



SOFTWARE DESIGN & DEVELOPMENT

Shaking up the Payment Services through Platform Integration

Jack Hakim, President

San Francisco*Seattle*Cheng Du, China

www.ecwise.com

EC Wise Background

- Senior team with > 20 years experience on data and business integration
 - Data management systems
 - Data modeling
 - Business analytics
 - Simple user interface design
- Wide range of clients addressing 3 primary areas
 - Fortune 1000 non-technology companies
 - Large Software companies
 - Venture backed startups

Sampling of Our Clients- Big and Small

- **Amazon.com**
- American Society Mech. Eng.
- Apollo Enterprise
- BIPT
- **Business Objects**
- Captura
- ClassroomConnect
- CoolerMaster
- Dasein Design
- DominyLevin Psych.
- iSayso
- iTango
- Joe Average
- **Kodak**
- Lehman Brothers
- Media Net Link
- Micromega Systems
- **Microsoft**
- Mindport
- **Oracle**
- Onrain
- Pharmacon
- **Regulus**
- SBT Accounting Software
- **Solar Turbines (Caterpillar)**
- Trilogy Resources
- **US Veterans Administration**
- **Vulcan Media Services**

Case Study – The Customer Environment

- Industry Leading Print, Remittance and ePresentment Processing Company
 - Many Acquisitions
 - Many facilities
 - Distributed Printing
 - Remittance Processing
 - IT processing environments
 - Site-by-site application development
 - 1000 independent applications and data extraction modules

The Business Challenge

- Provision many applications vs custom code
- Reuse target data mapping and processing components
- Improve operational efficiency
- Improve quality and provability
- Support financial, health and security regulatory standards
- Distribute across all Remote Printing and Remittance sites (postal needs)
- Reduce the cost of upgrades across clients
- Provide clients with a view across lines of business
- Expand BI services back to end users

The BI Challenge: Need for a Data Collection and Aggregation Platform

- Support efficient centralized application development and processing
- Enable efficient data aggregation from customer's client as mergers take place (e.g. large telecoms)
- Create vertical products by combining billing presentment and remittance processing data
- Enable the merging and analysis of paper and electronic processes
- Enable BI Services via data mining and OLAP through a common data repository (e.g. Cash management)

ETL Build or Buy – BUY!

- Reduce development time/expense of creating and maintaining ETL processes with a minimum of procedural code
- Decrease development/QA time cost
- Eliminate costs in developing
 - Front-end provisioning environment
 - Back-end transformation capability
 - Data drivers for multiple data sources/targets
 - Logging/audit
 - Maintenance support/ongoing enhancements
 - Shared meta data repository

Critical Criteria for ETL Evaluation

- Automatic generation of data extraction programs using drivers provided by vendor
 - Generate test data (also support QA needs)
 - Reuse 70%-80% of all data mappings (using existing string/date/number /predefined functions)
- Regulus has thousands of applications with at least 5-30 processes, 15-20 steps, 5-30 transforms per app.
 - Extensive operational reports (statistics, diagnostic messages) for all data load processes, steps, maps
 - Call ETL process from Java, C++, shell scripts
 - Reuse DI maps used prior, when migrating to new platform (Data sources are not changing)
- Extensive platform needs-profiling, adapters, complex business transforms

Critical Criteria for ETL Evaluation (cont'd)

- Support most data formats (flat files, a variety of RDBM's, EBCIDC, XML, etc.) DI supports Regulus customer's connectors
- Provide a complete development environment: engine profiler, run-time debugging, version control, automatic code generation for lookups/commonly used functions (date, string, logic control)
- Graphical front end development environment
- Platform independent transformation engine (Unix/Linux/Windows)

