OpenBI

where strategy meets intelligence

Open Source Business Intelligence in the Cloud

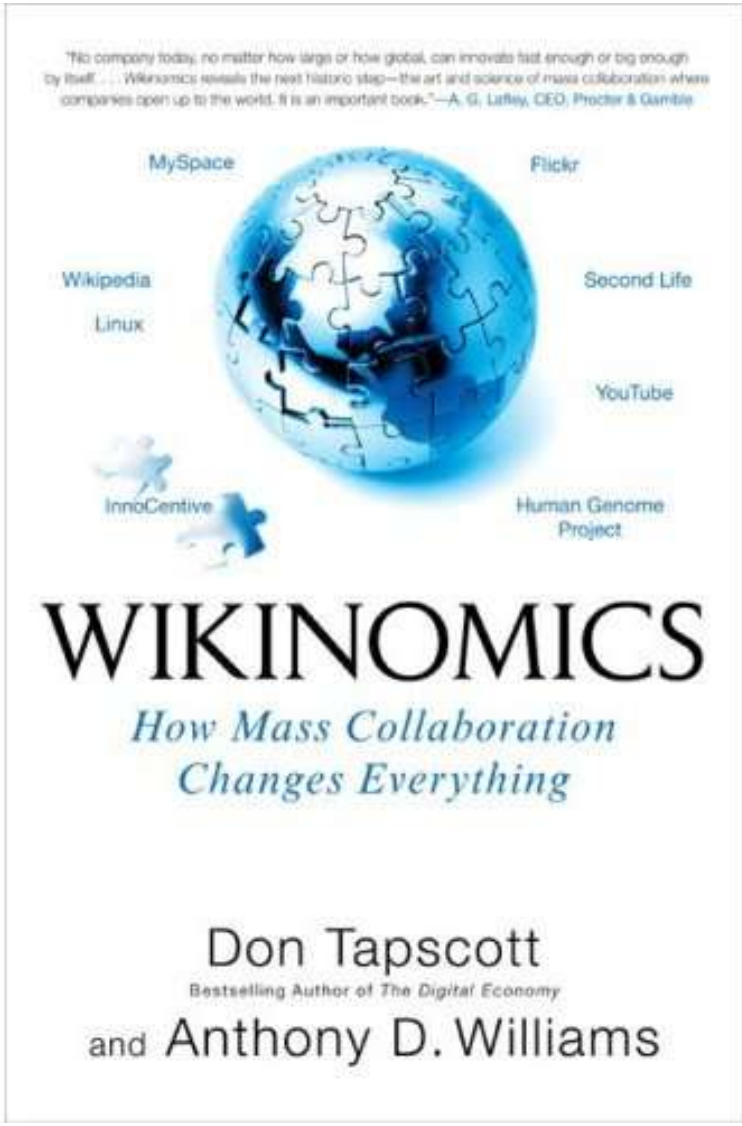
July 2009

Business Intelligence in the Age of Collaboration

"**Billions of connected individuals can now actively participate** in innovation, wealth creation, and social development in ways we once only dreamed of.

And **when these masses of people collaborate** they collectively can advance the arts, culture, science, education, government, and the economy in **surprising but ultimately profitable ways**."

Don Tapscott and
Anthony D. Williams



Why Open Source Software?

Commercial Open Source is changing the rules

- Customer Control
  - Free, global access to software
  - Removal of license fee amortization
  - Annual “proof-of-value” for vendors
- Lower Costs
  - < 50% of the software cost of proprietary alternatives
- Better Technology
  - Modern, open architectures
  - Global innovation engine

Sampling of OSBI Related Technologies

Open Source Business Intelligence Technologies

* WEKA is part of the Pentaho BI Suite

Free Open Source vs. Commercial Open Source

Free Open Source differs from Commercial Open Source in a number of ways.

