

# PORTAL DE PROCESSOS E ETL PARA INTEGRAÇÃO DE DADOS PROCEDENTES DE BASES DISTRIBUÍDAS E HETEROGÊNEAS<sup>1</sup>

*Turíbio Tanus Salis*<sup>2</sup>

*Cadson Alexandre Alves*<sup>3</sup>

*Marcelo Vale Ferreira Menezes*<sup>4</sup>

*Sérgio Cordeiro*<sup>5</sup>

*Glauber Luis Pereira Lima*<sup>6</sup>

*Cláudio Magno do Carmo*<sup>7</sup>

*Cleudir José De Castro Junior*<sup>8</sup>

*Andre Ribeiro Almeida*<sup>9</sup>

## Resumo

O desenvolvimento de sistemas industriais por diferentes fornecedores para atender áreas produtivas distintas, utilizando ferramentas diversas em diferentes eras tecnológicas, acaba por gerar um grande número de repositórios de dados distribuídos, heterogêneos que são de difícil gerenciamento e exploração. A situação agrava-se ainda mais com transações corporativas como aquisições de novas unidades produtivas e fusões. Torna-se então interessante para as empresas, integrar e organizar os dados, buscando redução de custos de gerenciamento e facilidade na exploração dos mesmos. Neste contexto, o presente trabalho busca apresentar o desenvolvimento de um Portal de Processos e a utilização de uma ferramenta ETL para integração de dados heterogêneos e distribuídos. A ferramenta ETL empregada foi o ODI (Oracle Data Integrator) da Oracle e o Portal de Processos foi desenvolvido utilizando-se ferramentas Web da Microsoft como o MS Sharepoint e o Silverlighth, além de bibliotecas específicas de comunicação com o PIMS, bibliotecas de acesso á bancos de dados diversos e também web services.

**Palavras-chave:** Integração de dados; ETL; Portal de processos; Armazém de dados.

## DATA INTEGRATION COMING FROM HETEROGENEOUS AND DISTRIBUTED DATA BASES BY USING PROCESS PORTAL AND ETL

### Abstract

The development of Industrial Systems by different suppliers to meet the necessities of different production areas, using tools from diverse technological ages, produces a large number of distributed and heterogeneous data repositories that are difficult to manage and explore. This situation is intensified with corporative transactions such as acquisitions and mergers of new units. Therefore, it becomes interesting for companies to integrate and organize its data in a way that saves management costs and decreases exploring efforts. In this context, this study aims to present the development of a Process Portal and the use of an ETL tool for heterogeneous and distributed data integration. The ETL tool used was ODI (Oracle Data Integrator). The Process Portal was developed using Microsoft Web tools such as MS Sharepoint and Silverlighth. It was also used some communication libraries for PI PIMS System, libraries for several databases access and also webservices.

**Key words:** Data integration; ETL; Processes Portal; Datawarehouse.

<sup>1</sup> *Contribuição técnica ao 17º Seminário de Automação e TI Industrial, 24 a 27 de setembro de 2013, Vitória, ES, Brasil.*

<sup>2</sup> *Analista de Sistemas de Informação. Msc. Analista de Automação. Vallourec e Sumitomo Tubos do Brasil. Jeceaba, MG, Brasil.*

<sup>3</sup> *Eng. de Controle e Automação. MBA. Vallourec e Sumitomo Tubos do Brasil. Jeceaba, MG, Brasil.*

<sup>4</sup> *Eng. de Controle e Automação. PMP. Gerente de Sistemas Industriais. Vallourec e Sumitomo Tubos do Brasil. Jeceaba, MG, Brasil.*

<sup>5</sup> *Eng. de Controle e Automação. Techsteel Sistemas Especiais. Belo Horizonte, MG, Brasil.*

<sup>6</sup> *Eng. de Controle e Automação. Techsteel Sistemas Especiais. Belo Horizonte, MG, Brasil.*

<sup>7</sup> *Eng. de Controle e Automação. Techsteel Sistemas Especiais. Belo Horizonte, MG, Brasil.*

<sup>8</sup> *Eng. de Controle e Automação. MBA. Vallourec e Sumitomo Tubos do Brasil. Jeceaba, MG, Brasil.*

<sup>9</sup> *Eng. de Controle e Automação. MBA. Vallourec e Sumitomo Tubos do Brasil. Jeceaba, MG, Brasil.*

## 1 INTRODUÇÃO

O problema da integração de dados em ambientes industriais, principalmente de dados oriundos do chão de fábrica, não é recente, mais ainda representa uma questão a ser resolvida em boa parte das organizações. O desenvolvimento de sistemas industriais por diferentes fornecedores utilizando diferentes ferramentas em diferentes eras tecnológicas, somados às transações corporativas como aquisições de novas unidades e fusões, acabam por gerar um grande número de repositórios heterogêneos e distribuídos, que são de difícil gerenciamento e exploração. Degan<sup>(1)</sup> destaca tais problemas e salienta que é interessante para as empresas integrar os dados que estão persistidos em bases distribuídas e heterogêneas, visto que a integração possibilita redução de custos no desenvolvimento de novas aplicações, facilita a exploração e a distribuição da informação no ambiente interno, além de permitir a integração com parceiros, fornecedores e clientes. Uma base de dados que reúna informações das diversas etapas do processo industrial facilita a criação de ferramentas de estudos estatísticos, de otimização e da qualidade, viabilizando assim a melhoria dos processos e redução de custos de produção.

