

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

REFONTE DE L'ENTREPÔT DE DONNÉES INSTITUTIONNEL DE BANQ

RAPPORT DE PROJET
PRÉSENTÉ
COMME EXIGENCE PARTIELLE
À LA MAITRISE EN GÉNIE LOGICIEL

PAR
PASCAL LAFORCE

FÉVRIER 2019

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de ce document diplômant se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.10-2015). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

REMERCIEMENTS

Je tiens à commencer mon rapport en prenant le temps de remercier les personnes qui ont contribué à la réalisation de ce projet ainsi qu'à toutes celles qui m'ont supporté durant ces longues années d'études. Mon parcours universitaire débuta en 2009 avec la simple intention de décrocher un certificat en informatique pour faire reconnaître mes compétences de développeurs, mais s'est finalement prolongé jusqu'à la maîtrise et de développer une expertise de pointe en architecture logicielle et devenir éventuellement un agent du changement des pratiques informatiques en entreprise.

Ce long parcours nécessitât un nombre incalculable de soirées passées en classe et à étudier à la maison, tout en tentant de minimiser les impacts sur le travail et la famille. Cet accomplissement n'aurait pas été possible sans le soutien de mes proches et tout particulièrement de ma femme qui fut présente dès le début. Marie-Ève, merci de m'avoir encouragé à poursuivre mes études et d'avoir su me remonter le moral lorsque je rencontrais des difficultés. Merci infiniment !

Je tiens également à remercier mes collègues de bureau qui ont participé à mes divers projets universitaires que je tentais d'appliquer systématiquement au travail depuis ces cinq dernières années. Paulo, Philippe, Jean-François et Gilles, merci d'avoir supporté toutes mes initiatives, de parfois m'avoir servi de cobaye et de m'avoir permis de transformer mon travail. Merci de votre participation !

Finalement, je souhaite remercier ma directrice de recherche, Mme Sylvie Trudel, pour son soutien durant mon parcours académique, pour son mentorat pendant la préparation de mon projet et la rédaction de ce rapport. Merci pour vos conseils !

RÉSUMÉ

Ce projet technique vise à rehausser l'entrepôt de données institutionnelles de Bibliothèque et Archives nationales du Québec (BAnQ) qui ne répond plus aux besoins des directions d'affaires en raison de l'obsolescence des technologies utilisées. Le projet consiste à mettre à niveau l'infrastructure technologique, à remplacer l'outil d'extraction-transformation-chargement (ETL) par une plateforme développée à l'interne ainsi qu'à migrer tous les processus d'affaires vers la nouvelle plateforme en plus ou moins 300 heures d'effort réparties sur une période de 6 mois. Sont exclues de ce projet toutes les activités d'analyse des besoins et des processus avec les directions d'affaires qui seront traitées ultérieurement. Le projet s'appuie sur une méthodologie reconnue de développement incrémental et itératif. Les activités consistent à la planification, à l'analyse, à la conception, au développement et au déploiement de la plateforme ETL et à la migration des traitements. Le bilan du projet montre que la méthodologie s'applique parfaitement aux projets de réalisation d'entrepôt de données conçus sous le style d'architecture en bus de données. La plateforme développée répond aux attentes de l'équipe interne et tous les traitements ont été migrés avec succès dans les délais convenus avec les clients. Je dois reconnaître que plus d'effort qu'anticipé fut nécessaire pour le développement de la plateforme ETL ainsi que pour le nettoyage des données, ce qui força un rajustement de la portée initiale, mais les objectifs sont en très grande partie atteints à la satisfaction des parties prenantes. Finalement, la réalisation du projet permettra à l'équipe interne de se concentrer davantage sur l'ajout de nouveaux processus d'affaires et de fonctionnalités analytiques ajoutant ainsi de la valeur aux données de l'institution.

Mots clés : Entrepôt de données (DW), Extraction-Transformation-Chargement (ETL), Intelligence d'affaires (BI), Analyse de données

ABSTRACT

This technical project aims to enhance the corporative data warehouse of *Bibliothèque et Archives nationales du Québec* (BANQ), which no longer meets the needs of business management due to the obsolescence of the technologies used. The project consists of upgrading the infrastructure, replacing the Extract-Transform-Load (ETL) tool with an internally developed platform, and migrating all business processes to the next-generation platform in more or less 300 hours of work spread over 6 months. This project excludes all business and process analysis which will be dealt with later. The project relies on a recognized methodology of incremental and iterative development cycle. Activities include planning, analysis, design, development and deployment of the ETL platform, and the migration of every existing ETL process. The project's results show that the methodology is perfectly applicable to data warehouse implementation projects designed in the data bus architecture style. The developed platform meets the expectations of the internal team and all the ETL processes have been successfully migrated within the agreed time frame. I must recognize that more effort than anticipated was required for the development of the ETL platform, as well as for data cleansing which forced the adjustment of the scope, but the objectives are largely met to stakeholders' satisfaction. Ultimately, the project completion will allow the team to focus more on adding new business processes and analytic capabilities that are adding value to the institution's data.

Keywords: Data Warehouse (DW), Extract-Transform-Load (ETL), Business Intelligence (BI), Data Analysis

TABLES DES MATIÈRES

	Page
INTRODUCTION	1
CHAPITRE I PROBLÉMATIQUE ET OBJECTIFS	3
1.1 Mise en contexte.....	3
1.2 Problématique.....	6
1.3 Portée du projet	8
1.4 Objectifs	10
1.5 Exigences de haut niveau	12
1.6 Facteurs de succès	13
CHAPITRE II ÉTAT DE L'ART.....	17
CHAPITRE III MÉTHODOLOGIE.....	21
3.1 Processus et activités.....	21
3.2 Présentation des résultats.....	26
3.2.1 Planification et gestion du projet.....	26
3.2.2 Définition des besoins d'affaires	28
3.2.3 Définition de l'architecture technique.....	31
3.2.4 Définition des données et de leur transformation.....	40
3.2.5 Définition des besoins d'intelligence d'affaires	44
3.2.6 Déploiement, amélioration et maintenance du système	46
CHAPITRE IV INTERPRÉTATION DES RÉSULTATS	51
4.1 Évaluation des objectifs.....	51
4.2 Interprétation des résultats.....	53
4.3 Qualité de service (QoS)	55

CHAPITRE V DISCUSSION DE L'EXPÉRIENCE	57
CONCLUSION	61
RÉFÉRENCES.....	63
ANNEXE A Charte de projet.....	65

LISTE DES FIGURES

Figure	Page
Figure 3.1 Vue d'ensemble de la méthode Data Warehouse Lifecycle	21
Figure 3.2 Analyse budgétaire des activités du projet	27
Figure 3.3 Extrait de la matrice en bus de données réalisée pour le projet.....	30
Figure 3.4 Architecture logicielle avant le projet.....	33
Figure 3.5 Architecture logicielle après le projet.....	34
Figure 3.6 Vue d'ensemble du processus ETL	36
Figure 3.7 Activité d'extraction des données.....	37
Figure 3.8 Activité de chargement initial des données	37
Figure 3.9 Activité de transformation des données.....	38
Figure 3.10 Activité de chargement final.....	39
Figure 3.11 Modélisation du processus d'affaires sur les emprunts de documents	41
Figure 3.12 Extrait de la modélisation de la table « sfait_ci_evenement »	43
Figure 3.13 Exemple du processus de transformation	44
Figure 3.14 Nouvelle section « Documents numériques » dans l'application BI	45
Figure 3.15 Rapport quotidien sur l'exécution du traitement « BN »	49
Figure 4.1 Résultat d'amélioration obtenu pour chacun des objectifs	51

LISTE DES TABLEAUX

Tableau	Page
Tableau 3.1 Processus d'affaires inclus dans la phase 1	29
Tableau 3.2 Déploiements et itérations	47

LISTE DES ABRÉVIATIONS, DES SIGLES ET DES ACRONYMES

3FN	Base de données normalisée à la 3 ^e forme normale.
BAnQ	Bibliothèque et Archives nationales du Québec.
BI	Intelligence d'affaires ou <i>Business Intelligence</i> .
CIF	Modèle <i>Corporate Information Factory</i> de William H. Inmon.
CPI	<i>Cost Performance Index</i> .
DaaS	Service de données ou <i>Data as a Service</i> .
DBA	Administrateur de bases de données.
DGTIT	Direction générale des technologies de l'information et des télécommunications.
DW	Entrepôts de données ou <i>Data Warehouse</i> .
DWL	<i>Data Warehouse Lifecycle</i> de Ralph Kimball.
ERD	Diagramme entité-relation ou <i>Entity-Relation Diagram</i> .
ELT	Extraction, Chargement et Transformation des données.
ETL	Extraction, Transformation et Chargement de données.
ÉTS	École de technologie supérieure.
EVM	Gestion de la valeur acquise ou <i>Earned Value Management</i> .
GB	Grande Bibliothèque.
KPI	Indicateur clé de performances ou <i>Key Performance Indicator</i> .
MDM	Gestion des données maîtresses ou <i>Master Data Management</i> .
MEP	Mise en production.
NoSQL	Base de données non relationnelle.
ODS	Voute de données opérationnelles.
OWB	<i>Oracle Warehouse Builder</i> .
PMBok	<i>Project Management Body of Knowledge</i> .
PMI	<i>Project Management Institute</i> .
ROI	Rendement ou <i>Return on investment</i> .
SCD-#	<i>Slowly changing dimension</i> de type 1, 2 ou 3.
SGBD	Système de gestion de base de données.
SOA	Architecture orientée service.
SPI	<i>Schedule Performance Index</i> .
SQL	<i>Structured Query Language</i> .
TI	Technologie de l'information.
UQAM	Université du Québec à Montréal.

INTRODUCTION

Ce rapport de projet technique est réalisé dans le cadre de l'activité de synthèse du programme de maîtrise en génie logiciel et fait état d'un projet de refonte d'un entrepôt de données (DW) réalisé chez Bibliothèque et Archives nationales du Québec (BAnQ) de mai à novembre 2018.

D'une manière générale, les entrepôts de données (DW) mettent en commun les données provenant d'un vaste ensemble de systèmes opérationnels et permettent de bâtir des indicateurs de performances (KPI) pour mesurer et analyser la performance globale de l'entreprise. Les entrepôts offrent des fonctionnalités d'analyse de données avancées permettant aux spécialistes en intelligence d'affaires (BI) de faire des recherches très pointues et des corrélations qu'il serait impossible de faire autrement. BAnQ exploite un entrepôt de données depuis 2005, mais qui n'a que très peu évolué au fil des ans, qui engendre désormais des coûts importants de maintenance et qui ne couvre plus les besoins de la haute direction de l'institution.

Mon rôle sera de concevoir, de développer et de mettre en place une nouvelle solution d'extraction, de transformation et de chargement de données (ETL) pour l'entrepôt qui permettra de corriger les problèmes latents et de faire en sorte que les composantes technologiques ne soient plus un frein à l'évolution de l'entrepôt de données.

Le rapport est structuré en fonction des chapitres suivants :

- I. Une description du contexte, des problématiques, de la portée, des objectifs et des exigences de haut niveau ;
- II. Une revue de l'état de l'art quant à l'évolution des entrepôts de données ;
- III. Une description de la méthodologie employée et des résultats obtenus ;
- IV. Une interprétation des résultats quant à l'atteinte des objectifs ;
- V. Une discussion sur l'expérience et les leçons apprises.

CHAPITRE I

PROBLÉMATIQUE ET OBJECTIFS

1.1 Mise en contexte

Bibliothèque et archives nationales du Québec (BAnQ)

BAnQ est une société d'État relevant du Ministère de la Culture et des Communications du Gouvernement du Québec et qui regroupe plus de 800 employés répartis sur tout le territoire québécois. BAnQ fut créé en 2002 lors de la fusion avec la Bibliothèque nationale du Québec (BnQ) et la Grande Bibliothèque du Québec (GB). Une deuxième fusion en 2006 avec les Archives nationales du Québec (AnQ) est venu compléter sa mission.

Le plan stratégique 2015-2018 définit la mission de BAnQ comme suit :

« BAnQ rassemble, traite, conserve et met en valeur le patrimoine documentaire québécois et une vaste collection universelle au bénéfice des générations présentes et futures. Elle fournit également les services d'une bibliothèque publique d'envergure à toute la population du Québec.

Grâce au professionnalisme de son personnel, à ses collections, à ses activités, à ses lieux physiques répartis sur l'ensemble du territoire québécois et à sa présence dans l'univers virtuel, cette institution de mémoire et de savoir offre aux citoyens un accès démocratique à la culture et à la connaissance (Site web de BAnQ). »

BAnQ offre des services aux citoyens (bibliothèque, archives, patrimoine), mais également aux milieux documentaires (bibliothèques publiques, archives publiques, éditions, etc.). BAnQ souhaite utiliser des indicateurs clés de performance (KPI) pour évaluer la performance de chacune de ces trois directions générales de mission qui sont l'acquisition, la conservation et la diffusion de ressources documentaire, d'où l'importance d'avoir un entrepôt de données institutionnelles permettant de consolider toutes les sources d'informations pertinentes. Un nouveau plan stratégique est en préparation et nécessitera une révision des KPI et des sources d'information retenus.

Direction générale des technologies de l'information et des télécommunications

La Direction générale des technologies de l'information et des télécommunications (DGTIT) regroupe un peu plus de 50 techniciens et analystes en informatique qui doivent concevoir, développer et exploiter plus de 200 systèmes informatiques afin de soutenir les directions d'affaires de BAnQ dans leurs opérations.

Le plan directeur RI 2018-2021 définit la mission de la DGTIT comme suit :

« La Direction générale des technologies de l'information et des télécommunications de BAnQ soutient l'institution dans la réalisation de ses missions en assumant un leadership en matière de gestion des ressources informationnelles que sont les processus d'affaires informatisés, les documents et toutes les données numériques, les ressources logicielles et le matériel informatique. Elle sauvegarde et sécurise les documents et données numériques de BAnQ et y donne accès sur un portail Internet en constante évolution. Elle dirige un ensemble de projets nationaux et internationaux qui permettent à BAnQ de maintenir sa position de pionnière en matière de création de ressources virtuelles et numériques (Plan directeur RI). »

La plupart des systèmes informatiques et de l'infrastructure technologique furent implantés avant l'ouverture de la Grande Bibliothèque en 2005 et plusieurs de ces composantes sont aujourd'hui en fin de vie utile et doivent être remplacées.

Entrepôt de données institutionnelles

Le service de l'entrepôt de données institutionnelles est défini comme suit dans le catalogue des services TI de BAnQ.

« Entrepôt corporatif de données, dans lequel certains systèmes d'information envoient leurs propres données. Production de statistiques pour les différentes missions de BAnQ et rapports annuels (Catalogue de services TI). »

L'équipe interne est composée de trois personnes.

