

Livre Blanc DALIBO #01

Migrer d'Oracle à PostgreSQL

18.10

Dalibo SCOP

<http://www.dalibo.com/>

Migrer d'Oracle à PostgreSQL

Livre Blanc DALIBO #01

TITRE : Migrer d'Oracle à PostgreSQL
SOUS-TITRE : Livre Blanc DALIBO #01

REVISION : 18.10
COPYRIGHT : © 2005-2018 DALIBO SARL SCOP

Le logo éléphant de PostgreSQL ("Slonik") est une création sous copyright et le nom "PostgreSQL" est marque déposée par PostgreSQL Community Association of Canada.

Remerciements : Ce livre blanc est le fruit d'un travail collectif. Nous remercions chaleureusement ici toutes les personnes qui ont contribué directement ou indirectement à cet ouvrage, notamment : Carole Arnaud, Damien Clochard, Léo Cossic, Adrien Nayrat, Thomas Reiss, Maël Rimbault. Nous remercions également la société DSIA qui a contribué au financement de ce livre blanc.

À propos de DALIBO :

DALIBO est le spécialiste français de PostgreSQL. Nous proposons du support, de la formation et du conseil depuis 2005.

Retrouvez toutes nos livres blancs sur <https://dalibo.com/>

Chers lectrices & lecteurs,

Nos livres blancs sont issus de plus de 12 ans d'études, d'expérience de terrain et de passion pour les logiciels libres. Pour Dalibo, l'utilisation de PostgreSQL n'est pas une marque d'opportunisme commercial, mais l'expression d'un engagement de longue date. Le choix de l'Open Source est aussi le choix de l'implication dans la communauté du logiciel.

Au-delà du contenu technique en lui-même, notre intention est de transmettre les valeurs qui animent et unissent les développeurs de PostgreSQL depuis toujours : partage, ouverture, transparence, créativité, dynamisme... Le but premier de nos livres blancs est de vous aider à mieux exploiter toute la puissance de PostgreSQL mais nous espérons également qu'ils vous inciteront à devenir un membre actif de la communauté en partageant à votre tour le savoir-faire que vous aurez acquis avec nous.

Nous mettons un point d'honneur à maintenir nos productions à jour, avec des informations précises et des exemples détaillés. Toutefois malgré nos efforts et nos multiples relectures, il est probable que ce document contienne des oublis, des coquilles, des imprécisions ou des erreurs. Si vous constatez un souci, n'hésitez pas à le signaler via l'adresse contact@dalibo.com !

Table des Matières

La transition vers PostgreSQL est une création de valeur	9
L'état de l'art	11
Fonctionnalités	11
Stabilité	11
La sécurité	12
Extensibilité	12
Le Respect des standards	12
5 changements de paradigmes	12
Pas de licences payantes	13
Pas de freins à la virtualisation	13
Pas de Vendor Lock-in	13
Pas de ventes liées	14
Pas d'audits intrusifs	14
PostgreSQL, c'est aussi...	14
Des innovations régulières	14
Un projet en plein essor	15
Partitionnement Externe ("Scale-out")	16
Partie 2 - Cinq étapes pour réussir la transition	17
Étape 1 - Étude	17
Comment faire un inventaire de parc Oracle ?	18
Étape 2 - Formation	18
Quelle Certification ?	19
Étape 3 - Socle	19
Quels outils ?	20
Étape 4 - Migration	20
Et le code applicatif ?	21
Étape 5 - Support	21
Faut-il placer les instances de pré-production en support ?	22
Partie 3 - Enjeux	23
Évaluer l'investissement	23
Étape 4 - Migration	23
Étude de cas	24
Partie 4 - PostgreSQL : pourquoi Météo-France a parié sur l'open-source ?	27
PostgreSQL : pourquoi Météo-France a parié sur l'open-source ?	27

LA TRANSITION VERS POSTGRESQL EST UNE CRÉATION DE VALEUR

L'informatique moderne regorge d'exemples de "duels" entre logiciels libres et logiciels propriétaires : Linux face à Windows, Firefox versus Internet Explorer, etc. Parmi toutes ces oppositions emblématiques, c'est dans le domaine des bases de données que le contraste est le plus frappant ! D'un côté, Oracle avec sa position dominante et ses pratiques commerciales agressives. De l'autre, PostgreSQL avec sa communauté décentralisée et son modèle économique basé sur le partage.

Toutefois, malgré ces divergences évidentes, les deux SGBD sont extrêmement proches sur le plan technique et partagent énormément de concepts communs : transactions, vues matérialisées, procédures stockées, fonctions analytiques, etc. Au point que PostgreSQL est généralement présenté comme l'alternative n°1 à Oracle. Dès lors, migrer d'Oracle à PostgreSQL n'a rien d'un séisme technologique : au contraire, les DBA Oracle peuvent maîtriser sans difficultés un parc d'instances PostgreSQL.

Au final, le réel changement de paradigme se situe au niveau de la "relation fournisseur". Dans la communauté PostgreSQL, il n'y a pas d'éditeur unique, pas de vente liée et pas de "vendor lock-in".

À l'inverse, PostgreSQL offre aux entreprises des libertés nouvelles notamment : choisir un mode de virtualisation sans être pénalisé, pouvoir changer de prestataire technique sans remettre en cause le choix d'architecture ou encore allouer plus de CPU à une instance sans mettre son budget en péril.

Ces libertés nouvelles ont évidemment d'autres implications : au premier regard, la richesse de l'écosystème PostgreSQL peut apparaître comme une complexité. Pour chaque domaine du métier de DBA (sauvegarde, tuning, haute-disponibilité, etc.) la communauté PostgreSQL propose des dizaines d'outils open-source. Difficile de faire son choix lorsque l'on vient du monde Oracle où la stack applicative est monolithique !

Autre enjeu majeur de la migration : comment maîtriser les coûts cachés. Car bien sûr "logiciel libre" ne veut pas dire "logiciel gratuit" et là encore le mode de calcul du coût de possession d'une instance PostgreSQL est radicalement différent du modèle Oracle.

