

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

ESCOLA POLITÉCNICA

DEPARTAMENTO DE ELETRÔNICA E DE
COMPUTAÇÃO

**SISTEMA GLANCE: RECUPERAÇÃO E
PROCESSAMENTO DE GRANDES VOLUMES DE
DADOS**

Autor:

Felipe Fink Graef

Orientador:

Carmen Lúcia Lodi Maidantchik, D.Sc.

Orientador:

José Manoel de Seixas, D.Sc.

Examinador:

Érica Ribeiro Polycarpo Macedo, D.Sc.

Examinador:

Márcio Portes de Albuquerque, D.Sc.

Examinador:

Antônio Cláudio Gómez, M.Sc.

DEL

Dezembro de 2009

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

Escola Politécnica - Departamento de Eletrônica e de Computação

Centro de Tecnologia, bloco H, sala H-217, Cidade Universitária

Rio de Janeiro - RJ CEP 21949-900

Este exemplar é de propriedade da Universidade Federal do Rio de Janeiro, que poderá incluí-lo em base de dados, armazenar em computador, microfilmear ou adotar qualquer forma de arquivamento.

É permitida a menção, reprodução parcial ou integral e a transmissão entre bibliotecas deste trabalho, sem modificação de seu texto, em qualquer meio que esteja ou venha a ser fixado, para pesquisa acadêmica, comentários e citações, desde que sem finalidade comercial e que seja feita a referência bibliográfica completa.

Os conceitos expressos neste trabalho são de responsabilidade do(s) autor(es) e do(s) orientador(es).

Agradecimento

Agradeço, em primeiro lugar, à minha família. Aos meus pais, Christiane e Oscar, pelo carinho e apoio incondicional e sempre presente. À minha avó Walma e minha tia Luciane. Seus ensinamentos, incentivos e companhia foram fundamentais para mim durante essa caminhada.

À Carmen Maidantchik, por toda orientação, paciência, oportunidades e valiosos ensinamentos que levarei por toda a vida. Além de excelente orientadora, é uma grande amiga.

Ao professor José Seixas, pelo apoio, incentivo e orientação ao longo de todo esse tempo.

À Kathy Pommès e Marzio Nessi, pela oportunidade e acolhida no CERN. Pelo tempo de trabalho extremamente agradável, que me proporcionou um grande crescimento profissional e pessoal.

Aos amigos, e também colegas de trabalho, Kaio Karam, Fernando Ferreira, Andressa Sivoella, Amanda Alves e Rodrigo Torres. Obrigado pelos incentivos, críticas, sugestões e conhecimentos que enriqueceram tanto o trabalho quanto a mim. A todos os amigos do CERN, do Laboratório de Processamento de Sinais, e da UFRJ. Obrigado por tornarem prazeroso o dia-a-dia da universidade, que de outra forma seria estressante. Agradeço muito a todos os amigos, dentro e fora da universidade, pelos excelentes momentos e pelo constante incentivo, pois me motivaram a continuar a trilhar esse caminho.

Agradeço aos meus professores, excelentes professores que tive a vida toda, pelas lições tanto acadêmicas quanto pessoais, sem as quais não teria chegado onde cheguei.

Muito obrigado!

Resumo

O presente trabalho desenvolve o Glance, um sistema *Web* de recuperação e operação de grandes massas de dados armazenados em bancos de dados dispersos. A motivação para o sistema veio das necessidades da colaboração internacional que construiu, e atualmente opera, o detector de partículas ATLAS, no CERN (*European Organization for Nuclear Research*).

O ATLAS foi construído por um dos maiores esforços colaborativos do meio científico, envolvendo 172 institutos de 37 países e é composto por centenas de milhares de componentes. Nessa colaboração, cada equipamento foi construído por grupos de colaboradores em seus respectivos institutos de origem e, em seguida, levado ao CERN para ser montado e testado em sua posição definitiva. Os dados que são gerados sobre os equipamentos são armazenados em repositórios gerenciados pelos próprios colaboradores que constroem os equipamentos, estando, portanto, geograficamente dispersos e não usam uma mesma modelagem, tecnologia ou não necessariamente uma mesma terminologia.

O Glance é genérico pelo fato de não depender de uma modelagem ou tecnologia específica. Para isso o sistema usa o conceito de interfaces de recuperação, que são descritas em uma linguagem independente da tecnologia de recuperação. As interfaces de buscas apresentadas pelo Glance são paramétricas, ou seja, o usuário pode especificar critérios de busca envolvendo os atributos da interface e um valor fornecido pelo usuário. Após recuperados, o sistema pode processar os dados, tais como a extração de estatísticas ou geração de gráficos, a fim de exibi-lo em um formato que seja útil aos colaboradores.

O sistema está instalado nos servidores do CERN e em uso pela colaboração. Dentre as aplicações do sistema, podem-se destacar a recuperação, pela Coordenação Técnica do ATLAS, do estado de instalação dos equipamentos e conectividade deles com os cabos, e a monitoração dos sensores de nível, pelo grupo que controla o alinhamento entre os componentes do ATLAS. Além disso, o sistema é utilizado como fonte de dados para o sistema que monitora temperaturas e voltagens das fontes de alimentação de um dos sub-detectores.

Palavras-Chave: banco de dados, web, colaboração internacional, CERN, ATLAS.

Abstract

This document describes the development of Glance, a Web system for retrieval and processing of big amounts of data stored in different databases. The motivation for this system has come from the necessities of the international collaboration that built, and currently operates, the ATLAS detector, at CERN (European Organization for Nuclear Research).

Built by one of the largest collaborative efforts in the scientific medium, ATLAS involves 172 institutes of 37 countries, and is made of hundreds of thousands of components. Each equipment was designed and built by a group of collaborators in their home institute, and then taken to CERN to be assembled in its final position, and then tested. The data generated for each equipment during this process is stored in repositories managed by the same group that built it. The information for the detector equipments, therefore, is geographically scattered and is not stored in a standard technology, modeling and might even use different terminologies.

Glance is a generic system for data retrieval, as it does not depends the modeling or technology of the repository. In order to achieve this, the system uses the concept of search interfaces, described in a technology independent language. The search interfaces handled by Glance are parametric, i.e., the user specifies search criteria by choosing an attribute, a comparison operator and a value. Once retrieved from the database, the data can be processed in order to be presented to the user in a useful format. Examples of processing are the generation of plots, and the calculation of means and standard deviation.

The Glance system is currently installed at CERN's Web servers and being used by the collaboration. Among the applications of the system are the retrieval of equipment installation status by the ATLAS Technical Coordination, and the monitoring of hydrostatic level sensors by the ATLAS Alignment Survey group. Glance is also used as a data source for the DCS Web System, which monitors temperature, voltage and electric current sensors of the Detector Control System from one of the sub-detectors.

Key-words: database, web, international collaboration, CERN, ATLAS.

Sumário

Resumo	iii
Abstract	iv
1 Introdução	1
1.1 Objetivo	3
1.2 Organização do documento	3
2 O CERN e o ATLAS	4
2.1 O CERN e o Colisionador de Partículas LHC	4
2.2 O Detector de Partículas ATLAS	6
2.3 Armazenamento de dados do ATLAS	6
2.3.1 Dados de instalação de equipamentos	7
2.3.2 Sistema de controle do detector	7
2.3.3 Alinhamento e Posicionamento do Detector	8
3 Especificação do sistema Glance	10
3.1 Análise dos problemas	10
3.2 Análise dos requisitos	11
3.3 Proposta do Sistema Glance	12
3.4 Projeto do Glance	13
3.5 Tecnologias	14
3.5.1 A linguagem de programação C++	15
3.5.2 A linguagem Python	15
3.5.3 Banco de dados Oracle e a OCCI	15
3.5.4 XML e XSLT	16

4	O Sistema Glance	17
4.1	Descrição de uma interface de busca	17
4.2	Criação de interfaces de busca	19
4.3	Interface de busca	22
4.4	Recuperação de dados	24
4.5	Operação sobre dados	25
4.5.1	Mecanismo para processamentos não parametrizados	25
4.5.2	Operações genéricas e parametrizadas	27
4.5.3	Operações específicas e parametrizadas	29
5	Aplicações do Glance	31
5.1	Dados de instalação de equipamentos	31
5.2	Sistema de Controle do Detector	32
5.3	ATLAS Survey	33
6	Conclusão	35
	Referências Bibliográficas	38
A	Publicações	40

Lista de Figuras

2.1	Túnel subterrâneo do LHC e seus quatro experimentos. No destaque da esquerda, uma foto do acelerador. Na direita, o detector ATLAS. Extraídas de [3]	5
3.1	Arquitetura do sistema Glance	13
4.1	Descrição XML de uma interface de busca	20
4.2	O Glance apresentando estrutura do banco de dados e gerando uma interface adequada	21
4.3	Personalização de uma interface de busca durante sua criação.	21
4.4	Página inicial do Glance	22
4.5	Lista de interfaces de busca existentes	23
4.6	Interface de busca gerada pelo Glance, destacando os operadores de busca sensíveis ao tipo do atributo.	23
4.7	Resultado de uma busca mostrado na forma de uma tabela	24
4.8	Realização de processamento sobre dados recuperados por um programa externo.	26
4.9	Diagrama conceitual de classes do mecanismo genérico para transformações não parametrizadas.	26
4.10	Processo de recuperação de dados realizando operações definidas pelo usuário.	28
4.11	Diagrama conceitual de classes do módulo que realiza operação sobre os dados.	28
4.12	Processo de recuperação de dados realizando operações definidas pelo usuário.	30
5.1	DCS Web System exibindo médias e RMS calculados pelo Glance	32

5.2	Interface para os dados do Survey. Na esquerda, a interface do Glance com os parâmetros para as operações parametrizadas. Na direita, o resultado na forma de gráfico.	34
-----	--	----

Capítulo 1

Introdução

Empresas e organizações em diversas áreas do mercado coletam dados sobre suas atividades e equipamentos, a fim de avaliar a qualidade de seus processos e auxiliar nas tomadas de decisões. Frequentemente, os dados coletados são armazenados em diferentes bancos de dados, seja porque os processos e equipamentos que geram esses dados ficam geograficamente dispersos, ou porque as atividades da empresa são muito complexas para serem totalmente contempladas em uma única modelagem de dados.

A existência de diferentes repositórios em uma mesma organização faz com que a recuperação das informações seja dificultada, pois frequentemente esses repositórios foram projetados por pessoas diferentes analisando necessidades diferentes. Além disso surgem novos desafios relacionados à construção das interfaces de recuperação de dados e sua posterior manutenção, pois nesses ambientes, as pessoas que necessitam das informações não são as mesmas que projetaram o repositório e podem até necessitar analisar os dados de diversas maneiras.

Todas essas características estão presente no contexto da colaboração internacional que construiu e atualmente opera o detector de partículas ATLAS, no CERN. O CERN (Centro Europeu para Pesquisas Nucleares) é o maior centro de pesquisas em física de partículas do mundo, situado na fronteira da Suíça com a França. Atualmente, o CERN está envolvido na construção e operação do LHC (*Large Hadron Collider*), um acelerador de prótons circular de 27km de comprimento. Em torno do acelerador, são instalados detectores de partículas que observam o subproduto das colisões dos prótons acelerados pelo LHC. O ATLAS é um dos detectores do

LHC, sendo construído e operado por uma colaboração internacional envolvendo 169 institutos de 37 países. O detector ATLAS é dividido em 4 sub-detectores e mais alguns sistemas necessários para a operação, tais como eletrônica de aquisição de dados, criogenia e magnetos. Durante a construção, cada um desses subsistemas foi dividido em partes menores e cada parte foi delegada a uma instituição da colaboração. As dezenas de milhares de equipamentos construídos foram levados ao CERN para serem instalados e testados.

Os equipamentos que compõem o detector passaram por algumas fases, tais como a de construção, a de instalação, a de comissionamento, onde os equipamentos foram testados em conjunto, e, atualmente, estão na fase de operação. Durante cada uma dessas fases, uma série de dados são gerados, tais como componentes usados no equipamento, resultados de testes de qualidade, informação de posição e conectividade dentro do detector e monitoração do funcionamento com sensores, tais como de posição, tensão elétrica ou temperatura. Os equipamentos construídos por um dado grupo de colaboradores, ao serem levados ao CERN, se integram com partes feitas por outras instituições. Por isso, os colaboradores precisam recuperar informações geradas por outros grupos que, frequentemente, estão armazenadas em servidores diferentes dos seus, usando outras tecnologias e técnicas de modelagem. Além disso, as pessoas podem ter diferentes interesses sobre o mesmo conjunto de dados. Tomando, como exemplo, o banco de dados que armazena as conexões entre cabos e equipamentos, observamos que os responsáveis por um dos sub-detectores estão interessados em saber quais cabos se ligam aos seus componentes, enquanto que os responsáveis pela instalação da infra-estrutura do detector se interessam pelo tipo de cabo e a rota que ele deve percorrer.

Alguns tipos de dados precisam ainda ser processados para fornecerem informações úteis. Os processamentos podem ser operações comuns, tais como cálculo de médias ou desvio padrão ou traçar um gráfico de uma série temporal, ou computações específicas preparadas por um especialista no equipamento. Ao avaliar o funcionamento de certas partes da eletrônica, por exemplo, é necessário extrair estatísticas e gerar gráficos a partir das amostragens de sensores de corrente e tensão elétrica que foram armazenadas em bancos de dados.