- Un architecte de données (moi) qui est responsable de l'analyse des besoins en intelligence affaires, de la modélisation de l'entrepôt, du développement des nouveaux processus ETL et du support de niveau 3.
- Un administrateur de base de données applicatif qui est responsable des tâches planifiées, de la mise à jour du système de gestion de la base de données, du support des anciens processus ETL et d'*Oracle Warehouse Builder*, et du support de niveau 2.
- Un administrateur de base de données infrastructure qui est responsable de vérifier le fonctionnement général du serveur et de la base de données et d'effectuer le support de niveau 1.

L'essentiel du développement de l'entrepôt de données et des traitements ETL a été réalisé en 2005 en marge de l'ouverture de la Grande Bibliothèque par des consultants ayant terminé leur mandat quelques mois plus tard. Il n'y a pas eu depuis de projet de refonte majeure des processus d'affaires de l'entrepôt faisant en sorte que

la couverture de l'entrepôt de données s'attarde quasi exclusivement aux processus d'affaires en lien avec la Grande Bibliothèque.

Les technologies ont beaucoup évolué depuis 2005 et les exigences des utilisateurs de l'entrepôt de données aussi. Les tendances vont vers les outils de présentation BI, l'intégration de l'intelligence artificielle et l'obtention de résultats près du temps réel. L'équipe interne souhaite un outil ETL facile à utiliser et produisant des traitements performants tout en offrant une grande maintenabilité.

1.2 Problématique

L'entrepôt a été conçu bien avant la fusion avec les archives nationales, l'intégration du centre de conservation et l'augmentation de l'importance des collections numériques. La couverture de l'entrepôt est devenue insuffisante au fil du temps et force depuis les directions à récolter les statistiques manuellement sans outil spécialisé pour l'analyse de l'information. De plus, les directions ne compilent pas toutes leurs données de la même façon depuis la fin du comité sur les statistiques institutionnelles.

L'entrepôt existant a été développé avec la suite logicielle *Oracle Warehouse Builder* (OWB) par quelques consultants sans transmission des connaissances à l'équipe interne. Le logiciel est spécialisé dans la conception et la réalisation des traitements ETL, mais celui-ci n'a pas été mis à niveau au fil des années par l'équipe interne de BANQ. De plus, OWB a été abandonné par Oracle il y a quelques années au profit de logiciels plus polyvalents. Le manque de temps accordé au service de l'entrepôt a conduit l'équipe interne à modifier directement le code source produit par l'outil, ce qui a rendu son utilisation très difficile. Aucun nouveau traitement n'a été ajouté depuis trois ans et l'obsolescence des outils complique la tâche pour la maintenance et l'évolution du système. Un projet de mise à niveau des bases de données

institutionnelles est d'ailleurs en attente du remplacement d'OWB qui est incompatible avec la prochaine version majeure du système de gestion de base de données (SGBD).

La modélisation multidimensionnelle de l'entrepôt de données est inadéquate, ce qui cause régulièrement des problèmes importants de performance dans le rafraichissement des jeux de données et la production des rapports. Plusieurs rapports récurrents sont produits manuellement et nécessitent beaucoup de temps de l'équipe interne. La lourdeur du travail fait en sorte que les utilisateurs ne peuvent parfois obtenir leurs informations qu'une fois par an ou avec plusieurs semaines de délai. Le suivi du plan stratégique, la production des rapports annuels et l'analyse de plusieurs indicateurs s'effectuent encore à l'aide de classeurs Excel échangés sur le réseau interne et nécessitent beaucoup d'effort à maintenir. Les activités opérationnelles de l'entrepôt de données tel que le suivi de l'exécution des traitements, l'analyse des données rejetées et la réalisation des demandes particulières des utilisateurs, accaparent beaucoup de temps de l'équipe interne, l'empêchant ultimement de consacrer du temps à l'amélioration du service. Les directions souhaiteraient pouvoir consulter, analyser et produire leurs analyses directement à partir des applications BI alors que l'équipe interne souhaiterait quant à elle réduire le temps passé en soutien et exploitation du service.

L'entrepôt de données a originalement été développé avec un budget raisonnable d'un millions de dollars (1 M \$), mais presque qu'aucun budget n'a été consacré à la maintenance et à l'évolution du service depuis plusieurs années. Le service souffre d'une dette technique suffisante pour nécessiter une refonte majeure. La haute direction souhaite que le service de l'entrepôt de données soit revalorisé à court et moyen terme afin de faciliter la reddition de compte et la vérification de l'atteinte des objectifs du plan stratégique de l'institution, mais elle ne dispose toutefois pas de

suffisamment de ressources humaines et monétaires pour rebâtir l'entrepôt à partir d'une feuille blanche.

1.3 Portée du projet

Programme de projet

En raison de la complexité et de l'ampleur importante du projet de refonte de l'entrepôt, il fut convenu avec les directions d'affaires que la refonte de l'entrepôt serait structurée en un programme de projets échelonnés sur une période de trois ans. La première phase priorisera la documentation des processus actuels et le remplacement de l'outil ETL. La deuxième phase ajoutera des processus d'affaires, des applications BI et réévaluera les KPI.¹

Échéance

Le projet a débuté dès mai 2018 et a été complété en novembre 2018 afin de répondre aux exigences académiques des projets techniques. L'échéance a permis également à la direction de m'affecter à un projet différent en janvier 2019.

Budget

La DGTIT a accepté d'investir un premier montant de 18 000 \$ avec une marge d'erreur acceptable de 15 % pour la réalisation de la première phase de la refonte. Considérant un taux horaire approximatif, incluant les bénéfices marginaux, à 60 \$/heure, l'effort alloué initialement au projet a été d'environ 300 heures (excluant la rédaction de ce rapport). Ce nombre d'heures correspond également à la contrainte académique d'un minimum de 270 heures nécessaires pour la réalisation d'un projet technique.

¹ Pour la suite de ce rapport, le terme « projet » sera utilisé pour couvrir les activités de la première phase de la refonte de l'entrepôt alors que le terme « refonte de l'entrepôt » servira à décrire le programme de projets dans son ensemble.

Ressources humaines

Il fut convenu avec mon directeur que j'étais pour être la seule personne dédiée au projet et que la banque d'heures planifiée était pour mon usage exclusif. Un effort approximatif de 20 h/semaine pendant 15 semaines devait permettre de consacrer suffisamment de temps au projet sans trop nuire à mes autres responsabilités.

Mes collègues DBA ont été consultés à l'occasion, mais leur mandat consistait essentiellement à la mise en place de la nouvelle infrastructure technologique (nouveaux serveurs, système d'exploitation, SGBD, migration, etc.) et au soutien de l'entrepôt de données actuel pendant que je travaille sur la nouvelle plateforme. Leur temps a donc été imputé à d'autres projets que celui-ci.

Contenu

En raison des contraintes importantes quant au budget et à l'échéance du projet, la portée du projet s'est limitée aux éléments suivants :

- Une révision des processus d'affaires BI pour bien évaluer les opportunités d'amélioration à l'entrepôt et permettre aux directions de bénéficier d'outils analytiques. Certains processus d'affaires à prioriser ont déjà été identifiés et le projet tentera d'incorporer ces nouvelles sources si le temps le permet.
- Le remplacement d'OWB par une nouvelle plateforme ETL compatible avec l'architecture d'entreprise de BANQ et qui inclut la conversion de tous les traitements ETL existants.
- L'analyse de la modélisation multidimensionnelle et des outils de présentation afin de diminuer les délais de traitements des demandes provenant des directions d'affaires.

1.4 Objectifs

Le projet a pour objectifs de diminuer le temps de collecte et d'analyse des données, de réduire les coûts de maintenance de l'entrepôt de données et d'améliorer la qualité et la cohérence des données. Chaque objectif est composé d'une cible à atteindre, d'une justification et de moyens pour les réussir.

Objectif 1 : Diminuer le temps de collecte et d'analyse des données

- Cible : Diminuer le temps total annuel de 25 %.
- Justification : La collecte de l'information ne se fait bien souvent qu'une fois par jour ou par mois en raison du temps de traitement important que nécessite le processus ETL. Il en va de même pour la production de rapports qui ne se fait bien souvent qu'une fois par trimestre ou par année en raison des délais importants dans la préparation des données.
- Moyens :
 - Les demandes *ad hoc* récurrentes seront davantage automatisées.
 - Les jeux de données et rapports inutiles seront retirés.
 - Les rapports nécessitant plus de cinq minutes seront optimisés.
 - De nouveaux rapports sur les sources existantes seront ajoutés.
 - Les temps d'exécution des traitements ETL seront réduits.
 - Les temps réponses seront journalisés.
 - L'indexation et les options de stockage seront optimisées.
 - Les calculs complexes seront automatisés la nuit.

Objectif 2 : Réduire les coûts de maintenance de l'entrepôt de données

- Cible : Diminuer les coûts annuels de maintenance de 50 %.
- Justification : L'architecture actuelle de l'entrepôt de données est très complexe et nécessite des compétences multiples dont ne dispose pas toute l'équipe. L'outil

Oracle Warehouse Builder n'est plus utilisé, mais l'équipe interne doit continuer à maintenir le code source qu'il a généré.

- Moyens :
 - OWB sera remplacé.
 - L'architecture logicielle sera simplifiée.
 - L'infrastructure technologique sera mise à niveau.
 - La journalisation des traitements sera améliorée.
 - Un mécanisme de relance des traitements ETL sera mis en place.
 - L'effort en soutien et maintenance sera diminué.
 - La maintenabilité des traitements ETL sera améliorée.
 - Le temps nécessaire pour ajouter une source de données sera diminué.

Objectif 3 : Améliorer la qualité et la cohérence des données

- Cible : Diminuer le nombre d'incohérences de résultat de 50 %.
- Justification : De nombreux rapports ne balancent pas et certaines données historiques bougent dans le temps, engendrant une perte de confiance des utilisateurs. Les données rejetées par les traitements ETL actuels ne sont pas vérifiées et corrigées.
- Moyens :
 - Les requêtes de service seront comptabilisées.
 - Les problèmes d'intégrité de données seront corrigés.
 - Des contraintes d'intégrité seront ajoutées.
 - Les résultats des traitements ETL seront comparés aux anciens.
 - Des validations seront ajoutées dans les traitements ETL.
 - Les données rejetées seront remontées au système source.
 - Un mécanisme de suivi des données rejetées sera implanté.
 - Les rejets et les validations seront journalisés.

1.5 Exigences de haut niveau

Kimball fait mention de six objectifs de haut niveau que doit respecter tout entrepôt de données (Kimball et al, 2008). Les objectifs de Kimball sont considérés comme étant des exigences de haut niveau pour ce projet. Chaque exigence est constituée de sa justification et ses implications pour le projet.

L'entrepôt de données doit fournir un accès aux informations stratégiques de l'entreprise (Accès aux données).

- Justification : L'entrepôt doit être appuyé sur toutes les sources de données ayant une incidence sur les objectifs stratégiques de l'entreprise.
- Implication : Dans le cadre de ce projet, cet objectif ne sera pas atteint et devra attendre la phase 2 du projet de refonte.

Les données de l'entrepôt de données doivent être cohérentes dans le temps (Cohérence des données).

- Justification : L'entrepôt doit être en mesure de produire une image instantanée des données à un moment précis.
- Implication : Dans le cadre de ce projet, la gestion des changements (SCD) de l'information utilisée devra permettre de conserver l'historique des modifications.

Les données d'un entrepôt de données doivent être modélisées de sorte qu'elles peuvent être utilisées pour tout besoin d'analyse (Dimensions conformes).

- Justification : L'entrepôt utilise une modélisation multidimensionnelle qui permet de réutiliser les dimensions conformes et de faire des croisements de données.
- Implication : Dans le cadre de ce projet, l'information sera systématiquement modélisée dans un schéma en étoile en utilisant le niveau de grain le plus approprié.

L'entrepôt de données n'est pas constitué que des données, mais aussi d'un ensemble d'outils permettant d'interroger, d'analyser et de présenter l'information (Gestion des données).

- Justification : Le plus gros de l'effort d'un projet est consacré à la gestion des données (ETL) plutôt qu'à l'ajout d'outil d'analyse (BI).
- Implication : Dans le cadre de ce projet, la plus grande partie de l'effort sera consacré aux activités de développement ETL plutôt qu'au BI.

L'entrepôt de données publie des données utiles et exploitables (Qualité des données).

- Justification : La qualité des données est plus importante que la quantité de données contenues dans l'entrepôt.
- Implication : Dans le cadre de ce projet, la structure et la pertinence de tous les jeux de données de l'entrepôt seront évaluées afin d'en déterminer leur utilisation et leur pertinence. Les jeux de données inutiles ou inutilisables seront retirés.

La qualité des informations produites doit être un moteur pour la transformation des processus de l'entreprise (Aide à la décision).

- Justification : Les informations obtenues de l'entrepôt de données permettent d'améliorer la prise de décision.
- Implication : Dans le cadre de ce projet, aucune amélioration significative ne sera réalisée dans la phase 1 du projet de refonte de l'entrepôt.

1.6 Facteurs de succès

Kimball fait également mention de trois conditions essentielles pour le succès d'un projet d'entrepôt de données (Kimball et al, 2008). Chaque facteur clé de succès est composé des risques de ne pas le satisfaire en plus de ses implications pour le projet.

Parrainage de la haute direction

- Justification : Parrainage de la haute direction qui recherchera le succès à long terme et qui sera en mesure de prioriser les processus d'affaires et faire participer les différentes directions au projet.
- Risque : Sans le parrainage du projet, il est probable que le financement du projet soit refusé ou qu'il soit stoppé à tout moment advenant que survienne un problème impactant une autre direction d'affaires.
- Implication : Le projet devra être approuvé par le directeur général des technologies de l'information et des télécommunications ainsi que présenter à une majorité de directeurs de différentes directions.

Motivation et légitimité du changement

- Justification : Motivation et légitimité du changement en raison d'un besoin urgent d'améliorer l'accès à de l'information de qualité et en raison de la dette technologique importante qu'a accumulée le service au fil des ans.
- Risque : La pertinence du projet pourrait être compromise si des données essentielles à la prise de décision ne sont pas disponibles pour analyse dans l'entrepôt.
- Implication : Le projet devra démontrer que l'entrepôt de données sera prêt à temps pour l'ajout des indicateurs clés de performance du prochain plan stratégique de BANQ afin d'être réellement utile pour le suivi du prochain plan.

Faisabilité technique

- Justification : La faisabilité technique en fonction de la disponibilité de données sources de qualité et d'un effort raisonnable pour traiter les données en information pertinente.