A integração dos dados é vantajosa para as empresas. Porém, não é uma tarefa trivial, principalmente nos ambientes industriais onde é comum encontrar repositórios de dados pertencentes a diversos equipamentos que estão localizados em diferentes locais físicos da planta. Mesmo uma empresa nova como a VSB encontra este tipo de problema, visto que os equipamentos e sistemas que compõem a planta foram desenvolvidos por fornecedores diferentes de diferentes países que adotaram diferentes conceitos, ferramentas e técnicas em suas soluções.

A tarefa de integração de dados em ambientes industriais possui alguns problemas típicos relacionados à extração, ao processamento e à gravação dos dados no concentrador destino. Tais problemas podem ser referentes à infra-estrutura de rede, aos hardwares utilizados ou mesmo aos próprios dados. A

**Tabela 1** exibe os principais problemas referentes à integração de dados.

**Tabela 1.** Problemas relacionados à integração de dados<sup>(2)</sup>

<b>Problemas encontrados na integração de dados</b>	
Problemas de acesso aos dados.	Topologia da rede e regras e segurança da informação que por vezes impedem ou restringem consideravelmente o acesso aos dados.
Problemas de concorrência da leitura de dados com os sistemas de tempo real.	As bases origens podem pertencer a um sistema de tempo real cujo desempenho pode ser afetado devido à concorrência em sua base de dados.
Necessidade de substituição de dados.	Necessidade de substituição de alguns caracteres desconhecidos ou tradução de palavras.
A padronização de abreviações e siglas.	Siglas e palavras abreviadas devem ser tratadas da mesma forma no concentrador.
Conflitos de representação de dados.	Atributos com o mesmo significado sejam de tipos diferentes de dados.
	Correção do formato e número de atributos de uma mesma entidade.
	Por exemplo, número da ordem de produção representado como string em um repositório, e como inteiro em outro.
Conflitos de semântica (nomes)	Correção dos conflitos semânticos de modelagem. Os conflitos semânticos são todos aqueles que envolvem o significado do dado. Por exemplo, mesmo nome utilizado para diferentes entidades ou diferentes nomes para a mesma entidade.
Conflitos de precisão	Diferenças de precisão dos algarismos significativos. Ocorre quando a precisão numérica adotada varia de um ambiente para o outro (por exemplo: o resultado de uma análise química é armazenada com duas casas decimais "0,25" ou com quatro casas decimais "0,2536")
Diferenças da codificação ou uso de expressões.	Quando o código utilizado difere um do outro (por exemplo: tipo de material representado por A ou B e por 0 ou 1)
Conflitos de valores padrão (Default)	Pode ocorrer de dois dados semanticamente iguais (mesmo significado) possuírem valores default diferentes em repositórios diferentes. Por exemplo, valor previsto de bitola (400 em um banco e 500 em outro...)
Diferenças de abstração	A forma de modelar uma mesma informação segue padrões distintos (por exemplo: número do pedido armazenado em um único atributo, ou subdividido em pedido e item-pedido).
Sincronia da execução da carga.	Alguns dados para serem carregados para o concentrador necessitam que outros dados já tenham sido carregados anteriormente.
Formação de novas entidades.	Criar destinos a partir dos atributos de diferentes entidades origem.

## 1.1 O Ambiente dos Sistemas Industriais da VSB

Os sistemas e demais elementos que compõem o ambiente de automação industrial são segmentados em divisões conceituais chamadas Níveis. Cada Nível é criado em função do conjunto de responsabilidades que possui no processo produtivo. Os níveis se iniciam no chão de fábrica e vão até o ambiente corporativo da empresa. Apesar da classificação padrão proposta pela ISA (Instrumentation Society of America), cada empresa pode adotar uma definição própria. A seguir apresenta-se a divisão adotada na VSB.

No Nível 0 encontra-se os elementos que possuem interface direta com as grandezas físicas do chão de fábrica. Possuem funções de aquisição e atuação no processo, grosso modo, trata-se da Instrumentação. São considerados como sendo Nível 0 os equipamentos de atuação e sensores, destacando-se os elementos responsáveis por acionamentos e instrumentos de campo como sensores, atuadores, transdutores, válvulas, transmissores de temperatura, medidores laser, encoders, entre outros. A principal responsabilidade dos elementos do Nível 0 está associada à aquisição de dados para o Nível de Controle (Nível 1) e a atuação direta no chão de fábrica.

No Nível 1 encontramos os elementos que possuem interface com o Nível 0 e são responsáveis por funções de controle e supervisão do chão de fábrica. São

considerados como Nível 1 os CLPs, as IHMs e os sistemas SCADA (Supervisory Control and Data Acquisition). As principais funções dos elementos do Nível 1 são: aquisição e tratamento das informações disponibilizadas pelos elementos de Nível 0, atuação no processo associado às lógicas de processo, Controle PID, Supervisão das variáveis de processo para apresentação mais amigável aos operadores nas telas gráficas de IHMs ou SCADAs, armazenamento de tendências de variáveis disponibilizadas pelos CLPs nos SCADAs, alarmes relacionados aos elementos do Nível 0, interface entre o chão de fábrica e a operação assim como a interface com os sistemas de Nível 2.