Depuis plus de 10 ans, Dalibo assiste des sociétés en France et en Europe dans leur transition vers l'Open-source. De cette expérience unique, nous avons tiré le bilan et défini une méthode en 5 étapes pour réussir le passage à PostgreSQL. Nous proposons cette méthode dans ce livre blanc, non pas comme une recette qu'il faudrait suivre à la lettre, mais comme une feuille de route qui aidera le lecteur dans sa démarche pour moderniser son infrastructure et reprendre le contrôle de ses données. Car au-delà des réductions des

Migrer d'Oracle à PostgreSQL

coûts de licence, la transition vers PostgreSQL est une véritable dynamique d'innovation et de création de valeurs au cœur des systèmes d'informations.³

L'ÉTAT DE L'ART

Conçu dans les années 80 à l'université de Berkeley et publié sous licence open-source en 1995, PostgreSQL est le Système de Gestion de Bases de Données Relationnel (SGBDR) open-source de référence. Depuis 20 ans, ce projet a traversé les époques et a assis sa réputation sur des principes simples et forts : stabilité, ouverture, performances et respect des standards.

À tel point qu'aujourd'hui, PostgreSQL (également appelé "Postgres") est désormais considéré comme l'alternative numéro 1 à Oracle Database. L'État français lui-même préconise PostgreSQL dans son Socle Logiciels Libres (SILL) et, en 2015, Jacques Marzin, alors Directeur de la Direction Interministérielle des Systèmes d'Information et de Communication (DISIC) déclarait :

« L'équipe de pilotage du SILL publie désormais des guides d'accompagnement basés sur des retours d'expérience. Le premier concerne la migration vers le logiciel de gestion de bases de données Postgres. »

Pour aller plus loin :

<http://references.modernisation.gouv.fr/sites/default/files/SILL-2015-socle-interministeriel-logiciels-libres.pdf> <http://www.modernisation.gouv.fr/le-sgmap/organisation/la-disic> <http://www.silicon.fr/logiciels-libres-secteur-public-aller-plus-loin-115810.html#4>

FONCTIONNALITÉS

Aux fondements de PostgreSQL se trouvent 4 valeurs essentielles qui définissent ce logiciel et en font le SGBD open-source de référence.

STABILITÉ

Depuis la fin des années 2000 et l'apparition des bases NoSQL, les fondements des bases de données sont remis en cause... Pour obtenir plus de performances, les bases NoSQL s'affranchissent partiellement des 4 principes d'ACID (Atomicité, Cohérence, Isolation et Durabilité) qui sont pourtant essentiels pour garantir la sécurité des données et des transactions. À l'inverse, PostgreSQL avance et améliore ses performances chaque année sans faire de compromis sur la sécurité des données. Grâce à cette exigence et ses 20 ans d'existence, PostgreSQL a acquis une solide réputation de stabilité, de pérennité et de robustesse.

LA SÉCURITÉ

La protection des données est un enjeu majeur des années à venir. Dans ce contexte, PostgreSQL propose un arsenal de défenses à la fois puissant et varié. Plusieurs méthodes d'authentification sont disponibles (LDAP, GSSAPI, ...) et l'accès aux données peut être filtré de manière fine avec les permissions par colonnes et les politiques Row-Level Security (RLS). Face aux attaques extérieures, il est possible de chiffrer les communications avec des certificats SSL, de chiffrer les données avec l'extension `pg_crypto` et il est possible de renforcer la sécurité en activant SE-Postgres (équivalent de SE-linux).

EXTENSIBILITÉ

Créé initialement dans un laboratoire à l'université de Berkeley (voir encadré) et donc conçu comme un objet, PostgreSQL est une base de données extensible dans le sens où il est très simple d'ajouter des objets spécifiques : procédures stockées, opérateurs, extensions... Autant d'éléments qui permettent d'implémenter rapidement de nouvelles fonctionnalités. Le logiciel PostGIS (voir plus loin) en est la meilleure illustration.

LE RESPECT DES STANDARDS

PostgreSQL implémente la majeure partie des préconisations du standard SQL :201 1 et à ce titre il est l'un des SGBD le plus proche de la norme. PostgreSQL

Pourquoi ce nom ?

Aussi imprononçable soit-il le nom "PostgreSQL" contient depuis son origine un état d'esprit de dépassement et d'innovation. En 1986, Michael Stonebraker le fondateur du logiciel Ingres (et de nombreux autres SGBD) décida de lancer une nouvelle génération qu'il nomme Postgres pour "Post-Ingres". Le suffixe SQL fut ajouté au milieu des années 90, lorsque le support du standard SQL fut implémenté.

5 CHANGEMENTS DE PARADIGMES

Au niveau technique, PostgreSQL est le SGBD le plus proche d'Oracle tout en proposant un modèle économique radicalement différent. Choisir PostgreSQL ce n'est pas seulement faire un choix technique, c'est surtout opter pour un virage à 180° dans la relation client-fournisseurs.

PAS DE LICENCES PAYANTES

Opter pour PostgreSQL, c'est tout d'abord quitter le monde des licences propriétaires pour une licence open-source de type BSD. Cette licence donne le droit d'utiliser gratuitement PostgreSQL sans restrictions et sans redevance. Pour autant, cela ne signifie pas que le coût de possession de PostgreSQL est nul : Logiciel Libre ne veut pas dire logiciel gratuit. Il existe un TCO incompressible qui devra être évalué (voir plus loin). Cependant, d'une manière générale, il est largement inférieur au coût de possession des SGBD propriétaires.