- Risque : La faisabilité sera compromise si les données essentielles à la prise de décision ne sont pas disponibles ou ne sont pas exploitables par l'entrepôt de données.
- Implication : Il n'y a pas de risque pour la première phase du projet de refonte, car les jeux de données sont déjà bien connus, mais il faudra s'assurer que les outils sont mis en place pour couvrir les besoins futurs de l'entrepôt.

CHAPITRE II

ÉTAT DE L'ART

Cette section présente une brève revue de la littérature sur l'évolution des entrepôts de données, depuis leur création jusqu'à aujourd'hui, sur les différents styles d'architecture et sur l'avènement de la gestion des données maitresses et des mégadonnées.

En 2018, Keith D. Foote nous rappelle que l'entrepôt de données conventionnel a été conçu vers la fin des années 80 en réponse à un problème d'éparpillement de l'information provoqué par l'utilisation du style d'architecture client-serveur au détriment des ordinateurs centraux (Foote 2018). Son utilisation dans l'industrie s'est démocratisée dans les années 90 à mesure que les organisations réalisaient qu'il devenait de plus en plus nécessaire d'obtenir une vision globale et unifiée de l'information. Au cours des années 2000, l'accent fut davantage mis sur les applications BI et les indicateurs clés de performance (KPI) de leurs processus d'affaires. Depuis les années 2010, les concepts de mégadonnées, de lac de données, de fouille de données et d'intelligence artificielle préoccupent les chercheurs qui cherchent des moyens de réunifier des masses d'information non nettoyée, non structurée et disséminée sur de plus en plus d'appareils différents.

En 2004, Margy Ross du groupe Kimball faisait une comparaison des deux grands courants d'architecture en entrepôt de données de l'époque, soit l'architecture en bus de Ralph Kimball et l'entrepôt de données corporatif de Bill Inmon (Ross, 2004).

Kimball introduit les concepts de modélisation en étoile, de dimensions conformes, de faits, de développement itératif et d'agilité appliqués au développement des entrepôts de données. Le modèle propose que l'entrepôt de données consolide les données clés de l'entreprise dans un schéma multidimensionnel favorisant la performance tout en utilisant une méthodologie qui favorise le développement itératif et la réutilisation des dimensions. Inmon proposa plutôt un modèle où toutes les données transactionnelles de l'entreprise seraient centralisées dans une seule base de données normalisée (3FN) qui servirait ensuite de point unique pour le partage de l'information entre systèmes et pour alimenter des magasins de données de l'entrepôt (Inmon et al, 2002). Selon Ross, le modèle de Kimball a jusqu'ici connu davantage de succès en raison de sa simplicité et sa rapidité de mise en œuvre. Le modèle de Inmon nécessite un effort considérable de préparation en amont, mais a l'avantage de s'intégrer facilement à une architecture orientée service (SOA) et au référentiel de données maîtresses (MDM). Devant la complexité de la mise en place de son modèle, Inmon revint à la charge en 2008 avec une approche plus souple afin de remplacer la modélisation 3FN par une modélisation en voute de données (ODS) (Inmon et al, 2008). Ross mentionne que les deux approches ne sont pas complètement incompatibles et qu'il est tout à fait possible de commencer par l'approche Kimball puis d'y ajouter éventuellement un ODS qui servirait alors d'unique système source pour l'entrepôt de données (Ross, 2004).

En 2017, Stuart Birkbeck se questionne si les deux approches ne sont pas devenues toutes les deux désuètes avec l'avènement des mégadonnées, des bases de données *NoSQL* et de la modélisation sous forme de lacs de données (*data lake*). Bickbeck fait état des inconvénients des approches traditionnelles qui nécessitent beaucoup d'effort de préparation et de nettoyage de données. Les lacs de données permettent d'injecter toutes les données brutes dans l'entrepôt sans tenter de les nettoyer ou de les transformer au préalable. La transformation des données sera gérée plus tard par les applications BI (Birkbeck, 2017). Christopher Tozzi définit les données massives non

pas par leur taille, mais plutôt par le niveau de leur transformation (brute ou raffinée) (Tozzi, 2018). Birkbeck juge qu'il est préférable de prévoir un certain effort de consolidation en amont afin d'éviter les problèmes de performance, le désordre potentiel et la complexité inutile dans les applications BI. Finalement, Birkbeck croit que les deux approches vont tendre à se compléter en permettant à l'entrepôt de données de stocker ou de référer des données semi-structurées et semi-raffinées (Birkbeck, 2017). Slater Victoroff du groupe *Indico Data Solutions* pense quant à lui qu'un grand volume de données ne pose pas réellement de problème pour un entrepôt de données conventionnel du moment qu'un effort minimal pour nettoyer les données a été effectué sur la source (Victoroff, 2015). Le groupe Gartner a révisé ses prévisions à la baisse quant à l'adoption des technologies liées aux mégas données en raison de sa complexité et des nouveautés offertes par les solutions infonuagiques pour les SGBD conventionnels (Meulen, 2017).

Depuis quelques années, le principal sujet de discussion en architecture de données concerne la gestion des données maitresses (MDM) et les systèmes pour les gérer. Le groupe *Talend* définit le MDM comme étant un processus pour assurer à l'organisation qu'elle pourra toujours prendre ses décisions sur la base d'une seule et « vraie » version de l'information appelée *golden record* (Talend, 2018). Les MDM s'inspirent de l'architecture de Inmon et centralisent les entités les plus importantes de l'entreprise (ex. les employés, les produits, les adresses, etc.) à un seul endroit en utilisant une modélisation canonique favorisant la réutilisation maximale de l'information. Dans sa forme la plus pure de MDM transactionnel, le modèle fait en sorte que plus aucun autre système n'a à modéliser et à stocker lui-même les données maitresses et qu'il devrait plutôt faire appel au MDM pour lui fournir l'information sous la forme *Data as a Service* (DaaS). La mise en place d'un entrepôt de données ne devient alors qu'une simple formalité, car le MDM se charge déjà d'assurer que l'information soit unique, globale et cohérente pour tous les systèmes informatiques de l'entreprise. L'approche la plus répandue pour le moment, et qui se rapproche du

modèle d'ODS de Inmon, est l'utilisation du style MDM consolidé où l'information demeure dans chaque système source, mais est systématiquement répliquée et consolidée dans la MDM. Dans tous les cas, l'implémentation d'un entrepôt de données est généralement le premier jalon pour la mise en place d'une MDM dans une organisation. Les principaux fournisseurs d'outils ETL tels que *IBM*, *Informatica* et *Talend* proposent tous des suites logicielles MDM puisque ces systèmes ont beaucoup utilisé les fonctionnalités des outils ETL en complément. Les MDM nécessitent toutefois une maturité importante en architecture d'entreprise et beaucoup de temps à mettre en place. Selon le cycle de maturité produit par le groupe *Gartner* en 2017, la technologie MDM était au plus creux du gouffre des désillusions et nécessitera encore quelques années avant d'être suffisamment mature pour devenir une pratique répandue dans l'industrie (Meulen, 2017).

CHAPITRE III

MÉTHODOLOGIE

3.1 Processus et activités

La méthodologie employée pour ce projet est calquée sur la méthode *Data Warehouse Lifecycle* conçue par Ralph Kimball depuis les années 80 et décrite dans l'ouvrage *The Data Warehouse Lifecycle Toolkit* (Kimball et al, 2008). La figure 3.1 présente la vue d'ensemble du processus traduite par l'ÉTS (Chafky et Desrosiers, 2011).

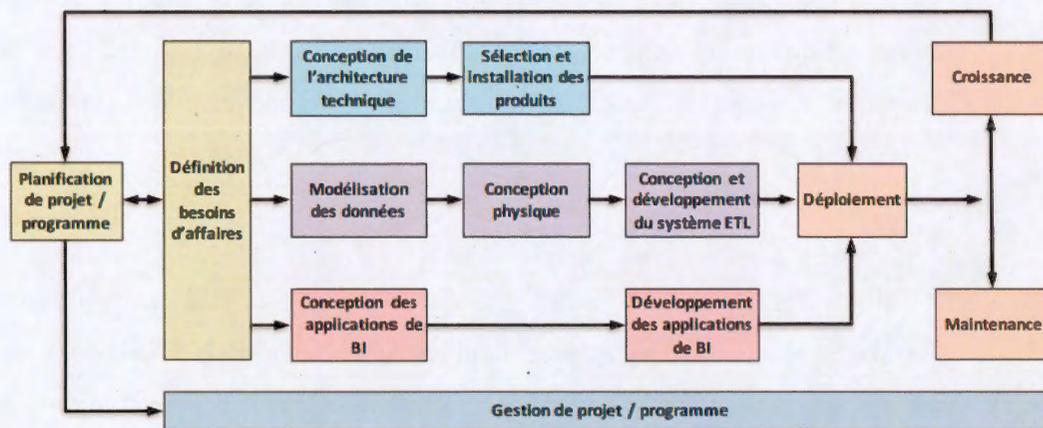


Figure 3.1 Vue d'ensemble de la méthode *Data Warehouse Lifecycle*

Il s'agit d'un processus de développement itératif qui a pour objectif de couvrir toutes les activités d'analyse, de conception, de réalisation et d'implantation d'un entrepôt de données au sein d'une organisation. La méthode vise à implanter graduellement de nouveaux processus d'affaires, de nouvelles sources de données et de nouvelles applications BI en priorisant ceux qui donneront le plus de valeur ajoutée à l'entreprise.

1. Planification et gestion du projet

1.1. Planification du projet : Définir les objectifs, la portée, les exclusions, les parties prenantes, les coûts, les risques, les bénéfices, le rendement (ROI) et les principales tâches du projet dans une charte de projet de quelques pages. Pour ce projet, il est prévu de produire la charte sous forme d'une courte présentation remise et présentée aux membres du comité de direction DGTIT.

1.2. Gestion du projet : Contrôler le projet et l'ajuster si un écart se creuse en cours de route entre l'avancement réel et celui planifié. Il est prévu d'utiliser le système de feuilles de temps interne de BANQ afin de suivre l'effort réel et l'évolution des coûts sur une base mensuelle en la comparant avec le budget.

2. Définition des besoins d'affaires

2.1. Définition des besoins d'affaires : Établir une liste priorisée des processus d'affaires documentant les attentes du client, la granularité des données et les dimensions d'analyses. La méthode Kimball repose sur les entrevues réalisées avec les parties prenantes de l'entrepôt pour collecter leurs besoins. Dans ces entrevues, on discute, entre autres, des processus d'affaires, de la granularité attendue, des besoins analytiques, des attentes du projet et des problématiques actuelles de l'entrepôt. Pour ce projet, il n'est toutefois pas

prévu de réaliser d'entrevues dans le cadre de la phase 1 de la refonte puisque la portée se limitera au remplacement de l'infrastructure et des traitements ETL. Le projet inclura toutefois la réalisation d'une matrice en bus des données afin de documenter les processus existants.

3. Définition de l'architecture technique

3.1. Conception de l'architecture technique : Évaluer les besoins non fonctionnels de l'entrepôt (performance, maintenabilité, etc.), les types de systèmes ETL pertinents, les systèmes sources, les technologies utilisées, l'intégration des données et la conformité à l'architecture d'entreprise. Les besoins non fonctionnels et l'architecture logicielle seront documentés dans l'outil de gestion des connaissances de BANQ et représentés par des diagrammes UML afin de présenter l'architecture actuelle, l'architecture cible et le *delta*.

3.2. Sélection et installation des produits : Comparer et évaluer les différents produits disponibles sur le marché et autres solutions maison en fonction de l'architecture choisie. Pour ce projet, les outils et technologies présentement disponibles chez BANQ ainsi que les suites commerciales se trouvant dans le cadran magique du groupe *Gartner* seront étudiées (Beyer et al, 2017). Un prototype de la solution retenue sera mis en place pour en vérifier ses qualités.

4. Définition des données et de leur transformation

4.1. Modélisation des données : Détailler chacun des magasins de données à intégrer à l'entrepôt afin de déterminer quels seront les faits à analyser, leurs granularités, les dimensions conformes et leur hiérarchie, les stratégies de modélisation en étoile ou en nuage, et la gestion du changement des données.

Pour ce projet, des diagrammes d'entité-relation (ERD) par domaine d'affaires seront conçus.

4.2. Conception physique : Modéliser physiquement les tables, les contraintes d'intégrité, les clés de remplacement, les index, les options de stockage, le partitionnement, etc. en fonction du SGBD. Dans le cadre de ce projet, une analyse approfondie de la conception physique existante sera effectuée, car elle présente plusieurs défauts d'intégrité et de performance. La structure de la base de données sera modifiée au besoin.

4.3. Conception et développement du système ETL : L'activité consiste à établir les méthodes d'extraction de données, à archiver les données extraites, à établir les normes pour la sélection et le nettoyage de données, à gérer les SCD, à gérer la séquence des événements, à utiliser un système de journalisation, et à préparer l'environnement de transformation des données (*staging area*). Pour ce projet, les traitements existants seront réécrits avec les mêmes entrées et les mêmes sorties autant que possible. Le livrable sera le code source des traitements ETL.

5. Définition des besoins d'intelligence d'affaires

5.1. Conception des applications BI : Concevoir les applications BI et les rapports en fonction du niveau de détail et des fonctionnalités attendues par les utilisateurs selon leur type d'intérêt. Certains utilisateurs n'ont besoin que de tableaux de bord alors que d'autres veulent effectuer des fouilles de données très complexes. Pour ce projet, les applications BI actuelles seront testées après chaque itération pour en assurer le fonctionnement et une analyse d'impact sur l'ajout d'un nouveau processus d'affaires quant à la consultation des documents numériques sera effectuée.

5.2. Développement des applications BI : Développer les rapports, les graphiques, les tableaux, les cubes de données, les filtres et la navigation. Pour mon projet, un nouveau jeu de données sera ajouté pour l'analyse du nombre de consultations des documents numériques et nécessitera de nouveaux rapports et de nouveaux cubes de données.

6. Déploiement, amélioration et maintenance du système

6.1. Déploiement : Réaliser les tests d'intégration et de déploiement à la fin de chaque itération, livrer les artefacts dans l'environnement de production et retirer les anciens traitements ETL. Pour ce projet, un environnement de préproduction sera monté afin de permettre une exécution côte-à-côte des nouveaux et des anciens traitements afin de déceler les anomalies.

6.2. Maintenance : Vérifier quotidiennement le bon déroulement des traitements ETL à partir des fonctionnalités de surveillance et de gestion des erreurs. La performance globale des traitements, le nombre d'enregistrements traités et le nombre de rejets seront analysés par moi et l'équipe interne. J'effectuerai les correctifs au besoin.

6.3. Croissance : Évaluer et prioriser les demandes d'amélioration continue de l'entrepôt de données reçues par les utilisateurs du système. Pour ce projet, aucune amélioration continue n'est prévue en cours de projet comme tel, mais les itérations seront effectuées en fonction de la priorisation des processus d'affaires et des traitements ETL jugés prioritaires par les directions d'affaires.