1.1 Objetivo

O objetivo deste trabalho é desenvolver um sistema genérico, capaz de recuperar informações de qualquer banco de dados da colaboração ATLAS e realizar os processamentos necessários para fornecer o resultado no formato adequado para o usuário. O sistema deve apresentar a estrutura do repositório e permitir que o usuário selecione o conjunto de dados de interesse. A partir disso, o sistema reconhece a modelagem desses dados e gera automaticamente uma interface de recuperação, sem que o usuário tenha que realizar nenhum tipo de programação. As informações sobre a modelagem do dado a ser recuperado e as operações a serem realizadas devem ser descritas em um formato de alto nível, independente da tecnologia do repositório, de forma que o sistema possa manipular diversas interfaces de busca. Para conciliar diferentes terminologias, o sistema deve permitir que as interfaces de recuperação sejam personalizadas, agregando valores à descrição dos dados. O sistema deve ainda fornecer a possibilidade de realizar processamentos sobre os dados recuperados, tanto usando cálculos comuns como médias e desvio padrão, ou com processamentos específicos para um certo tipo de dado.

1.2 Organização do documento

O capítulo 2 apresenta o contexto de colaboração internacional no qual o projeto foi desenvolvido, dando destaque às necessidades apresentadas por alguns grupos da colaboração. No capítulo 3, são realizadas as análises dos problemas e requisitos e é apresentada a proposta do sistema. As tecnologias utilizadas para implementar a proposta são discutidas no capítulo 4. O sistema desenvolvido é então apresentado em detalhes no capítulo 5, e as formas como o sistema é usado pela colaboração no capítulo 6. Por fim, o capítulo 7 apresenta as conclusões do trabalho e os possíveis desdobramentos.

Capítulo 2

O CERN e o ATLAS

Experimentos em física de altas energias envolvem a construção de experimentos complexos e de grande porte. Esse trabalho está inserido no contexto da colaboração que constrói o detector ATLAS, o maior dos detectores do LHC. As duas primeiras sessões desse capítulo apresentam o contexto de colaboração internacional para construção e operação do ATLAS, no CERN. A seguir são apresentados os bancos de dados utilizados pela colaboração para gerência de equipamentos.

2.1 O CERN e o Colisionador de Partículas LHC

O CERN (*Centre Européene pour la Rechèreche Nucleaire*) é o maior laboratório de física de partículas do mundo, situado na fronteira da Suíça com a França [1]. Ele conta com colaboração de cientistas vindos de institutos do mundo todo. Atualmente, o CERN está engajado no projeto LHC (Large Hadron Collider), um acelerador de partículas circular de 26,7 quilômetros de circunferência, que irá colidir feixes de prótons a uma energia de 14TeV, o que equivale a 99.9998% da velocidade da luz.

O LHC se situa em um túnel que fica 100 metros abaixo da superfície. Há quatro pontos de colisão distribuídos ao longo do acelerador, em torno dos quais são instalados detectores de partículas para analisar os subprodutos das colisões.

Os dois maiores experimentos são o ATLAS e o CMS, que consistem em detectores de partícula de propósito geral. Os outros são o LHCb, ALICE, TOTEM e LHCf, que são detectores especializados no estudo de certos fenômenos.[2]

O ATLAS e o CMS, como são de propósito geral, são capazes de estudar diversos fenômenos físicos. Dentre os fenômenos estudados, pode-se destacar o bóson de Higgs, que explicaria a origem da massa das partículas, a matéria negra, e dimensões extras. O fato de existirem dois detectores projetados de forma diferente, mas para o mesmo fim, é importante para que um possa confirmar o resultado do outro. O LHCb é um detector especializado no estudo do quark *beauty* com o intuito de investigar a diferença entre matéria e antimatéria. O ALICE, por sua vez, estudará o plasma quark-gluon que será formado quando o LHC fizer colisões de íons de chumbo, recriando as condições que existiram instantes após o Big Bang.

Os dois menores detectores são o LHCf e o TOTEM. O LHCf, que se situa na mesma caverna do ATLAS, estuda o comportamento de raios cósmicos, utilizando partículas que não chegaram a colir no ponto de colisão, mas foram desviadas. O TOTEM, que está instalado perto do CMS, estuda o feixe de prótons produzidos pelo LHC, bem como a estrutura interna do próton.

A Figura 2.1 mostra uma ilustração do túnel subterrâneo, com destaque para o detector ATLAS, que é o detector junto ao qual esse trabalho foi desenvolvido.

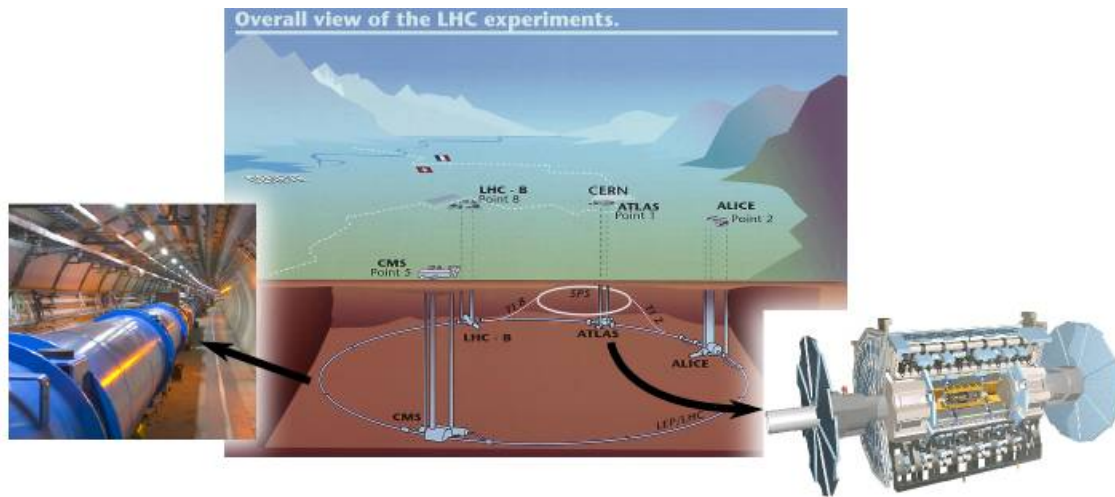


Figura 2.1: Túnel subterrâneo do LHC e seus quatro experimentos. No destaque da esquerda, uma foto do acelerador. Na direita, o detector ATLAS. Extraídas de [3]

2.2 O Detector de Partículas ATLAS

O ATLAS (*A Toroidal LHC Apparatus*) é o maior dos detectores que operam no LHC, medindo 45 metros de comprimento e 25 metros de altura e largura. Como é um detector de propósito genérico, o detector registra dados sobre os eventos de colisões de partículas que podem ser usados para estudos em diversas áreas da física. Um dos estudos mais esperados é o que comprova ou rejeita a existência do Bóson de Higgs, confirmando a teoria que explica de onde vem a massa das partículas.

O experimento está na fase de operação desde setembro de 2008 [4], quando o primeiro feixe de partículas circulou no acelerador e os primeiros eventos de colisão dos prótons com moléculas de gás dentro do tubo do acelerador foram registrados. Para coletar dados suficientes para as análises físicas o experimento pode durar cerca de 20 anos [5].

O ATLAS foi construído e é operado por uma colaboração internacional envolvendo cerca de 2500 físicos de mais de 169 instituições e laboratórios de 37 países [6]. Esses números incluem a UFRJ, que participa através da COPPE, da Escola Politécnica e do Instituto de Física. O detector é composto por 4 sub-detectores: o Inner detector, detector de traço; o Liquid Argon, que é o calorímetro eletromagnético; o Tile Calorimeter, que é o calorímetro hadrônico, e o Muon Spectrometer responsável por realizar medições sobre os múons. Cada uma dessas partes teve seus componentes construídos por um grupo diferente, pertencente às instituições colaboradoras. Uma vez construídos e testados, os componentes foram levados ao CERN, onde foram instalados em seus lugares definitivos. Para coordenar essa montagem e fornecer a infra-estrutura necessária no local do experimento, existe o grupo da Coordenação Técnica (*ATLAS Technical Coordination*).

2.3 Armazenamento de dados do ATLAS

Durante as fases de construção e comissionamento do ATLAS um grande volume de dados sobre seus equipamentos foram armazenados em bancos de dados, tais como informações sobre a interligação entre os componentes, local de instalação, configuração e resultados de testes. Na fase de teste e na atual fase de operação também é comum armazenar amostras de sensores que monitoram os equipamen-

tos e eventuais modificações na instalação dos equipamentos devido a reparos ou substituições de peças defeituosas.

2.3.1 Dados de instalação de equipamentos

A Coordenação Técnica do ATLAS é o grupo responsável por prover a infraestrutura necessária para a instalação e funcionamento do detector [7]. Esse grupo coordena o acesso e a ordem de instalação dos equipamentos na caverna experimental, onde fica o detector, e nas cavernas de serviço, onde ficam a eletrônica de controle e leitura de dados.

No banco de dados ATLASIntegration, mantido pela Coordenação Técnica, cada componente instalado nas cavernas é registrado, junto com sua posição e parte do detector a que pertence. Também são cadastrados todos os cabos que interligam os equipamentos, junto com os lugares por onde passam e onde devem ser conectados.

O sistema MTF (*Manufacturing and Testing Folder*), mantido pelo grupo de Tecnologia da Informação do CERN, é usado pelo LHC e pelo ATLAS para manter um inventário de todos os equipamentos instalados na área experimental, e permite que seja registrado o histórico de onde um determinado equipamento esteve, a fim de avaliar ao quanto de radiação ele foi exposto.

Esses bancos de dados são usados pela Coordenação Técnica para planejar as atividades de instalação dos equipamentos e da infraestrutura necessária para eles, tais como energia e sistema de refrigeração. Esse banco de dados é usado por colaboradores ligados aos sub-detectores, porém com interesses diferentes, tais como verificar o estado da instalação da infraestrutura requerida por seus componentes ou depurar a conexão entre os equipamentos.

2.3.2 Sistema de controle do detector

O DCS (Detector Control System) [8] é o sistema que monitora e controla os sub-detectores e os equipamentos de infra-estrutura comum, tais como refrigeração, criogenia, distribuição elétrica e segurança.

Os dispositivos que monitoram o detector são controlados pelo software comercial PVSS (*Prozessvisualisierungs- und Steuerungs-System*) [9]. O PVSS amos- tra periodicamente os sensores controlados pelos dispositivos de monitoração e arma-

zena seus valores em bancos de dados Oracle, para serem posteriormente recuperados por outros sistemas. Com o objetivo diminuir o volume de informações inseridas no banco de dados, é realizado o processo de *smoothing*, que consiste em só armazenar o valor no banco de dados caso a diferença entre a leitura atual e a anterior exceda um certo limiar ou caso tenha se passado um intervalo definido de tempo desde o último dado registrado.

Os sensores monitorados pelo DCS são definidos de acordo com as necessidades dos sub-detectores. Como o TileCal é o sub-detector com o qual a equipe deste projeto tem contato mais próximo, suas necessidades específicas foram analisadas primeiro.

O TileCal (Tile Calorimeter, ou calorímetro de telhas) mede a energia das partículas que o penetram por absorção total, usando aço como material absorvedor de energia e telhas cintilantes como material ativo. O TileCal é constituído de uma estrutura cilíndrica com um raio externo de 4,2 metros. O sub-detector é dividido em um barril central (barrel) e dois barris estendidos (extended barrels), cada um composto por 64 módulos no formato de cunhas. Ao atravessar o detector, as partículas liberam energia na forma de luz, produzidas nas telhas cintilantes e transformadas em sinal elétrico pelos fotomultiplicadores.

No caso do TileCal, o DCS monitora tensões, correntes e temperaturas das fontes de baixa tensão (LVPS) e de alta tensão (HVPS), bem como o sistema de refrigeração desses componentes para cada um dos 256 módulos. A partir da análise dessas tensões e correntes, é possível identificar problemas nas fontes ou na conectividade com a eletrônica dos módulos. A análise consiste em observar o valor médio e seu desvio padrão durante intervalos de tempo, variando de um dia a um mês.

Em alguns casos também é necessário analisar como se comporta, no tempo, a diferença de tensão entre dois sensores. Calcular essa diferença, no entanto, envolve reverter o processo de *smoothing*, pois, frequentemente, o PVSS não armazena as medições dos dois sensores simultaneamente.

2.3.3 Alinhamento e Posicionamento do Detector

Apesar de possuir em torno de 44 metros de comprimento e 25 metros de diâmetro, o experimento observa partículas de dimensões subatômicas. Por isso, é

necessário que a diferença entre as dimensões reais dos equipamentos e as previstas em projeto seja muito pequena, e que o alinhamento dos equipamentos entre si, e em relação ao feixe de partículas, seja muito preciso.

Para garantir a conformidade dos equipamentos, o correto alinhamento e até decidir a ordem nos quais os sub-detectores e seus equipamentos foram instalados na caverna experimental, o grupo *ATLAS Survey* acompanhou a evolução do detector desde a fabricação das partes em seus institutos de origem. Esse grupo mede pontos definidos em projeto para cada equipamento usando instrumentos de precisão e, para os equipamentos instalados, a posição desses pontos dentro das coordenadas da caverna.

A estabilidade do chão da caverna é de interesse especial. De acordo com cálculos realizados pelos engenheiros civis durante a escavação da caverna, o chão se movimentaria cerca de 6mm devido a acomodação do solo e subiria cerca de 1mm por ano devido à pressão hidrostática [10]. Isso pode afetar diretamente o alinhamento dos equipamentos com o ponto onde ocorrerá a colisão.

Para monitorar esse movimento do solo, o grupo *ATLAS Survey* implementou um sistema de sensores de nível hidrostático (HLS, na sigla em inglês para *Hydrostatic Levelling Sensors*) usando sensores capacitivos interligados por tubos de água criando um sistema de vasos comunicantes. Os sensores são dispostos paralelamente à direção do feixe e dentro das estruturas sobre as quais é montado o detector, chamadas de *bed-plates*. Esse sistema consegue medir variações no nível da água na ordem de micrômetros, e consegue observar, inclusive, reflexos de abalos sísmicos tais como o que provocou um tsunami no oceano Índico em 2004.