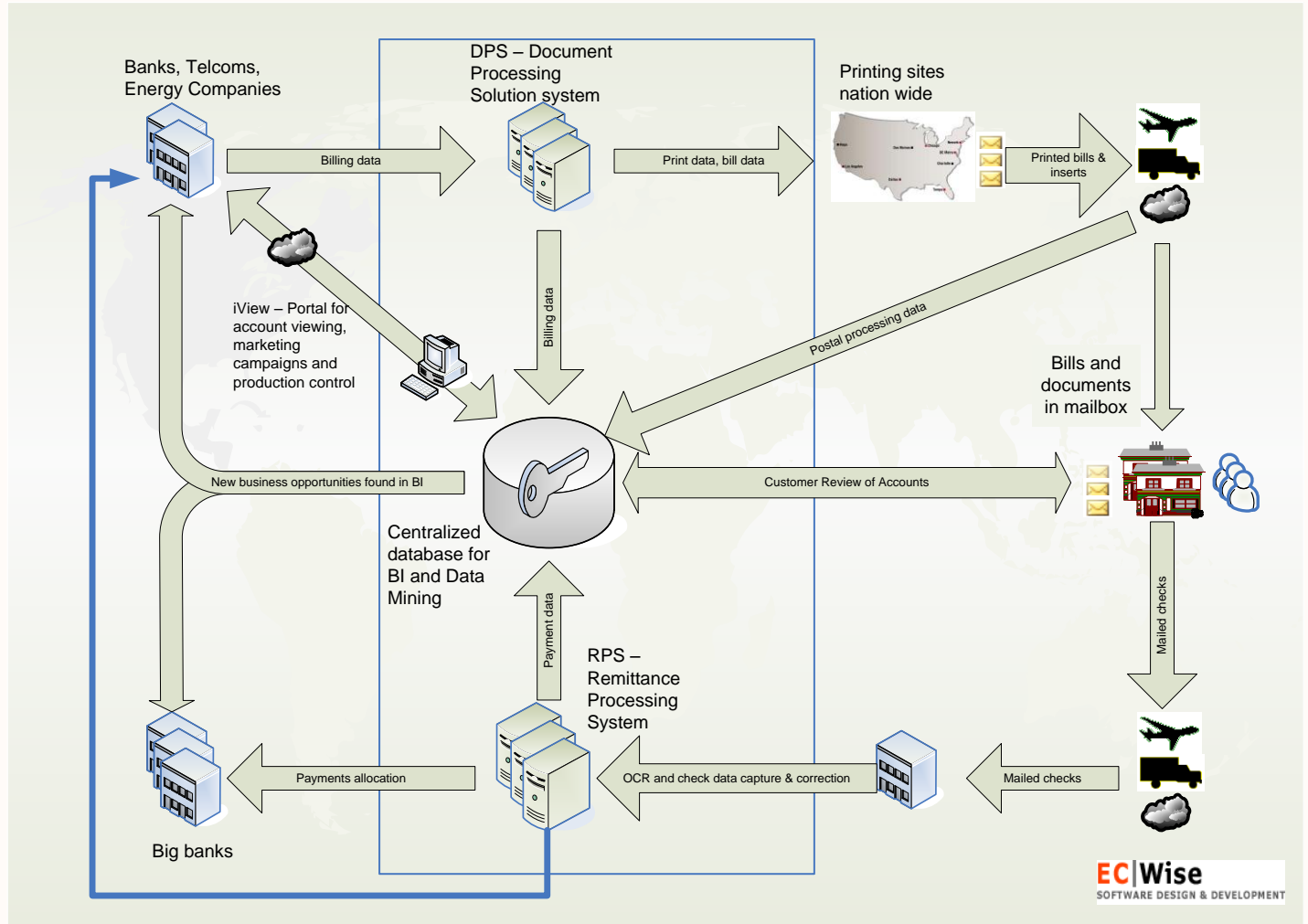
Why Pervasive Data Integrator?

- Multi-thread
 - Hundreds of time compressed SLA's
 - Billions of monthly source records
 - Nightly aggregation/data feed to BI DW
 - Potentially 1000s of applications running at same time
- Pass engine resource parameters at run-time
- Integrate conceptual, physical and logical models with DI (import ERWIN and Power designer schemas)
- Exceptional value for price (UI vs. \$)