3.2 Présentation des résultats

3.2.1 Planification et gestion du projet

3.2.1.1 Planification du projet

Une présentation de 20 minutes du projet en avril 2018 fut préparée à l'attention du comité de direction de la direction générale des technologies de l'information et des télécommunications (DGTIT) afin d'y présenter l'état de la situation actuelle de l'entrepôt de données, des problématiques vécues par les utilisateurs, de l'architecture du service et des outils désuets. Le comité a accepté un premier projet à temps partiel de mai à novembre 2018 qui minimise les impacts sur la disponibilité du service pour les utilisateurs et limite les impacts sur les autres systèmes informatiques. Les membres du comité ont imposé comme contrainte que le projet soit limité à des ressources internes et qu'il n'augmente pas les coûts d'exploitation du service. Un représentant d'une direction d'affaires a demandé que soit tout de même inclus dans le projet l'ajout d'un nouveau processus d'affaires sur la consultation des documents numériques dès la phase 1 du projet puisque toute l'analyse avait déjà été complétée. La définition de la portée permet de faciliter les choix d'architecture et de concentrer mes énergies sur le développement ETL plutôt que sur les applications BI. La charte de projet s'est résumée en une présentation PowerPoint incluse à la fin de ce rapport dans l'annexe A.

3.2.1.2 Gestion du projet

Le projet avait initialement été estimé à 300 heures, mais un effort supplémentaire de 77 heures fut nécessaire pour réaliser la portée initiale. Un contrôle de la performance du projet fut réalisé mensuellement en utilisant les techniques d'évaluation de la valeur acquise provenant du *Project Management Body of Knowledge* (PMBoK) (Battendier, 2009). La figure 3.2 présente le portrait de l'avancement et des coûts réels du projet à la fin des activités de développement.

No	Norm	Plan value (PV)	Mai	Juin	Juillet	Août	Septembre	Octobre	Actual Cost (AC)	Earned Value (EA)	Schedule Variance (SV)	Cost Variance (CV)	Schedule Performance Index (SPI)	Cost Performance Index (CPI)
1.	Définition du projet	10	5	1	1	1	1	1	10	10,0	0,0	0,0	1,00	1,00
1.1.	Planification de projet	5	4						4	5,0	0,0	1,0	1,00	1,25
1.2.	Gestion de projet	5	1	1	1	1	1	1	6	5,0	0,0	-1,0	1,00	0,83
2.	Définition des besoins d'affaires	20	13	0	0	1	2	0	16	15,0	-5,0	-1,0	0,75	0,94
2.1.	Définition des besoins d'affaires	20	13			1	2		16	15,0	-5,0	-1,0	0,75	0,94
3.	Définition de l'architecture technique	30	23	23	2	0	0	0	48	30,0	0,0	-18,0	1,00	0,63
3.1.	Conception de l'architecture technique	10		10					10	10,0	0,0	0,5	1,00	1,05
3.2.	Sélection et installation des produits	20	23	14	2				39	20,0	0,0	-18,5	1,00	0,52
4.	Définition des données	200	0	49	86	35	62	13	245	200,0	0,0	-44,8	1,00	0,82
4.1.	Modélisation des données	50		19	22	5	7	1	54	50,0	0,0	-3,5	1,00	0,93
4.2.	Conception physique	50		6	32	7	20	2	65	50,0	0,0	-15,0	1,00	0,77
4.3.	Conversion et dév. du système ETL	100		24	33	24	36	10	126	100,0	0,0	-26,3	1,00	0,79
5.	Définition des besoins BI	10	0	0	0	8	0	0	8	10,0	0,0	2,5	1,00	1,33
5.1.	Conception des applications BI	5				3			3	5,0	0,0	2,0	1,00	1,67
5.2.	Développement des applications BI	5				5			5	5,0	0,0	0,5	1,00	1,11
6.	Définition des besoins post-dév.	30	0	0	0	21	23	7	51	30,0	0,0	-20,8	1,00	0,59
6.1.	Déploiement	5				4	2	2	7	5,0	0,0	-2,0	1,00	0,71
6.2.	Croissance et maintenance	25				17	22	5	44	25,0	0,0	-18,8	1,00	0,57
	Total	300	41	73	89	66	88	21	377	295,0	-5,0	-82,1	0,98	0,78

Figure 3.2 Analyse budgétaire des activités du projet

À noter que seules mes heures ont été comptabilisées dans le tableau et que l'effort de l'équipe interne (environ 35 h) a été exclu volontairement. La figure montre que l'indicateur de performance *Cost Performance Index* (CPI) affiche une valeur de 0,78 indiquant un dépassement de cout important. L'indicateur *Schedule Performance Index* (SPI) affiche toutefois une valeur de 0,98 indiquant que l'essentiel de la portée initiale fut livré dans les temps planifiés.

Une sous-évaluation des activités de « sélection et installation des produits » et de « croissance et maintenance » ajoutée à des complications mineures survenues dans les activités de « développement ETL » ont nécessité une réévaluation de la portée du projet en milieu de projet. Les activités « d'analyse d'affaires » et « d'intelligence d'affaires » ont ainsi été partiellement reportées à la phase 2 afin de concentrer les activités sur le développement ETL.

3.2.2 Définition des besoins d'affaires

L'essentiel du temps passé dans cette activité a été consacré à la documentation des processus d'affaires et des traitements ETL déjà présents dans l'entrepôt de données actuel. Les traitements ETL n'avaient aucune documentation écrite et il a fallu analyser le code source des traitements ETL existants pour la rebâtir. Pour ce faire, j'ai produit une documentation préliminaire par un processus d'ingénierie-inverse à partir des principales étapes des anciens processus en interrogeant le référentiel de l'outil *Oracle Warehouse Builder*. Ensuite, j'ai complété manuellement la documentation en faisant une revue minutieuse de l'ancien code source afin de m'assurer qu'aucune règle de validation et de transformation n'ai été oublié.

Au terme du projet, les 12 processus d'affaires existants furent analysés et un 13^e fut ajouté. Les travaux de la phase 2 permettront de réévaluer les besoins analytiques des directions et d'ajouter les processus d'affaires manquants les plus importants. La liste

des processus d'affaires inclus dans la phase 1 du projet est détaillée dans le tableau 3.1.

Tableau 3.1 Processus d'affaires inclus dans la phase 1

SIGLE	DESCRIPTION	AMPLEUR
AUTH	Données sur l'utilisation des systèmes informatiques en fonction de la journalisation de l'authentification aux systèmes.	0 dim. 1 fait
BN	Données sur les documents numériques consultés sur la plateforme BAnQ Numérique. Il s'agit d'un nouveau traitement priorisé dans le cadre de ce projet.	3 dim. 1 fait
C2	Données sur la disponibilité des systèmes informatiques, des incidents et des pannes saisies dans le système de demandes de services informatiques.	1 dim. 1 fait
FREQ	Données sur la fréquentation des espaces physiques provenant des portillons et des activités d'évaluation effectuées par le personnel.	0 dim. 2 faits
GRC	Données sur les abonnés, les institutions et les demandes de service provenant du système de gestion de la relation client.	2 dim. 2 faits
GU	Données sur les téléchargements de notices bibliographiques du service québécois de traitement documentaire.	5 dim. 8 faits
PEB	Données sur les prêts entre bibliothèques.	9 dim. 1 fait
RH	Données sur la structure organisationnelle, les postes et les emplois à temps complet provenant du système de gestion des ressources humaines.	6 dim. 1 fait
SAM	Données sur l'utilisation des postes publics.	1 dim. 2 faits
SIGB	Données sur le catalogue, les acquisitions et la circulation provenant du système intégré de gestion de bibliothèque.	20 dim. 2 faits
SV	Données sur la fréquentation du portail BAnQ.	0 dim. 1 fait
TEL	Données sur les appels reçus à partir du système téléphonique.	0 dim. 1 fait.
WIFI	Données sur l'utilisation du réseau sans-fil public des espaces de consultation.	0 dim. 2 faits.

Une matrice en bus de données a ensuite été produite en croisant les informations sur la modélisation multidimensionnelle des différents processus de l'entrepôt de données. La matrice soulève un problème au niveau d'un nombre trop élevé de dimensions conformes en raison d'une modélisation en flocons un peu trop étendue. Un extrait de cette matrice est présenté dans la figure 3.3 et la matrice complète est disponible sur demande.

	dimensions conformes	date (temps)	heure (time)	région postal (rta)	usager	localisation	fournisseur	sous-localisation	catégorie documentaire	politique de circulation	état de disponibilité	organisation depositaire	raison desherbage	type note exemplaire	budget	notice d'autorité	notice bibliographique	caract. d'exemplaire	exemplaire	année d'édition	type d'élagage	type de note	info dépôt légal
processus d'affaires		56	1	1	20	14	2	5	5	0	2	1	1	0	0	0	7	0	12	1	1	1	1
exemplaire acquis (ac)	8	1				2	1										1		1				1
facture (ac)	5	2				1	1																
réquisition (ac)	6	4				1											1						
dépôt temporaire (ca)	8	3				1		1	1		1								1				
desherbage (ca)	5	1										1	1						1		1		
note d'exemplaire (ca)	3	1																	1			1	
suppression (ca)	3	2																					
disponibilité (ci)	3	1						1												1			
événement de circulation (ci)	10	1	1		1	1		1	1								1		1				
frais (ci)	6	1			2	1													1				
perception (ci)	5	1			1	1																	
prêt contesté (ci)	4	1			1	1													1				
réservation (ci)	9	3			1	2			1								1		1				
transfert (ci)	7	3				3																	
état de la collection	3	1															1		1				
taux de roulement	3	1																					

Figure 3.3 Extrait de la matrice en bus de données réalisée pour le projet

La documentation des processus et la réévaluation des besoins avec les directions d'affaires n'ont pas été entièrement réalisées lorsqu'il fut évident que le temps manquerait pour réaliser toutes les activités du projet. Un espace de travail a été préparé dans l'outil de collaboration, mais il reste à réévaluer les objectifs et les raisons d'être des processus à l'aide d'entrevues avec les utilisateurs.

3.2.3 Définition de l'architecture technique

3.2.3.1 Conception de l'architecture technique

L'activité consista à documenter l'architecture existante de l'entrepôt de données afin de déterminer précisément les problèmes logiciels et les impacts potentiels d'une modification sur l'infrastructure. La liste suivante présente les principales composantes ayant nécessité une réflexion :

- Serveur OWB : Le logiciel OWB a été abandonné depuis plusieurs années par Oracle qui recommande son remplacement. Après une analyse des outils disponibles sur le marché et une discussion avec les DBA, il fut convenu qu'une solution développée en interne en utilisant des outils déjà disponibles suffirait à l'équipe interne pour les prochaines années. Le groupe Kimball recommande d'utiliser un outil ETL commercial autant que possible sauf si les besoins sont bien connus et que les moyens de l'organisation sont très limités (Kimball et al, 2008).
- Client OWB : L'outil graphique de modélisation des processus n'était plus utilisé depuis plusieurs années par l'équipe interne qui préférait modifier directement le code source généré par l'outil. Il fut convenu qu'une plateforme ETL légère serait développée sans outil graphique. Une application web facilitant le paramétrage et le suivi des traitements sera développée plus tard.
- Programmation *Shell* : Le fonctionnement des traitements reposait sur une multitude de scripts (*Shell script*) pêle-mêle localisés sur le serveur de l'entrepôt. Ces scripts étaient très difficiles à maintenir alors que l'expertise de l'équipe interne est surtout dans les bases de données. Il fut convenu de remplacer tous ces scripts et de les intégrer dans la plateforme ETL.
- SGBD : La base de données de l'entrepôt résidait sur le même serveur et utilisait les mêmes ressources que les bases de données opérationnelles de

l'institution, ce qui pouvait poser des problèmes de performance importants aux systèmes de mission. Les besoins des systèmes transactionnels et des entrepôts de données sont très différents et il arrivait à l'occasion que l'entrepôt nuise au fonctionnement des systèmes opérationnels. Il fut convenu de placer l'entrepôt de données sur son propre serveur avec ses propres licences où elle pourra utiliser à fond les ressources matérielles (CPU, mémoire, disque) sans impacter les opérations.

- Infrastructure technologique : Les capacités de la machine physique, la version du système d'exploitation et la version de la base de données de l'entrepôt sont vieillissantes, mais les mises à niveau sont bloquées par nos outils ETL. Cela expose le serveur à des failles de sécurité importantes en plus de la perte du support des fournisseurs. Il fut convenu que le retrait des composantes vulnérables était prioritaire afin de mettre à niveau l'infrastructure.
- *Reporting* : Il a été convenu que la solution de rapport sera remplacée dans la phase 2 de la refonte de l'entrepôt. Il reste à déterminer les besoins des utilisateurs à savoir si l'on demeure avec des rapports personnalisés développés à l'interne ou si l'on acquiert une solution commerciale tel que Tableau Server.

Une fois les problématiques identifiées, une architecture cible et une carte routière des modifications furent élaborées afin de simplifier l'architecture technologique et de ramener l'essentiel des composantes à l'intérieur de la base de données de l'entrepôt. Les figures 3.4 et 3.5 présentent les diagrammes d'architecture logicielle avant et après le projet.