Os sensores do HLS são lidos periodicamente por um sistema eletrônico e os dados são armazenados em um banco de dados. Esses dados devem ser analisados posteriormente pelos responsáveis pelo Survey, a fim de identificar como a elevação do chão irá impactar nas leituras e calibrações do detector e planejar correções no alinhamento, caso necessário. Para isso é necessário calcular estatísticas sobre os dados dos sensores e exibir a evolução das médias de cada sensor com o tempo na forma de gráficos.

Capítulo 3

Especificação do sistema Glance

Neste capítulo são apresentadas as análises dos problemas relacionados a recuperação e operação de dados, e dos requisitos identificados. Em seguida são apresentados a proposta do sistema Glance, o projeto de sua arquitetura e as tecnologias a serem usadas.

3.1 Análise dos problemas

A colaboração que construiu e atualmente opera o detector ATLAS é composta por pessoas de diferentes especialidades, tais como físicos, engenheiros e informáticos, e provenientes de diversos países. Por isso seria necessário um esforço inviável para definir uma única modelagem de dados que atenda às necessidades de armazenamento de todos os colaboradores. É necessário, portanto, que cada um crie seus próprios meios de armazenamento e recuperação para os dados que necessitam. No entanto, para que um colaborador crie sua própria interface é necessário que ele tenha domínio sobre ferramentas de programação e sobre as tecnologias e modelagens usadas nos repositórios, que, frequentemente, foram projetados por colaboradores de outros institutos. Vemos, então, que não é razoável exigir conhecimentos de todas as tecnologias de armazenamento de dados utilizadas no experimento, principalmente quando o objetivo de estudo da maioria dos integrantes do ATLAS é a física, e não as ferramentas computacionais. Também não há garantias de que a pessoa que criou uma determinada interface de recuperação participe da colaboração durante todo o tempo no qual ela será útil, visto que o ATLAS vem sendo construído ao longo dos

últimos 20 anos e deve operar por, pelo menos, 10 anos.

Além disso, os colaboradores possuem diferentes idiomas nativos e, por isso, há o risco de um mesmo equipamento ou fenômeno ser conhecido por diferentes termos. Por exemplo, tubos por onde passam líquido para refrigeração podem ser chamados de *pipes* ou *tubes*, e até mesmo o lugar onde o ATLAS está instalado pode ser chamado de *pit*, *cavern* ou *underground*. Isso faz com que a busca por conjuntos de dados, para um certo termo, não produza resultados, caso os registros dos repositório tenham outra denominação.

Alguns dados ainda necessitam de processamento para serem úteis. No caso do DCS do TileCal, por exemplo, as amostras dos sensores de temperatura e voltagem que são armazenadas em bancos de dados precisam ser conformadas e depois transformadas em um gráfico. A conformação dos dados envolve reverter o processo de *smoothing*, que é uma operação específica para o DCS do TileCal. Uma vez conformados, os valores podem ser desenhados em um gráfico. Essa transformação de desenhar um gráfico a partir de um conjunto de valores pode ser aplicada em outros contextos diferentes do DCS. Os colaboradores do TileCal também precisam observar a média e o desvio padrão das medições desses mesmos sensores, para analisar o comportamento das fontes de tensão ao longo de um grande período de tempo. Isso mostra que um mesmo conjunto de dados pode precisar ser processado de diferentes formas, e ainda que o processamento pode ser específico à natureza do dado.

3.2 Análise dos requisitos

A partir do conhecimento do ambiente, como apresentado na seção 2.3, e dos problemas identificados, é possível extrair os seguintes requisitos:

- Fácil acesso por qualquer colaborador, independente de onde esteja.
- Permitir que o usuário final crie interfaces de recuperação sem precisar de nenhuma linguagem de programação ou de saber detalhes sobre modelagem e tecnologia do repositório.
- O sistema deve poder acessar diferentes servidores de bancos de dados, mesmo que sejam de diferentes tecnologias.

- Centralizar acesso aos dados da colaboração.
- Permitir que um mesmo conjunto de dados seja recuperado sob diferentes perspectivas, acomodando diferenças de interesses e terminologias.
- Permitir que o usuário especifique filtros para a recuperação dos dados.
- Processar dados depois de recuperados conforme necessário.
- Permitir que programas especialistas de processamento sejam aplicados aos dados recuperados.

Para ser útil durante todo o tempo de vida do experimento, é necessário que o sistema seja de fácil manutenção e que possa ter novas funcionalidades agregadas caso necessário.

3.3 Proposta do Sistema Glance

Para atender a esses requisitos, propõe-se criar um sistema de recuperação de dados acessível pela Web. O sistema apresenta ao usuário a organização dos bancos de dados, independente da tecnologia em que está armazenado, permitindo que ele escolha os conjuntos de dados de interesse. Uma vez escolhido, um processo automático deve reconhecer a modelagem desses dados e gerar uma interface de recuperação correspondente, obedecendo seus tipos e possíveis relacionamentos.

As interfaces de recuperação devem ser descritas em alto nível, de forma a abstrair especificidades da tecnologia por trás dos bancos de dados. Essa descrição deve poder ser armazenada para que possa ser acessada posteriormente por outros colaboradores. Ao editar a descrição de uma interface, pode-se adaptá-la para necessidades particulares, tais como a associação de sinônimos e descrição dos dados (meta-dados), de forma a facilitar a compreensão e acomodar diferenças de terminologias.

Ao manter as interfaces independentes, a manutenção do sistema é simplificada, pois um único sistema é capaz de gerenciar diversas interfaces. Por isso, qualquer mudança na estrutura de um banco de dados implica somente na adaptação da descrição das interfaces relacionadas, sem interferir nas outras interfaces nem necessitar de mudanças no código fonte.

Uma interface de recuperação deve apresentar uma busca paramétrica, identificando o tipo dos atributos e oferecendo operadores de comparação adequados, ou permitindo que o usuário escolha o valor a partir de uma lista de valores existentes, quando aplicável. Os resultados da busca podem ser disponibilizados para o usuário de diferentes formas: mostrados na tela na forma de tabela, gráficos ou salvos em arquivo.

Uma vez recuperados, os dados podem sofrer múltiplas etapas de transformação, de acordo com a configuração da interface e com as entradas do usuário. Algumas transformações podem ser genéricas, tais como traçar gráficos ou gerar arquivos para programas de planilha eletrônica, mas o sistema deve também permitir que programas externos específicos à natureza da informação recuperada sejam utilizados.

3.4 Projeto do Glance

Para implementar a proposta, o sistema Glance foi projetado usando um conjunto de componentes. Cada componente deve ser um conjunto de classes em C++ que realizam uma funcionalidade do sistema, e interagem entre si através de chamadas de métodos das suas classes dentro do programa. Os principais componentes são: Interface Web, Motor de Interfaces, Conector de Banco de Dados e operação sobre dados. As interações entre esses componentes são mostradas na figura 3.1.

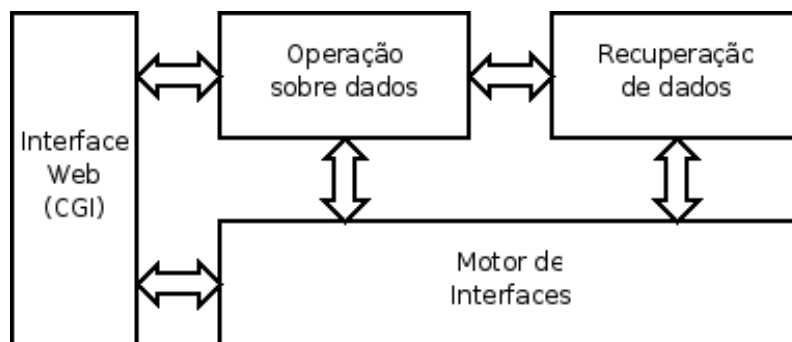


Figura 3.1: Arquitetura do sistema Glance

O componente “Interface Web” é o responsável por se comunicar com o servidor Web utilizando o padrão *Common Gateway Interface* (CGI). O conector de banco de dados é o responsável por reconhecer a estrutura de um repositório, bem

como recuperar os dados selecionando-os de acordo com os parâmetros especificados na descrição da interface de recuperação e pelo usuário. Esse componente encapsula todo o código que depende de uma tecnologia de banco de dados, desacoplando-o do sistema. Dessa forma, novos bancos podem ser acessados pelo sistema implementando um novo conector. Atualmente, o sistema recupera dados de bancos de dados Oracle, usando para isso a biblioteca OCCI.

O motor de operações recebe os dados vindos do conector de banco de dados e os processa de acordo com a definição da interface de busca. Os processamentos podem ser desde cálculos simples, tais como a soma entre dois dos atributos recuperados, até processamentos específicos, como a conformação de leituras de sensores do DCS.

O motor de interfaces, que é o principal componente do sistema, integra os demais componentes. Ele analisa as definições de interfaces de busca e coordena a recuperação e a operação dos dados, de acordo com os parâmetros recebidos do usuário pela Web. As interfaces de busca são descritas em um arquivo XML, que especifica como recuperar um determinado conjunto de dados, processá-los e exibi-los ao usuário.

3.5 Tecnologias

As tecnologias foram escolhidas levando em consideração que o sistema deve continuar operacional durante toda a vida útil do experimento, que pode chegar a 20 anos, e que futuramente outras pessoas poderão assumir a responsabilidade de manutenção e desenvolvimento do sistema. Para que o sistema permaneça disponível durante a vida útil do experimento, é necessário que as tecnologias tenham suporte da equipe que mantém a infra-estrutura de computação do CERN. Além disso, para facilitar manutenções e desenvolvimentos futuros, é necessário que sejam usados padrões abertos apoiados pela comunidade e bibliotecas ativamente mantidas por seus desenvolvedores.

Esta seção descreve as tecnologias aplicadas no projeto.

3.5.1 A linguagem de programação C++

C++ é uma linguagem de propósito geral baseada na linguagem C. Ela oferece orientação a objetos e algumas características, tais como sobrecarga de operador e tratamento de erro [11]. É uma das linguagens mais populares na indústria de softwares atualmente, o que faz com que ela seja amplamente documentada.

O núcleo do sistema Glance foi feito em C++. A modelagem orientada a objetos do sistema permite que ele seja mais facilmente estendido. Além disso, é uma linguagem com suporte técnico e amplamente utilizada no CERN.

3.5.2 A linguagem Python

Python é uma linguagem de scripts de propósito geral [12]. É uma linguagem dinâmica e oferece recursos como orientação a objetos e introspecção. A sintaxe é simples e sua biblioteca padrão é bastante completa, fazendo com que o desenvolvimento e a manutenção sejam mais rápidos. Além disso, por se integrar facilmente com bibliotecas em C e C++, a linguagem Python é muito popular no CERN, pois permite o rápido desenvolvimento de programas de análise de dados, enquanto os cálculos custosos são delegados à ferramenta ROOT, criada no CERN, para análise de dados [13], que é feita em C++.

Os programas para processamento de dados que serão utilizados no Glance são feitos em Python.

3.5.3 Banco de dados Oracle e a OCCI

O Oracle é o sistema gerenciador de bancos de dados (SGBD) que, segundo o grupo Gartner, foi líder do mercado em 2005 e 2006 [14]. Suporta os paradigmas de bancos de dados relacionais, orientados a objetos e objeto-relacionais. O CERN usa tecnologia Oracle nos seus bancos de dados, e essa é a única tecnologia de banco de dados que é mantida pela equipe de Tecnologia de Informação.

A OCCI (Oracle C++ Call Interface) é a biblioteca oficial do Oracle para desenvolvimento de programas em C++ que se comuniquem com seus bancos de dados. O Glance usa essa biblioteca para recuperação dos dados.

3.5.4 XML e XSLT

O XML (eXtensible Markup Language) [15] é uma linguagem de marcação de texto simples, muito parecida com o HTML. Enquanto o HTML dá ênfase à apresentação de dados, o XML permite descrever os dados concentrando-se na estrutura destes. Esta linguagem é utilizada na troca de dados entre sistemas computacionais, já que permite a formatação dos dados num padrão que pode ser entendido por qualquer sistema, e na publicação de dados, com o auxílio de linguagens e programas que apresentam dados armazenados em XML de maneira amigável ao usuário final.

A linguagem utilizada para a apresentação de dados é o XSL (XML Stylesheet Language). É uma linguagem baseada em XML que define transformações sobre os documentos XML que resultam em novos documentos XML ou em documentos HTML.

O sistema Glance utiliza o XML para descrever as interfaces de busca, bem como na troca de dados com outras aplicações. Para isso, utiliza a biblioteca Xerces-C++ [16].

Capítulo 4

O Sistema Glance

A primeira seção desse capítulo apresenta o conceito de interfaces de buscas e mostra em detalhes como se dá a descrição de uma interface. A criação de uma interface de busca e seu funcionamento são descritos, respectivamente, na segunda e terceira seção. A quarta seção mostra como o Glance faz a recuperação de dados de acordo com os parâmetros da interface de busca. Por fim, a última seção mostra como os dados recuperados são transformados, antes de serem exibidos para o usuário.

4.1 Descrição de uma interface de busca

A principal inovação do sistema Glance é o conceito de interfaces de busca descritas em uma linguagem independente do banco de dados. A descrição é feita usando XML, de forma que novas funcionalidades podem ser adicionadas ao sistema mantendo compatibilidade com as interfaces existentes. Cada interface é descrita em um documento XML separado e isolado, que pode ser armazenado pelo sistema para acesso futuro.

