



USP  
REITORIA  
Departamento  
de Informática

## **Projeto de Distribuição de Dados Corporativos para Unidades da USP**

Março/2009



## Objetivo

Permitir às Unidades, através do recebimento sistemático de dados, desenvolver esforços locais de informatização, complementares aos Sistemas de Informação Corporativos, para fins de apoio aos processos gerenciais bem como estimular a elaboração de relatórios internos para aferição de índices de atividade e de produtividade.

## Tecnologia

Em operação desde fevereiro de 1999, a tecnologia adotada é a de Replicação, para disponibilizar os dados para os sites locais das Unidades, a partir do site Corporativo da USP.

A Replicação de Dados é um processo assíncrono, que assegura uma reduzida contenção de recursos computacionais, garantindo, entretanto, uma contínua atualização de dados para os sites locais que os requisitem, com baixa latência.

O recebimento dos dados corporativos pelas Unidades demanda a aquisição de um Sistema de Gerenciamento de Banco de Dados para cada site local. Este Sistema de Banco de Dados deve ser compatível com o **Sybase Replication Server**, sistema gerenciador encarregado da distribuição dos dados, a partir da base corporativa da Universidade.

## Crítérios de Segurança, Técnicos e Operacionais

Um site local é entendido como um "nó" físico de rede, numa Unidade da USP, capacitado a receber dados provenientes de Replicação.

A habilitação de um site local, para participar da estratégia de distribuição de dados, deve satisfazer um conjunto de requisitos que viabilizem tecnológica e administrativamente a distribuição física dos dados.

Foram estabelecidos alguns critérios com o objetivo de minimizar impactos de segurança e de desempenho junto ao site corporativo; e outros, por sua vez, voltados à garantia das restrições de uso e de visibilidade dos dados nos ambientes locais.



## ***Diretrizes Operacionais e Técnicas***

Devido ao caráter corporativo e integrado do projeto de distribuição de dados, seguem algumas regras e procedimentos a serem seguidas:

1. **IMPORTANTE:** O início de recebimento dos dados por qualquer site local, pressupõe a ciência e concordância da instância central responsável pelos mesmos. Para isso o Diretor da Unidade deve encaminhar uma solicitação à CODAGE (no caso de dados do Marte\* e/ou Mercúrio\*) e às Pró-reitorias responsáveis (Graduação, Pós-graduação e/ou Cultura e Extensão, conforme o caso); \* vide item 9
2. O site local será de responsabilidade exclusiva e total das Unidades, desde os investimentos para aquisição de softwares e equipamentos, e a operação e manutenção destes, até o desenvolvimento de quaisquer aplicativos sobre os dados recebidos do ambiente corporativo.
3. O Departamento de Informática terá a responsabilidade e a gerência sobre os critérios de distribuição de dados, e a definição e administração da tecnologia envolvida para a distribuição dos dados;
4. Haverá uma camada intermediária, entre o site corporativo e os sites locais, composta por um Sistema Gerenciador de Banco de Dados, operando numa máquina específica e distinta, para organização, avaliação e aplicação de critérios, de maneira a normatizar os processos de replicação de dados;
5. O sentido da distribuição de dados será único: do site corporativo para os sites locais;
6. Cada site local possuirá acesso restrito aos dados que compõem seu domínio de atuação e de trabalho, e total aos dados considerados de domínio público;
7. As cláusulas de requisição de dados ("Subscriptions") pelos ambientes locais serão construídas e definidas no site central, com as respectivas restrições lógicas, para garantir uso e acesso adequados aos dados;
8. O Departamento de Informática disponibiliza os Modelos de dados que compõem o esquema de replicação, através de diagramas e descrições dos objetos envolvidos para facilitar ao máximo a compreensão das informações enviadas;
9. O conjunto de objetos replicados, oferecidos às Unidades, são agrupados de acordo com os contextos dos sistemas: Fênix, Júpiter, Apolo (Cultura e Extensão Universitária), parcial do Marte (somente dados pessoais e funcionais atuais), parcial do Mercúrio (patrimônio e estoque do almoxarifado) e Lattes, sempre respeitando a restrição dos



dados da própria Unidade. Além disso, também são disponibilizados dados de catraca para controle de acesso local.