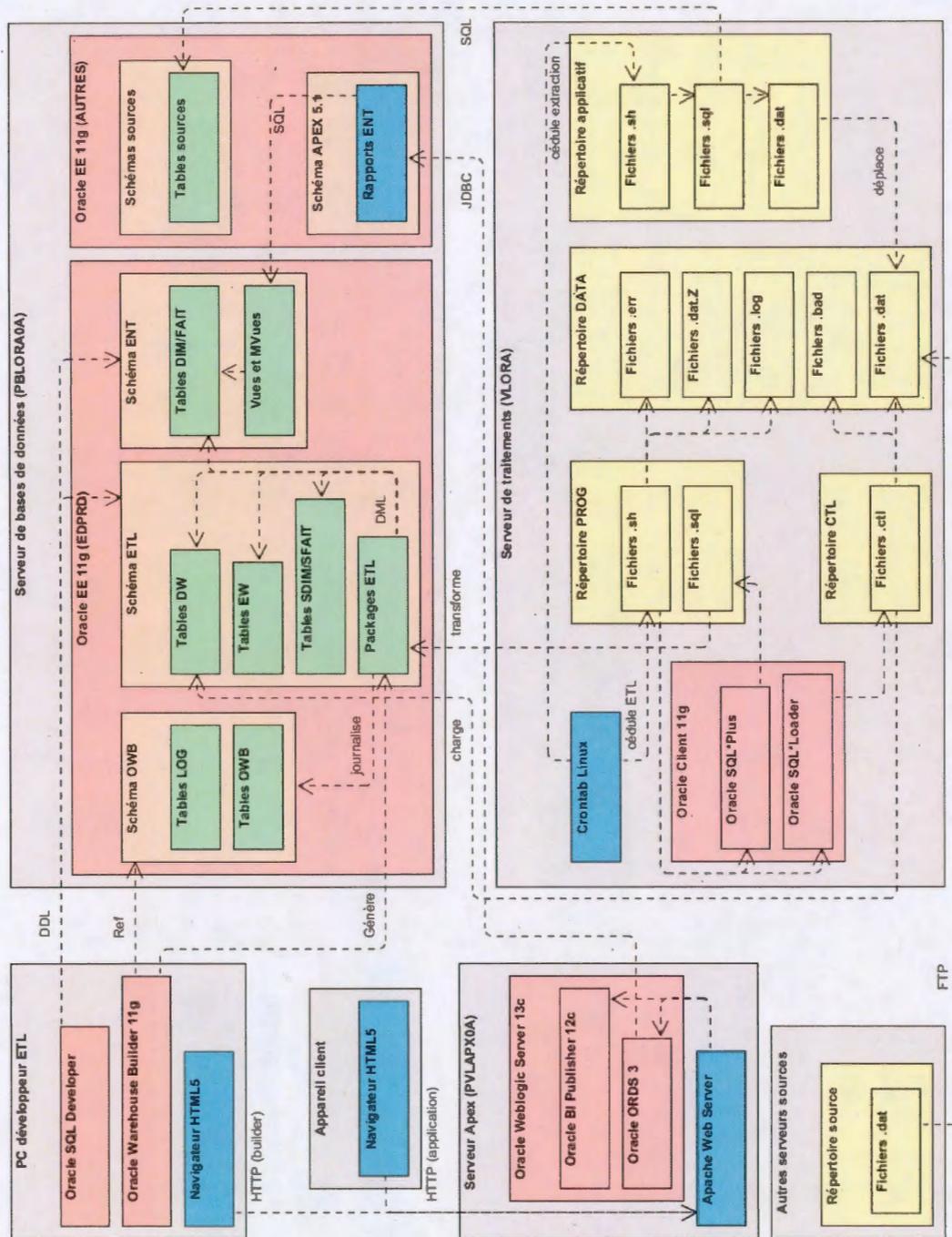


Figure 3.4 Architecture logicielle avant le projet

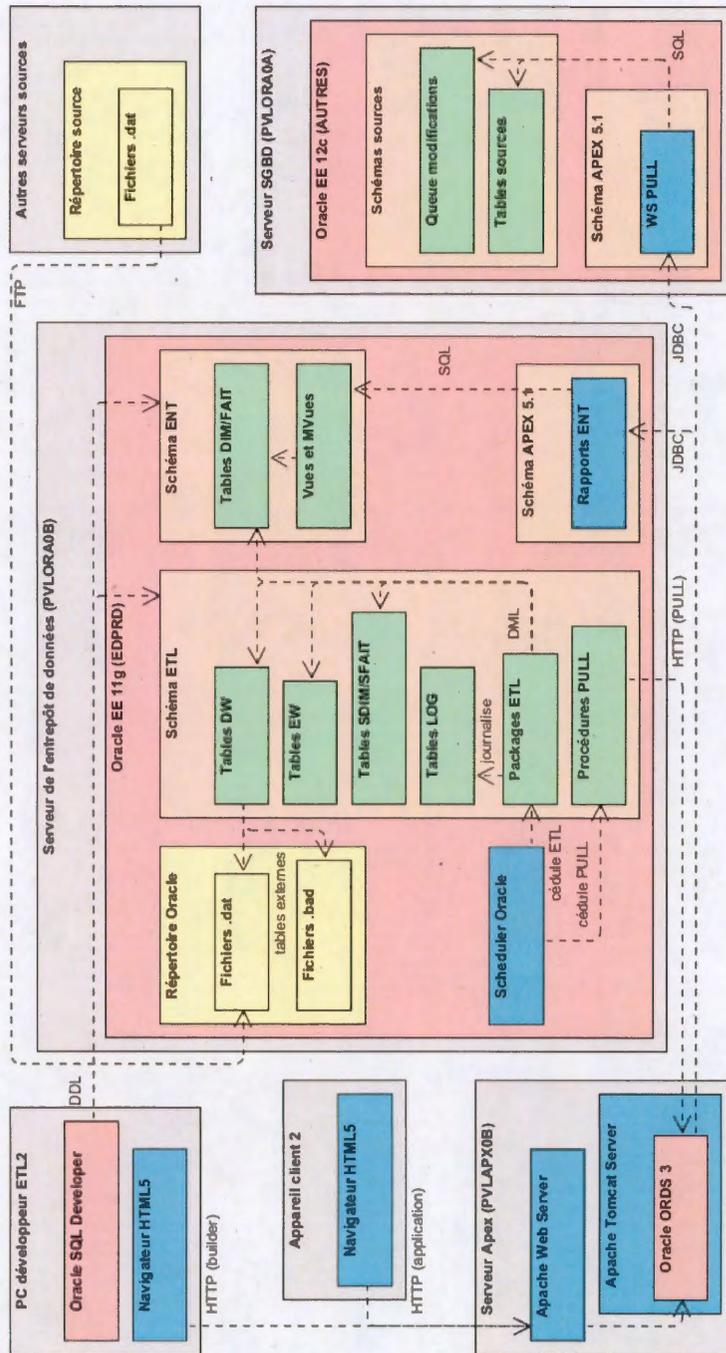


Figure 3.5 Architecture logicielle après le projet

3.2.3.2 Sélection et installation des produits

Cette activité consistait à développer un outil qui servira à concevoir les traitements ETL. Le groupe Kimball a inventorié les fonctionnalités typiques de ces outils (Kimball et al, 2008). La liste ci-dessous présente une traduction de la liste de Kimball.

- Extraction des données
 - o Système d'extraction
 - o Système de détection des changements
 - o Système d'analyse de données
- Transformation des données
 - o Système de nettoyage de données
 - o Système de validation de la conformité de données
 - o Gestionnaire de dimension d'audit
 - o Système de déduplication
 - o Système de gestion de la qualité de données
- Chargement des données
 - o Gestionnaire de *Slowly Changing Dimension* (SCD)
 - o Système de gestion des clés de substitution
 - o Gestionnaire des dimensions hiérarchiques
 - o Gestionnaire des dimensions particulières
 - o Système de chargement des tables de faits
 - o Gestionnaire des *lookup*
 - o Gestionnaire de dimensions multivaluées
 - o Gestionnaire des données arrivant en retard
 - o Système de gestion des dimensions
 - o Système de gestion des tables de faits
 - o Gestionnaire d'agrégation

- Gestionnaire de cubes multidimensionnels
- Gestionnaire d'extraction de données
- Gestion de l'environnement
 - Ordonnanceur des processus ETL
 - Système de sauvegarde
 - Système de recouvrement et de reprise
 - Système de contrôle des versions
 - Système de déploiement
 - Système de surveillance du flux des processus ETL
 - Système de tri
 - Système d'analyse de correspondance et de dépendance
 - Système de gestion des erreurs
 - Gestionnaire de parallélisme et de pipelines
 - Système de sécurité
 - Gestionnaire de conformité aux règles
 - Gestionnaire de référentiel des métadonnées

Fonctionnement de la plateforme ETL

Puisque la solution développée préconise que l'activité de transformation s'effectue à l'intérieur de la base de données de l'entrepôt plutôt que par un logiciel commercial sur un serveur intermédiaire de traitement, il est commun de nommer ce type d'architecture comme étant un modèle *Extract-Load-Transform* (ELT) plutôt qu'*Extract-Transform-Load* (ETL). La figure 3.6 présente la liste des activités du processus ETL tel que développé dans la plateforme maison.

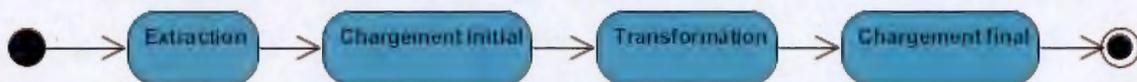


Figure 3.6 Vue d'ensemble du processus ETL

Extraction

L'activité consiste à extraire les données modifiées des systèmes sources depuis la dernière exécution du traitement et de les déposer sur le serveur de l'entrepôt. La figure 3.7 présente les tâches de l'activité d'extraction.

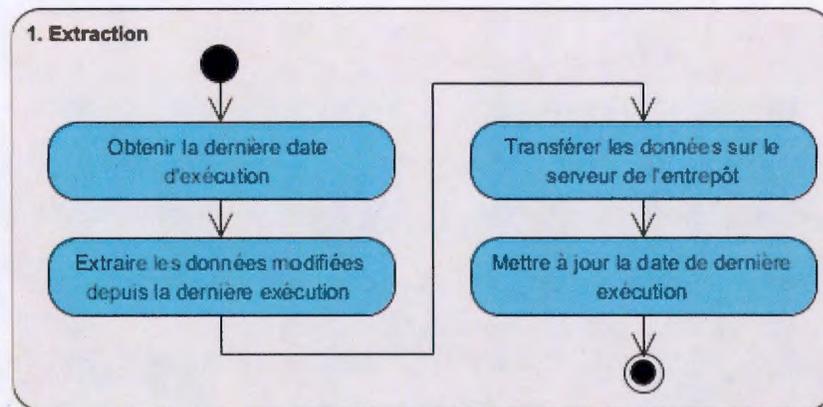


Figure 3.7 Activité d'extraction des données

Chargement initial

L'activité consiste à charger les données brutes vers la zone de travail (*staging area*) située à l'intérieur de la base de données de l'entrepôt. La figure 3.8 présente les tâches de l'activité de chargement initial.

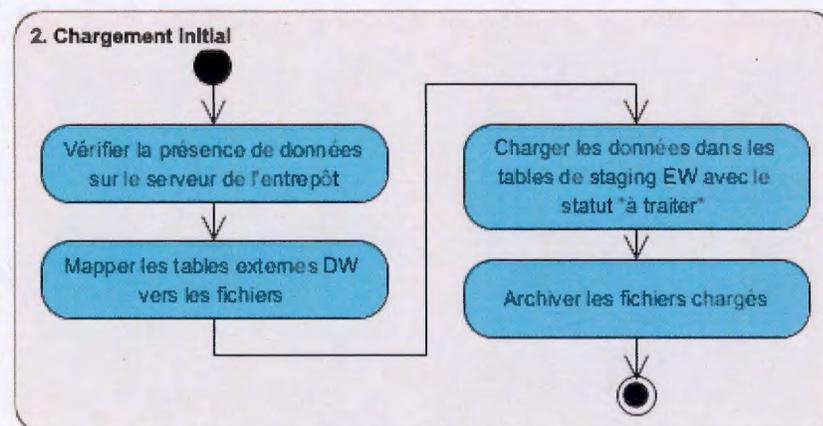


Figure 3.8 Activité de chargement initial des données

Transformation

L'activité consiste à nettoyer, transformer et valider les données brutes vers le modèle multidimensionnel dans des tables temporaires qui permettront de valider que le traitement fonctionne correctement avant le chargement final. La figure 3.9 présente les tâches de l'activité de transformation.

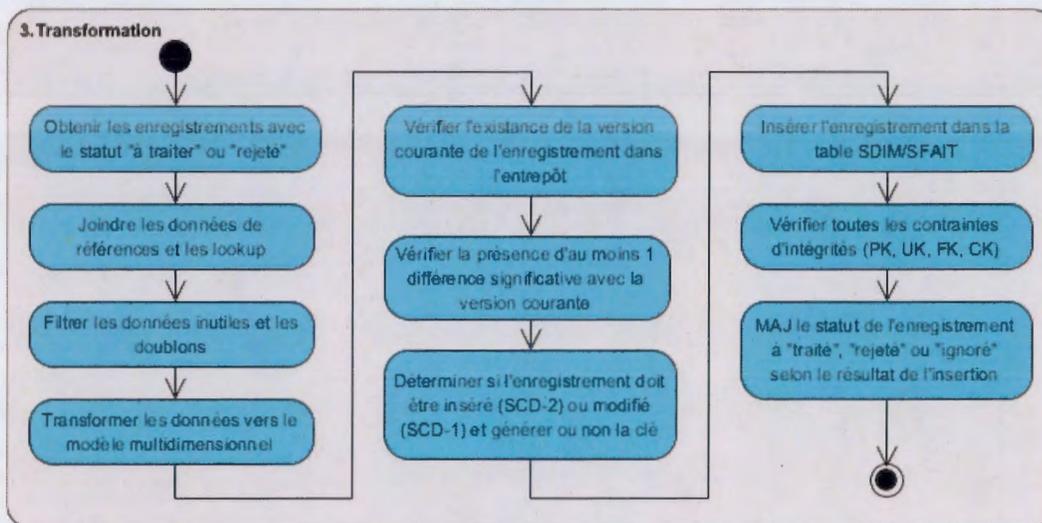


Figure 3.9 Activité de transformation des données

Chargement final

L'activité consiste à charger les données transformées vers les tables de destination de l'entrepôt en plus de mettre à jour les index, les vues et les rapports. La figure 3.10 présente les tâches de l'activité de chargement final.

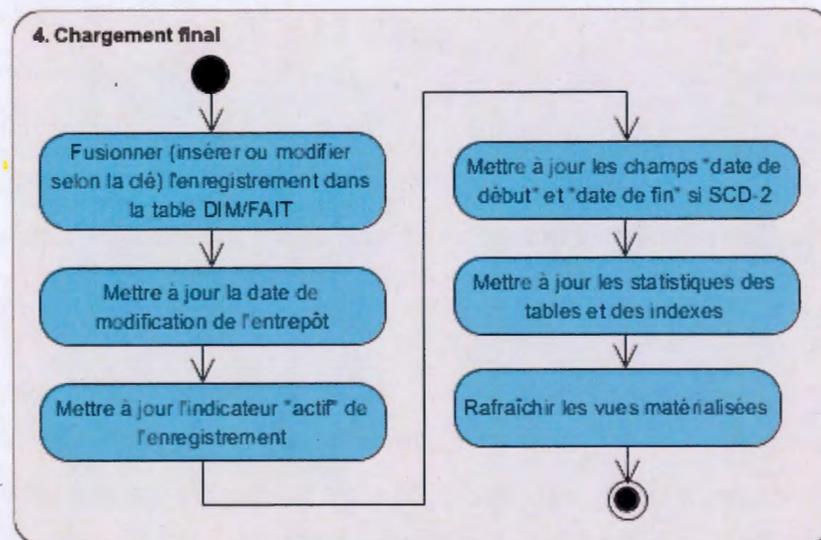


Figure 3.10 Activité de chargement final

Prototypage

Le développement de trois prototypes fut nécessaire avant de trouver une solution offrant le compromis idéal entre la performance des traitements ETL et la maintenabilité du code source. La liste suivante présente une description de chacune des tentatives.