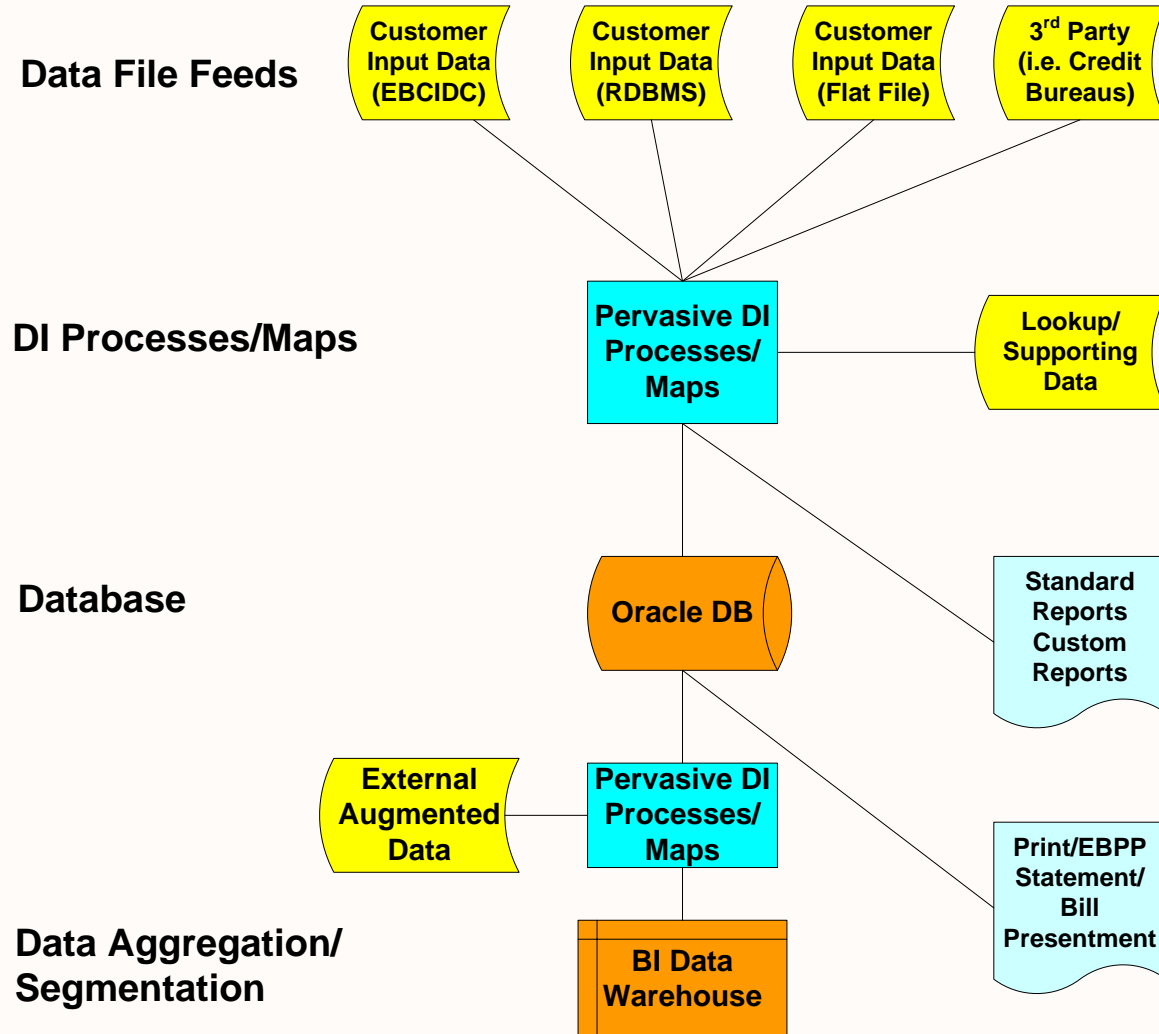
Why use Pervasive's Data Integrator for DW and BI?

- \$ - Data mapping and cleansing across 1000s of applications using custom code would be **more than 10 times the total budget** for data warehouse construction !
- Facilitated DW Implementations
 - Need to populate dimensional structures for analysis and use either direct table load or flat file for later bulk data load
 - Need to populate star schemas at either detail or summary levels
 - Need the ability to provide bulk data movement as well as transactional data/message processing

Unified Platform Print Remit an EBPP



ETL Solution: A Bird's Eye View



High Volume -> Optimization

- 10 files of EBCIDC Data per day – Average 4 Gig total
- Fixed record length – 1200 bytes
- Peak 1 million statements p/day
 - 250,000,000 inserts
 - Volumes Increasing

Time to load Label, Value and Section tables	
Multimode	2 hrs+
Optimized Multimode	1 hr 20 min
SQL Loader - .dat	26 min
SQL Loader - sql.log	13 min
SQL Loader - Multi Instance	6 min

Our Approach

- Single integrated Print, Remit and EBBP Platform
- Utilizing Pervasive Data Integrator:
 - Isolate a canonical standard from variable client data feeds
 - Convert custom rule sets and legacy client applications to “configuration sets”
 - Enables merging of paper (print and remittance processing) services with electronic Bill Presentment and payment (eBPP) services
 - Accelerate creation of a data warehouse to enabling vertical services based on integrated Remittance and Presentment data
 - Enables corporate-wide data mining and business intelligence processes