PAS DE FREINS À LA VIRTUALISATION

Comme pour la plupart des SGBD propriétaires, le coût de possession des bases Oracle est directement indexé sur le nombre de processeurs des serveurs hébergeant les bases. Pire, dans les environnements virtualisés de type VMware, le tarif reste indexé sur le nombre de processeurs physiques et non pas sur le nombre de processeurs virtuels. Un mode de calcul volontairement pénalisant pour les sociétés qui modernisent leur parc en virtualisant les environnements de production. Avec PostgreSQL, ces freins n'existent pas. PostgreSQL fonctionne parfaitement sur des serveurs physiques ou virtuels, sans que cela n'impacte les coûts de possession. Ce n'est d'ailleurs pas un hasard si de grands acteurs du Cloud Computing ont opté pour PostgreSQL comme base de données par défaut.

Pour aller plus loin :

http://docs.aws.amazon.com/redshift/latest/dg/c_redshift-and-postgres-sql.html

PAS DE VENDOR LOCK-IN

La "révolution PostgreSQL" réside aussi dans le choix et la liberté rendus aux utilisateurs. Depuis plus de 20 ans, PostgreSQL est développé par une communauté d'entreprises, d'universitaires et d'utilisateurs répartis dans le monde entier. Comme pour Linux, il n'y a pas d'éditeur unique contrôlant le projet. Dès lors pour le support, la formation et les autres services associés, les utilisateurs ont le choix entre différents acteurs de différentes tailles et entre différents niveaux de prestation. Et à tout moment, une société peut changer de prestataire, mais garder ses bases PostgreSQL. En dissociant le choix technologique et le choix du prestataire, PostgreSQL permet aux entreprises de changer de prestataire de service sans pour autant remettre en cause les fondements de leurs Systèmes d'Information.

PAS DE VENTES LIÉES

Le projet PostgreSQL est totalement indépendant et fonctionne sur tous les environnements de productions modernes. Dès lors, avec le choix de PostgreSQL, chacun est libre de choisir la solution de stockage, le constructeur matériel, le système d'exploitation, les serveurs d'applications et le langage de développement qui conviennent en fonction du contexte.

PAS D'AUDITS INTRUSIFS

Profitant de leurs positions de quasi-monopoles, les grands éditeurs de bases de données ont fait évoluer leurs pratiques pour mettre une pression plus forte sur leurs utilisateurs, en imposant notamment des audits de parcs pour recenser les instances déployées et revaloriser les contrats de licences. Oracle est la tête de pont de ce mouvement avec des batailles juridiques qui font rage avec plusieurs de leurs grands clients. Pour la majorité des entreprises, le risque d'un audit Oracle est devenu un enjeu majeur dans leur politique de gestion de coûts de licence. À l'inverse, par son modèle décentralisé, PostgreSQL élimine de facto toute menace de ce type. Les entreprises sont libres de déployer autant d'instances que nécessaire et ne risquent pas une explosion de leurs coûts de possession.

Plus d'informations sur :

<http://www.cio-online.com/actualites/lire-audits-oracle%C2%A0-carrefour-a-obtenu-de-ne-pas-executer-de-scripts-de-contrôle-7400.html>

POSTGRESQL, C'EST AUSSI...

DES INNOVATIONS RÉGULIÈRES

Depuis 10 ans, la communauté PostgreSQL a adopté un cycle de vie à la fois clair et extrêmement dynamique. Chaque année, une nouvelle version majeure est publiée et introduit des nouvelles fonctionnalités et des améliorations de performances. Ce rythme exceptionnel fait de PostgreSQL le SGBD le plus innovant et le plus évolutif du marché.

Pour preuve, suite à l'arrivée sur le marché des bases NoSQL, la communauté Postgres s'est mise en ordre de marche pour intégrer et développer les fonctionnalités de ces bases d'un nouveau type (format JSON, stockage clé-valeur, etc.)

Le même mouvement est visible dans de nombreux domaines : pour les bases de données Géographiques (SIG), PostgreSQL est devenu la référence de l'industrie. Pour l'inter-

opérabilité des données, c'est également le SGBD le plus puissant au point que l'on parle désormais de PostgreSQL comme une plate forme d'intégration de données.

Plus d'informations sur :

http://blog.taadeem.net/english/2015/03/15/70_Shades_of_Postgres/

Les Foreign Data Wrappers : vers une intégration de données directe et sans ETL

Basés sur la norme SQL/MED, les Foreign Data Wrappers (FDW) sont des connecteurs externes qui permettent à PostgreSQL de lire et écrire des données sur tout type de stockage externe (base Oracle, annuaire LDAP, cluster Hadoop, fichier CSV, web services, etc.) Dès lors, il devient possible de fédérer des données hétérogènes sans avoir à utiliser un outil d'intégration spécifique (ETL). Le chargement des données se fait directement par la base PostgreSQL elle-même et la transformation est spécifiée en langage SQL.

SQL + NoSQL : le meilleur des deux mondes

L'opposition entre bases SQL et NoSQL est dépassée ! L'avenir appartient aux bases hybrides, capables de gérer à la fois les données relationnelles et non structurées. Des bases capables de faire de justes compromis entre constance, disponibilité et performances. PostgreSQL est à la pointe de ce mouvement : avec le format JSON binaire, le stockage clef-valeur hstore, les index GIN, le langage PL/V8, Postgres n'a rien à envier aux moteurs NoSQL tel que MongoDB ou Cassandra.

PostGIS : La référence pour les données géo-localisées

PostGIS est la "cartouche spatiale" : une extension qui ajoute des outils et types de données géographiques supplémentaires pour qualifier et manipuler les données. Créé en 2004, ce projet est aujourd'hui la référence incontournable dans le monde des SIG : Map-py, l'IGN ou encore le projet Open Street Map reposent sur PostGIS et donc PostgreSQL.

UN PROJET EN PLEIN ESSOR

Difficile de prédire les prochaines avancées de PostgreSQL, la nature décentralisée de ce projet rend obsolète la notion de "roadmap" telle qu'on l'entend chez les éditeurs de logiciels propriétaires. On peut toutefois entrevoir 2 grands domaines d'innovations majeurs dans un proche avenir :

Parallélisme

Les bases de données OLAP n'ont que peu d'utilisateurs, mais des traitements très longs. Avec l'introduction du traitement parallèle des requêtes SQL, Postgres va pouvoir lancer plusieurs processus pour exécuter une requête et donc mobiliser plusieurs processeurs en simultané pour un traitement donné. Cela va réduire de manière importante les temps de calcul.