1. Approche procédurale : Ce prototype fut développé en utilisant quasi exclusivement un langage procédural (PL/SQL) avec un découpage important de fonctions réutilisables. Le prototype utilise des boucles et des blocs de code simples et minimisant les requêtes SQL complexes. La solution est facile à maintenir et à déboguer, mais n'est pas du tout performante et nécessite beaucoup d'effort de développement.
2. Approche SQL : Ce prototype fut développé en utilisant quasi exclusivement un langage de 4e génération (SQL) avec de très grosses requêtes monolithiques qui traitent tous les enregistrements d'un seul coup. La solution

est très performante lorsque le nombre d'enregistrements est modéré, mais devient rapidement problématique lorsqu'il y a de gros chargements avec de nombreuses jointures. La solution est également très difficile à maintenir et à déboguer, car les requêtes sont très complexes et qu'il est impossible de suivre l'avancement du traitement pendant les longues minutes d'exécution de chaque requête.

3. Approche hybride (solution retenue) : Ce prototype fut développé en utilisant une combinaison de SQL et PL/SQL ainsi que plusieurs fonctionnalités propriétaires d'Oracle qui offrent plusieurs avantages : une gestion des rejets plus efficace, des fonctions analytiques très performantes, des opérations en lot, un suivi de l'avancement détaillé, un code source compact, etc. La solution offre un excellent compromis entre performance et maintenabilité au détriment d'une solution moins portable en raison de l'utilisation de fonctionnalité propriétaire à Oracle.

La production des trois prototypes a nécessité à elle seule un peu plus de 40 h faisant en sorte que l'activité de sélection des produits a nécessité beaucoup plus de temps que prévu initialement.

3.2.4 Définition des données et de leur transformation

3.2.4.1 Modélisation des données

L'activité consistait surtout à effectuer une rétro-ingénierie de la modélisation des processus d'affaires dans des diagrammes d'entité-relation (ERD) afin d'identifier les dépendances fonctionnelles. Pour chaque entité, la granularité des données, la stratégie de gestion du changement ainsi que l'utilisation des dimensions ont été identifiées afin de planifier les travaux pour l'étape de réalisation des traitements ETL. L'analyse des dépendances a permis de constater que la modélisation existante était

très fragmentée et souvent loin d'être optimale. La modélisation idéale étant d'utiliser un modèle en étoile où les faits pointent directement sur toutes les dimensions conformes afin d'éliminer les jointures inutiles.

La réalisation de cette activité a permis d'identifier plusieurs tables candidates pour des changements qui feront toutefois partie d'une phase ultérieure d'évolution de l'entrepôt. La réalisation de tous ces changements dès la phase 1 aurait provoqué trop de changements sur les applications BI existantes et haussé grandement le niveau de risque du projet. La figure 3.11 présente un aperçu des dimensions utilisées par le processus d'affaires d'analyse des événements de prêts, de prolongations et de retours de documents.

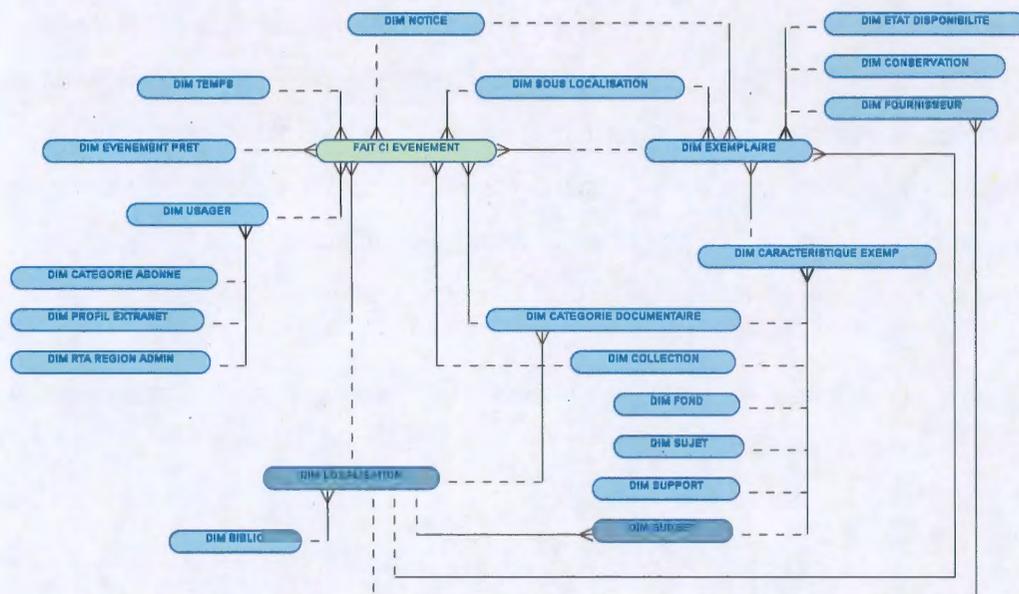


Figure 3.11 Modélisation du processus d'affaires sur les emprunts de documents

3.2.4.2 Conception physique

L'activité consista à créer le schéma relationnel pour l'environnement de transformation des données (*staging area*) pour chacun des traitements ETL. Ce

schéma est alimenté par le chargement des données brutes provenant des systèmes sources et permet d'y stocker les enregistrements qui y sont graduellement transformés et validés à mesure que le traitement ETL avance. La modélisation est conçue pour que les traitements ETL puissent reprendre le travail là où ils s'étaient arrêtés advenant une panne majeure du processus ou du serveur. Le résultat de l'étape consiste en un ensemble de scripts SQL permettant de créer les tables de transformation.

En plus de la modélisation de l'environnement de transformation, de nombreux scripts de redressement de données sur les tables de destinations finales furent nécessaires afin de régler les nombreux problèmes d'intégrité de données détectés pendant l'analyse des processus d'affaires. En effet, la plupart des tables de l'entrepôt n'avaient pas de contraintes d'intégrité pour renforcer la qualité et la cohérence des données. Les scripts de redressement ont permis, entre autres, de nettoyer les données en doubles ou manquantes, d'ajouter des contraintes, de diminuer l'espace disque et d'améliorer la performance des requêtes des applications BI. Cette étape a toutefois nécessité beaucoup plus d'effort qu'initialement planifié, car je ne m'attendais pas à trouver autant de problèmes de qualité de données remontant parfois jusqu'en 2005.

La figure 3.12 présente un extrait de la conception physique pour le processus de gestion des événements de circulation.

```

-- Create table
create table SFAIT_CI_EVENEMENT
(
  ETL_NOM_FICHER          VARCHAR2(4000) not null,
  ETL_NO_LIGNE            NUMBER not null,
  cp_id                   NUMBER(22) not null,
  date_maj_entrepot      DATE not null,
  id_no_abonne            NUMBER(22) not null,
  id_no_exemplaire        NUMBER(22) not null,
  id_no_evenement_pret    NUMBER(22) not null,
  id_no_localisation      NUMBER(22) not null,
  id_no_sous_localisation NUMBER(22) not null,
  id_no_categorie         NUMBER(22) not null,
  id_no_collection        NUMBER(22) not null,
  id_temps                NUMBER(22) not null,
  heure_transaction       CHAR(6) not null,
  id_usager               NUMBER not null,
  id_dim_usager           NUMBER not null,
  id_no_seq               NUMBER not null
);

-- Create/Recreate primary, unique and foreign key constraints
alter table SFAIT_CI_EVENEMENT add constraint SFAIT_CI_EVENEMENT_PK
  primary key (CP_ID);
alter table SFAIT_CI_EVENEMENT add constraint SFAIT_CI_EVENEMENT_UK1
  unique (ID_DIM_USAGER, ID_NO_EXEMPLAIRE, ID_NO_EVENEMENT_PRET, ID_TEMPS, HEURE_TRANSACTION);
alter table SFAIT_CI_EVENEMENT add constraint SFAIT_CI_EVENEMENT_FK1
  foreign key (ID_NO_EXEMPLAIRE) references ENT.DIM_EXEMPLAIRE (ID_NO_EXEMPLAIRE);
alter table SFAIT_CI_EVENEMENT add constraint SFAIT_CI_EVENEMENT_FK2
  foreign key (ID_NO_EVENEMENT_PRET) references ENT.DIM_EVENEMENT_PRET (ID_NO_EVENEMENT_PRET);

```

Figure 3.12 Extrait de la modélisation de la table « *sfait_ci_evenement* »

3.2.4.3 Conception et développement du système ETL

L'ensemble des processus ETL a été migré vers la nouvelle plateforme en utilisant le prototype sélectionné comme point de départ. Pour ce faire, les traitements ont été analysés scrupuleusement afin d'y déceler tous les cas particuliers. La majorité des transformations se sont avérées plutôt simples à réécrire, mais environ 10 % des tâches ont nécessité un remodelage du code de la plateforme afin de gérer des cas particuliers qui n'avaient pas été prévus au moment de la réalisation du prototype. L'analyse des anciens traitements a permis d'estimer qu'environ la moitié du code source était probablement mort (c.-à-d. qu'il n'était plus utilisé par aucun système ou par les utilisateurs). La figure 3.13 présente un extrait simplifié du mécanisme de transformation des données sources vers les tables de destination. Les cas les plus complexes ont nécessité parfois plus de 500 lignes de code source.

```

-- Transformer les données...
insert into sdim_sous_localisation
select sdim.etl_nom_fichier,
       sdim.etl_no_ligne,
       '0' as actif,
       sysdate as date_maj_entrepot,
       sdim.no_sous_localisation,
       case
         -- Redressement
         when dim.id_no_sous_localisation is not null then dim.id_no_sous_localisation
         -- Création
         else DIM_SOUS_LOCALISATION_SEQ.nextval
       end as id_no_sous_localisation,
       sdim.desc_sous_localisation
from (
  -- mapping avec les tables de destination...
  select rank() over (partition by ew.rowid order by ew.etl_nom_fichier desc, ew.etl_no_ligne desc) as ew_rnum,
         ew.etl_statut,
         ew.etl_msg_erreur,
         ew.etl_nom_fichier,
         ew.etl_no_ligne,
         upper(ew.code_table) as no_sous_localisation,
         ew.desc_id as desc_sous_localisation
  -- ew_pf_sous_localisation...
  from   EW_CA_TABLES ew
  where  ew.rowid = r_ew.rowid
  and    ew.table_1 = 'TABSUBLO'
) sdim
-- dim_sous_localisation...
left join dim_sous_localisation dim
on      (dim.no_sous_localisation = sdim.no_sous_localisation)
-- filtres...
where  1=1
and    sdim.ew_rnum = 1
-- vérifier la présence d'au moins 1 différence significative...
and (   nvl(to_char(dim.no_sous_localisation), ' ') != nvl(to_char(sdim.no_sous_localisation), ' ')
or     nvl(to_char(dim.desc_sous_localisation), ' ') != nvl(to_char(sdim.desc_sous_localisation), ' ')
)
log errors into err$_sdim_sous_localisation (1_nom_tache) reject limit unlimited;

```

Figure 3.13 Exemple du processus de transformation

3.2.5 Définition des besoins d'intelligence d'affaires

3.2.5.1 Conception des applications BI

Les travaux d'analyse BI consistent à la réévaluation de tous les processus d'affaires avec les utilisateurs feront partie de la deuxième phase du projet de refonte de l'entrepôt. Peu d'effort fut donc nécessaire sur cette activité dans ce projet hormis pour l'ajout d'une nouvelle source de données et la conception d'un nouveau rapport concernant le nombre de consultations de documents numériques. Il aurait été difficile de coordonner tous les ateliers de travail avec les groupes d'utilisateurs pendant l'été alors que la plupart des collaborateurs sont en vacances.

3.2.5.2 Développement des applications BI

Une nouvelle section fut tout de même ajoutée à l'application BI afin de permettre l'analyse du nombre de consultations de documents numériques par mois et par collection. Cette nouvelle section facilitera le travail des directions d'affaires qui doivent analyser et présenter l'information dans le cadre des activités reliées au plan stratégique de BAnQ. Les nouveaux rapports interactifs remplacent ainsi des classeurs Excel qui étaient maintenus à la main. La figure 3.14 présente un aperçu de l'un des nouveaux rapports.

Entrepôt de données | pascal.laforce | Déconnexion

Page d'accueil / Doc. numériques /

Nombre de consultations de documents numériques par mois

Ouvrir dans Excel

OK 5. Total par communauté, par année Actions

Type de ligne Communauté

Communauté : 2. Toutes les communautés

Année fiscale	Avril	Mai	Juin	Juillet	Août	Sept.	Oct.	Nov.	Déc.	Janvier	Février	Mars	Total
2016-2017									5 336	13 351	12 080	16 128	46 895
2017-2018	18 187	19 895	15 607	23 246	20 263	25 553	50 263	43 144	303 112	503 889	443 670	485 506	1 952 335
2018-2019	469 750	392 833	368 043	477 270	542 334	579 247	122 101						2 951 578
Toutes les années fiscales	487 937	412 728	383 650	500 516	562 597	604 800	172 364	43 144	308 448	517 240	455 750	501 634	4 950 808

Communauté : 1. Revus

Année fiscale	Avril	Mai	Juin	Juillet	Août	Sept.	Oct.	Nov.	Déc.	Janvier	Février	Mars	Total
2017-2018		60	142	148	526	1 545	2 377	1 566	10 813	15 339	15 850	16 843	65 209
2018-2019	14 501	15 295	12 500	11 635	10 408	12 765	2 292						79 396
Toutes les années fiscales	14 501	15 355	12 642	11 783	10 934	14 310	4 669	1 566	10 813	15 339	15 850	16 843	144 605

Figure 3.14 Nouvelle section « Documents numériques » dans l'application BI

3.2.6 Déploiement, amélioration et maintenance du système

3.2.6.1 Déploiement

Les traitements ETL ont été développés, testés et déployés itérativement tout au long du projet. Les traitements étaient déployés tout d'abord dans l'environnement d'acceptation qui rechargeait les mêmes données sources qu'en production et permettait de comparer les résultats des anciens et des nouveaux traitements. Une fois les tests complétés, les traitements étaient déployés et programmés dans l'environnement de production, les systèmes sources étaient modifiés pour rediriger leurs données vers la nouvelle infrastructure et les anciens traitements étaient désactivés dans l'ancienne infrastructure. Une fois tous les nouveaux traitements déployés en production, des travaux ont été réalisés pour retirer l'ancienne infrastructure et OWB. Le code source des anciens traitements a toutefois été conservé pour de futures références. La décision d'avoir limité les modifications à la modélisation multidimensionnelle et aux processus d'affaires a facilité grandement la comparaison des résultats, car les résultats des traitements devaient être sensiblement les mêmes. Le tableau 3.2 présente la liste des déploiements effectués en production durant le projet.

