyzing huge data volumes, beyond what statistics and DSS could handle
- But: You're going to do what with my database!?
- Expert users, extensive preprocessing required

Data Mining & Data Warehouse: Not Always Obvious

- Transformation layer for reporting and OLAP
- Priority of aggregation and summarization; data mining likes granularity
- Solving the distributed database/replication problem
- Competition from OLAP, ROLAP, MOLAP

Defining a Process

- Knowledge Discovery: “The process of identifying valid, novel, potentially useful, and ultimately understandable structure in data” (Fayyad & Simoudis, 1998)
- Closed loop of business improvement: Track, Analyze, Use Predictive Models, Adjust...and then do it again
- Using OLAP to direct data mining, and vice versa

IDC Top Five, 1997-1998

- Data Mining Adoption: Banking, Insurance, Other Finance, Health Care, Education (*then* Retail, Telecomm)
- Data Warehouse Adoption: Other Finance, Communications, Banking Wholesale/retail, Education, Insurance
- Fraud detection well publicized, but customer relationship, sales/marketing of higher interest

Winnowing Out: Top Statistics and Data Mining Vendors (IDC, 1999)

- SAS Institute (29.3% market share and increasing)
- SPSS (13.5)
- IBM (6.1)
- Silicon Graphics (3.2)
- Mathsoft, Hitachi, Unica Tech., Angoss, Cognos

Out of the Spotlight: First the Web, then CRM

- Silver bullets are always most dangerous to the gunslinger
- Acquisition and consolidation into suites, led by SAS, IBM, and SPSS (Clementine)
- Process integration hits the front office: Trying to make customer relationship management (CRM) “like” ERP--with all the lock-in potential

CRM: Bonanza Days Return

- Analytics and business intelligence becoming the competitive differentiation among suites
- Stunning round of acquisitions (Xchange buys Knowledge Stream Partners, \$50+ m)
- Tools and services headed for the solutions market

DM/CRM: Software or Service?

- Happening in the midst of the ASP/hosting trend
- Service from Data Providers (Axciom, etc.)
- Advertising Agencies, Direct Marketers
- Major vendors (Intel, Sun, IBM) providing data mining over the Web

CRM: The Establishment

- Siebel Systems (18.4% market share, IDC)
- Oracle (6.7)
- Peoplesoft (4.9)
- Clarify (4.8)
- Trilogy, Group 1, Broadvision, Goldmine, Onyx, etc.

CRM: Gunning for E-Business

- Siebel eBusiness 2000 (personalization, marketing automation)
- Blue Martini (E-Business with deep analytics and data mining)
- Unica Corp. (Affinium Suite: marketing automation, data mining, predictive modeling, “Universal Dynamic Indicator” transformation/access layer)

CRM: BI Vendors in the Fray

- SAS Institute: Expanding suite to handle e-BI, CRM
- Hyperion Solutions: Vertical suite integration
- MicroStrategy, Cognos, Business Objects: Analytic Applications for CRM

CRM/DM: Waking Up the Database Industry?

- Oracle: Acquisition of Thinking Machines, but more an 11i thought
- IBM: Data mining support with extenders; SQL integration coming
- Microsoft: OLAP Services now Analytical Services; OLE DB for DM
- Sybase: Service Solution, support with DBMS

Data Mining: Defining One-to-One

- Personalization means granularity
- Data mining: On the precipice between mass marketing (huge data sets) and mass customization (market of one)
- Finding share of customer, not just market share

Opening Up Analysis

- Data mining as catalyst for interaction; customers teach models
- “The Kitty Hawk Era” in personalization (Jeff Bezos, Amazon.com)
- Intelligent Agents
- Search engines (music minus one)

CRM: Focusing Technology Trends

- Pattern recognition
- Profiling
- Visualization (GIS, spatial integration)
- Text mining (since so much customer data, advertising is not structured)

Visualization and Text Mining: Areas to Watch

- Frustration with search engines
- Data battles: Search engines scuffle to control results
- Text mining: Is it mining?
- Visualization: Visions of interactive multimedia