No Nível 2, foco deste estudo, encontra-se os elementos responsáveis pela monitoria e rastreamento, em tempo de execução, da produção. Representam os sistemas que fazem o acompanhamento e a orquestração dos processos produtivos em tempo real. Nas linhas de Acabamento de barras, Laminação de tubos de aço sem costura, Tratamento térmico, Rosqueamento, Chanfragem e Fábrica de luvas, o Nível 2 é responsável pelo rastreamento da produção, peça a peça, em tempo de execução. Tal rastreamento online e peça a peça é realizado pelos sistemas denominados QTS, Quality and Traceability System, em cada uma das linhas. Salis, et al.<sup>(3)</sup> descrevem em detalhes o sistema QTS. Na VSB considera-se como sendo elementos do Nível 2 os sistemas de rastreamento QTS, o sistemas LIMS, o PIMS (Plant Information Management System) e o WMS, Warehouse Management System, que representa o sistema de gerenciamento do estoque intermediário de tubos. As principais responsabilidades dos elementos de Nível 2 são: O rastreamento em tempo de execução do processo produtivo, a regulagem (parametrização) do Nível 1 de acordo com o produto a ser fabricado, aquisição das informações dos equipamentos do chão de fábrica, a melhoria no desempenho do processo, geração automática de paradas operacionais e a contabilização da utilização das ferramentas correntemente em uso. Também faz parte do Nível 2 da VSB um repositório de dados de longa duração. Tal base de dados de histórico busca armazenar dados do Nível 2 das diversas áreas da planta. Outra responsabilidade do ambiente de Nível 2 é a interface com os sistemas de Nível 1 (supervisão e controle) e também com o Nível 3.

Na VSB os elementos de Nível 3 consistem nos diversos módulos do sistema MES, que faz o rastreamento da produção entre as diversas linhas produtivas. Algumas das responsabilidades dos elementos deste Nível são: a consolidação das informações geradas pelos sistemas de Nível 2, a interface com os sistemas de Nível 2 e Nível 4, a apresentação de relatórios gerenciais que envolvam grandes massas de dados e a sumarização de informações enviadas para o Nível 4.

No Nível 4 encontramos os elementos responsáveis pelo gerenciamento de todo o negócio da VSB de forma corporativa e integrada. São considerados como sendo Nível 4 o sistema SAP R/3 e os sistemas de planejamento de produção. As principais funcionalidades são o controle de vendas, controle financeiro, o planejamento de produção e controle de custos.

## **1.2 O Ambiente dos sistemas de Nível 2 na VSB**

O foco deste trabalho é o processo de integração dos dados no ambiente de automação da VSB, especificamente o Nível 2. Cada linha de produção, desde o pátio de matéria primas, passando pela redução, aciaria, laminação, têmpera, linhas acabadoras (rosqueamento e chanfragem) até o produto final, possui os seus próprios sistemas de Automação/Nível 2. Cada um destes sistemas de Nível 2

possuem ricas e importantes massas de dados acerca da execução dos respectivos processos. É consideravelmente interessante integrar os dados dos sistemas de Nível 2 das várias plantas em um único servidor, buscando integrá-los, correlacioná-los e mantê-los armazenados por longos períodos. A Figura 1 mostra o contexto dos sistemas de Nível 2 na VSB.

Os sistemas de Nível 2 de cada área possui sua própria base de dados local aonde armazenam os dados que foram coletados durante a execução dos processos. Como os sistemas de Nível 2 são sistemas de tempo real e de missão crítica, suas bases de dados locais devem estar sempre enxutas para não comprometerem o desempenho das aplicações. Principalmente no que diz respeito à concorrência de leitura efetuada por outros sistemas. Buscando deixar os bancos locais sempre enxutos, os dados mais antigos, que pertencem aos materiais já produzidos, são removidos dos bancos locais e levados para o Servidor de Histórico do Nível 2, aonde serão integrados com os dados oriundos das bases locais de outras áreas. Os sistemas de Nível 2 também acessam alguns bancos de dados locais pertencentes à sistemas legados que foram fornecidos juntamente com os diversos equipamentos que compõe as linhas de produção. Trata-se principalmente das bases de dados dos sistemas de supervisão e operação dos equipamentos adquiridos pelas áreas de produção. Alguns dados existentes nestas bases de dados também podem ser carregados para o Servidor de Histórico do Nível 2.

**Figura 1** - O ambiente dos sistemas de Nível 2 na VSB.

O Servidor de Histórico do Nível 2 ou Servidor de Fronteira e Dados consiste em um grande repositório de dados de histórico no qual os dados coletados durante a execução dos processos produtivos e também os dados referentes à qualidade do produto serão armazenados por um longo período. A Arquitetura da Rede de

Automação da VSB permite que todos os sistemas de Nível 2 das diversas áreas possam enxergar o Servidor de Banco de Dados de Histórico do Nível 2.