Tableau 3.2 Déploiements et itérations

NO	SIGLE	DATE MEP	COMMENTAIRES
1	GRC	2018-06-26	Gestion de la relation client. Ce processus fut utilisé pour les prototypes. Plusieurs redressements nécessaires sur la dimension « abonné ».
2	GU	2018-07-09	Gestion du service québécois du traitement documentaire. Consolidation de 10 traitements en 1.
3	BN	2018-07-24	Gestion de la consultation des documents numériques.
4	RH	2018-08-16	Gestion des ressources humaines. Ajout de la dimension « entité administrative ».
5	C2	2018-08-21	Gestion des incidents informatiques.
5	TEL	2018-08-21	Gestion des appels téléphoniques.
5	WIFI	2018-08-21	Gestion de l'utilisation du sans-fil.
6	SV	2018-08-28	Gestion des accès au portail institutionnel.
6	FREQ	2018-08-28	Gestion de la fréquentation physique.
7	AUTH	2018-09-06	Gestion de l'authentification logicielle.
8	SAM	2018-09-10	Gestion de la réservation des postes publics. Fusion de 3 dimensions en 1. Redressement nécessaire sur la dimension des « postes publics ».
9	PEB	2018-09-18	Gestion du prêt entre bibliothèques. Difficulté pour déterminer la granularité des faits.
10	SIGB	2018-10-15	Gestion de la circulation des documents physiques. Il s'agit du traitement le plus ancien (2005) et celui avec le plus grand nombre de sources (62).
11	OWB	2018-10-29	Retrait de OWB et de l'ancienne infrastructure.

3.2.6.2 Maintenance

Un suivi quotidien des nouveaux traitements déployés en production fut nécessaire afin d'évaluer les enregistrements rejetés par les traitements et documenter les cas particuliers. Pour y arriver, j'ai intégré à mes traitements ETL un mécanisme de surveillance et de journalisation compatible avec le SGBD Oracle. Les règles de validation sont conçues principalement à l'aide de contraintes d'intégrité SQL qui génèrent des rejets journalisés par le mécanisme de surveillance. Ces rejets sont analysés par l'équipe de support et envoyés aux analystes des différents systèmes sources qui corrigeront les données dans les jours qui suivent. Les exceptions non contrôlées remontent quant à elles jusqu'au SGBD et sont par la suite journalisées comme étant des erreurs critiques au bon fonctionnement du traitement par le mécanisme de surveillance. Les erreurs critiques nécessitent une intervention rapide et doivent être analysées et corrigées par l'équipe de support.

Les mécanismes de journalisation des traitements, de gestion des enregistrements rejetés et de notification des erreurs critiques ont grandement facilité la surveillance de la plateforme. La majorité des rejets proviennent de données de mauvaise qualité que transmettaient les systèmes sources à l'entrepôt (ex. doublons, données manquantes, etc.). Peu d'anomalies furent retrouvées dans les traitements ETL eux-mêmes en raison des tests effectués dans l'environnement d'acceptation. Il reste toutefois à développer une interface web pour faciliter le suivi de l'exécution des traitements et l'analyse des rejets. L'équipe interne se contente pour le moment de quelques rapports envoyés quotidiennement par courriel tel que présenté dans la figure 3.15.

De : entrepot@banq.qc.ca [entrepot@banq.qc.ca]
 Envoyé : 24 septembre 2018 22:33
 À : Laforce Pascal
 Objet : Entrepot - BANQ Numerique

Type	ID	Code	Debut	Fin	Statut	Lus	Traites	Ignorees	Rejetes	Message
PRO	5648	CHARGER_BN	02:30:00		ENCOURS	223983	136618	87302	63	
ACT	14164	EPURER_TABLES	02:30:00	02:30:00	SUCCES	0	0	0	0	
ACT	14169	MAP_EW_BN_COLLECTION	02:30:01	02:30:02	SUCCES	838	838	0	0	
ACT	14176	MAP_EW_BN_ITEM	02:30:02	02:30:02	SUCCES	85	85	0	0	
ACT	14184	MAP_EW_BN_BITSTREAM	02:30:02	02:30:04	SUCCES	1826	1826	0	0	
ACT	14195	MAP_EW_BN_DOC_CONSULTE	02:30:04	02:30:35	SUCCES	101004	101004	0	0	
ACT	14200	MAP_SDIM_BN_COLLECTION	02:30:35	02:30:36	SUCCES	838	0	838	0	
ACT	14201	MAP_DIM_BN_COLLECTION	02:30:36	02:30:36	SUCCES	0	0	0	0	
ACT	14202	MAP_SDIM_BN_ITEM	02:30:36	02:30:42	SUCCES	85	62	23	0	
ACT	14204	MAP_DIM_BN_ITEM	02:30:42	02:30:51	SUCCES	62	62	0	0	
ACT	14205	MAP_SDIM_BN_BITSTREAM	02:30:51	02:31:00	SUCCES	1826	118	1708	0	
ACT	14206	MAP_DIM_BN_BITSTREAM	02:31:00	02:31:17	SUCCES	118	118	0	0	
ACT	14207	MAP_SFAIT_BN_DOC_CONSULTE	02:31:17	02:32:56	SUCCES	101048	16252	84733	63	
ACT	14213	MAP_FAIT_BN_DOC_CONSULTE	02:32:56	02:33:17	SUCCES	16253	16253	0	0	
ACT	14214	RAFRAICHIR_MVIEWS	02:33:17	02:33:17	SUCCES	0	0	0	0	

Figure 3.15 Rapport quotidien sur l'exécution du traitement « BN »

3.2.6.3 Croissance

Très peu d'effort fut consacré à la croissance de l'entrepôt et à l'amélioration des processus d'affaires puisque la priorité était de compléter la migration de tous les traitements ETL avant de leur apporter des modifications. La deuxième phase de travaux prévue dans la prochaine année permettra d'analyser les opportunités de croissance et d'amélioration.

CHAPITRE IV

INTERPRÉTATION DES RÉSULTATS

4.1 Évaluation des objectifs

Les objectifs initiaux avaient été fixés pour une base annuelle, mais l'exercice fut ajusté en fonction d'une période d'évaluation de 6 mois pour les fins de ce rapport. Pour chacun des objectifs, j'ai donc comparé les résultats obtenus pour la période de mai à novembre 2018 avec la moyenne obtenue des cinq dernières années pour les mêmes mois.

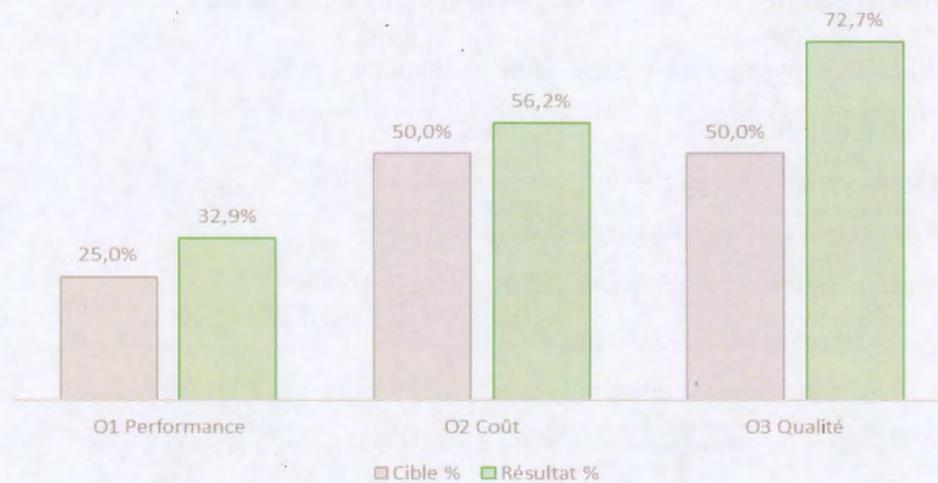


Figure 4.1 Résultat de l'amélioration obtenue pour chacun des objectifs

Objectif 1 : Diminuer le temps de collecte et d'analyse des données

- Cible : Diminuer le temps total annuel de 25 %.
- Moyens mis en œuvre :
 - ✗ Les demandes *ad hoc* récurrentes seront davantage automatisées.
 - ✓ Les jeux de données et rapports inutiles seront retirés.
 - ✗ Les rapports nécessitant plus de cinq minutes seront optimisés.
 - ✓ De nouveaux rapports sur les sources existantes seront ajoutés.
 - ✓ Les temps d'exécution des traitements ETL seront réduits.
 - ✓ Les temps réponses seront journalisés.
 - ✓ L'indexation et les options de stockage seront optimisées.
 - ✓ Les calculs complexes seront automatisés la nuit.
- Résultat obtenu : Réduction du temps total annuel de 32,9 % (735 contre 1095 heures de traitement). Pour obtenir ce résultat, la durée totale d'exécution de tous les traitements et des rapports de juin 2018 à novembre 2018 ont été comparés à la moyenne des cinq dernières années pour les mêmes périodes.

Objectif 2 : Réduire les coûts de maintenance de l'entrepôt de données

- Cible : Diminuer les coûts annuels de maintenance de 50 %.
- Moyens mis en œuvre :
 - ✓ *Oracle Warehouse Builder* sera remplacé.
 - ✓ L'architecture logicielle sera simplifiée.
 - ✓ L'infrastructure technologique sera mise à niveau.
 - ✓ La journalisation des traitements sera améliorée.
 - ✓ Un mécanisme de relance des traitements ETL sera mis en place.
 - ✓ L'effort en soutien et maintenance sera diminué.
 - ✓ La maintenabilité des traitements ETL sera améliorée.
 - ✓ Le temps nécessaire pour ajouter une source de données sera diminué.

- Résultat obtenu : Réduction des coûts annuels de maintenance de 56,2 % (95 contre 217 heures). Pour obtenir ce résultat, l'effort de l'équipe interne de juin 2018 à novembre 2018 a été comparé à l'effort moyen des cinq dernières années.

Objectif 3 : Améliorer la qualité et la cohérence des données

- Cible : Diminuer le nombre d'incohérences de résultat de 50 %.
- Moyens mis en œuvre :
 - ✓ Les requêtes de service seront comptabilisées.
 - ✓ Les problèmes d'intégrité de données seront corrigés.
 - ✓ Des contraintes d'intégrité seront ajoutées.
 - ✓ Les résultats des traitements ETL seront comparés aux anciens.
 - ✓ Des validations seront ajoutées dans les traitements ETL.
 - ✓ Les données rejetées seront remontées au système source.
 - ✓ Un mécanisme de suivi des données rejetées sera implanté.
 - ✓ Les rejets et les validations seront journalisés.
- Résultat obtenu : Diminution du nombre d'incohérences détectés de 72,7 % (3 contre 11 billets). Pour obtenir ce résultat, le nombre de requêtes de service impliquant des données incohérentes de juin 2018 à novembre 2018 a été comparé à la moyenne des cinq dernières années pour la même période.

4.2 Interprétation des résultats

Les gains quant à la performance de la collecte de l'information sont principalement dus à la réécriture des traitements ETL, à l'élagage des données inutiles et à l'optimisation des index qui a ainsi permis de réduire substantiellement la durée nécessaire à l'exécution des traitements ETL et des rapports automatisés.

Les gains quant au cout de maintenance sont principalement dus au remplacement de OWB et à la simplification du processus ETL ainsi qu'à une réduction importante du nombre total de ligne de code de chaque traitement, ce qui nécessite beaucoup moins d'effort d'analyse par l'équipe de support. La taille totale du code source est ainsi passée de 780 000 lignes de codes à seulement 35 000 tout en offrant minimalement les mêmes fonctionnalités. À titre d'exemple, l'effort requis pour ajouter la nouvelle source de données sur la consultation des documents numériques a été moindre (3 jours) que l'effort nécessaire en 2017 pour ajouter une autre source (7 jours) pourtant de moindre complexité.

Les gains quant à la qualité des données sont principalement dus au redressement des données existantes, à l'ajout de contraintes d'intégrité et de validations dans les traitements. À titre de comparaison, un total de 1 800 règles de validation étaient présentes dans l'ancien code source alors que le nouveau en contient plus de 25 000. En identifiant les données erronées et les erreurs automatiquement dans les traitements ETL plutôt que manuellement dans les rapports, le nombre d'incohérences et d'incidents majeurs rapportés par les utilisateurs de l'entrepôt a chuté de manière importante et leur satisfaction globale tend à augmenter avec le temps.

Les résultats obtenus sur une année complète pourraient être légèrement inférieurs en raison d'une utilisation plus importante de l'entrepôt de données par les directions d'affaires au moment de la clôture de l'année financière au printemps. Le serveur est alors plus sollicité et une dizaine de rapports annuels plus exigeants en ressources doivent être exécutés. Je n'ai toutefois pas de doute que les activités de fin d'année seront elles aussi améliorées puisque j'avais inclus la simulation de ces activités lors des tests d'acceptation.

4.3 Qualité de service (QoS)

Selon le chroniqueur Sylvestre Marceau de Magellan Consulting, la qualité de service (QoS) est un outil de management important, mais souvent négligé par les directions TI (Marceau 2016). L'auteur définit la QoS comme suit :

« Selon plusieurs référentiels (ITIL, ISO 9000...), il est possible de définir la QoS comme la capacité pour un service informatique de répondre à des besoins métiers clairement définis dans l'objectif de satisfaire les utilisateurs. »

Pour les besoins de ce projet, les indicateurs suivants ont été utilisés :

- Le délai de réponse pour l'exécution d'un processus métier dans l'outil ;
- Le délai de traitement d'une demande utilisateur ;
- Le délai de résolution d'un incident selon la criticité.

Bien que l'entrepôt de données soit opérationnel en production chez BAnQ depuis une quinzaine d'années, aucune entente de niveau de service n'a jamais été conclue avec les directions d'affaires. Une activité d'analyse de la qualité du service attendue avec les parties prenantes sera ajoutée à la phase 2 du programme afin que les indicateurs utilisés à l'avenir pour évaluer le service aient été le résultat d'une entente de service entre les directions d'affaires et la DGTIT.

CHAPITRE V

DISCUSSION DE L'EXPÉRIENCE

Le grand nombre de problématiques et leur haut niveau de complexité ont fait en sorte qu'il n'était pas possible d'adresser toutes les demandes initiales en aussi peu de temps. Dès les premières semaines, il fut convenu avec les directions d'affaires de scinder le projet en deux phases distinctes afin de s'attarder plus spécifiquement sur la problématique la plus urgente qui était la désuétude des technologies, puisqu'elles empêchaient tout ajout ou amélioration de quoi que ce soit depuis de nombreuses années. Cette décision lors des activités de planification de projet fut avantageuse, car elle me permit de consacrer davantage d'effort dans les activités d'architecture et de développement ETL afin d'offrir des livrables de meilleure qualité. Après coup, il aurait été impossible de repartir d'une feuille blanche et d'inclure la révision complète des processus d'affaires dans ce projet.