Lessons Learned

Many, many, many, many lessons, I'll mention a few

Lesson 1: Data Profiling

- We used our standard data analysis techniques to profile data
- We didn't fully validate the provided content, sample data files and test data truly represented the data

We now would use tools to:

- Fully profile everything
- Document extreme outliers
- Log invalid data to separate files
- Forward only valid data to transformation/loading steps

Take away -> **Use Data Profiling tools as the first step**

Lessons Learned (cont'd)

Lesson 2: Performance in the large

Post implementation performance tuning is very expensive

- Make sure that every component of the ETL/Database environment architecture is scalable
- Splitting across 5 Transformations (when performing sequentially)... is much slower than multithreading

Take away -> **More splits across transformation objects decreases performance**

Aside: Keeping the data flowing through the transformation objects rather than branching around them is always preferable for performance.

Take away -> **Create multithread architecture from project inception**

Lessons Learned (cont'd)

Lesson 3: Performance in general

- Running mixed-mode (inserts and updates) within the same stream causes performance slowdown by orders of magnitude

Take away -> **Rule of thumb with ETL: Always separate your inserts from your updates from your deletes**

Lessons Learned (cont'd)

Lesson 4: WOW Building acceptance in the organization

Take away ->

- Prove the value (i.e. DI vs. tool of choice)
- Upgrades drive opportunities for improvement
- Best practices
 - Share BP techniques (tuning, data parsing)
 - Share tricks/workarounds/shortcuts
- Pair “programming”

Lessons Learned (cont'd)

Lesson 5: Use DI to prepare data; write to ascii files, bulk loading of data into tables

Take away ->

- Make use of RDBMS bulkloaders
- Apply business rules to individual fields/records within DI
- Apply business rules to sets of records in SP after target is loaded

Lessons Learned (cont'd)

Lesson 6: Specify metrics that define your performance expectations

- Data extract times
- Data cleansing and transformation times
- Load times
- Query response times
- Report-generation times, etc.

Without the right set of metrics there is no valid baseline to compare and monitor performance improvement

Take away -> **Build sufficient performance metrics into the system from the start**

Recipe for Leveraging Data Integrator in N Way Integrations

- Profile both the data and meta data; understand and define the metadata, then pull it into new processes
- Scope the data references and processes
- Define all use cases, scenarios, paths, rules, workflows
- Explore approaches to add value and reduce cost
- Prototype, proof of concept, and optimization plan before building
- Use data modeling to define conceptual and logical models; abstract before defining common components
- Data simulation and testing, **high volume -> edge cases**
- Final performance tuning after implementation
- Leverage strong database expertise ... critical asset!
- The ease of use of DI and it's fast ramp-up time lets our clients be productive in a short period of time

Questions ?

For more information please contact :

**EC Wise Headquarters
and California
Development Center**

101 Glacier Point Road,
Suite D
San Rafael, CA 94901

Voice:
415-355-WISE (9473)

Fax:
415-436-WISE (9473)

**EC Wise Research and
Development Center,
Chengdu**

Qing Jiang Dong Road #1
Vancouver Square, 13th Floor
Suite GH
Chengdu, P.R. China 610072

Voice:
87710503 , 87710504

Fax:
87710503-802

www.ecwise.com

EC Wise and ETL

- **Pervasive partner with certified trainers used by Pervasive to train people on Pervasive's tools**
- **More than 12 certified Pervasive developers**
- **Hands-on experience in solving large, high volume, complex ETL problems**
- **In depth knowledge of available ETL tools**
- **Experience in designing ETL tools for ETL vendors**
- **Experience in evaluating, recommending and applying ETL tools for clients and vendors**
- **ETL modeling skills include the ability to:**
 - **Identify the client's ontology and set standards to represent the relevant domain and universe of their core data**
 - **Developing and defining the terms, attributes, use cases and workflows appropriate for the industry domain**
 - **Develop efficient models and processes for mapping data**
 - **Best practices for ETL and DW**

Industry Expertise

- Banking and financial
 - Banking, Automated payments processing, Securities ...
- Business management applications
 - Accounting software, CRM, Knowledge management ...
- Entertainment Industry
 - Media-on-demand: product ideation to delivery
 - Digital management, selection and delivery of multi-media asset
- Ecommerce
- Process control systems
 - Pipeline control and monitoring
 - Production defect tracking and analysis
- Telecommunications
 - Provisioning, Analytics ...