### 1.3 O Servidor de Histórico e Integração de Dados do Nível 2

O ambiente computacional do Nível 2 na VSB, possui a figura de um servidor de dados históricos. Este servidor busca armazenar dados por longos períodos (5 a 7 anos), obedecendo à questões legais, contratuais e requisições de clientes. O Servidor de Dados Históricos está localizado na DMZ da rede Nível 2, podendo ser acessado da rede corporativa (Nível 3, ambiente de TI). Conceitualmente este servidor está localizado entre os sistemas de Nível 2 das linhas de produção e os sistemas de Nível 3 no ambiente de TI. As principais razões para adoção do Servidor de Histórico no ambiente de Nível 2 da VSB são:

- *Integração de dados:* Integração, agrupamento e arquivamento das informações produzidas nas diversas linhas do processo produtivo da usina em um mesmo servidor. A integração dos dados de diversas linhas em um único repositório facilitará a manipulação, a exploração e a distribuição dos dados, beneficiando principalmente as áreas de qualidade e de produção da empresa.
- *Garantia de desempenho nos sistemas de produção:* A inclusão deste servidor desonera os bancos de dados de produção. Desta forma, o requisito de armazenamento por vários anos, não é atendido diretamente pelos sistemas ligados à produção e sim por este servidor de histórico. Um servidor que busca armazenar massas de dados por longos períodos e que recebe grandes massas de dados diariamente, possui uma dinâmica de manutenção, principalmente do próprio banco de dados, muito diferente dos sistemas ligados diretamente à produção, visto que estes últimos devem ter alta disponibilidade e responderem à solicitações rapidamente.
- *Separação de sistemas:* Os sistemas de Nível 3, no ambiente de TI, não devem fazer leituras ou acessos custosos diretamente nos sistemas ligados à produção. As redes estão completamente segmentadas, física e logicamente, aumentando a segurança e estabilidade dos sistemas de produção. Havendo necessidade de consultar os dados no Nível 2, os sistemas de nível 3 fazem consultas no Servidor de Histórico, que está na DMZ, e não diretamente nos sistemas de Nível 2 ligados diretamente à produção;

### 1.4 Objetivos

Os principais objetivos deste trabalho são:

- descrever o estudo realizado acerca das formas de integração dos dados locais, pertencentes aos sistemas que compõe o Nível 2 de cada área, no Servidor de Fronteira e Dados;
- descrever as principais técnicas e ferramentas adotadas no processo de integração de dados no ambiente industrial da VSB.

## 2 MATERIAL E MÉTODOS

Busca-se agora comentar as principais técnicas, ferramentas e métodos utilizados no trabalho de integração de dados provenientes do ambiente industrial da VSB. Inicialmente descreve-se o levantamento realizado acerca das principais formas de integração de dados praticadas atualmente, destacando-se a forma selecionada. Em seguida, apresenta-se o desenvolvimento das soluções que propiciaram a integração, a manipulação e exploração dos dados.

## **2.1 Comparação entre as Principais Formas de Integração de Dados**

Avaliou-se os principais métodos de integração de dados a fim de elencar qual possui a melhor relação custo-benefício. Os principais métodos de integração avaliados foram:

### **2.1.1 Método 1 - Conexão repositório a repositório**

O primeiro método avaliado se refere a utilização de recursos dos próprios Sistemas Gerenciadores de Bancos de Dados (SGBD) para efetuar a transferência de dados entre as várias bases origens e o servidor destino. Geralmente utiliza-se entidades executáveis criadas no banco de dados, como as stored procedures, funções, tasks e triggers, que manipulam os dados por meio de comandos SQL (Structured Query Language). Nem todos os sistemas gerenciadores de banco de dados utilizados na planta possuem tais recursos. A utilização da comunicação banco a banco requer ainda um grande esforço de programação, um elevado nível de conhecimento sobre a linguagem SQL e conhecimento sobre as estruturas de indexação dos bancos de dados para não causar perdas de desempenho. Tais perdas de desempenho devem ser fortemente evitadas, pois, se ocorrerem nas bases de origem, podem prejudicar os sistemas de Nível 2 que estão fortemente ligados à produção, efetuando a monitoria dos processos. Em contrapartida, a utilização deste método possibilitará um maior controle de como cada dado está sendo tratado. Este método propicia uma maior flexibilidade na manipulação dos dados. Já o tratamento de erros é mais trabalhoso e a geração de Logs detalhados também pode não ser trivial. O custo de desenvolvimento e manutenção da solução é de baixo a médio.

### **2.1.2 Método 2 - Aplicação lendo dados de um repositório e gravando em outro repositório de dados**

Nesta opção utiliza-se um aplicativo executável compilado que seria responsável por ler os dados diretamente dos repositórios origem e os concentraria no Servidor de Fronteira. A aplicação deveria ler dados das diversas fontes, portanto, a linguagem de programação utilizada deveria dar suporte ao acesso à muitos SGBDs e outros repositórios, possibilitando ler dados desde arquivos texto até banco de dados mais complexos. Adotando esta opção pode-se detalhar e formatar os arquivos de Log de forma personalizada. Por outro lado, a utilização de um aplicativo executável necessita da recompilação do programa a cada alteração realizada. O custo de desenvolvimento e manutenção da solução é médio. O desenvolvimento não é complexo, porém é trabalhoso, visto que, para se obter maior flexibilidade, deveria haver uma rotina específica para cada elemento a ser copiado.