A descrição XML básica de uma interface de busca contém as credenciais para conexão ao banco de dados, a relação de tabelas e colunas que constituem os atributos de busca da interface e informações para correlacionar as tabelas (*Join*), caso esteja-se usando mais de uma. Cada atributo tem registrado seu tipo, que pode ser texto, número, data ou enumeração.

Para especificar os processamentos que devem ser realizado sobre os dados,

a descrição também pode conter formatos de saída diferentes, particulares à interface em questão. Para cada formato de saída, deve ser especificada a sequência de transformações que devem ser executadas.

O XML que descreve a interface deve ter o elemento raiz chamado *searchInterface*. Dentro desse elemento, o elemento *name* contém o nome que será usado para identificar a interface. O elemento *database* especifica o banco de dados no qual a informação está armazenada. O conteúdo desse elemento pode variar de acordo com a tecnologia do repositório. Como atualmente o sistema conecta-se a Oracle, são aceitos os elementos *connectString*, *username* e *password*, contendo o nome do banco de dados Oracle, o nome do usuário e a senha, respectivamente.

Para cada tabela de onde o sistema deve recuperar dados, deve existir um elemento *table*. Dentro desse elemento, devem ser especificados nos elementos *tableName* e *schemaName*, o nome da tabela e do esquema, respectivamente. Além disso, para cada coluna a ser recuperada da tabela deve ser especificado um elemento *column*. Dentro de cada *column*, deve ser especificado um identificador único no elemento *colId* dentro da interface para uso pelo sistema, o nome da coluna em *colName*, seu tipo em *colType* e sua descrição, que é o nome que será exibido na interface de busca, em *colDescr*. Caso seja necessário recuperar uma expressão envolvendo colunas de uma tabela, em vez de uma coluna simples, deve-se especificar o elemento *colExpr* contendo a expressão desejada, em vez do *colName*.

Os tipos de uma coluna podem ser *text*, *integer*, *float*, *datetime* ou *enumeration*. Caso o tipo seja *enumeration*, é necessário fornecer elementos *element* que representarão os possíveis elementos desse campo enumerado. Em cada *element*, deve haver um *name* e um *value*, que conterão, respectivamente, o nome do elemento que será mostrado na interface, e o correspondente valor no banco de dados.

Outros possíveis elementos filhos do *searchInterface* são o *whereClause* e o *outputFormat*. O primeiro elemento especifica cláusulas fixas de seleção de dados, incluindo as cláusulas para correlacionar duas tabelas (*join*). O *outputFormat*, por sua vez, permite especificar formatos de saída particulares à interface. Este elemento deve conter um *formatId* contendo um identificador único na interface para o formato de saída, *formatDescription* com a descrição que será vista pelo usuário na interface, e elementos *transformation* especificando quais operações devem ser executadas e a

ordem.

A Figura 4.1 mostra um exemplo da descrição de uma interface em XML. O nome da interface é “*Search Interface for Tile DCS Data*”. Ela recupera dados das tabelas *ELEMENTS* e *EVENTHISTORY* do esquema *ATLAS_PVSSTIL*, no banco de dados *ATLR*. Ela contém três atributos: “System”, “Time Stamp” e “Value”. O primeiro atributo corresponde a uma coluna da primeira tabela, enquanto os outros dois são colunas da segunda tabela. O atributo “System” está declarado como sendo um campo enumerado, então na interface o usuário poderá escolher entre “LBA” e “LBC” para os valores desse campo, que são os nomes de duas das quatro subdivisões principais do subdetector TileCal. Já o atributo “Value” não corresponde diretamente a uma coluna mas à expressão “TO_NUMBER(VALUE_NUMBER)” e, por isso, foi declarado com o elemento *colExpr*. Esta interface possui o formato de saída “Plot”, que é obtido passando os dados recuperados para serem operados pelas transformações *dcsUnsmoother* e, em seguida, *pyPlotter*.

4.2 Criação de interfaces de busca

Para que um colaborador possa criar facilmente uma interface de busca, o sistema Glance cria automaticamente uma interface para o conjunto de dados escolhido pelo usuário. Para isso o sistema usa o conector com banco de dados para reconhecer a estrutura do repositório, e a exhibe para o usuário, aumentando o nível de detalhes sucessivamente sobre os objetos selecionados. Uma vez escolhido o conjunto de dados de interesse, o conector reconhece quais os atributos envolvidos e seus tipos e, a partir disso, o sistema gera uma descrição XML de uma interface de busca.

No caso de um banco de dados Oracle, o sistema inicialmente pede o endereço do banco e as credenciais para iniciar a conexão. Em seguida, a lista dos Schemas disponíveis é exibida. Quando um *schema* é escolhido, as tabelas contidas nesse são exibidas. Por fim, deve ser escolhido o método de criação entre rápido ou completo. Este processo está ilustrado na Figura 4.2.

A partir daí o Glance descreve essa tabela e cria uma interface de busca onde os atributos correspondem às colunas da tabela. O tipo de cada atributo é escolhido

```

<searchInterface>
  <name>Search Interface for Tile DCS Data</name>
  <database>
    <connectString>ATLR</connectString>
    <username>ATLAS_PVSS_READER</username>
    <password>MY_DCS_PASS</password>
  </database>
  <outputFormat>
    <formatId>dcsPlot</formatId>
    <formatDescription>Plot</formatDescription>
    <transformation>dcsUnsmoother</transformation>
    <transformation>pyPlotter</transformation>
  </outputFormat>
  <table>
    <tableName>ELEMENTS</tableName>
    <schemaName>ATLAS_PVSSTIL</schemaName>
    <column>
      <colName>SYS_ID</colName>
      <colId>ATLASPVSSSTILEELEMENTSSYSID</colId>
      <colDescr>System</colDescr>
      <colType>enumeration</colType>
      <element><name>LBA</name><value>88</value></element>
      <element><name>LBC</name><value>87</value></element>
    </column>
  </table>
  <table>
    <tableName>EVENTHISTORY</tableName>
    <schemaName>ATLAS_PVSSTIL</schemaName>
    <column>
      <colName>TS</colName>
      <colId>TS</colId>
      <colDescr>Time Stamp</colDescr>
      <colType>datetime</colType>
      <colResolution>second</colResolution>
    </column>
    <column>
      <colExpr>TO_NUMBER(VALUE_NUMBER)</colExpr>
      <colId>VALUE</colId>
      <colDescr>Value</colDescr>
      <colType>float</colType>
    </column>
  </table>
  <whereClause>elements.element_id = eventhistory.element_id</whereClause>
</searchInterface>

```

Figura 4.1: Descrição XML de uma interface de busca

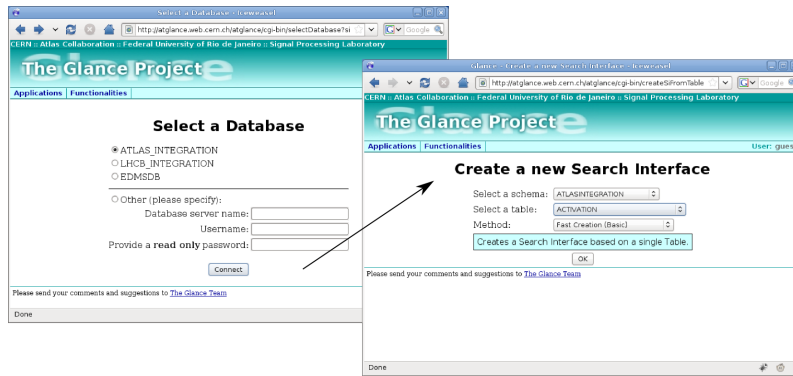


Figura 4.2: O Glance apresentando estrutura do banco de dados e gerando uma interface adequada

de acordo com o tipo da coluna, para que os operadores de busca sejam os adequados. A diferença entre o método rápido e o completo é que o último apresenta uma etapa intermediária onde o usuário pode personalizar a interface. As mudanças de trocar a descrição de um atributo ou torná-lo visível somente na busca ou no resultado pode ser feita em uma interface simplificada, conforme mostrada na tela da esquerda na Figura 4.3. O sistema também permite que um usuário avançado edite diretamente a descrição em XML para personalizações mais complexas. Essa edição é mostrada na tela da direita, na mesma figura.

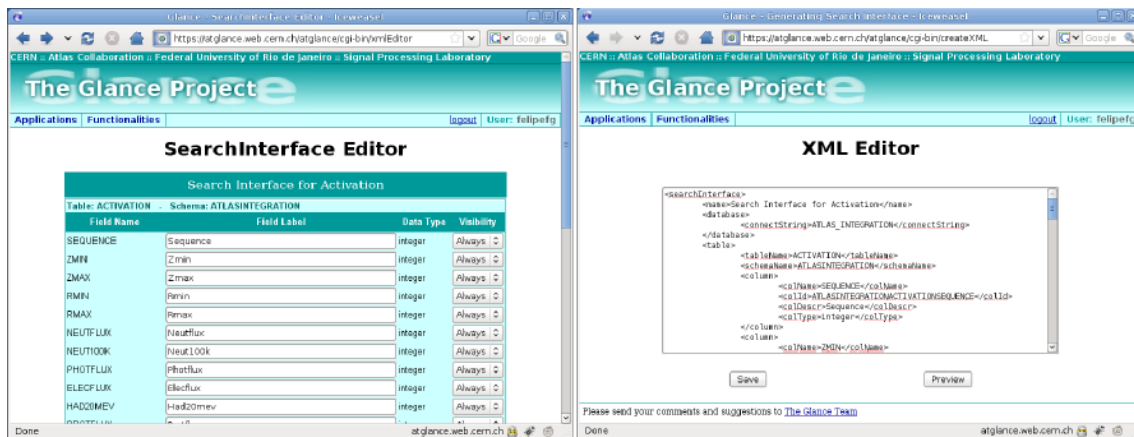


Figura 4.3: Personalização de uma interface de busca durante sua criação.

Após os ajustes, a interface criada é exibida e pode ser usada imediatamente para realização de buscas. A interface pode também ser armazenada pelo sistema para ser acessada posteriormente. O funcionamento de uma interface de busca é descrito na próxima sessão.

4.3 Interface de busca

A partir da página inicial do sistema, mostrada na Figura 4.4, é possível selecionar entre a visualização de interfaces de busca existentes, ou a criação de uma nova interface.

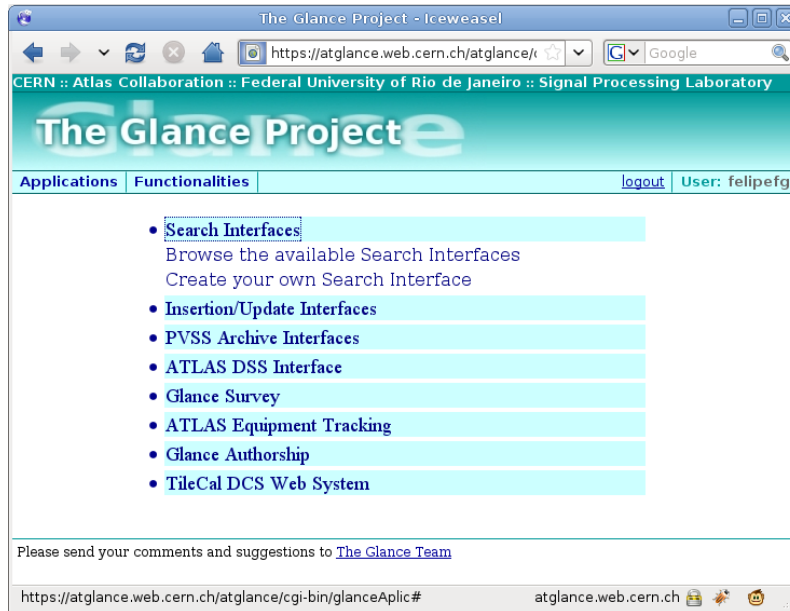


Figura 4.4: Página inicial do Glance

As interfaces de busca existentes estão agrupadas por categorias e aplicações. Ao escolher uma categoria, uma lista de interfaces é mostrada, conforme a Figura 4.5.

Quando uma interface nova é criada, de acordo com o procedimento apresentado na Seção 4.2, ou escolhida na lista da Figura 4.5, sua descrição em XML é passada para o motor de interfaces, que apresenta, para o usuário, uma interface conforme exemplificado na Figura 4.6.