Pour en savoir plus :

<http://rhaas.blogspot.fr/2015/03/parallel-sequential-scan-for-postgresql.html>

PARTITIONNEMENT EXTERNE ("SCALE-OUT")

Face à l'explosion des volumes de données, les stratégies sont multiples. Le partitionnement "classique" est déjà possible avec PostgreSQL et il permet de "découper" une table de plusieurs To et de répartir les données sur de multiples "tables filles" selon un critère sélectif. Le partitionnement externe étend cette possibilité

en permettant d'exporter de manière transparente ces tables filles sur d'autres instances. On pourra ainsi mettre en place des stratégies de scale-out et répartir les données sur différents serveurs.

Pour en savoir plus :

<https://momjian.us/main/writings/pgsql/sharding.pdf>

PARTIE 2 - CINQ ÉTAPES POUR RÉUSSIR LA TRANSITION

Les bases de données sont des pièces essentielles des Systèmes d'Information modernes. Remplacer une brique aussi essentielle nécessite de la méthode et un plan de route précis. Les étapes exposées ici sont issues de 10 ans d'expérience dans ce domaine.



Fig. 1 : Processus de transition

L'ordre proposé est donné à titre indicatif et peut bien sûr être adapté à votre contexte. À vous de trouver la formule gagnante pour passer à PostgreSQL en toute sérénité

ÉTAPE 1 - ÉTUDE

À l'origine de toute migration réussie, il y a une analyse précise des enjeux, des différences techniques et de l'effort à produire.

Certaines bases sont éligibles à la migration vers PostgreSQL sans problèmes. D'autres demandent plus d'attention. On classe généralement les instances Oracle en 3 catégories selon leur degré de compatibilité avec PostgreSQL :

- Type A : les bases qui peuvent être migrées via un processus automatique, c'est à dire avec un ETL ou un outil de migration dédié (Ora2PG notamment) ;
- Type B : les bases qui peuvent être transférées de manière semi- automatique. Dans ces cas-là, la majeure partie de l'effort de transition est prise en charge par un processus automatique (comme pour les bases Type A) mais il reste quelques modifications "manuelles" qui devront être réalisées par un DBA maîtrisant à la fois Oracle et PostgreSQL. Le temps nécessaire pour ces opérations est généralement inférieur à 10 jours/homme ;
- Type C : les bases qui ne peuvent pas être migrées de manière automatique. Il s'agit de bases Oracle contenant un grand volume de procédure stockée PL/SQL qui implique un travail de réécriture important.

La proportion des différents types de base dépend de différents facteurs. Pour une société qui utilisent massivement des ORM, les bases de Type A seront prédominantes. À l'inverse si une société possède beaucoup d'applications métier qui reposent sur des procédures stockées alors les bases de Type B et C seront nombreuses.

Ci-dessous un exemple "classique" d'inventaire de parc :

Classification des bases

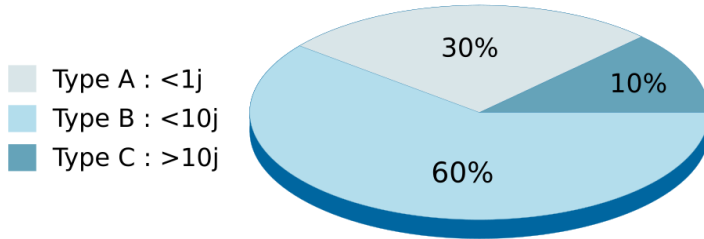


Fig. 2 : Typologie d'un parc d'instance

Certaines bases sont éligibles à la migration vers PostgreSQL sans :

COMMENT FAIRE UN INVENTAIRE DE PARC ORACLE ?

Ora2PG est un outil open-source de migration d'Oracle vers PostgreSQL qui prend en charge la migration des tables, des données, des droits et d'une partie du code des procédures stockées.

En 2011 via un partenariat avec le Ministère des Finances et Atos, DALIBO a développé un module d'évaluation qui analyse si une instance Oracle est éligible. Ce module produit un rapport complet sur le nombre de jours/homme nécessaire pour la migration.

À partir de cet outil, il est possible de connaître rapidement l'éligibilité des instances Oracle d'un parc entier et ainsi se focaliser sur les instances de "classe A" dans un premier temps, puis financer les suivantes avec les premières économies dégagées.⁹

ÉTAPE 2 - FORMATION

L'élément clef de la migration est le transfert des compétences des équipes DBA en place. Le succès d'une transition est étroitement lié à la montée en compétence des équipes d'exploitation du logiciel.

Les similarités entre Oracle et PostgreSQL sont importantes et des administrateurs Oracle peuvent rapidement monter en compétences sur PostgreSQL. Toutefois le changement

PARTIE 2 - CINQ ÉTAPES POUR RÉUSSIR LA TRANSITION

peut engendrer des inquiétudes et des réticences. Certains administrateurs de SGBD ayant une longue et riche expérience avec Oracle peuvent vivre la transition comme une certaine dévalorisation de leur travail et une perte de leurs acquis techniques.

Les sessions de formation sont des moments importants pour évoquer ces réticences et désamorcer les a priori. Les équipes techniques peuvent librement poser toutes leurs questions et travailler avec le formateur pour comprendre les variations de concepts, les similarités et les nouvelles possibilités.

La construction d'un savoir-faire autour de PostgreSQL s'appuie nécessairement sur les connaissances des autres SGBD. Il est donc important de choisir un organisme de formation spécialisé et un formateur qui connaîtra PostgreSQL mais également Oracle, pour aider les équipes techniques à acquérir la double-compétence.

QUELLE CERTIFICATION ?