**Esclarecimento importante:** o contexto dos dados de uma Unidade inclui todos os dados que a afetam de alguma forma. Por exemplo, quando se trata dos dados acadêmicos, recebe tanto os dados de turmas, docentes, alunos, matrículas, etc, da própria unidade e das outras unidades onde 'seus' alunos tem algum vínculo, assim como recebe os dados de todos os alunos de outras unidades que cursam as 'suas' disciplinas.

Existem também as tabelas consideradas públicas, como por exemplo Disciplinas, Cursos, Localidade, Colegiado e outras que são distribuídas integralmente para todos.

10. As tabelas replicadas deverão ser usadas apenas para consulta, pois a atualização do seu conteúdo ficará, exclusivamente, por conta do processo de replicação de dados. Além disso, essas tabelas não poderão ter seus esquemas alterados arbitrariamente, sob pena de não funcionamento das estratégias de replicação e de conseqüente interrupção dos serviços de distribuição para a Unidade;

11. Os esquemas (scripts) de criação e atualização dos objetos distribuídos serão enviados pelo Departamento de Informática, através do endereço eletrônico [replicacao@usp.br](mailto:replicacao@usp.br), para um mail-list ([di\\_repunidade@listas.usp.br](mailto:di_repunidade@listas.usp.br)) assinado pelos administradores dos sites locais, que deverão retornar e-mail confirmando sua aplicação.

**Observação:** o endereço [di\\_repunidade@listas.usp.br](mailto:di_repunidade@listas.usp.br) foi criado também com a finalidade de lista de discussão, permitindo aos usuários das unidades compartilharem conhecimentos, dúvidas ou experiências referentes à replicação.

12. Sempre que o Departamento de Informática solicitar ao pessoal técnico do Site Local, a aplicação de procedimentos sobre algum objeto, haverá interrupção do fornecimento de dados para o respectivo site, até que seja efetuada a manutenção solicitada;

13. O Banco de Dados local poderá ter novos objetos (tabelas, índices,...) criados pelos responsáveis locais, para atender as necessidades específicas da Unidade, desde que não infrinjam qualquer dos itens anteriores, e respeitem as seguintes exigências:

- não criar referências explícitas (*foreign key constraints*) a chaves candidatas das tabelas replicadas e,
- não criar índices com cláusula de unicidade sobre qualquer atributo das tabelas replicadas

14. Dúvidas relativas ao modelo de dados corporativo, aos critérios de uso e de visibilidade de dados e procedimentos técnicos operacionais sobre o banco de dados local, devem ser endereçoadas ao Serviço de Administração do Ambiente Distribuído, sempre através do e-mail [replicacao@usp.br](mailto:replicacao@usp.br);



## Requisitos Técnicos Básicos

Antes de proceder à escolha e a instalação de uma solução tecnológica para o Ambiente do Site Local, recomenda-se consultar o pessoal técnico do Departamento de Informática para uma avaliação precisa do custo, da complexidade e da viabilidade da alternativa apontada.

### 1. Plataforma de Hardware

A configuração específica da máquina (processador, memória, disco,...) deverá ser estimada em função do volume dos dados a serem recebidos, da perspectiva de uso local, dos recursos complementares, e também das exigências dos serviços de dados e de distribuição adquiridos pela Unidade.

Preferencialmente, esta plataforma deve estar dedicada apenas aos serviços de recebimento dos dados corporativos, via Replicação, e de fornecimento de dados para os sistemas locais.

Quanto à conexão de rede, a tecnologia de Replicação, dado seu caráter assíncrono, pode conviver com interrupções de comunicação com o site local, desde que eventuais e de curta duração. Idealmente, o site local deve estar preparado para uso ininterrupto.