A interface de busca gerada é paramétrica, ou seja, o usuário pode especificar parâmetros de busca que serão aplicados sobre os dados para a consulta. Ao especificar um parâmetro, o usuário escolhe qual atributo deve ser comparado, o operador e o valor. Os operadores disponibilizados dependem do tipo do campo. Por exemplo, um campo numérico permite os operadores “igual”, “maior que”, “menor que”, etc. Um campo de texto, por sua vez permite operadores como “contém”, “igual a” e “diferente de”. Mais de um parâmetro pode ser especificado, e é possível escolher se todos os resultados devem atender ao conjunto de parâmetros ou a pelo menos

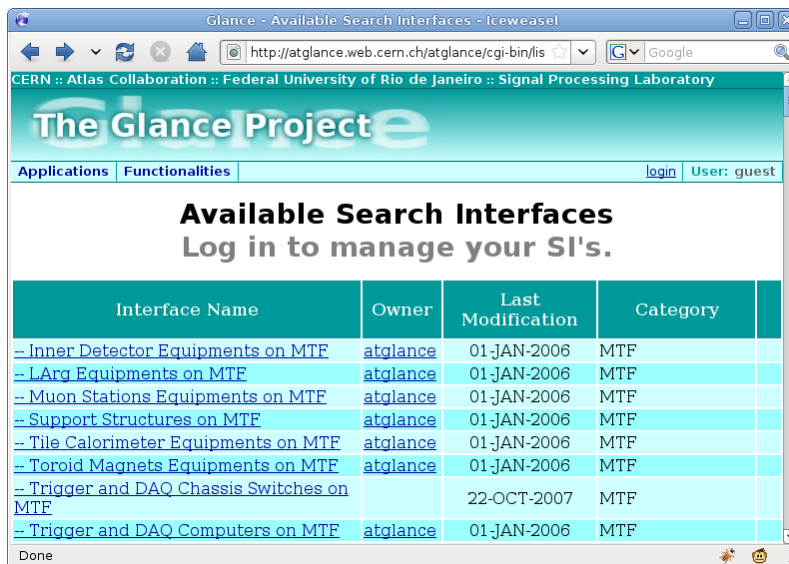


Figura 4.5: Lista de interfaces de busca existentes

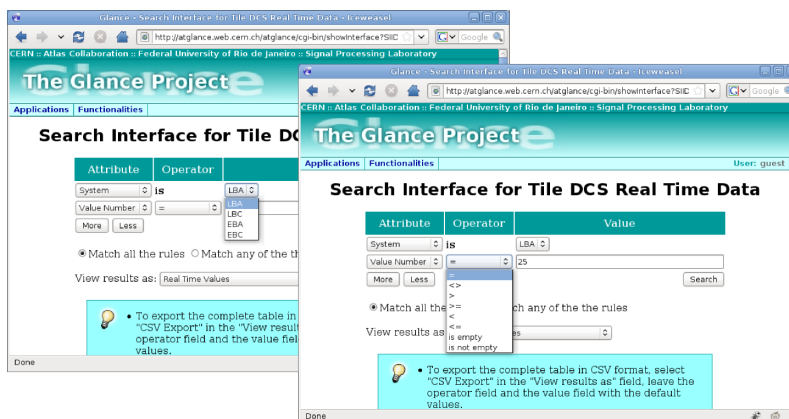


Figura 4.6: Interface de busca gerada pelo Glance, destacando os operadores de busca sensíveis ao tipo do atributo.

um deles (i.e. combinação “e” ou “ou”).

Ainda na interface, o usuário pode escolher o tipo de visualização dos dados recuperados. Os resultados podem ser exibidos no próprio navegador, na forma de uma tabela, ou salvos em arquivo, no formato CSV (*Comma Separated Values*), que pode ser usado por diversos programas como, por exemplo, um processador de planilhas como o Excel, ou em XML. Uma interface pode ter definidos formatos de saída personalizados, que serão disponibilizados entre esses formatos padrão. Os formatos de saída farão com que o sistema exiba, para o usuário, os dados após serem processados pelo mecanismo de operações, descrito na Seção 4.5.

No caso da escolha por uma visualização dos resultados no próprio navegador

Web, o sistema formata a saída em uma tabela HTML (ver Figura 4.7). O sistema também fornece uma URL que contém todas as informações necessárias para realizar essa busca. Toda a vez que essa URL for acessada, o sistema realizará a busca e mostrará os resultados. Dessa forma, um usuário pode armazenar nos favoritos, e sempre que acessar terá as informações mais recentes vindas do banco de dados.

System	Module Number	Brick Name	Channel Name	Time Stamp	Value Number
88	25	ATLTILLV01.ELMBILVCAN2 ILVPS_25	ATLTILLV01.ELMBILVCAN2 ILVPS_25	2009/08/19 00:04:10	3.4623596270054855
88	25	ATLTILLV01.ELMBILVCAN2 ILVPS_25	ATLTILLV01.ELMBILVCAN2 ILVPS_25	2009/08/19 01:04:11	3.4621057155048223
88	25	ATLTILLV01.ELMBILVCAN2 ILVPS_25	ATLTILLV01.ELMBILVCAN2 ILVPS_25	2009/08/19 02:04:22	3.4627388347041892
88	25	ATLTILLV01.ELMBILVCAN2 ILVPS_25	ATLTILLV01.ELMBILVCAN2 ILVPS_25	2009/08/19 03:04:23	3.4628657904545213
88	25	ATLTILLV01.ELMBILVCAN2 ILVPS_25	ATLTILLV01.ELMBILVCAN2 ILVPS_25	2009/08/19 04:04:23	3.4628657904545213
88	25	ATLTILLV01.ELMBILVCAN2 ILVPS_25	ATLTILLV01.ELMBILVCAN2 ILVPS_25	2009/08/19 05:04:24	3.4628657904545213
88	25	ATLTILLV01.ELMBILVCAN2 ILVPS_25	ATLTILLV01.ELMBILVCAN2 ILVPS_25	2009/08/19 06:04:25	3.4626127087300036
88	25	ATLTILLV01.ELMBILVCAN2 ILVPS_25	ATLTILLV01.ELMBILVCAN2 ILVPS_25	2009/08/19 07:04:26	3.4622326712551539
88	25	ATLTILLV01.ELMBILVCAN2 ILVPS_25	ATLTILLV01.ELMBILVCAN2 ILVPS_25	2009/08/19 08:04:36	3.4627388347041892

Figura 4.7: Resultado de uma busca mostrado na forma de uma tabela

4.4 Recuperação de dados

O conector de banco de dados é responsável por tratar as particularidades de cada tecnologia de banco de dados com o qual o sistema pode se conectar. Para isso, é necessário que o conector seja capaz de descrever a estrutura de um banco de dados e realizar a recuperação correspondente.

Durante o processo de criação de uma interface de busca o conector é usado para reconhecer a estrutura do banco de dados. Para um banco de dados relacional, por exemplo, o conector deve ser capaz de listar os esquemas existentes e, em seguida, listar as tabelas e *views*. Para o conjunto de tabelas escolhido pelo usuário, o conector deve então reconhecer os atributos envolvidos e seus tipos. O sistema então usa essas informações para gerar automaticamente uma interface de busca.

Durante a recuperação de dados, o conector recebe, do motor de interfaces, a relação de atributos a serem recuperados, bem como os critérios de busca preenchidos pelo usuário. A partir dessas informações, o conector deve realizar a consulta no

banco de dados, gerando os comandos adequados, conforme necessário.

O principal conector utilizado pelo sistema é para bancos de dados Oracle. Ao descrever os bancos de dados, são listados os esquemas e, em seguida, as tabelas para um dado esquema. Uma vez selecionada a tabela, o conector gera uma descrição XML baseada nas colunas e seus tipos.

4.5 Operação sobre dados

Esta seção descreve o processamento dos dados que ocorre sobre o resultado de uma recuperação realizada por uma dada interface de busca que tem definida em sua descrição formatos de saída personalizados. Cada formato de saída escolhido pode especificar diversas etapas de processamento, que serão executadas em sequência. Cada etapa pode ser de um tipo diferente. A primeira subseção descreve processamentos que podem ser executados somente com os dados recuperados, não recebendo nenhuma informação do usuário. A segunda subseção discute um mecanismo para que a operação a ser realizada sobre os dados seja especificada pelo usuário ao realizar a busca. A última subseção, então, apresenta o mecanismo que permite processamentos complexos que frequentemente são específicos à natureza dos dados, e recebem parâmetros especificados pelo usuário durante a busca.

4.5.1 Mecanismo para processamentos não parametrizados

Esta funcionalidade tem como objetivo permitir que dados recuperados pelo Glance sejam processados de acordo com regras estabelecidas no momento da criação de uma interface de busca. Os processamentos, no entanto, dependem fortemente da natureza desses dados, podendo haver casos em que um certo tipo de processamento somente se preste a um conjunto específico de informações. Por isso, foi decidido delegar o processamento a um programa externo, e definir uma forma de integração do Glance com esse programa.

Após executar a consulta no banco de dados, o Glance transforma os resultados em um arquivo XML que então é repassado ao programa externo. O programa, então, deve retornar a saída de forma adequada a apresentação para o usuário. A Figura 4.8 ilustra esse conceito.

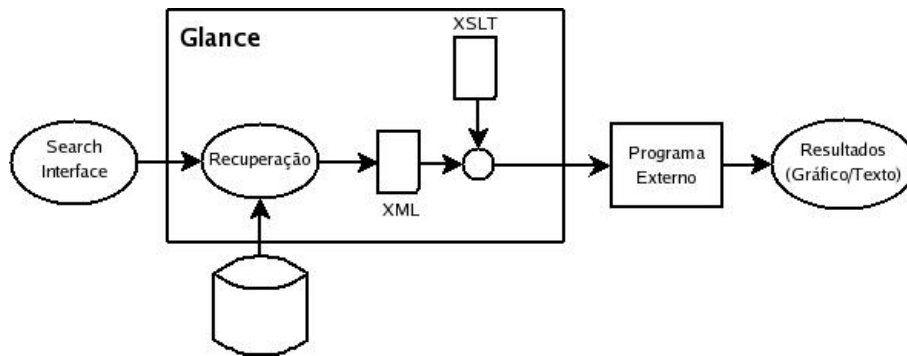


Figura 4.8: Realização de processamento sobre dados recuperados por um programa externo.

No mecanismo desenvolvido, cada interface de busca tem a descrição de formatos de saída e, para cada formato, é descrita a sequência de transformações que devem ser realizadas. A classe *OutputFormat* é a responsável por gerenciar e aplicar as transformações para um dado formato. Para isso ela guarda uma lista de objetos da classe *Transformation*. Essa classe é abstrata, de forma que cada tipo de transformação (XSLT, programa externo, etc) é uma especialização dessa. O diagrama da Figura 4.9 mostra conceitualmente as classes envolvidas e suas relações.

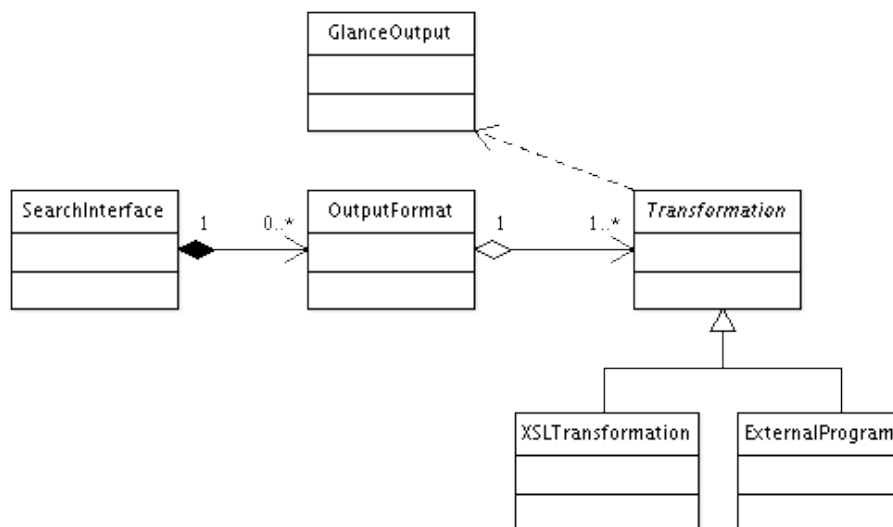


Figura 4.9: Diagrama conceitual de classes do mecanismo genérico para transformações não parametrizadas.

Ao recuperar o resultado de uma busca, um objeto do tipo *GlanceOutput* contendo os dados é gerado e submetido, através da *OutputFormat* adequado, pelas transformações. O resultado das transformações será também um objeto do tipo

GlanceOutput, que deve ser retornado ao usuário.

Esse mecanismo atende todo tipo de processamento que não necessita de parâmetros fornecidos pelo usuário para realizar seu cálculo. Como cada programa externo está encapsulado por um objeto do tipo *Transformation*, um procedimento genérico, tal como transformar uma série de números de qualquer natureza em um gráfico, pode ser reutilizado por diversas interfaces. No entanto, as transformações necessitam ser especificadas no momento da criação da interface de busca, e os programas externos precisam ser feitos por um especialista nos dados.

As subseções a seguir tratam do desenvolvimento dos mecanismos que tratam o caso no qual parâmetros podem ser especificados pelo usuário diretamente na interface de busca.

4.5.2 Operações genéricas e parametrizadas

O mecanismo descrito na subseção anterior atende bem ao caso no qual o usuário que define uma interface de busca sabe que tipo de processamento os dados recuperados devem sofrer, de forma que sejam úteis à aplicação. No entanto, é difícil prever todos os casos nos quais um determinado dado deve ser acessado a nível de colaboração internacional. Para manter a flexibilidade do sistema, faz-se necessária uma funcionalidade que permita que o usuário final especifique as operações a serem realizadas no momento da busca.

Para realizar essa funcionalidade, foi inicialmente especificada uma extensão do mecanismo de transformações descrito na subseção anterior. A extensão se baseia em uma transformação especial, que recebe não só os resultados da busca realizada pelo Glance, mas também uma descrição de como devem ser operados, de acordo com o que foi especificado pelo usuário na interface de busca. O programa que implementa essa transformação especial aplica as transformações especificadas para cada registro vindo do resultado do pré-processamento das informações recuperadas. O resultado pode então passar por um pós-processamento a fim de formatá-lo adequadamente. A Figura 4.10 ilustra esse processo.