- Free Open Source
  - Informal support and broad services providers
  - Uneven velocity of change
  - Community-directed roadmap
  - Functional gaps
  - Challenging licensing provisions

- Commercial Open Source
  - Formal support with service level agreements (SLAs)
  - Indemnification
  - Professional services and partnerships
  - Product management and roadmaps, and advisory boards
  - Business-friendly subscription models
  - Reference accounts, cases studies, and user groups

Source: Pentaho

Differing Philosophies on Open Source

OpenBI

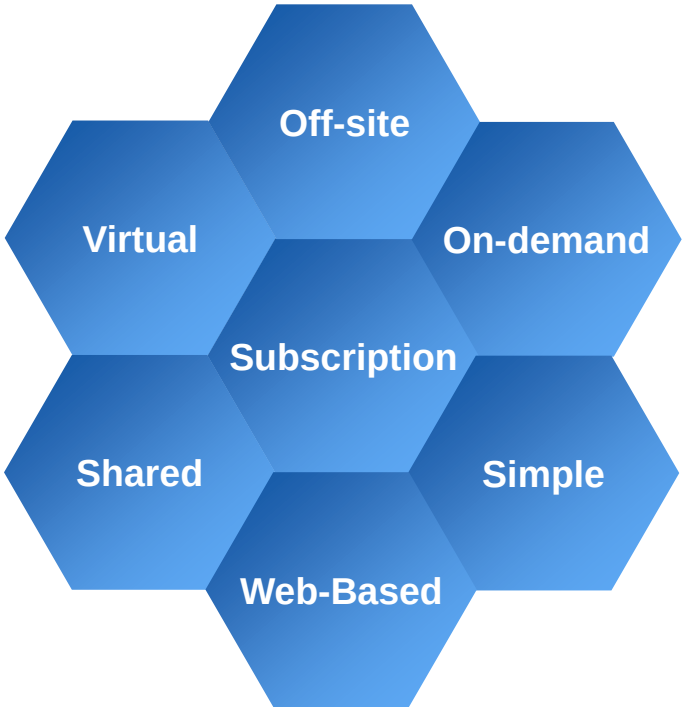
> I think it addresses a niche market for high-end data analysts that want free, readily available code. We have customers who build engines for aircraft. I am happy they are not using freeware when I get on a jet.
>
> Anne H. Milley, director of technology product marketing at SAS

> It's interesting that SAS Institute feels that non-peer-reviewed software with hidden implementations of analytic methods that cannot be reproduced by others should be trusted when building aircraft engines.
>
> Dr. Frank Harrell, Professor of Biostatistics and Department Chair at Vanderbilt University and R Community Member

What Is “Cloud Computing?”

- “Cloud computing is on-demand access to virtualized IT resources that are housed outside of your own data center, shared by others, simple to use, paid for via subscription, and accessed over the Web.”
  - –John Foley, Information Week, September 2008

Seven Principles:



…or, use how Larry Ellison described it:
"idiocy," "crap," "gibberish," "crazy," and "stupidest"

What isn't Cloud Computing (our def)

- Google Apps
  - Maybe a "Cloud Application"
    - End User cannot determine WHAT they run
- IBM Computing on Demand
  - Not truly "on demand"
  - Activate physical processors already within a Box
  - Not utility based (yet)
- Microsoft Azure
  - Limited to the Azure "technology"
- Etc.

The Amazon Elastic Computing Cloud
Powered By Open Source

"If an economic downturn cools IT capital spending, some business technology managers may turn to rent-by-the-hour cloud computing resources…

If they turn to Amazon EC2, they're tapping into open source Linux, Apache, and a tweaked Xen open source hypervisor that powers much of the company's cloud's operation."