La scission du projet en deux phases fut également profitable aux directions d'affaires puisque celles-ci avaient besoin de plus de temps pour la révision de leurs processus en vue de l'arrivée en cours de route du nouveau président-directeur général. Ce dernier demanda aux grandes directions de démarrer la préparation d'un nouveau plan stratégique pour l'institution d'ici la fin de l'année, rendant les ressources nécessaires au projet non disponibles. Cela aura également permis d'éviter de gaspiller trop d'effort sur des processus d'affaires dont l'importance et la priorité changeront possiblement dans le prochain plan.

L'effort supplémentaire ajouté aux activités de conception de l'architecture technologique a permis de livrer une plateforme ETL plus robuste, plus performante et plus polyvalente qu'il n'aurait été possible de faire en gardant tout dans un seul projet. Tel que mentionné dans la littérature, ce sont les activités de réécriture des traitements ETL vers la nouvelle plateforme maison qui auront nécessité le plus de travail. Il n'avait toutefois pas été anticipé qu'autant de problèmes d'intégrité de données auraient été découverts en cours de route. Les activités de soutien et de maintenance ont été nettement sous-évaluées puisqu'il n'avait pas été prévu de consacrer autant de temps au redressement des données et à la correction d'anomalies dans les règles d'affaires.

L'utilisation de la méthodologie *Data Warehouse Lifecycle* fut un franc succès, car elle a grandement simplifié les activités de planification du projet et permis de s'assurer qu'aucune activité importante ne fut oubliée. La méthodologie est utilisée depuis de nombreuses années dans l'industrie, elle est bien documentée dans la littérature et est facile à mettre en place. De plus, il s'agissait pour moi d'une méthodologie déjà connue puisqu'elle est enseignée dans le cours «MTI820 — Entrepôts de données et intelligence d'affaires» offert par l'École de Technologie Supérieure (ÉTS) aux étudiants des programmes d'étude de maîtrise en génie logiciel et de maîtrise en technologie de l'information.

Le choix d'opter pour le développement d'une plateforme ETL maison avec les outils déjà disponibles dans l'organisation et déjà bien maîtrisés par l'équipe interne, comparativement à l'achat d'un outil commercial, a permis de produire une solution taillée sur mesure pour les besoins de BANQ dans des temps très courts. L'effort investi dans les activités d'architecture et de conception de la plateforme, mon expérience en développement ETL et les connaissances déjà acquises sur les processus d'affaires ont permis de réduire grandement les risques du développement. L'utilisation d'un outil commercial aurait sans doute offert plus de possibilités pour la

conception des traitements ETL, mais aurait toutefois nécessité une courbe d'apprentissage beaucoup plus grande pour moi et l'équipe interne, et occasionné des délais importants pour la mise en place de l'outil. Kimball conseille d'ailleurs de commencer par du développement maison lorsque les moyens sont limités et les besoins bien maîtrisés (Kimball et al, 2008), ce qui était notre cas.

L'idée d'utiliser une suite logicielle commerciale de gestion des données maîtres (MDM) dans le projet fut rapidement écartée lors de la planification du projet en raison de la grande complexité de la mise en œuvre de ces solutions et du manque de maturité de l'organisation en architecture d'entreprise. Le projet permettra toutefois de préparer BAnQ à la pratique d'architecture d'entreprise par la réalisation d'un premier référentiel des données de référence qui sera intégré à l'entrepôt de données en inventoriant les principaux actifs informationnels de l'institution. L'entrepôt permet ainsi de lier les processus d'affaires les plus importants aux principaux systèmes informatiques.

Le projet a également respecté les facteurs clés de succès établis par le groupe Kimball. Le soutien de la haute direction et des directions d'affaires a permis de trouver un équilibre dans la portée du projet. L'entrepôt est important pour l'institution, puisqu'il s'agit bien souvent du seul endroit permettant de faire des corrélations pouvant améliorer la prise de décision des directeurs. L'architecture de l'entrepôt nécessitait une intervention urgente sans quoi le service risquait sérieusement de se dégrader. Ce projet répond bien au principe répandu en informatique qu'il vaut mieux éviter d'effectuer des modifications importantes à la fois à l'architecture et aux fonctionnalités d'un système dans un même projet en raison de l'explosion des facteurs de risque. Le projet fut facile à vendre à l'équipe interne puisque cela faisait plus de cinq ans qu'elle souhaitait travailler sur l'entrepôt, mais qu'elle manquait toujours de temps pour s'y consacrer pleinement. La réussite de ce projet permet ainsi à l'équipe interne de se consacrer davantage à des tâches

d'analyse, d'architecture et de conception logicielle plutôt qu'à des tâches répétitives de correction de données moins intéressantes.

Le projet s'inscrit parfaitement dans la pratique du génie logiciel puisqu'il m'a permis d'intégrer l'ensemble des activités du développement logiciel dont notamment : la gestion de projet ; l'analyse d'affaires (surtout BI), l'architecture de solution, la conception détaillée, le développement ETL, l'implantation d'outils informatiques, la modélisation de données, la gestion des déploiements et la recherche bibliographique en informatique. En plus des activités liées au développement logiciel, les cours axés sur les bases de données, le forage de données, l'intelligence d'affaires et l'architecture d'entreprise m'ont également été très utiles pour la réalisation de ce projet spécifique.

CONCLUSION

Ce projet visait à résoudre les problèmes les plus importants de l'entrepôt de données institutionnel de BAnQ qui empêchaient tout changement ou l'ajout de fonctionnalités longuement souhaitées par les directions d'affaires. Étant donné l'ampleur du projet initial, il fut convenu avec la direction que ce projet technique de 300 heures ne s'attarderait qu'aux problèmes liés à l'architecture logicielle, au remplacement de l'outil ETL en place et à la conversion de tous les traitements vers le nouvel outil. Les principales activités du projet consistèrent en l'analyse, l'architecture, la conception détaillée, le développement et le déploiement d'une toute nouvelle plateforme ETL respectant au mieux la méthode *Data Warehouse Lifecycle* de Ralph Kimball.

La méthode proposée s'appliqua parfaitement au type de projet et fut suivie plutôt rigoureusement grâce aux activités de contrôle de projet. Quelques imprévus mineurs sont survenus en cours de route en lien avec des particularités de quelques traitements ETL et d'activités de maintenance plus importantes que prévu. Ces activités nécessitèrent des ajustements à la planification et finirent par engendrer un dépassement de cout d'environ 25 % sans toutefois dépasser la date de livraison planifiée. Le résultat final du projet est le déploiement d'une toute nouvelle infrastructure basée sur une architecture simplifiée, plus performante et plus polyvalente. Tous les traitements ETL ont été réécrits sans nécessiter de modifications du côté des systèmes sources. Les applications BI et les utilisateurs de l'entrepôt n'ont été que très légèrement impactés. Le projet n'a pas réussi à aller aussi loin que souhaité dans la réévaluation des processus d'affaires et des applications BI, mais résout tout de même les points bloquants qui permettront la réalisation

d'améliorations importantes pour les utilisateurs à court terme. La prochaine phase d'amélioration sera consacrée à l'ajout de fonctionnalités en fonction du prochain plan stratégique de l'institution.

Finalement, ce projet démontre qu'il n'est pas nécessaire de dépenser une fortune pour moderniser un entrepôt de données ou réaliser un projet de développement en mode itératif lorsque les ressources affectées au projet connaissent bien les besoins et les processus d'affaires des clients, qu'ils possèdent les compétences techniques, qu'ils utilisent une méthodologie appropriée, qu'ils utilisent les bonnes technologies et qu'ils ont l'appui indéfectible de la haute direction. Réunir toutes ces qualités n'est pas toujours facile, mais il est évident que la formation générale combinée à l'expérience sur le terrain confère un atout important à la réalisation de tout projet lié au génie logiciel.

RÉFÉRENCES

Battandier, A. (2009). "Gestion de la valeur acquises (EVM)". Collectif Sarka-SPIP. 2009. Récupéré le 28 octobre 2018, de <http://alain.battandier.free.fr/spip.php?article12>

Beyer, M., Thoo, E., Selvage, M.Y. et Zaidi, E. (2017). Magic Quadrant for Data Integration Tools. Gartner Group. 3 aout 2017.

Birkbeck, S. (2017). "Is Kimball Still Relevant. Datametrics". 29 juin 2017. Récupéré le 28 octobre 2018, de <https://datametrics.co.nz/insights/article/is-kimball-still-relevant>

Chafky, S. et Desrosiers, C. (2011). "Le cycle de vie d'un projet en intelligence d'affaires". Département de génie logiciel et des TI. École de Technologies Supérieures. 2011. Récupéré le 28 octobre 2018, de https://cours.etsmtl.ca/mti820/public_docs/acetates/MTI820-Acetates-CycleDeVieProjetBI_1pp.pdf

Duncan, A.D. (2015). How to Establish a Data-Driven Culture in the Digital Workplace. Gartner Group. 9 juin 2015.

Foote, K.D. (2018). "A Brief History of the Data Warehouse". Dataversity. 19 avril 2018. Récupéré le 28 octobre 2018, de <http://www.dataversity.net/brief-history-data-warehouse/>

Kimball, R., Ross, M., Mundy, J. et Becker, B. (2008). The Data Warehouse Lifecycle Toolkit: Second Edition. John Wiley & Sons, 2008.

Inmon, W.H., Imhoff, C. et Sousa, R. (2002). Corporate Information Factory. John Wiley & Sons. 14 mars 2002.

Inmon, W.H., Strauss, D. et Neushloss, G. (2008). DW 2.0: The Architecture for the Next Generation of Data Warehousing. Elsevier. 2008.

Marceau, S. (2016). "Management de la qualité des services IT". JournalDuNet. 5 décembre 2016. Récupéré le 21 février 2019, de <https://www.journaldunet.com/solutions/expert/65798/management-de-la-qualite-des-services-it.shtml>

van der Meulen, R. (2017). Gartner Reveals the 2017 Hype Cycle for Data Management. Gartner Group. 28 septembre 2017.

Ross, M. (2004). "Differences of Opinion. Kimball Group". 3 mars 2004. Récupéré le 28 octobre 2018, de <https://www.kimballgroup.com/2004/03/differences-of-opinion/>

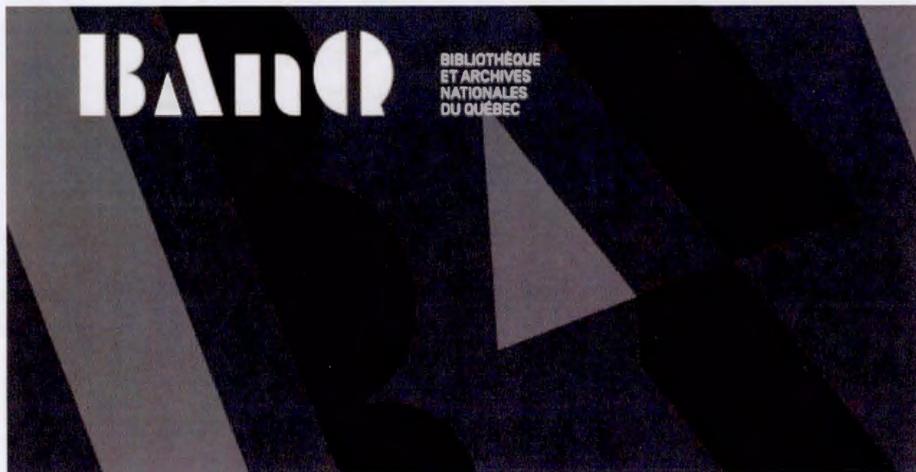
Talend Team. (2018). "What is Master Data Management". Talend Group. Récupéré le 28 octobre 2018, de <https://www.talend.com/resources/what-is-master-data-management/>

Tozzi, C. (2018). "Big Data vs Traditional Data: What Defines Big Data?". Syncsort. 6 mars 2018. Récupéré le 28 octobre 2018, de <http://blog.syncsort.com/2018/03/big-data/big-data-vs-traditional-data/>

Victoroff, S. (2015). "Big Data Doesn't Exist. Techcrunch". 10 septembre 2015. Récupéré le 28 octobre 2018, de <https://techcrunch.com/2015/09/10/big-data-doesnt-exist/>

ANNEXE A

CHARTRE DE PROJET



Refonte de l'entrepôt de données

Pascal Laforce
Architecte de données
Direction de l'architecture et de la conception

Bibliothèque
et Archives
nationales
Québec 

Processus Extract-Transform-Load

- Développé avec Oracle Warehouse Builder 10g et script Linux.
- L'outil n'est plus supporté depuis plusieurs années (remplacé par ODI).
- Nous modifions le code généré par OWB à la main depuis plusieurs années.
- La journalisation et la gestion des erreurs est déficiente.
- La modélisation dimensionnelle est parfois déficiente
ex. les notices bibliographiques
- Les extractions de Portfolio et Pivotal devront être réécrits prochainement.
- Règle du pouce : ETL = 70% des coûts d'un entrepôt de données.

Analyse BI

- Rapport statique et aucun tableau de bord.
- La plupart des indicateurs de performance ne proviennent pas de l'entrepôt.
- Beaucoup de demandes reçues pour ajouter des rapports.
- Beaucoup de demandes adhoc à chaque année (1 mois de travail / an).
- Le comité des statistiques institutionnelles avait commencer à prioriser les travaux de l'entrepôt à la demande de la PDG et de la vérificatrice.
- Quelques opportunités :
 - Qte. d'activités et de participation (plan stratégique).
 - Qte. de prêts numérique (plan stratégique).
 - Qte. de projets réalisés et de demandes reçus à l'informatique.
 - Qte. de numérisation à la conservation.
 - Qte. de réquisitions aux archives.

Proposition

- Projet de fin d'étude de 270h minimum sur une période de 3 à 6 mois.
- Temps : Début du projet mai 2018 et fin en décembre 2018.
- Ressources : Pascal (50%), Philippe (20%) et autres (quelques heures).
- Budget : Évolution des services corporatifs et communs.

Bibliothèque et Archives nationales du Québec

Principales activités du projet

- Remplacer complètement OWB, les scripts Linux et les transferts FTP.
- Évaluer les outils ETL du marché (ex. Talend) VS développement maison.
- Ajouter un processus de journalisation des erreurs.
- Réécrire l'extraction de Portfolio et de Pivotal.
- Régler les problèmes de performance.
- Automatiser les demandes récurrentes (rapport annuel).
- Ajouter de nouvelles sources de données et de nouveaux rapports.
- Améliorer les outils de présentation.
- Créer un tableau de bord institutionnel pour le CA.

Bibliothèque et Archives nationales du Québec
