

C. 525



RECHERCHES
INFORMATIQUES

PROCEDURES DE REPRISE DANS LES SYSTEMES DE GESTION
DE BASES DE DONNEES DISTRIBUEES

BY

P. BOUCHET

JUIN 1978

Rapport de Recherche n° 15

EQUIPE DE RECHERCHE ASSOCIEE AU C.N.R.S. n° 452

"AL KHOWARIZMI"

LABORATOIRE DE RECHERCHE EN INFORMATIQUE, Bât. 490

Université de Paris-Sud - Centre d'Orsay - 91405 ORSAY

Université de Paris-Sud

Département de Mathématique

Bât. 425

91405 **ORSAY** France

PROCEDURES DE REPRISE DANS LES SYSTEMES
DE GESTION DE BASES DE DONNEES DISTRIBUEES (*)

P. BOUCHET (**)

RESUME

Une procédure de reprise basée sur le transfert des copies d'une base de données dans un réseau d'ordinateurs est proposée. Un modèle mathématique du système de gestion des copies de la base est élaboré pour analyser sa disponibilité et la comparer à celle résultant de l'utilisation de la procédure de reprise traditionnelle dans les cas d'une répartition de deux, trois et quatre copies. L'efficacité de la procédure apparaît au travers des résultats numériques dans certaines conditions d'exploitation de la base de données.

(*) Cette recherche a été effectuée dans le cadre de la convention de recherche IRIA-SIRIUS n° 77.077

(**) Université Paris-Nord - IUT de Villetaneuse - Département Informatique - Avenue Jean-Baptiste Clément
93400 - Villetaneuse

Equipe de Recherche Associée au CNRS n° 452 "AL KHOWARIZMI"

1 - INTRODUCTION

L'automatisation des systèmes d'information impliquant des usagers largement répartis géographiquement ne se conçoit plus, depuis quelques années, de façon uniquement centralisée. Le développement des facilités de télécommunications, l'abaissement des coûts du matériel informatique, permet dans certaines conditions de distribuer les ressources en information logiciel et matériel et de les utiliser avec un meilleur rapport coût-efficacité. C'est dans cette perspective que se situent les travaux de CHU [5,6], CASEY [3,4] WHITNEY [12], LEVIN & al [10] sur l'implantation optimale des copies d'une base de données en différents noeuds d'un réseau d'ordinateurs et de MAHMOUD & al [11] sur l'implantation optimale des copies et le choix des capacités de transmission. Les modèles mathématiques que ces auteurs ont étudiés permettent de déterminer le nombre de copies de la base de données et les noeuds du réseau auxquels elles sont affectées en tenant compte des paramètres caractéristiques du système d'information et du réseau, ainsi que d'un certain nombre de contraintes.

La fiabilité des systèmes informatique de gestion des copies est un paramètre important de la répartition qui n'a pas été pris en compte dans les modèles analysés dans [3,4,5,6,10,12]. Le modèle analysé dans [11] suppose une fiabilité parfaite des systèmes informatiques et une fiabilité limitée des noeuds et des voies de communication qui intervient dans l'expression de la contrainte de disponibilité du système informatique.

L'influence de la multiplication des copies sur la fiabilité du système de gestion de la base de données est due au fait qu'elle accroît la probabilité de dégradation d'une copie ou d'immobilisation du système informatique permettant d'en assurer la gestion. Comme la reprise de l'activité du système ou la remise en état de la copie dégradée passe par l'exécution éventuelle d'un processus de réparation et dans tous les cas par l'exécution de procédures de restauration des informations en mémoires centrale et auxiliaire et de reprise des traitements, que nous appellerons dans la suite plus brièvement procédures de reprise, pendant la durée desquelles, le système informatique global est partiellement immobilisé, sa disponibilité sera diminuée dans certaines conditions d'exploitation.

Dans cette note, une procédure de reprise, basée sur le transfert des copies d'une base de données distribuées dans un réseau d'ordinateurs, est proposée. Les systèmes informatiques de gestion des copies réparties dans le réseau, supposé parfaitement fiable, sont soumis à des fautes intermittentes de deux types, nécessitant, selon que l'un ou l'autre type de faute se présente, l'application soit de la procédure connue



sous le nom de procédure de reprise à chaud, soit de la procédure de reprise à froid utilisant le transfert des copies. Un modèle mathématique du système informatique réparti est élaboré pour analyser sa disponibilité et comparer l'efficacité des procédures de reprise à chaud et à froid traditionnelles à celles des procédures à chaud traditionnelle et à froid avec transfert de copies, conjointement utilisées. Dans certaines conditions d'exploitation du système informatique réparti, une meilleure disponibilité de ces dernières procédures apparaît au travers des résultats numériques.

Cette étude se situe dans le prolongement des travaux effectués dans [1] qui montrent que dans un système de gestion de données centralisées, le fait de doubler la base de données, améliore l'efficacité des procédures de reprise. Le gain de disponibilité qui en résulte se traduit par une économie de coût d'utilisation, qui est, dans certaines conditions d'exploitation de la base de données, supérieur à l'investissement nécessaire pour doubler la capacité physique de celle-ci.

2 - PROCEDURES DE REPRISE DANS LES SYSTEMES DE GESTION DE DONNEES

2.1 - Principe des procédures de reprise "à froid" et "à chaud" dans les systèmes de gestion de données centralisées

Malgré les moyens mis en oeuvre pendant les étapes de conception et de réalisation du matériel et du logiciel d'un système informatique, des fautes se produisent inévitablement pendant son exploitation. Une faute se traduit par l'impossibilité pour le système de gestion de données d'exécuter correctement les tâches qui lui sont assignées. Les fautes sont causées par les défaillances du matériel, les erreurs du logiciel, les erreurs dans la conduite du système et par les données erronées introduites par les usagers du système et qui n'ont pas été interceptées par les programmes de contrôle. Les fautes se caractérisent par la détérioration des données stockées dans la mémoire auxiliaire et la mémoire centrale ou exclusivement dans cette dernière.

Dans le cas où la qualité des données de la mémoire auxiliaire est affectée, il est nécessaire d'exécuter une procédure de reprise dite "à froid" précédée éventuellement d'un processus de réparation. Lorsque les informations de la mémoire centrale sont uniquement détériorées, l'exécution d'une procédure de reprise "à chaud" précède la reprise de l'exécution des traitements, le processus de réparation étant rarement nécessaire.

2.1.1 - Principe de la reprise à froid

La restauration des données de la mémoire auxiliaire fait intervenir :