### **2.1.3 Método 3 – Ferramenta ETL para fazer o gerenciamento da integração dos dados**

Existe ainda a opção de usar uma ferramenta pronta de mercado para efetuar a leitura dos dados nos bancos origem e gravá-los no Servidor de Dados Históricos. A proposta é utilizar uma aplicação ETL (Extraction, Transformation, and Loading) para extrair os dados dos repositórios de origem, e gravá-los de forma centralizada no Servidor de Dados Históricos. A sigla ETL vem do termo em inglês “Extract, Transformation, Load (Extração, Transformação, Carga)”. As ferramentas ETL são aplicativos desenvolvidos para extrair dados de diferentes repositórios, transformar (manipular, converter) esses dados conforme regras de negócios pré-estabelecidas e por fim efetuar a carga (persistência) dos dados de forma integrada em um repositório destino (usualmente um datawarehouse, concentrador de dados ou servidor de dados históricos). As ferramentas ETL encontradas no mercado possibilitam integrar dados oriundos de diferentes repositórios, desde arquivos texto e planilhas até os bancos de dados mais estruturados conhecidos no mercado. A utilização de uma ferramenta ETL para efetuar a integração dos dados não necessita de conhecimentos aprofundados sobre banco de dados e linguagem SQL. O tempo de desenvolvimento é menor do que escrever código SQL, pois as ferramentas ETL fornecem interfaces gráficas amigáveis que auxiliam a programação e possuem a capacidade de acessar repositórios de diferentes plataformas espalhados pelas redes. Tais características ajudam a reduzir consideravelmente o esforço de trabalho necessário para o desenvolvimento e gerenciamento das rotinas de migração dos dados. A grande desvantagem desta opção se deve ao custo das soluções mais conhecidas e já consolidadas no mercado que ainda são caras comparando com as outras soluções.

## 2.2 Comparação e Escolha do Método mais Indicado para o Ambiente Industrial da VSB

Elencou-se os quesitos de maior importância ao tratar de ferramentas de desenvolvimento a fim de comparar as três propostas comentadas anteriormente e escolher a mais atrativa entre elas. A Tabela 2 mostra os quesitos e a avaliação das três opções de solução levantadas.

**Tabela 2.** Avaliação das opções de ferramentas para integração dos dados

Item Avaliado	Conexão repositório a repositório	Programa executável compilado	Ferramenta ETL
Produtividade	Média	Baixa	Alta
Nível de Conhecimento Necessário	Alto	Alto	Médio
Portabilidade	Média	Média	Alta
Dificuldade Manutenção	Média	Alta	Média Baixo
Controle da execução	Médio	Alto	Baixo
Custo	Baixo	Baixo	Alto

Avaliando os benefícios e desvantagem dos métodos propostos, optou-se por uma ferramenta ETL para efetuar a concentração dos dados distribuídos ao longo das linhas de produção no Servidor de Dados Históricos. As aplicações ETL mais conhecidas no mercado possuem interfaces amigáveis que permitem aos usuários construir, gerenciar e monitorar graficamente a execução de cada caso de exportação, transformação e carga dos dados. Por exemplo, as ferramentas permitem os usuários escolherem o campo, tabela, banco de dados e o servidor de

origem e conectá-los ao campo, tabela, banco de dados e o servidor de destino utilizando somente o mouse, sem digitar nenhuma linha de código. A

Tabela 3 exibe a lista dos principais fornecedores e ferramentas ETL de integração de dados.

**Tabela 3.** Lista de fornecedores de ferramentas ETL<sup>(4)</sup>

Ferramenta ETL	Versão corrente	Fornecedor
Oracle Warehouse Builder (OWB)	11gR1	Oracle
Data Services	XI 4.0	SAP Business Objects
IBM Information Server (Datastage)	8.1	IBM
SAS Data Integration Studio	4.4	SAS Institute
PowerCenter Informatica	9.5	Informatica
Elixir Repertoire	7.2.2	Elixir
Data Migrator	7.7	Information Builders
SQL Server Integration Services	10	Microsoft
Talend Studio for Data Integration	5.2	Talend
DataFlow Manager	6.5	Pitney Bowes Business Insight
Pervasive Data Integrator	10.0	Pervasive Software
Open Text Integration Center	7.1	Open Text
Oracle Data Integrator (ODI)	11.1.1.5	Oracle
Data Manager/Decision Stream	8.2	IBM (Cognos)
Clover ETL	3.1.2	Javlin
Centerprise	5.0	Astera
DB2 Infosphere Warehouse Edition	9.1	IBM
Pentaho Data Integration	4.1	Pentaho
Adeptia Integration Suite	5.1	Adeptia
DExpress	5.5	Syncsort
Expressor Data Integration	3.7	QlikTech

O relatório etltool.com<sup>(5)</sup> faz um comparativo entre as principais ferramentas ETL disponíveis no mercado. Os autores fazem a comparação avaliando cerca de noventa critérios técnicos e comerciais. Trata-se de um interessante guia para auxiliar a escolha de uma ferramenta ETL que atenderá uma determinada necessidade. Vale destacar ainda que também existem ferramentas ETL Open Source como o Talend Open Studio, CloverETL e Pentaho.