Information Week, November 2008

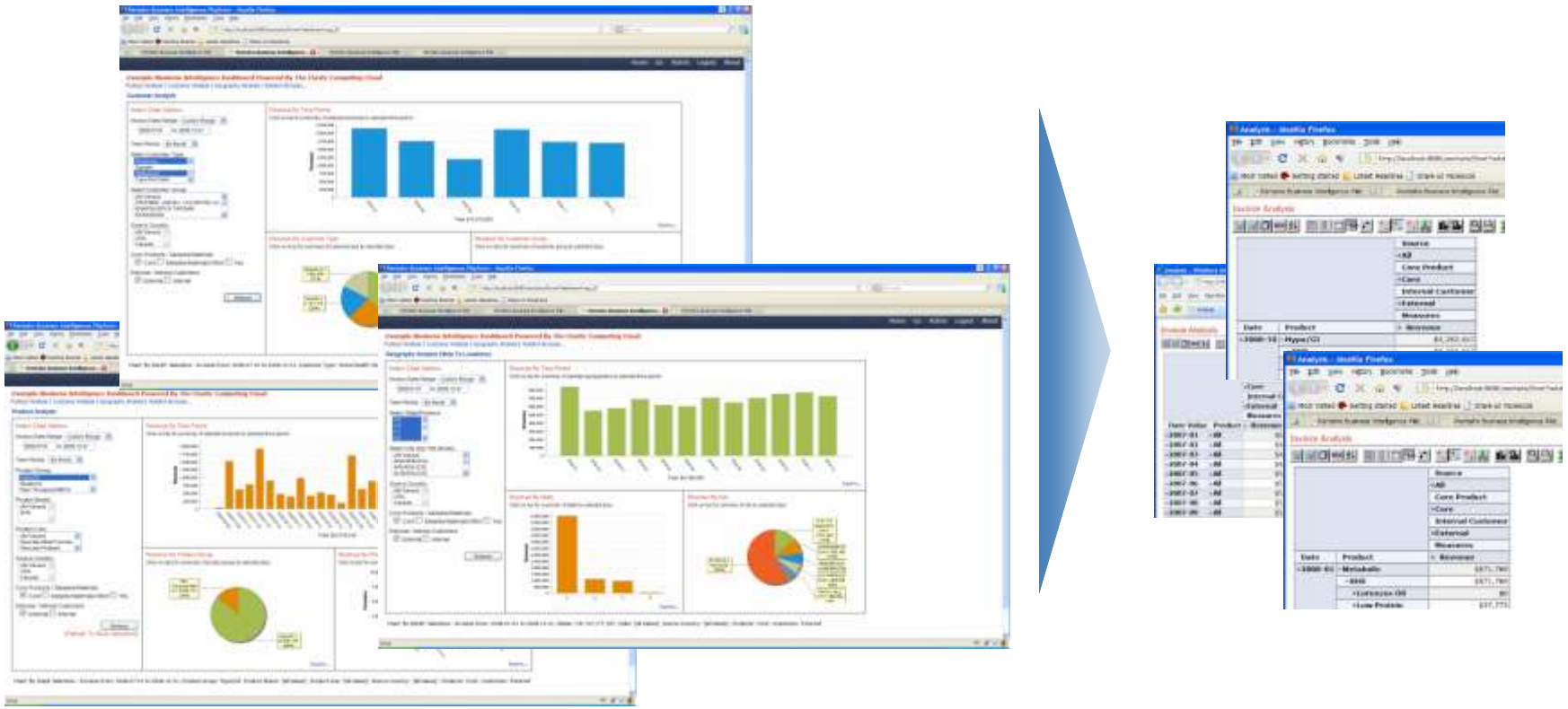
Amazon Elastic Computing Cloud (EC2)

- Most well known “Cloud Computer”
- Allows customization of Amazon Machine Images that can be started and run on demand
- Different instance sizes from small 32-bit (1 CPU, 1.7GB RAM equivalent) to extra large 64 bit (8 CPU, 15GB RAM) or extra large 64 bit, high CPU (20 CPU, 7GB RAM)
- Runs varied operating systems (Linux, Windows) and charged on an hourly basis (Windows is 25-50% more expensive)
- Can attach persistent storage to an instance, charged by the GB
- Accessed via command line or web interface
- Some data charges apply for transfer in and out of Amazon
- Competitors:
  - IBM (Computing On Demand), Google (App Engine), AT&T (Synaptic), Microsoft (Azure), Sun (OpenCloud)
  - Rackspace, Flexiscale, GoGrid