#### Configuração Mínima de Referência

- Processador – 1.5 GHz (mínimo)
- Memória RAM – 1 GB (mínimo)
- Disco rígido – 4 GB (mínimo). Sugestão: 2 HDs para separar o armazenamento de Log (1 GB mínimo) e de Dados (3 GB mínimo)

### 2. Software Básico e Gerenciador de Banco de Dados Relacional

Para o Sistema Operacional, os softwares compatíveis são:

- Windows (XP, Vista ou Server)
- Unix (Solaris)
- Linux

Os Gerenciadores de banco de dados compatíveis são:

- Sybase Adaptive Server Enterprise (versão mínima: ASE 12.5.3)
- Sybase ASE Express Edition for Linux (Free) - <http://www.sybase.com/linuxpromo>

Obs: A versão Free do ASE para Linux é limitada para uso de uma CPU, 5Gb de área em disco e 2Gb de RAM por instância (sem suporte técnico da Sybase).

- Sybase Adaptive Server Anywhere (versão mínima: ASA 9) e respectivo Gateway
- Microsoft SQLServer 2005 ou SQLServer 2008 e respectivo Gateway
- ORACLE e respectivo Gateway



Para outros gerenciadores, recomenda-se consultar o Departamento de Informática.

Obs: Os Gateways's para conexão com as Unidades são de responsabilidade do DI.

### 3. Recomendações iniciais para instalar o Gerenciador e criar o Banco de Dados

Na instalação do gerenciador, é necessário fazer alguns ajustes na configuração default do servidor:

1. Sort order tem que ser case-sensitive, compatível com ISO 8859-1(Latin-1)
2. Language = US English
3. Aumentar o número de conexões
4. Aumentar o número de locks (valor mínimo sugerido: 30.000)

Criar o Banco de Dados com no mínimo 3Gb para área de dados e 1Gb para área de log, em discos diferentes. Sugere-se adotar a convenção de nome do banco de dados igual à Sigla da Unidade.

É preciso criar um usuário com nome **dbmaint**, com privilégio de criar tabelas no Banco de Dados (alias de dbo), pois todos os scripts enviados pelo Departamento de Informática deverão ser submetidos através do usuário **dbmaint**.

Deve-se configurar uma profile para que o Banco possa ser aberto e utilizado. Sugere-se colocar o Banco de Dados recém-criado como **default**, no *statement* de inicialização do Servidor.

Uma vez configurado o ambiente do site local, informar à Seção de Banco de Dados do Departamento de Informática (através do e-mail [replicacao@usp.br](mailto:replicacao@usp.br)) os seguintes dados:

nome do Servidor,  
nome do Banco de Dados,  
IP,  
Port (TCP/IP),  
senha do usuário dbmaint  
e outras características do Site Local

Observação: Após criar o banco de dados, recomenda-se ativar a opção de *truncate log on checkpoint*, até o término da carga inicial da replicação. Depois que o processo de replicação inicial terminar, convém retirar essa opção, e implementar a política de backup's tanto de dados quanto de transações.



#### 4. Recomendações para *backups* nos sites locais

É recomendável manter uma política de ‘backup’ diário do Banco de Dados (*dump database*) e um ‘backup’ de transações (*dump tran*) pelo menos duas vezes ao dia: no início e no final.

E se ocorrer um problema, dependendo da situação, podem ser adotadas várias soluções para a reconstrução e sincronização dos Bancos replicados, mas é essencial que a Unidade entre em contato com o Departamento de Informática, para orientação sobre o melhor procedimento para restaurar e sincronizar o banco de dados local.

Quando ocorrer algum problema no servidor local, e a replicação ficar parada, as transações ficarão na fila de replicação do Banco de Dados central.

Se assim orientada, a Unidade deverá fazer a restauração do seu ambiente, aplicando os ‘backups’: do banco inteiro (*load database*) e das transações (*load tran*) na ordem cronológica.

A sincronização será realizada pelo DI, que poderá ser parcial ou total, de acordo com o caso.

#### 5. Pessoal dedicado a administrar o site local e os serviços de dados e de distribuição

Como a gerência e a responsabilidade pelo site local são da própria Unidade, deve haver pessoal qualificado, capaz de conduzir as diretrizes tecnológicas do projeto de distribuição e assistir os usuários locais em suas necessidades de Serviços de Dados.

Muitas tarefas, decorrentes da forte interação entre o site local e o Departamento de Informática, na administração do ambiente distribuído, pressupõem qualificação técnica e conhecimentos de procedimentos básicos de administração de máquina e administração de banco de dados.