O módulo que realiza operações deve receber sua entrada no mesmo formato de XML que é gerado pelo Glance. Esse formato é geral o suficiente para poder ser usado em qualquer contexto. Os dados são analisados pela classe *GlanceParser*, que

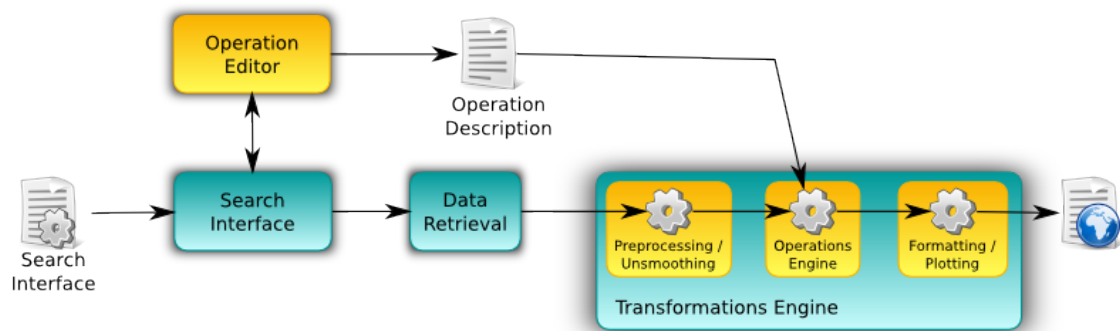


Figura 4.10: Processo de recuperação de dados realizando operações definidas pelo usuário.

implementa um analisador SAX de XML. A descrição das operações é interpretada pela classe *OperationDescriptionParser*, que monta uma árvore de operadores para cada cálculo realizada. Os cálculos podem dar origem a uma nova coluna na tabela, ou gerar um sumário, tal como uma média. Cada operador é uma realização da interface *Operator*, de forma que novos operadores podem ser incluídos conforme necessário. Sobre cada registro lido pelo *GlanceParser*, as operações são aplicadas e uma saída é gerada. Após processar todas as entradas, os valores sumários são apresentados. A Figura 4.11 mostra o diagrama conceitual de classes para essa especificação.

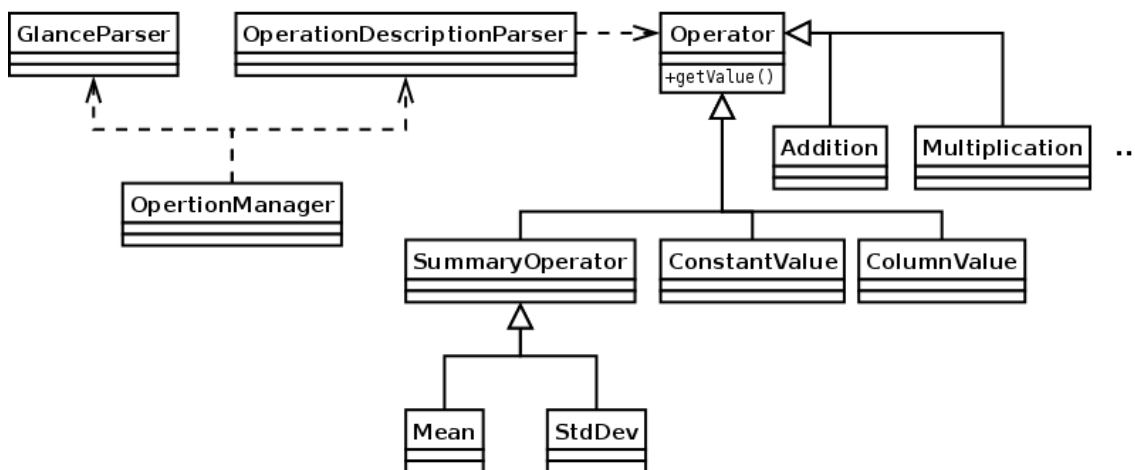


Figura 4.11: Diagrama conceitual de classes do módulo que realiza operação sobre os dados.

Esse programa foi implementado usando a linguagem de programação Python, e recebe da interface de busca um XML que define as operações a serem realizadas

sobre os dados. A primeira versão do programa admite os operadores matemáticos de soma, subtração, divisão e multiplicação de números e, como operações de sumário, realiza somatório, média e desvio padrão. Como os dados são interpretados pelo método SAX, cada entrada é processada assim que é lida. Dessa forma o uso de memória não cresce à medida que o volume de dados de entrada crescem e, portanto, pode ser aplicado sobre qualquer quantidade de dados.

Os dados da entrada são arranjados em uma pilha FIFO (*First In, First Out*) de 3 posições. Ao chegar um registro novo da entrada, o programa descarta o elemento da última posição e insere os dados novos no topo da pilha. As operações descritas pelo usuário são então calculadas sobre o elemento do meio. Dessa forma, as operações podem se dar não só por atributos do registro atual, mas também por atributos do registro anterior no arquivo de entrada e do próximo.

O mecanismo desenvolvido é genérico, pois pode ser usado para qualquer modelagem de dados. A subseção a seguir trata de programas especializados em algum tipo de modelagem, que precisa receber parâmetros fornecidos pelo usuário.

4.5.3 Operações específicas e parametrizadas

O mecanismo relatado na Seção 4.5.2 permite que operações simples sejam especificadas pelo usuário na hora da recuperação e pode ser usado para qualquer tipo de modelagem de dados. No entanto, as operações são simples demais para solucionar problemas como o condicionamento dos dados do ATLAS Survey. Além disso, criar uma função nesse sistema genérico que resolvesse o problema do ATLAS Survey iria contra o propósito desse sistema, pois a operação dependeria fortemente da modelagem dos dados.

Para solucionar esses problemas, foi projetada uma extensão ao mecanismo original de operações de forma que cada processo de transformação, seja ele genérico ou não, possa receber parâmetros especificados pelo usuário no momento da consulta. A Figura 4.12 mostra conceitualmente o processo de realização de operações.

Com essa extensão, assim que o usuário submete sua consulta, um arquivo XML é gerado não só com as operações a serem executadas pelo mecanismo da subseção anterior, mas também com valores de campos que obedecem uma nomenclatura específica. Esse arquivo, então, é disponibilizado para todas as transformações a

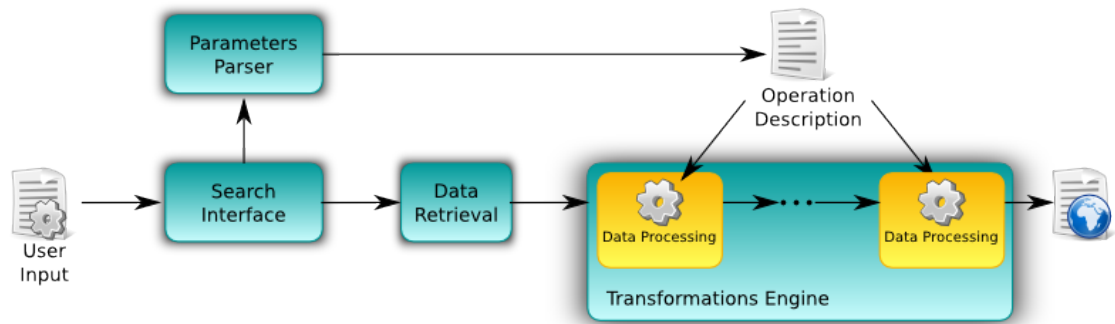


Figura 4.12: Processo de recuperação de dados realizando operações definidas pelo usuário.

serem executadas para aquela busca. O programa de transformação, por sua vez, deve recuperar as opções do usuário, realizando validações.

Como as opções são passadas na forma de campos do formulário HTML da busca, é possível publicá-las na URL, de forma que a busca possa ser executada novamente simplesmente acessando essa URL. A busca pode, então, ser guardada nos favoritos do usuário, ou executada de dentro de um programa externo.

Capítulo 5

Aplicações do Glance

O sistema desenvolvido foi usado para sanar a necessidade de recuperação de diversos grupos da colaboração do ATLAS. Este capítulo explica como o Glance foi usado nas suas três principais aplicações.

5.1 Dados de instalação de equipamentos

O banco de dados ATLAS Integration, mantido pela Coordenação Técnica do ATLAS, possui dados que interessam a diferentes partes da colaboração. Portanto, foram criadas diferentes interfaces de busca do Glance, cada uma atendendo a um interesse em particular.

Os dados de cabeamento, por exemplo, deram origem a cinco interfaces de busca. Uma delas recupera todos os cabos, junto com o tipo do cabo, por onde ele passa e seu estado de instalação. Essa interface é utilizada pela Coordenação Técnica para auxiliar a instalação dos cabos. Outras duas interfaces listam cabos que conectam os canais de dados do detector às placas que conformam e registram os dados. Como essa interface é usada por pessoas trabalhando nos sub-detectores, ela mostra posição dos conectores de cada cabo. Além disso, duas interfaces são usadas para recuperar informações de tubos para refrigeração do detector, que são armazenados no banco de dados como se fossem cabos. Além disso, existem mais seis interfaces para recuperar dados de instalação das placas eletrônicas de aquisição de dados e controle dos sub-detectores e dos racks onde elas são montadas.

5.2 Sistema de Controle do Detector

O DCS Web System é um sistema que foi desenvolvido por Fernando Guimarães Ferreira, aluno de iniciação científica do mesmo grupo de trabalho. O sistema centraliza o acesso aos dados coletados pelo DCS do TileCal. Em sua página inicial, um esquema gráfico do detector é apresentado com os módulos coloridos de acordo com o estado de funcionamento (funcionando, valores abaixo do limiar, uma fonte com problema ou mais de uma fonte com problema) de suas fontes de tensão. O estado é determinado a partir das médias e RMS das variáveis adequadas durante o período de um dia. Quando clica-se em um módulo, o sistema apresenta detalhadamente seus valores e permite a geração de gráficos com dados específicos. Atualmente o sistema está instalado no CERN e sendo ativamente usado pelos colaboradores do TileCal, e pode ser acessado no endereço <http://cern.ch/tcws/DCS>.

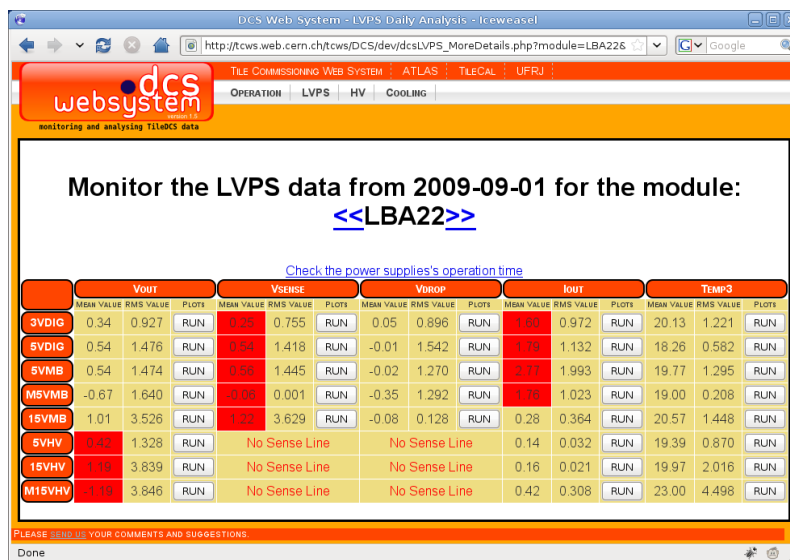


Figura 5.1: DCS Web System exibindo médias e RMS calculados pelo Glance

A figura 5.1 apresenta a análise das fontes de baixa tensão para o módulo LBA22. As médias e RMS usadas pelo DCS Web System vêm dos valores armazenados no banco de dados do DCS que são recuperados pelo Glance e operados pelo sistema de processamento de dados descrito na Seção 4.5. Para tanto, foram desenvolvidas duas novas transformações para serem utilizadas pelo DCS Web System. Uma das transformações desenvolvidas é aplicada sobre as medidas de tensões e correntes, e serve para recuperar e agregar as informações sobre o estado de funcionamento do módulo geradas pelo DCS. A outra transformação se aplica a qualquer

tipo de dado onde uma das colunas é uma marcação de tempo, e gera um arquivo NTuple, que pode ser interpretado pelo Root, a principal ferramenta para análise de dados usada pela colaboração.

5.3 ATLAS Survey

Conforme apresentado na Seção 2.3.3 o grupo do ATLAS Survey mantém um conjunto de sensores de nível hidrostático, que são amostrados com um período de cerca de 10 segundos. As leituras para cada sensor são armazenadas em um banco de dados.

Para analisar os dados, os colaboradores precisam agrupar as leituras de cada sensor em períodos com um determinado parâmetro, que frequentemente é de uma hora. Para cada período deve ser calculada a média do sensor. A partir disso é necessário analisar através de um gráfico ao longo do tempo a diferença entre um dado sensor e um sensor de referência, ou a diferença entre o valor do sensor e um outro valor do mesmo sensor em um tempo de referência.

Para realizar esses cálculos, foi criado um programa especialista que recebe parâmetros do usuário, conforme descrito na Seção 4.5.3. Esses parâmetros são o tamanho do período no qual as entradas devem ser agrupadas, os períodos para os quais os dados devem ser descartados caso seja necessário, e se deve ou não calcular o desvio a partir do primeiro valor encontrado.

A interface de busca do Glance para essa aplicação ainda faz com que os dados gerados por esse programa especialista passem por uma etapa de operação genérica conforme a mostrada na Seção 4.5.2. Nessa etapa, o usuário pode, por exemplo, calcular a diferença entre o sinal de um sensor e outro de referência.

A Figura 5.2 apresenta a interface para os dados do Survey. Na parte da esquerda da figura, está a interface de busca do Glance com duas caixas onde os parâmetros para a operação específica do Survey e a genérica são especificadas. Ao submeter a busca, o Glance recupera os dados do banco de dados e passa para o programa de processamento específico ao Survey. O resultado desse é então passado para o programa de operações genéricas e, a seguir, para uma transformação não parametrizada para geração do gráfico.