A ferramenta ETL adotada na VSB foi a ODI (Oracle Data Integrator) devido à negociação e contratos já existentes com a Oracle e ao retorno positivo de outros usuários da ferramenta. Adotou-se a ferramenta ETL ODI para extrair as informações das bases locais dos sistemas de Nível 2 e carregá-los para o banco de dados de histórico. Deste modo, os dados provenientes das diversas linhas de produção da usina ficaram concentrados em um único servidor, facilitando assim o gerenciamento e a exploração.

### 2.3 O Portal de Processos

O Portal de Processos VS Factory é um conjunto de aplicações web que fazer a integração e a exploração dos dados provenientes do ambiente industrial da

VSB. As aplicações que compõem o Portal de Processos podem obter e exibir dados das principais fontes existentes no ambiente de automação da VSB (região da TA na Figura 1), tais como, PIMS, LIMS e bancos de dados, principalmente o banco de dados de histórico e integração de dados do Servidor de Histórico do Nível 2. A

Figura 2 mostra a arquitetura do Portal de Processos com seus principais módulos.

**Figura 2** - Arquitetura do Portal de Processos para integração de dados do ambiente de TA.

## **2.4 O Módulo de Relatórios**

O módulo de relatórios do Portal de Processos fica hospedado na plataforma de aplicações web da Microsoft denominada Sharepoint. A plataforma Sharepoint é geralmente adotada no desenvolvimento de portais colaborativos, intranets, sites de gestão de documentos e sites de gestão de conteúdos. Trata-se de uma ferramenta de desenvolvimento de soluções web que buscam atender à demandas relacionadas à gestão do conhecimento e à necessidades empresarias referentes á gestão de intranets. O Módulo de relatórios foi desenvolvido adotando-se o SSRS (SQL Server Report Service) no Visual Studio 2010. Elaborou-se uma grande diversidade de relatórios com informações gerenciais, estatísticas e de acompanhamento do processo. Os relatórios exibem ainda um considerável número de informações sobre os resultados de testes de qualidade dos produtos. A ferramenta de desenvolvimento permite ainda que os relatórios sejam exportados para alguns formatos de arquivos, por exemplo, pdf, xls, csv, tiff e xml. Trata-se de uma funcionalidade interessante, visto que os dados do relatório podem ser utilizados sem esforço em outras ferramentas como o Minitab, o Excel e o Matlab. As Figura 3 e 4 mostram exemplos de relatórios.

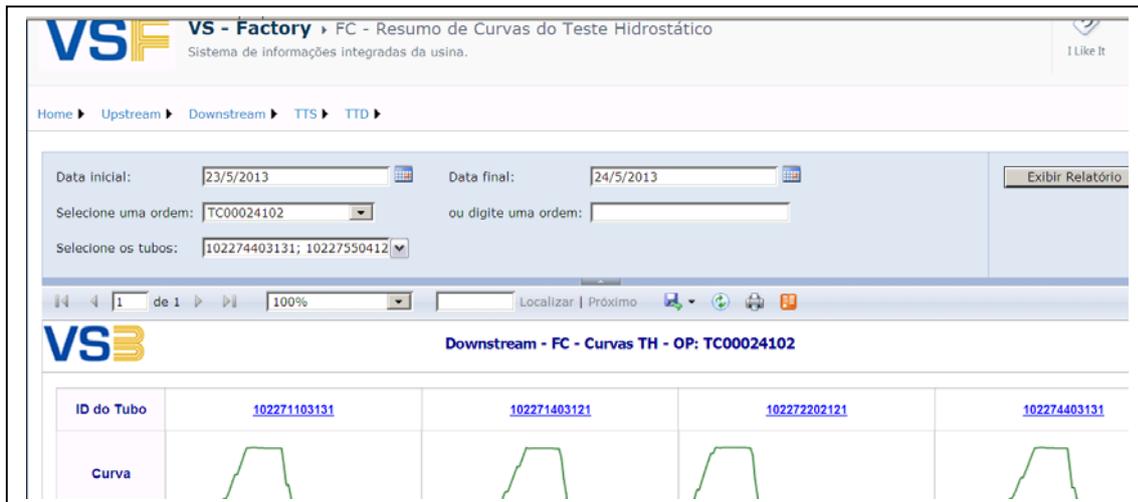


Figura 3 - Exemplo de relatório mostrando resultados do teste hidrostático para cada tubo.