Open BI in the Cloud: Danone/Nutricia

Client Profile	Nutricia, a division of Danone, specializes in Baby and Medical Nutrition products. They provide medical nutrition for the management of conditions such as milk protein allergy, inborn errors of metabolism (e.g., PKU), pediatric epilepsy, Alzheimer's & more. Nutricia markets its products across 19 countries.
Project Background	Internal order management system was limited in providing analytical insights on products, product groupings, time, customer, or geographic analysis in the aggregate.
Scope	Build a pilot analytical database and web-based business intelligence application to allow business users to see a high level snapshot of business performance, and be able to drill into detailed order and invoice activity to reveal performance trends.

Cloud BI In Action: Danone/Nutricia



- Web-based sales performance dashboard provides quick analysis on sales activity for products, customers, and sales regions
- Allows drilling from dashboard into interactive OLAP analysis sessions for a deeper look at sales activity

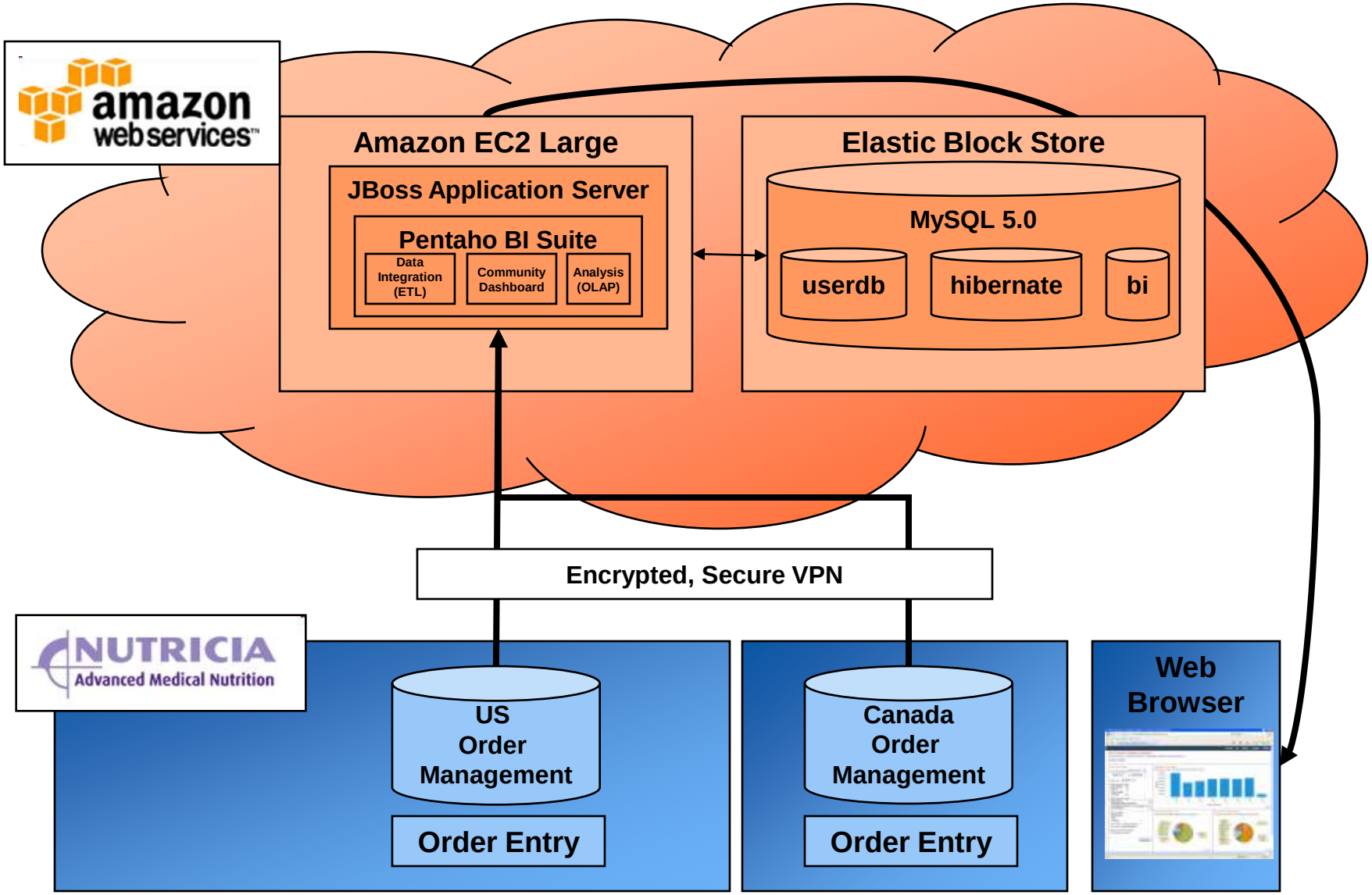
Community Dashboard Framework

- Demonstrates open source community development in action
  - Started by Ingo Klose and Pedro Alvares, first community releases were in 2008
- Provides a framework and templates for simple dashboard building
  - Includes basic selection/filtering objects, including text boxes, multi-select pick lists, calendar date selections, check boxes, etc.
- Uses the Pentaho platform's "guts" to provide data from databases, transforms, etc.
  - Allows Pentaho reports, charts, OLAP sessions and other objects to be embedded in the dashboard.
- Version 3.0 released Jan 2009

Environment Overview

- The pilot environment is hosted on an Amazon EC2 Large (Approx 8GB, 4CPU) Instance. This instance contains:
  - JBoss Web Server
  - Pentaho BI Suite
    - Includes custom web page templates, charts, and OLAP views
    - Community Dashboard Framework