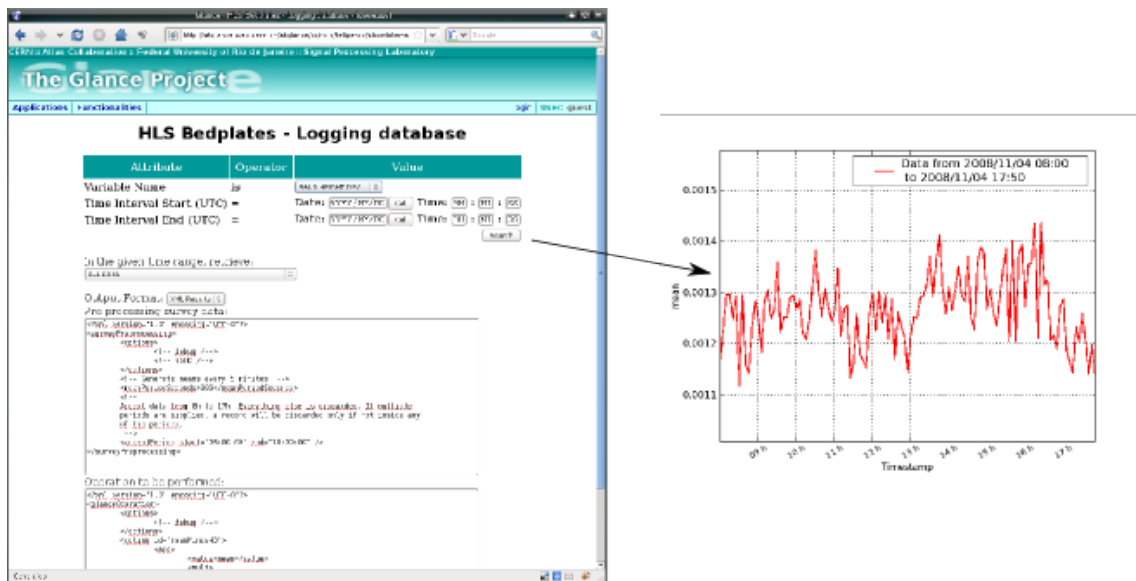


Figura 5.2: Interface para os dados do Survey. Na esquerda, a interface do Glance com os parâmetros para as operações parametrizadas. Na direita, o resultado na forma de gráfico.

Capítulo 6

Conclusão

O LHC, em operação no CERN, é o maior acelerador de partículas do mundo. O ATLAS, o maior dos seus detectores, foi construído e é operado por uma colaboração internacional envolvendo 172 institutos em 37 países. Nesse ambiente, dados referentes à instalação, resultados de testes ou monitoração do desempenho dos milhares de equipamentos, são armazenados em bancos de dados usando modelagens específicas para cada caso e até tecnologias de armazenamento diferentes, tais como banco de dados Oracle ou MySQL. Como a colaboração envolve pessoas inseridas em diferentes culturas, pode também acontecer de haver diferentes terminologias para um mesmo equipamento. Além disso, se uma busca é realizada sobre todos os dados da colaboração, é grande a probabilidade de não trazer resultados úteis por causa de possíveis erros de digitação ou pelo grande volume de dados retornados. Por outro lado, se cada grupo da colaboração criar um sistema de recuperação para os dados de interesse, será necessário um grande esforço de manutenção para que os sistemas continuem funcionando durante toda a operação do detector, que será de, pelo menos, 20 anos.

Para solucionar esses problemas de recuperação de dados, foi desenvolvido o Glance. O sistema é baseado em interfaces de buscas independentes, de forma que um único sistema pode manipular diversas interfaces. O custo de manutenção é, portanto, diminuído pois há somente um sistema para ser mantido. A descrição das interfaces de recuperação não é dependente de uma determinada tecnologia de armazenamento, e o sistema foi projetado de forma que possa recuperar dados de diferentes repositórios. Além disso, diferentes perspectivas sobre o mesmo conjunto

de dados podem ser exibidas, criando interfaces para cada caso de interesse.

O sistema Glance pode ser acessado através da Web, o que permite que seja usado pelos colaboradores de qualquer parte do mundo. O sistema auxilia a realização de buscas em bancos de dados. Para tal, apresenta opções para o usuário escolher, minimizando os problemas devidos a erros de digitação e terminologia dos atributos. Durante o processo de criação de uma interface, o sistema apresenta a estrutura do repositório, refinando sucessivamente os detalhes. Adicionalmente, as interfaces de recuperação criadas são paramétricas.

O mecanismo de operações do Glance permite que os dados recuperados sejam processados, e o resultado do processamento é exibido para o usuário final. As transformações a serem realizadas podem genéricas, no sentido de poderem ser usadas em diferentes contextos, tais como transformar uma série temporal de valores numéricos em um gráfico, ou específicas a uma determinada necessidade, tais como a realização do processo de *unsmoothing* do DCS. Dessa forma, os resultados da recuperação chegam ao usuário no formato adequado para análise.

O sistema Glance está instalado nos servidores Web do CERN e vem sendo usado ativamente pela colaboração. Dentre suas aplicações a casos reais da colaboração, destacam-se a interface ao banco de dados ATLASIntegration, o monitoramento do alinhamento dos componentes com o ATLAS Survey e a monitoração de sensores para o DCS do TileCal. Para o banco de dados ATLASIntegration, o sistema auxilia no processo de instalação dos componentes do detector recuperando dados sobre posição dos equipamentos, disposição dos cabos e os estados de conectividade entre cabos e eletrônicos. As aplicações para o ATLAS Survey e para o DCS do TileCal recuperam do banco de dados amostras de sensores e, através do mecanismo de processamento, calculam diferenças entre as leituras de sensores correlacionados, extraem médias e desvio padrão dos valores e gera gráficos. Os resultados são usados, no caso do ATLAS Survey, para detectar desníveis no chão da caverna experimental, e, no caso do DCS, para diagnosticar problemas nas fontes de alimentação do TileCal.

O sistema também foi aplicado ao banco de dados LHCb Integration, cujos requisitos são muito parecidos com os do ATLAS Integration, mas registra equipamentos do LHCb, que é outro detector de partículas do LHC. Isso mostra que,

devido ao fato do Glance ter sido projetado para ser genérico, atende aos requisitos de recuperação de outras colaborações dentro do CERN, ou até em contextos diferentes.

Atualmente existem no sistema mais de 400 interfaces de busca. Dessas, 150 são destinadas a recuperação de informações de equipamentos para aplicações do ATLAS Integration e LHCb Integration, enquanto 25 são utilizadas pelo sistema DCS do TileCal, e 2 pelo o ATLAS Survey. As demais interfaces são utilizadas por outros sistemas da Coordenação Técnica do ATLAS.

Dentre os possíveis futuros passos para o sistema, destacam-se: a integração de dados correlacionados, mas armazenados em repositórios distintos; melhorias no desempenho da recuperação e do processamento dos dados; criação de um isolamento entre as aplicações do Glance, de forma que novas funcionalidades possam ser introduzidas no sistema sem impactar as outras aplicações, caso haja alguma falha de programação.

O Glance também recebeu contribuições de outros alunos participantes da colaboração entre a UFRJ e o ATLAS, que desenvolveram funcionalidades além das apresentadas neste documento. Dentre elas, destacam-se a inserção de dados utilizando o conceito de interfaces de inserção, análogo ao da interface de busca, realizado por Kaio Karam, e o esquema de categorização hierárquica das interfaces armazenadas, para facilitar o acesso, realizado por Cimar Massulo.

Referências Bibliográficas

- [1] “CERN”, <http://cern.ch/>, acessado em novembro de 2009.
- [2] “The Large Hadron Collider”, <http://public.web.cern.ch/public/en/LHC/LHC-en.html>, acessado em novembro de 2009.
- [3] “ATLAS Photos”, <http://www.atlas.ch/photos/index.html>, acessado em novembro de 2009.
- [4] “LHC First Beam”, <http://cern.ch/lhc-first-beam>, acessado em Março de 2009.
- [5] *ATLAS detector and physics performance: Technical Design Report, 1*, Technical Design Report ATLAS. Geneva, CERN, 1999. Electronic version not available.
- [6] “The ATLAS Experiment”, <http://atlas.ch/>, acessado em novembro de 2009.
- [7] *ATLAS technical coordination: Technical Design Report*, Technical Design Report ATLAS. Geneva, CERN, 1999. Electronic version not available.
- [8] “Detector Control System”, <https://twiki.cern.ch/twiki/bin/view/Atlas/DetectorControlSystem>, acessado em novembro de 2009.
- [9] POY, A. B., BOTERENBROOD, H., BURCKHART, H. J., *et al.*, “The detector control system of the ATLAS experiment”, *Journal of Instrumentation*, v. 3, n. 05, pp. P05006, 2008.
- [10] LASSEUR, C., *Placement strategy and survey in Atlas*, Report ATL-TECH-PUB-2008-001. ATL-COM-TECH-2008-003, CERN, Geneva, Mar 2008.

- [11] SCHILDT, H., *C++: The Complete Reference, 4th Edition*. 4 ed. McGraw-Hill Osborne Media, novembro 2002.
- [12] “Python Programming Language”, <http://www.python.org/>, acessado em novembro de 2009.
- [13] “ROOT: A Data Analysis Framework”, <http://root.cern.ch/>, acessado em novembro de 2009.
- [14] “Gartner Says Worldwide Relational Database Market Increased 14 Percent in 2006”, <http://www.gartner.com/it/page.jsp?id=507466>, acessado em dezembro de 2009.
- [15] “Extensible Markup Language”, <http://www.w3.org/TR/xml11>, acessado em novembro de 2009.
- [16] “Xerces-C++ XML Parser”, <http://xerces.apache.org/xerces-c/>, acessado em novembro de 2009.

Apêndice A

Publicações

- Tile DCS Web System MAIDANTCHIK, C. L. L., FERREIRA, F. G., GRAEL, Felipe F. Computing in High Energy Physics, 2009, Prague.

The web system described here provides functionalities to monitor the Detector Control System (DCS) acquired data. The DCS is responsible for overseeing the coherent and safe operation of the ATLAS experiment hardware. In the context of the Hadronic Tile Calorimeter Detector, it controls the power supplies of the readout electronics acquiring voltages, currents, temperatures and coolant pressure measurements. The physics data taking requires the stable operation of the power sources.

The DCS Web System retrieves data automatically and processes it extracting the statistics for given periods of time. The mean and standard deviation outcomes are stored as XML files and are compared to preset thresholds. Further, a graphical representation of the TileCal barrels indicates the state of the supply system of each detector drawer. Colors are designated for each kind of state. This way problems are easier to find and the collaboration members can focus on them. The user can pick a module to see detailed information. It is possible to check the statistics and generate charts of the parameters over the time.

The DCS Web System also gives information about the power supplies latest status. The barrel colors green whenever the system is on. Otherwise it is colored red. Furthermore, it is possible to perform customized analyses. It provides search interfaces where the user can set the module, parameters, and

the time period of interest. Moreover the system produces the output of the retrieved data as charts, XML files, CSV and ROOT files according to the user's choice.

- Glance project: a database retrieval mechanism for the ATLAS detector In: Computing in High Energy Physics, 2007, Victoria. Maidantchik, C, Graef, F F, GALVÃO, K K, Pommès, K Journal of Physics: Conference Series. , 2008. v.119.

During the construction and commissioning phases of the ATLAS Collaboration, data related to the installation, testing and performance of the equipment are stored in relational databases. Each group acquires information and saves them in repositories placed in different servers, using diverse technologies. Both data modeling and terminology may vary among the storage areas. The development of retrieval systems for each data set would require too much effort and high maintenance cost. The goal of the Glance Project is to provide navigation mechanisms among the databases, which is independent of both the technology used to build them and their relationship. The browsing over the data sets results in hypertext tables and links on the Web. The user chooses the database and the system shows its structure. After selecting a partition of the repository, the system automatically creates a retrieval interface that allows the specification of search parameters. The interface can be customized for specific needs and saved in order to be easily accessed later. Glance corresponds to a single system that handles distinctive recovery mechanisms among diverse databases. Therefore, further knowledge about how data is organized and labeled is not required to perform queries. Maintenance costs are also minimized. This paper describes the Glance conception, its development and functionalities. The system usage is illustrated with some examples. Current status and future work are also discussed.

- Web system to support analysis of the Tile Calorimeter commissioning In: Computing in High Energy Physics, 2007, Victoria. Maidantchik, C, Faria, A, Graef, F F, Ferreira, F G, GALVÃO, K K, Dotti, A, Solans, C, Price, L Journal of Physics: Conference Series. , 2008. v.119.

During the ATLAS detector commissioning phase, installed readout electronics must pass performance standards tests. The resulting data must be analyzed to ensure correct operation. For the Tile Calorimeter, developers plug their code into a specific framework for physics data-processing. Collaboration members, taking shifts on commissioning work, interpret the results, in thousands of readout channels, to identify potential problems that may need correction during commissioning. The Tile Commissioning Web System (TCWS) facilitates the repetitive data analysis and quality control by encapsulating all necessary steps to retrieve information, execute programs, access the outcomes, register statements, and verify the equipment status. TCWS integrates different applications, each presenting a particular view of the commissioning process. The TileComm Analysis application stores plots and analysis results, provides equipment-oriented visualization, collects information regarding equipment performance, and summarizes its status. The Timeline application provides equipment status history in a chronological way. The Web Interface for Shifters application supports monitoring tasks by managing test parameters, graphical views of the calorimeter performance, and information status of all equipment that was used in each test. Finally, equipment quality control data can be filled, stored, modified, and retrieved as hypertext forms through the ATLASMonitor application. These applications are also connected with other commissioning programs that allow an automatic gathering of the commissioning data. This paper describes in detail the programs that compose the TCWS and how they are integrated within the Tile Calorimeter commissioning. Current status and future work are also discussed

- Management of Equipment Databases at CERN for the ATLAS Experiment
In: 10th ICATPP Conference on Astroparticle, Particle, Space Physics, Detectors and Medical Physics Applications, 2007, Villa Olmo. POMMES, K., GALVÃO, K K, MOLINA-PEREZ, J., Graef, F F, MAIDANTCHIK, C. L. L. 10th ICATPP Conference on Astroparticle, Particle, Space Physics, Detectors and Medical Physics Applications. , 2007.