PostgreSQL n'étant pas contrôlé par un éditeur unique, il n'existe pas de certification "officielle" pour valider l'acquisition d'un enseignement...

Néanmoins quelques sociétés de la communauté proposent leur propre programme de validation des compétences. Parmi ces sociétés, Dalibo est la seule à délivrer une certification en français qui s'adresse à toute personne exerçant une activité professionnelle impliquant PostgreSQL, ou à toute entreprise souhaitant voir les compétences de ses salariés ou collaborateurs certifiées.

Deux niveaux sont accessibles : "Dalibo Essential Postgres Professionnal" ou "Dalibo Advanced Postgres Professionnal", afin de valoriser les connaissances autour de PostgreSQL, que l'on soit DBA confirmé ou expert dans le domaine.

Bien qu'indépendante de la communauté, la certification Dalibo est le premier programme de certification PostgreSQL francophone, basée sur les 13 années d'expériences de Dalibo autour de PostgreSQL.

ÉTAPE 3 - SOCLE

À l'origine de toute migration réussie, il y a une analyse précise des enjeux, des différences techniques et de l'effort à produire.

PostgreSQL est un logiciel sophistiqué et versatile. Il existe de multiples manières de l'exploiter en production. Afin de garantir l'homogénéité du parc d'instances PostgreSQL, il est essentiel de construire un corpus de documents (appelé "socle technique" ou "référentiel"). Ce socle sera constitué le plus souvent d'un guide d'administration (manuel

Migrer d'Oracle à PostgreSQL

d'installation, mises à jours, opérations courantes, etc), d'un guide d'optimisation, d'une politique de sauvegarde, d'un protocole de supervision et de procédures d'urgence. Mais il ne s'agit pas là d'une liste exhaustive.

Là encore, les compétences Oracle peuvent être réutilisées et adaptées pour rédiger un socle PostgreSQL très proche de l'existant.

La rédaction d'un socle permet également de mieux encadrer l'utilisation des logiciels "satellites". En effet, à la différence d'Oracle, PostgreSQL n'est pas livré avec une pile applicative unique. Au contraire, pour chaque usage (sauvegarde, tuning, etc.) il existe plusieurs outils disponibles. Là encore il est nécessaire de référencer et qualifier les outils et ainsi construire une pile applicative interne répondant au besoin de l'entreprise.

QUELS OUTILS ?

Pour chaque outil de la sphère Oracle, il existe des alternatives open-source. Dans le cadre d'une transition vers PostgreSQL, il est important d'établir une liste des logiciels utilisés et leur remplacement. Voici une série d'équivalences donnée à titre d'exemple :

- Rman : Barman ou Pitrery
- Automatic Workload Repository : PoWA
- Grid Control : OPM
- Active Data Guard : Hot Standby
- Database Resident Connection Pool : PgBouncer
- DB LINK : Foreign Data Wrappers

ÉTAPE 4 - MIGRATION

La phase de migration elle-même est paradoxalement l'étape la plus triviale du processus de transition. Si les étapes précédentes ont été soigneusement préparées, la bascule d'Oracle vers PostgreSQL est parfois longue mais toujours simple.

- **Pour les bases de Type A** : la migration du schéma, des données, des triggers et des autres contraintes se fait en quelques minutes ;
- **Pour les bases de Type B** : les modifications non-automatiques doivent-être préparées en amont. Par exemple : adapter la méthode de partitionnement Oracle à PostgreSQL, retranscrire le DATABASE LINK avec le module ora_fdw, transformer les SYNONYMs en vues, placer les fonctions d'un PACKAGE au sein d'un schéma, etc. D'une manière générale, la durée de préparation peut varier de 1 à 10 jours ;

PARTIE 2 - CINQ ÉTAPES POUR RÉUSSIR LA TRANSITION

- **Pour les bases de Type C** : le travail de réécriture des procédures stockées constitue l'essentiel de l'effort de migration. Cette tâche nécessite des développeurs ayant une double compétence sur les langages PL/SQL (pour Oracle) et PL/pgSQL (pour PostgreSQL). La présence de tests unitaires originaux est évidemment un plus non négligeable car ces tests pourront garantir la validité des procédures transcrites et ainsi accélérer le processus de migration.

Quelle que soit la typologie d'une base, la durée de l'effort de migration est conditionnée au nombre de procédures et de lignes de code qui doivent être transcrites. Les objets Oracle non standards comme DBLINK, PACKAGE ou PARTITION sont également des points à surveiller mais il existe pour chacun des méthodes de contournement, quand un équivalent n'existe pas déjà.

ET LE CODE APPLICATIF ?

Même si le SQL a été conçu comme un langage universel, il n'est pas toujours implémenté de façon standard. Beaucoup de SGBDs, Oracle en tête, proposent des syntaxes qui leur sont propres et ne fonctionneront pas avec PostgreSQL.

Par exemple, le Query HINT d'Oracle est une entorse au standard. Si le code de l'application utilisant les données contient des requêtes SQL non standard, il faudra les réécrire et déployer une nouvelle version de l'application en parallèle de la migration.

Bien sûr, cette question ne se pose pas pour les logiciels utilisant un ORM (comme Hibernate) ou une couche d'abstraction (ADO.NET par exemple) ou lorsque l'éditeur du logiciel supporte plusieurs bases de données.

ÉTAPE 5 - SUPPORT

Toute base de données implique un contrat de support. Les SGBD open-source et PostgreSQL en particulier n'échappent pas à cette règle. La différence principale avec ce dernier, c'est qu'il n'y a pas un éditeur unique imposant un modèle de support monolithique.

Comme souvent avec l'open-source, les utilisateurs ont l'embarras du choix et dès lors pour ce qui concerne le support PostgreSQL, il est crucial de définir des critères de sélection précis pour choisir un prestataire de support capable de répondre aux besoins et attentes.

Afin de choisir une offre de support, les questions suivantes sont déterminantes :

- Est-ce que le prestataire de support est reconnu comme un sponsor officiel du projet PostgreSQL ?