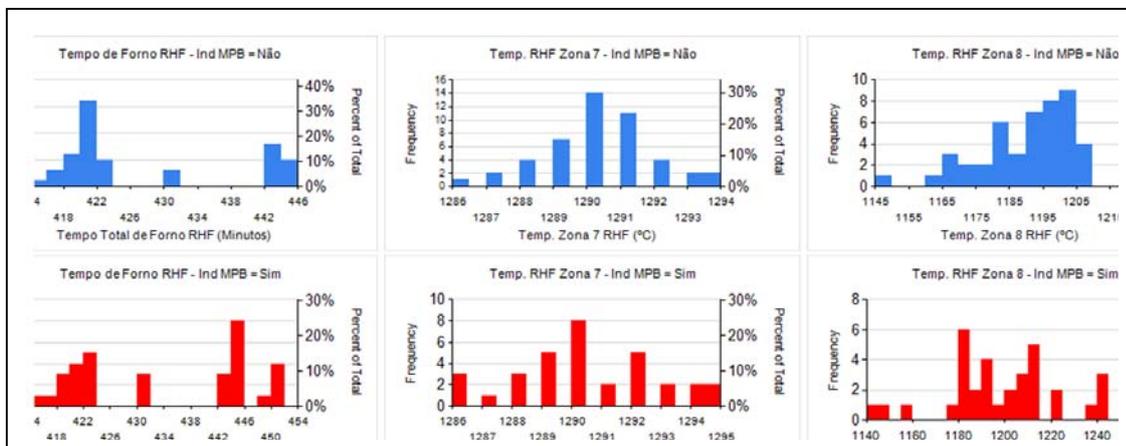


Figura 4 - Parte de um relatório que correlaciona as distribuições (histogramas) de variáveis de processo com a indicação de defeitos.

Os relatórios desenvolvidos permitem ainda a utilização de links para outros relatórios, a utilização de filtros e a utilização de funções de estatística e agregação, tais como desvio padrão, contagem, valores médios máximos, mínimos...

## 2.5 Aplicações que Compõe o Portal de Processos VS Factory

O Portal de Processos VS Factory é composto por diversas aplicações desenvolvidas em Microsoft Silverlight 5 por meio do Visual Studio 2010. Estas aplicações foram desenvolvidas buscando prover interfaces gráficas interativas e de fácil operação. O principal objetivo do projeto gráfico das aplicações foi a usabilidade e a riqueza gráfica das interfaces que devem facilitar a navegação e a exploração dos dados. Entendeu-se que, assim como o correto funcionamento das aplicações, as suas características gráficas também são fatores cruciais para o sucesso do projeto. Por isso, os detalhes da usabilidade e recursos gráficos foram tratados como fatores principais, tanto quanto a correta manipulação e exibição dos dados. Além dos componentes padrões disponibilizados pelo Silverlight 5, descritos em Snyder et al.,<sup>(6)</sup> desenvolveu-se novos componentes para exibição e manipulação dos dados, destaca-se aqui gráficos, componentes para seleção de datas-hora e seleção de alguns parâmetros de busca. As aplicações do portal de Processos VS

Factory também possui ferramentas para buscar dados do sistema PIMS (PI OSIssoft) e de SGBDs como o Oracle e o Microsoft SQL Server. As aplicações Silverlight podem ainda buscar informações do Sharepoint, o que permite adquirir informações relevantes acerca do controle de acesso de usuários, listas de conteúdos, links para outros documentos e links para determinados conteúdos. O acesso de aplicações Silverlight ao Sharepoint é provido por uma API denominada Client Object Model suportada pelas seguintes dlls:

- Microsoft.SharePoint.ClientSilverlight
- Microsoft.SharePoint.Client.Silverlight.Runtime

A comunicação com o PIMS (PI OSIssoft) é realizado de duas formas diferentes. Uma delas é por meio do PI Coresight (<http://picoresight.osisoft.com/>) e a outra é por meio do PI OLE DB fornecido com o pacote de produtos PI Data Access.<sup>(7)</sup> A comunicação com outros SGBDs (Sistemas Gerenciadores de Banco de Dados) é estabelecida por meio de OLE DB que é uma API desenvolvida pela Microsoft que permite acessar diferentes repositórios de dados. Os relatórios do módulo de relatórios são acessados diretamente por meio de links HTTP (url dos arquivos). A comunicação com os sistemas de Nível 2, principalmente os sistemas de rastreamento da produção, é estabelecida por meio de web services específicos. Devido aos recursos de comunicação disponibilizados para o Silverlight, é possível apresentar dados provenientes de várias fontes de dados no ambiente de TA da VSB. Estando os dados disponíveis para as aplicações do Portal de Processos, pode-se manipulá-los, correlacioná-los e apresentá-los de diversas formas, ainda que sejam provenientes de repositórios distintos. Os sistemas de rastreamento (Nível 2) de cada linha de produção armazenam as informações coletadas ao longo do processo vinculadas ao número de identificação de cada elemento rastreado. O número de identificação de um determinado elemento (pedido, ordem de produção, peça, ferramenta, amostra...) é o mesmo em todas as etapas de fabricação. As consultas pode ser feita em função de tais elementos. A mostra uma das aplicações do Portal de Processos.