  - Pentaho Data Integration
    - Includes custom ETL Routines to extract, transform and load data from the operational systems.
  - MySQL Database
    - Stores BI database, Pentaho Repository, and User Database

Environment Overview



Case Study - Conclusion

- Project took approx 6 weeks, including requirements, design, build and deploy to the cloud
- Has been operating since July 2008
- Users within and outside of the client's walls have secure access to performance metrics

Five Key Cloud Considerations

1. Everything is the same, and everything is different
   - OS is the same
   - Software installation and configuration is the same
   - Differences:
     - Connectivity
       - Secure connections over the internet
     - Persistence
       - “Native” file system is transient across “reboots”
         - » Internal Instance Storage vs. Elastic Block Store
       - Determine what is included in the instance image
         - » Isolate software versus “data”
         - » Use the Elastic Block Store for persistence

Five Key Cloud Considerations

OpenBI

2. Plan your sizing (at least a little)

- Pick your “bits”
  - Upfront decision between 32bit and 64bit
  - Starting with small…
    - Reduces upfront price but requires new AMI image creation for upsize
  - Consider your end platform need
- Pick your “up time”
  - 8-5, 6-6, WeekDays Only
  - Reserved Instances (16hr/day)
  - Remember “data” persistence

3. Plan your ETL extraction technique (pay by the byte)

- JDBC
  - Requires direct connection to source DB
- File transfer
  - S/FTP
  - Requires extracts

Five Key Cloud Considerations

4. Plan your security
   - MSAD utilized
     - Required VPN 24 x 7 (availability)
   - Mirror RDBMS on Cloud
     - Required ETL/Process
5. Your end-user functionality can be identical
   - Use DNS server aliasing
   - “Mercy” of internet
     - Has not been an issue, for us, but could for “data dumps”

Thank You!

Kevin Haas

kevin.haas@openbi.com