Migrer d'Oracle à PostgreSQL

- Le prestataire de support a-t-il une équipe d'experts PostgreSQL francophones compétents ?
- Combien de personnes font partie de l'équipe support ?
- Le support est-il en intégralité traité par l'équipe du prestataire ?
- Est-ce que les incidents de Niveau 1 ou les appels de nuits sont sous-traités à une société tierce ?
- Quelles sont les versions de PostgreSQL supportées ?
- Quelle est la durée de prise en charge de chaque version ?

Pour aller plus loin :

<http://www.postgresql.org/about/sponsors/>

FAUT-IL PLACER LES INSTANCES DE PRÉ-PRODUCTION EN SUPPORT ?

PostgreSQL offre la liberté de déployer autant d'instances que nécessaire, sans avoir à les déclarer et sans avoir à craindre un audit de parc. Sur le même principe, le support PostgreSQL est adaptatif : Chaque société peut contrôler le scope de ce support en dressant la liste des instances supportées et celle qui ne le sont pas.

Si le support des instances de production est une évidence, la question se pose pour les environnements de pré-production, de recette voire de développement. Différentes stratégies de couvertures support peuvent s'appliquer... Certaines sociétés limitent le champs du support aux instances PostgreSQL de production. À l'inverse d'autres sociétés optent pour une couverture intégrale, car elles préfèrent déclarer un incident dès qu'ils survient en pré-production ou en phase développement plutôt que d'attendre que les problèmes ne surviennent en production.

PARTIE 3 - ENJEUX

Open-source ne veut pas dire gratuit. PostgreSQL est dénué de coût de licence, ce qui rend bien sûr son coût de possession beaucoup plus faible que les autres SGBD propriétaires et notamment Oracle. Il n'en reste pas moins que toute technologie comporte des coûts associés, aussi bien en termes d'investissement (migrations, études et socle) qu'au niveau des frais de fonctionnement (support, formation et conseil). Mesurer les coûts induits et pouvoir déterminer le retour sur investissement d'une migration est un enjeu essentiel.

ÉVALUER L'INVESTISSEMENT

L'investissement lié à une migration dépend des choix opérés au niveau des étapes 1, 3 et 4 présentées précédemment. La phase d'étude, la rédaction d'un socle technique et la migration elle-même représentent les trois principaux postes de dépenses pour garantir une transition maîtrisée.

Comme vu précédemment, chacune de ces 3 étapes peut-être internalisée ou confiée à un ou plusieurs prestataires.

Nous proposons ci-dessous un chiffrage pour la migration de 40 instances de type A réalisées intégralement par un prestataire externe :

Étude	5 à 10 jours	10 k€
Socle	50 à 100 jours	60 k€
Migrations	40 à 60	50 k€

ÉTAPE 4 - MIGRATION

Déterminer le coût d'exploitation de PostgreSQL est une tâche difficile. Comme dit précédemment les logiciels libres comportent des coûts cachés et difficiles à évaluer. On peut néanmoins identifier 3 grandes catégories de dépenses liées au fonctionnement de PostgreSQL :

- La formation ;
- Le support ;
- Le conseil.

La formation est considérée comme un coût récurrent. Même si le coût de formation initial peut-être un palier important, il n'est reste pas moins que les équipes techniques

Migrer d'Oracle à PostgreSQL

doivent-être formées régulièrement pour s'adapter aux nouvelles versions de PostgreSQL, maintenir leur savoir-faire et développer de nouvelles compétences.

Pour 40 instances et une équipe de 3 DBAs :

Conseil	5 à 10 jours / an
Formation	3 à 8 jours / an
Support	5 à 12 jours / an
	+ 45 tickets ouverts

ÉTUDE DE CAS

Le scénario suivant représente un plan de migration classique pour une entreprise ayant un parc de 300 instances Oracle, dont 200 bases de Type A. Parmi ces instances, la société prévoit d'en migrer 1 50 sur 5 ans (de 2016 à 2020).

Le plan prévoit également que tous les nouveaux projets seront désormais basés sur PostgreSQL, ce qui représente 20 instances supplémentaires sur 5 ans.

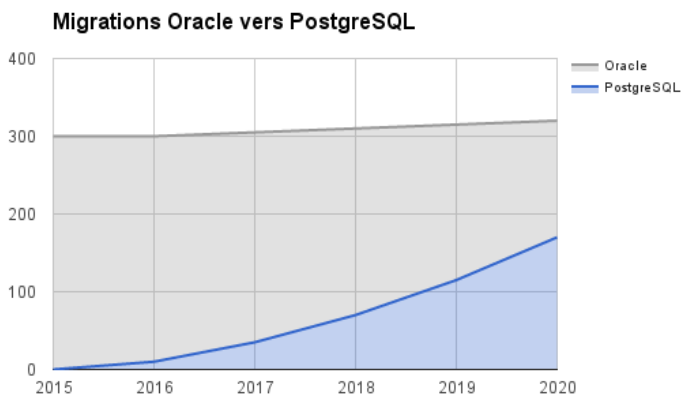


Fig. 3 : Plan de migration

Coté investissement, elle réalise une étude complète (étape 1) en 2015 pour concevoir son plan de migration, suivi de la rédaction d'un "socle" (étape 3) en 2016 afin d'intégrer PostgreSQL dans son parc logiciel.

Ces deux étapes représentent un coût de démarrage important, toutefois lorsque ces actions sont réalisées, l'investissement se stabilise et se concentre sur l'effort de migration en lui-même (étape 4).