The ATLAS experiment apparatus is divided into major components, composed by numerous single equipment devices and a back end electronics. A

complex cabling scheme connects the machinery. The detector has large dimensions and a complex geometry with components in several different shapes. Displacements and deformation are important in data analysis and calculus. Furthermore, nuclear and radiation control rules state the traceability of all the apparatus components. Graphical visualization systems were deployed to address and identify the several parts and manage displacements and deformation. Equipment Databases were designed to support the trace of objects composing the structure. User interfaces designed for portable devices facilitate data updates. A universal mechanism allows the management of the distinct information stored, providing means of data access and update by users and other applications. Various aspects of equipment installation and maintenance are addressed and supported by distinct software components, with heterogeneous information integrated by one universal tool.

- Sistema Glance para recuperação e operação de dados de repositórios heterogêneos para o ATLAS MAIDANTCHIK, C. L. L., GRAEL, Felipe F., GALVÃO, K K, EVORA, L. H. R. A. XXIX Encontro Nacional de Física de Partículas e Campos, 2008.

O detector ATLAS é construído por uma colaboração internacional heterogênea. Durante as etapas de construção, teste e integração dos componentes, dados sobre a construção e desempenho dos equipamentos foram gerados. No entanto, a recuperação desses dados envolve conhecer detalhes técnicos das tecnologias utilizadas para o armazenamento, a organização dos repositórios para que um método de recuperação seja programado.

Frente às dificuldades de recuperação de informações, foi desenvolvido na UFRJ, em colaboração com o ATLAS, o Glance, um sistema universal capaz de reconhecer a estrutura dos repositórios, auxiliar o usuário na criação de uma interface de recuperação e realizar buscas sobre os dados. As interfaces criadas são descritas em uma linguagem intermediária independente da tecnologia do repositório. Dessa forma um único sistema gerencia diversas interfaces independentes.

Ao longo da fase de comissionamento surgiu o requisito de processar os dados

recuperados para serem melhor interpretados pelos colaboradores. Para supor-
tar essa nova necessidade, foi proposto e implementado um mecanismo capaz
de realizar operações tais como calcular médias ou aplicar fórmulas envolvendo
atributos de uma pesquisa.

Com essa funcionalidade, o sistema reconhece o tipo dos dados que serão ope-
rados, que podem ser numéricos ou textuais, e apresenta operações adequadas.
Ao executar a busca, o Glance passa para um módulo intermediário os resulta-
dos, descritos em um formato intermediário, e a operação definida pelo usuário.
Pela existência do módulo intermediário, a operação é independente do tipo
de banco de dados. Os resultados são apresentados como uma nova coluna
ou como um sumário contendo, por exemplo, a média de uma coluna. Além
disso, os dados podem ser retornados na forma de uma tabela hipertextual,
como gráficos ou até em outros formatos como CSV ou XML.

Esse esquema está sendo aplicado ao Detector Control System (DCS) do ca-
lorímetro de telhas. O DCS monitora e registra em um banco de dados ten-
sões, correntes e temperaturas em determinados pontos da eletrônica do sub-
detector. Através da nova funcionalidade, o Glance calcula a diferença de
voltagem entre dois pontos e também a média e o RMS das medições sobre
um determinado período de tempo. Outra aplicação semelhante é junto ao
grupo que mede o posicionamento dos equipamentos no ambiente experimen-
tal e monitora desvios entre a posição teórica e a observada em determinados
pontos do detector.

Este trabalho apresenta a funcionalidade de operações sobre dados do Glance
e demonstra sua utilização. Também são discutidas em mais detalhes as apli-
cações existentes desse mecanismo.

- O sistema Glance para recuperação e inserção de dados em repositórios hetero-
gêneos MAIDANTCHIK, C. L. L., GRAEL, Felipe F., GALVÃO, K K XXVIII
Encontro Nacional de Física de Partículas e Campos, 2007.

Os componentes do detector de partículas ATLAS são construídos por uma
colaboração internacional geograficamente dispersa. A montagem de suas
dezenas de milhares de componentes está sendo finalizada no CERN, onde

entrará em operação. Uma vez integrados, os equipamentos são testados em conjunto. As informações de localização e conectividade de cada parte, bem como os resultados dos testes são armazenados em diferentes repositórios, utilizando diversas tecnologias, modelagens e terminologias.

Para solucionar os problemas de recuperação de dados heterogêneos, inerentes a esse ambiente, foi desenvolvido um sistema universal capaz de realizar buscas sobre os repositórios independentemente de suas tecnologias. Ao serem fornecidas as informações de como se conectar em um repositório, o Glance reconhece sua estrutura interna, e permite a seleção do conjunto de dados de interesse aumentando sucessivamente o nível de detalhes apresentados. O sistema então reconhece os atributos e seus tipos e gera automaticamente a descrição de uma interface de recuperação correspondente. A descrição é feita em uma linguagem de marcação baseada em XML e contém o nome dos atributos, as informações técnicas necessárias para localizá-los, seus tipos, descrições e possíveis valores. Baseado nessa descrição, o sistema gera uma interface de busca paramétrica, onde devem ser estabelecidos os atributos sobre o qual se quer realizar a pesquisa, os operadores e os valores de referência. Os operadores são sensíveis ao tipo do atributo, isto é, um atributo numérico possui os operadores “maior que” e “menor que”, enquanto um textual possui “contém”. Os resultados da busca são apresentados na forma de uma tabela hipertextual, ou como arquivos em formatos tais como CSV ou XML.

Ao instalar ou remover um componente do detector, a equipe de coordenação técnica precisa atualizar as informações de conectividade e localização nos bancos de dados. Para dar suporte a essa tarefa, o sistema permite também a criação de interfaces para inserção e modificação de dados. Antes de realizar qualquer alteração no repositório, o Glance verifica a integridade dos novos dados, evitando inconsistências. O sistema também é usado pelo grupo do calorímetro hadrônico, para recuperar dados do DCS (Detector Control System) que monitora tensões, correntes e temperaturas. Para isso, o Glance realiza operações pre-definidas sobre os resultados da busca, de acordo com o tipo de análise a ser realizada.

Este trabalho apresenta a concepção do Glance e demonstra sua utilização.

Também é mostrado em mais detalhes como o sistema é utilizado pela coordenação técnica e pelo calorímetro hadrônico.

- Sistema de Análise e Monitoração dos dados não-físicos do Calorímetro Hadrônico do detector ATLAS MAIDANTCHIK, C. L. L., FERREIRA, F. G., GRAEL, Felipe F. XXVIII Encontro Nacional de Física de Partículas e Campos, 2007.

O calorímetro hadrônico do detector ATLAS encontra-se em fase de comissionamento. Durante este período são testados os equipamentos e sistemas que o compõem a fim de garantir o seu futuro funcionamento. O Laboratório de Processamento de Sinais (LPS-UFRJ) é um dos laboratórios participantes desta colaboração internacional.

O DCS (*Detector Control System*), responsável pela monitoração das fontes de alta e baixa tensões que alimentam os componentes eletrônicos integrantes do sub-detector e pelo sistema de refrigeração que garante o funcionamento dessas fontes, é um desses sistemas. Para tal, adquire dados como tensões, correntes e temperaturas utilizando um programa desenvolvido dentro do framework do software proprietário PVSS, que é uma plataforma destinada especialmente para sistemas de controle. Após a aquisição tais informações são armazenadas numa base de dados relacional de tecnologia Oracle. Estes dados não são usados para análises físicas, mas são importantes para garantir a consistência delas.

Para realizar o monitoramento e análise desses dados foi proposto e implementado um sistema capaz de disponibilizar as informações via internet que possam garantir que as fontes estejam funcionando corretamente, e a este foi dado o nome de *DCS Web System*. Primeiramente os dados são recuperados da base de dados utilizando-se o sistema Glance, também desenvolvido pelo chamado grupo do Rio. Após a recuperação dos dados, o sistema *Glance* formata-os em arquivos XML, garantindo assim flexibilidade, e através desses gera médias e os desvio padrão das amostragens, dados suficientes para estudar a estabilidade dos equipamentos, para um dia ou um mês. Ainda são disponibilizados gráficos temporais que mostram o progresso de determinada

fonte. Ainda são previstos no projeto distribuições estatísticas dos dados em estudo.

Atualmente, o *DCS Web System* oferece o suporte à análise e ao monitoramento das fontes de baixa tensão e ao sistema de refrigeração, sendo o suporte às fontes de alta tensão uma atualização prevista e pelo sistema de refrigeração que garante o funcionamento dessas fontes, é um desses sistemas. Para tal, adquire dados como tensões, correntes e temperaturas utilizando um programa desenvolvido dentro do framework do software proprietário PVSS, que é uma plataforma destinada especialmente para sistemas de controle. Após a aquisição tais informações são armazenadas numa base de dados relacional de tecnologia *Oracle*. Estes dados não são usados para análises físicas, mas são importantes para garantir a consistência delas.

- Glance: a retrieval system for construction, installation and commissioning data for ATLAS equipments MAIDANTCHIK, C. L. L., Graef, F F, SEIXAS, J. M. XXVII Encontro Nacional de Física de Partículas e Campos, 2006.

The ATLAS detector is being constructed by an international collaboration for the LHC experiment at CERN. For each part a big amount of data related to its construction, installation and commissioning is being generated. Since each part is being constructed by different institute, the data are being stored at different places and modeled according to the specific needs of each part and institute.

During the current phase of installation and commissioning each working group needs to have access to the data related to their equipments, which can involve parts coming from different institutes. Different groups may be interested in different perspectives of the data. For example, the commissioning team of a sub-detector is interested in the installation and production data of their electronics and cables. Other group, like the cabling team, is interested in the data about installation and connectivity for all subsystems, but not about the construction.

In this context of distributed working groups it is very difficult to provide a single interface that delivers the information suitable for every group. In the

other hand, requiring each group to develop their own solution would need a great effort. This would also represent a high maintenance cost, since small changes to the repository of an institute can reflect into adjustments of multiple interfaces.

The goal of the project is to facilitate the retrieval of the data of construction, installation and commissioning for the equipments in an integrated way. The user selects the set of data of his interest. The system then recognizes the characteristics of the data and automatically generates a corresponding retrieval interface. This interface then can be stored in order to be accessed by the collaborators from the same group. Through the retrieval interface, the user can perform searches and filter the results successively to reach the desired data. The system stores the description of each interface, with the information needed to perform the retrieval, in a way that they can be added or modified as needed, without requiring changes to the main program.

The system is being actively used by the Technical Coordination group to retrieve data about cabling, positioning and physical installation of electronic crates and boards, and integrate these data with an inventory database, and with the production database of some of the sub-detectors. It is running on the CERN web servers, and is developed using C++ and XML technologies.

- Sistema WEB para Testes de Equipamentos em Física de Altas Energias MAI-DANTCHIK, C. L. L., SEIXAS, J. M., ALVES, A. M., FARIA, A., GRAEL, Felipe F., FERREIRA, F. G., GALVÃO, K K XXVII Encontro Nacional de Física de Partículas e Campos, 2006.

O detector ATLAS, acoplado ao acelerador de partículas LHC do CERN, encontra-se atualmente em fase de comissionamento. Os testes realizados geram uma enorme quantidade de dados, que são posteriormente analisados pelos colaboradores em diferentes países. A cada execução dos programas de análise, uma série de procedimentos e configurações deve ser realizada pelos pesquisadores. Os gráficos e histogramas resultantes se referem aos níveis de energia durante uma colisão de partículas, e dados para o controle de qualidade dos equipamentos também são gerados a partir das análises.

Este projeto apresenta o sistema *Tile Commissioning Web System* que apóia a manipulação e análise dos dados provenientes dos testes realizados no Calorímetro de Telhas (TileCal), um dos sub-detectores do ATLAS, e apresenta ferramentas para a recuperação dos resultados obtidos. O sistema é composto por três softwares com interface Web que possuem funções específicas: o *Web Interface for Offline Shifters*, o *Tilecomm Analysis* e o *AtlasMonitor*.

O *Web Interface for Offline Shifters* (WIS) automatiza o processo de análise, apresentando ao usuário uma tabela com todos os testes realizados e os tipos de análises que podem ser realizadas para cada um. Após a seleção do usuário, o WIS recupera o arquivo correspondente no sistema de armazenamento CASTOR (*Cern Advanced STORAGE*) e executa remotamente o programa de análise requerido, seguindo todos os procedimentos e configurações exigidos. Ao final do processo de análise, os resultados são disponibilizados na interface do sistema, apresentando gráficos de níveis de energia e dados de controle de qualidade. Os gráficos e os dados são automaticamente armazenados nas respectivas bases de dados dos sistemas *Tilecomm Analysis* e *Atlas Monitor*. O primeiro software recupera os gráficos, através da associação com o tipo e o identificador do teste e com a seção do sub-detector que foi testada. O *Atlas-Monitor* insere automaticamente os dados resultantes nas folhas de controle de qualidade. Posteriormente, o colaborador pode inserir informações adicionais ou até mesmo criar uma nova folha de controle de qualidade.

O sistema *Tile Commissioning Web System* está instalado no servidor do CERN e é utilizado pelos colaboradores do TileCal e seu desenvolvimento conta com a participação dos responsáveis tanto pelos testes dos equipamentos quanto pelo funcionamento do calorímetro.