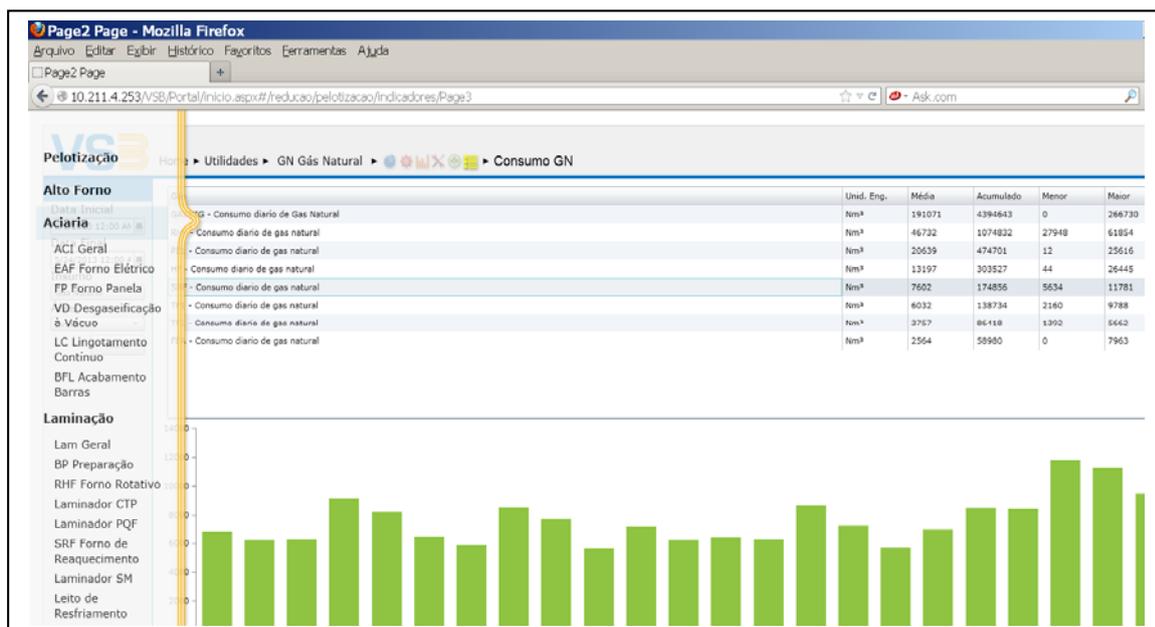


Figura 5 – Uma secção da interface do portal de processos VS Factory da VSB.

### 3 RESULTADOS

Existem cerca de 700 rotinas ETL de carga de dados das bases de dados dos sistemas de produção para o banco de histórico. Desde o início dos trabalhos referentes ao desenvolvimento do Portal de Processos VS Factory desenvolveu-se cerca de 290 visões de relatórios para os diversos setores produtivos da usina. Destas visões, 178 estão sendo usadas freqüentemente. O tempo de resposta de determinadas consultas aumentou acima do previsto, principalmente as consultas que envolvem campos do tipo blob. A adesão à utilização do Portal foi maior do que o esperado, o que levou à necessidade de expandir a memória RAM e os HDs na stored do servidor. A quantidade de usuários atual é 340 dos quais uma média de 97 acessam o sistema diariamente.

### 4 CONCLUSÃO

A integração dos dados por meio de uma ferramenta ETL agiliza o processo de preparação e gerenciamento das rotinas de carga dos repositórios origem para o concentrador. É interessante adotar alguma forma rotineira de conferência dos status das cargas de dados. Apesar das cargas de dados não influenciarem diretamente na produção, elas são politicamente importante, visto que falhas podem causar certo desconforto para os usuários que trabalham com o sistema de relatórios, principalmente no que tange às informações gerenciais. A persistência dos dados provenientes dos processos das diversas áreas em um mesmo servidor se mostrou consideravelmente vantajosa, tanto em termos de administração do sistema quanto em questões relacionadas á exploração e compartilhamento dos dados. Uma interface gráfica interessante com boa usabilidade é um fator motivador para os usuários utilizarem o sistema e explorarem seu potencial. Os próximos passos incluem a modelagem das estruturas para favorecerem o desenvolvimento de aplicações de BI (Bussiness Inteligence), Data Mining e KDD (knowledge-Discovery in Databases).

### REFERÊNCIAS

- 1 Degan, J. O. (2005). *Integração de dados corporativos:uma proposta de arquitetura baseada em serviços de dados*. Campinas, SP.: Universidade Estadual de Campinas.
- 2 etltool.com. (2013). *ETL tool Survey Report:ETL tools comparison* . <http://www.etltool.com/etl-tools-comparison/>.
- 3 Salis, T. T., Almeida, A. R., Menezes, M. V., Pereira, T. d., Alves, C. A., Lopes, V. d., et al. (2013). Um caso de sucesso no rastreamento online peça a peça nas linhas de laminação, ajustagem, tratamento térmico e linhas acabadoras da VSB. *17º Seminário de Automação e TI Industrial. 24 a 27 de setembro de 2013*. Vitória – ES – Brasil: ABM.
- 4 Lista de Ferramentas ETL (2013). <http://www.etltool.com/list-of-etl-tools/>. Acesso em 05 de 04 de 2013.
- 5 Passos, R. A. (2006). *Uma Arquitetura para Integração de Dados Baseada em Ontologia*. Recife, PE.: Universidade Federal de Pernambuco.
- 6 Snyder, T., Smith, J., Eden, J., & Duffield, M. (2012). *Microsoft Silverlight 5: Building Rich Enterprise Dashboards*. Birmingham, UK.: Packt Publishing Ltd.
- 7 OSisoft. (02 de 05 de 2013). *PI Data Access*. Acesso em 02 de 05 de 2013, disponível em PI OSisoft: [http://www.osisoft.com/software-support/products/PI\\_Data\\_Access.aspx](http://www.osisoft.com/software-support/products/PI_Data_Access.aspx)