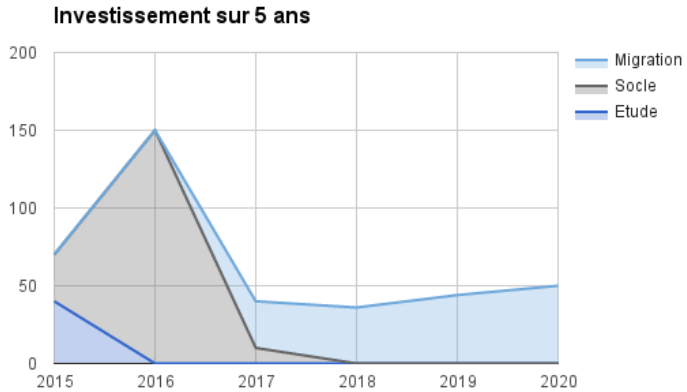


Fig. 4 : Plan d'investissement

Comme vu précédemment, le coût de possession (TCO) est composé essentiellement par l'assistance (conseil et support) et la formation, qui sont des besoins récurrents et indexés sur le nombre d'instances déployées.

Dès lors le TCO est assez simple à évaluer et il bénéficie d'un effet d'échelle évident avec la dégressivité progressive des coûts du support.

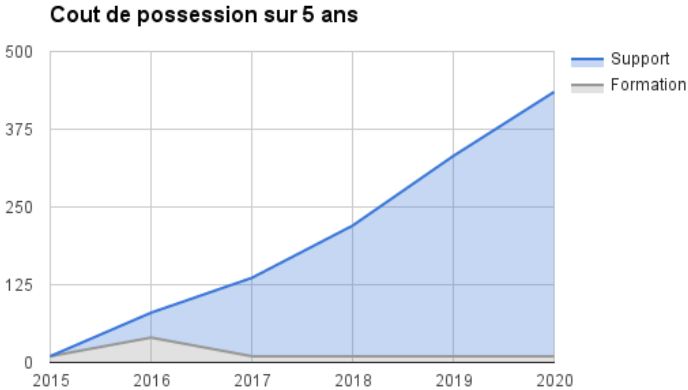


Fig. 5 : Evolution du TCO

PARTIE 4 - POSTGRESQL : POURQUOI MÉTÉO-FRANCE A PARIÉ SUR L'OPEN-SOURCE ?

La conférence sur le Climat de 2015 à Paris a confirmé qu'au sein d'une société confrontée au défi du développement durable, météorologie et climat sont l'objet d'une attention croissante. Dans ce contexte, une prévision au meilleur de l'état de l'art est une nécessité. Les choix logiciels et matériels effectués ces dernières années par l'organisme français de météorologie ont été déterminants pour atteindre ce niveau de qualité.

POSTGRESQL : POURQUOI MÉTÉO-FRANCE A PARIÉ SUR L'OPEN-SOURCE ?

« Pour valider a posteriori le choix stratégique en faveur d'une technologie, ainsi que l'investissement financier et humain qu'il a requis, l'analyse porte autant sur la pérennité du service rendu que sur la constance du niveau de qualité au fil des années en dépit même de l'inévitable accroissement des besoins. D'après ces critères, je peux affirmer que le choix effectué il y a plus de 10 ans par Météo-France d'un système de gestion de bases de données relationnel open-source en remplacement de systèmes "traditionnels" propriétaires était judicieux.

Dès les prémices de l'implémentation du système de gestion de bases de données relationnel open-source PostgreSQL à l'horizon 2001, l'ambition était claire : d'une part se libérer de la dépendance à un fournisseur technologique due aux verrous propriétaires et en limitant les coûts de licence, d'autre part conserver une maîtrise totale des coûts liés aux bases de données. La quantité et la qualité des données recueillies sont les piliers du métier de Météo-France. L'activité de Météo-France repose essentiellement sur l'acquisition massive et en temps réel de données d'observations issues d'une variété de systèmes et de techniques tels que les stations en surface, les satellites et les radars météorologiques, les radiosondages, les capteurs embarqués, etc. Une fois ces données recueillies, référencées à l'aide de métadonnées et insérées dans des bases, le travail de Météo-France consiste à effectuer une assimilation des données dans les modèles de prédiction. Pour y parvenir, plusieurs aspects essentiels sont à prendre en compte : la puissance de calcul mais aussi la quantité et la qualité des données à traiter. Il est impératif, pour affiner les prédictions et la précision géographique, de disposer de moyens de calcul conséquents. Météo-France s'est dotée en 2014 de deux supercalculateurs dont la tâche principale est de réaliser des prédictions en temps réel. Ils sont sollicités pour reconstituer les conditions climatiques passées à partir d'archives d'observations. Ces supercalculateurs sont également utilisés dans la recherche sur les phénomènes atmosphériques. Enfin ils effectuent régulièrement des projections climatiques qui participent aux travaux du GIEC (Groupe

d'experts intergouvernemental) sur l'impact du changement climatique. Par ailleurs, au-delà des prévisions, ces nouvelles infrastructures devraient s'avérer déterminantes pour la mise en place d'outils d'aide à la décision, notamment dans les contextes de vigilance pour lesquels il faut pouvoir quantifier les risques locaux. Sur le plan logiciel, sachant que Météo-France privilégie à ce jour une approche de type SQL (données structurées), s'applique le challenge assez classique "volume, variété et vitesse". Il faut pouvoir absorber et insérer rapidement dans la base, à des fins d'analyse en temps réel, les données issues des observations. Puisqu'il n'est pas question de supprimer les anciennes données car en météorologie les données historiques ont autant de valeur que les données actuelles, nous sommes soumis à une obligation d'augmenter régulièrement le volume de stockage disponible. Open-source : un choix financier et stratégique pertinent sur le long terme. Ces différents aspects ont été envisagés tôt et ont influé sur la réflexion long terme de Météo-France. Dès le début des années 2000, l'établissement a fait le pari des standards ouverts en optant pour des logiciels open-source de classe entreprise supportés par du matériel lui aussi standard (à savoir des serveurs x86). Ainsi les bases de données PostgreSQL sont associées à des systèmes d'exploitation Red Hat et CentOS. Ce choix repose largement sur des considérations économiques imposant de maîtriser les coûts sur le long terme sans compromis sur la qualité du service rendu. Dix ans après la mise en production des premiers serveurs PostgreSQL, on peut dire que non seulement l'objectif a été atteint, mais aussi que le choix reste pertinent. A vrai dire, il l'est plus que jamais puisque la densité des observations s'accroît pour "nourrir" les supercalculateurs et améliorer les prévisions. Les volumes de données ont augmenté ces dernières années et la tendance va se confirmer. Dans ce contexte, le choix de PostgreSQL met Météo-France à l'abri des effets de seuil propres aux fournisseurs de solutions propriétaires : nous sommes capables de calculer précisément les coûts qui seront engendrés par ces nouvelles volumétries, tant vis à vis de l'acquisition de matériel (serveurs, stockage), que des services de support. Météo-France conserve le contrôle de ces différents aspects et n'est pas à la merci des changements de licence, de tarification, des fins de support. Le respect des normes et des standards : un pas en avant vers une collaboration plus efficace. Si ces différents bénéfices étaient attendus dès le départ, d'autres avantages ont découlé de ce choix open-source. Le partage de connaissances fait partie des missions que s'est fixé Météo-France en accord avec son contrat d'objectif et la directive INSPIRE. Bien que le respect des standards ouverts ne soit pas l'apanage des solutions open-source, ces dernières encouragent une approche vertueuse en matière de qualité des données. Par exemple, l'approche consistant à intégrer des données géo-référencées dans le cadre d'une utilisation PostgreSQL associé à PostGIS (qui ajoute le support d'objets géographiques à la base de données PostgreSQL) et Quantum GIS (logiciel de SIG bureautique libre), s'avère très pertinente dans le cadre du respect de la Directive INSPIRE et du standard ISO 19115 :2003. Par ailleurs, la gratuité des logiciels favorise dans une certaine mesure la réutilisation des données de

PARTIE 4 - POSTGRESQL : POURQUOI MÉTÉO-FRANCE A PARIÉ SUR L'OPEN-SOURCE ?

Météo-France, dont certaines (on citera par exemple les données d'observation SYNOP) sont accessibles gratuitement sur le site de l'institut conformément à sa démarche Open Data. Cette réutilisation peut servir non seulement la collaboration internationale pour une meilleure compréhension des mécanismes globaux complexes qui régissent le climat, mais aussi dans des contextes de recherches universitaires, académiques ou dans une volonté de curation des données. Ainsi, l'ouverture tend à s'inscrire durablement dans l'ADN de Météo-France. Elle ne concerne pas seulement le code source des logiciels utilisés, mais s'inscrit dans une démarche plus large. L'ouverture y est aussi synonyme d'un décloisonnement visant à stimuler le partage de données, le partage de connaissances pour de meilleures prévisions et une compréhension plus fine des phénomènes météorologiques.

»

Michel Edwell est actuellement ingénieur DBA à la direction des systèmes d'information de Météo-France. À ce titre, il est responsable technique de la migration des bases de données d'Oracle vers PostgreSQL. De formation ingénieur météorologue, il a commencé sa carrière sur des frégates météorologiques dans le golfe de Gascogne. Il a également réalisé une mission en Terre Adélie lors de laquelle il a mis en oeuvre les sondages de la couche d'ozone. En 1994, il a obtenu un master informatique à SupAero Toulouse. En 1999, il a participé au développement du premier site web de Météo-France destiné au grand public.

NOTES

NOTES

NOTES

NOTES

NOTES

NOTES

NOTES

NOTES

NOS AUTRES PUBLICATIONS

FORMATIONS

- **DBA1 - PostgreSQL Administration** Révision 18.09 / ISBN : 979-10-97371-30-2
- **DBA2 - PostgreSQL Avancé** Révision 18.09 / ISBN : 979-10-97371-31-9
- **DBA3 - PostgreSQL Réplication** Révision 18.09 / ISBN : 979-10-97371-32-6
- **DBA4 - PostgreSQL Performances** Révision 18.09 / ISBN : 979-10-97371-33-3
- **DBA5 - PostgreSQL Sauvegardes avancées** Révision 18.09 / ISBN : 979-10-97371-34-0
- **SQL1 - SQL Conception et Mise en Œuvre** Révision 18.09 / ISBN : 979-10-97371-35-7
- **SQL2 - SQL Avancé** Révision 18.09 / ISBN : 979-10-97371-36-4
- **SQL3 - Développer en PL/pgSQL** Révision 18.09 / ISBN : 979-10-97371-37-1
- **SQL4 - PostgreSQL : L'état de l'art** Révision 18.09 / ISBN : 979-10-97371-38-8
- **SQL5 - Migrer d'Oracle vers PostgreSQL** Révision 18.09 / ISBN : 979-10-97371-39-5

LIVRES BLANCS

- **Migrer d'Oracle à PostgreSQL**
- **Industrialiser PostgreSQL**

TÉLÉCHARGEMENT GRATUIT

Les versions électroniques de nos publications sont disponibles gratuitement sous licence open-source ou sous licence Creative Commons. Contactez-nous à l'adresse contact@dalibo.com pour plus d'information.

DALIBO, L'EXPERTISE POSTGRESQL

Depuis 2005, Dalibo met à la disposition de ses clients son savoir-faire dans le domaine des bases de données et propose des services de conseil, de formation et de support aux entreprises et aux institutionnels.

En parallèle de son activité commerciale, DALIBO contribue activement aux développements de la communauté PostgreSQL avec notamment une implication forte sur les projets gAdmin, phpPgAdmin, pgpool et le code de PostgreSQL lui-même. Le société est également à l'origine de nombreux outils open-source de monitoring, de migration, de sauvegarde et d'optimisation.

Le succès de PostgreSQL démontre que la transparence, l'ouverture et l'auto-gestion sont à la fois une source d'innovation et un gage de pérennité. Dalibo a intégré ces principes dans son ADN en optant pour le statut de SCOP : la société est contrôlée à 100% par ses salariés, les décisions sont prises collectivement et les bénéfices sont partagés à parts égales.