

Dossier client

Laboratoire BioSeq — demande d'un outil de sauvegarde

Ressource R3.03 — BUT2 Informatique

Travail à la maison — À faire avant le premier cours — 30 min, individuel

1. Lire ce dossier en entier : le contexte, la demande, puis la transcription de l'entretien.
2. **Surligner de deux couleurs** : en jaune ce que Karim **demande**, en vert ce dont il a manifestement **besoin**. Ce ne sont pas les mêmes phrases — et il y en a beaucoup moins en vert.
3. Relever **trois faits chiffrés** de l'entretien. Ils serviront à justifier des exigences dès le TD1.
4. Arriver avec **une question** que vous auriez posée à Karim et qui n'a pas été posée dans cet entretien.

Ce travail est individuel. C'est la confrontation de vos lectures, au sein de votre groupe, qui ouvrira le TD1.

En parallèle, la logistique du groupe — constitution du groupe, création du dépôt Git, accès de l'enseignant, configuration de votre identité — est décrite dans le fascicule des séances, partie *Organisation et évaluation*. Elle doit être faite **avant le premier cours** : le TD1 et le TP1 se suivent sans intervalle.

1. Le contexte

Le laboratoire de recherche **BioSeq** est une unité mixte universitaire spécialisée dans le traitement de données génomiques. Il compte 18 chercheurs et 4 doctorants, répartis sur deux étages d'un même bâtiment.

Le parc informatique est hétérogène : une dizaine de postes de travail, dont sept sous Linux et trois sous macOS, plus deux serveurs de calcul sous Linux. Les données de travail sont centralisées sur un serveur de fichiers monté sur tous les postes, sous le point de montage /data. Un boîtier NAS, acheté il y a trois ans, sert de support de sauvegarde.

Le laboratoire n'a pas de service informatique. **Karim Berthier**, ingénieur d'études recruté sur un poste de bio-informatique, administre les machines « en plus » de son travail de recherche, à raison d'environ une demi-journée par semaine. Il est en déplacement sur des sites partenaires deux jours par semaine en moyenne.

C'est lui qui vous a contactés.

2. La demande, telle qu'elle est arrivée

Courriel reçu le 24 août

« Bonjour,

Suite à notre conversation téléphonique : il me faudrait un script qui fasse un `tar.gz` de /data toutes les nuits sur le NAS. Ça devrait être vite fait, non ? Dites-moi ce dont vous avez besoin comme accès.

Merci d'avance,
Karim »

3. Transcription de l'entretien

Entretien mené le 31 août, dans le bureau de Karim. Durée : 35 minutes. La transcription a été allégée des passages sans rapport avec le sujet.

Analyste : Merci de me recevoir. Avant de parler technique, j'aimerais comprendre le contexte. Racontez-moi la dernière fois où l'absence de sauvegarde vous a posé un vrai problème.

Karim : Alors ça, c'est frais. Il y a trois semaines. Une doctorante a lancé un script de nettoyage avec un mauvais chemin en argument — une erreur bête, elle voulait vider un dossier temporaire. Le script a écrasé six mois de résultats intermédiaires sur un projet de séquençage.

Analyste : Et vous vous en êtes aperçus quand ?

Karim : Quatre jours plus tard. C'est ça le pire. Elle était partie en conférence, personne n'a touché au dossier entre-temps. C'est en revenant qu'elle a vu que ses fichiers n'étaient plus là.

Analyste : Qu'auriez-vous voulu pouvoir faire à ce moment-là ?

Karim : Récupérer le dossier tel qu'il était le jeudi précédent. Juste ce dossier. Pas toute la machine, pas tout /data — un dossier, à une date donnée.

Analyste : Et vous n'avez pas pu ?

Karim : Non. J'avais bien une sauvegarde, mais de la veille. C'est-à-dire d'après l'incident. Le mal était déjà fait dedans.

Analyste : Vous n'en gardez qu'une ?

Karim : Deux. Celle de la veille et celle d'avant-veille. Je fais tourner.

Analyste : Pourquoi seulement deux ?

Karim : Parce que /data, c'est 400 Gio. Le NAS en fait 1,5 Tio. Vous faites le calcul : à 400 Gio par archive, je ne peux pas en empiler trente. Déjà à trois, je suis à l'étroit — et /data grossit d'environ 5 Gio par mois.

Analyste : Vous m'avez écrit que vous vouliez des tar.gz chaque nuit. Si vous en aviez trente, ça résoudrait le problème ?

Karim : En théorie oui, mais je viens de vous dire que je n'ai pas la place. Et puis même si je l'avais. . . vous imaginez ? Pour ressortir un dossier, il faudrait que je sache dans quelle archive chercher, que je la décompresse en entier — 400 Gio, ça prend un temps fou et il faut de la place pour le faire — juste pour en extraire trois fichiers. La dernière fois j'ai essayé, j'ai abandonné au bout d'une heure.

Analyste : Donc, si je résume : votre profondeur d'historique réelle est de deux jours.

Karim : (silence) Deux jours. Et l'incident a été découvert au bout de quatre. Dit comme ça, oui, ma sauvegarde ne sert à rien.

Analyste : Qui restaure, aujourd'hui ?

Karim : Moi. Enfin, moi quand je suis là. Je suis en déplacement deux jours par semaine, et je ne suis pas au bureau avant 9 h 30. Si quelqu'un a besoin d'un fichier un mardi matin, il attend.

Analyste : Est-ce que les chercheurs pourraient le faire eux-mêmes ?

Karim : Ce sont des biologistes, pas des informaticiens. Mais ils savent ouvrir un dossier et copier un fichier, ça oui. Ce qu'ils ne feront jamais, c'est apprendre une commande spécifique et retenir sa syntaxe pour l'utiliser deux fois par an.

Analyste : Vous aviez déjà essayé un autre outil ?

Karim : Le labo avait acheté un logiciel de sauvegarde il y a six ou sept ans. Quand j'ai voulu récupérer de vieilles données, l'éditeur avait disparu et les fichiers étaient dans un format

que rien ne savait ouvrir. On avait les données sur le disque, sous les yeux, et on ne pouvait rien en faire. J'ai encore le disque dans un tiroir, d'ailleurs.

Analyste : Comment savez-vous, le matin, que la sauvegarde de la nuit s'est bien passée ?

Karim : (*rire*) Je ne le sais pas. Je regarde de temps en temps. Il y a un fichier de log quelque part, mais je crois qu'il fait 200 Mo et je ne l'ai pas ouvert depuis un an.

Analyste : Donc si la sauvegarde échoue un soir, vous l'apprenez. . .

Karim : Le jour où j'en ai besoin. C'est-à-dire au pire moment possible.

Analyste : Parlons des contraintes pratiques. Quand la sauvegarde tourne-t-elle ?

Karim : La nuit, entre 2 h et 6 h. Après 7 h, les premiers arrivent et lancent des calculs ; si la sauvegarde tourne encore, tout rame et je reçois des plaintes. Donc il faut que ce soit fini avant.

Analyste : Combien de fichiers y a-t-il dans /data, à peu près ?

Karim : Beaucoup. J'ai lancé un `find` une fois par curiosité, c'était de l'ordre de 500 000. Mais la plupart ne bougent jamais : une fois qu'un run de séquençage est terminé, les fichiers sont figés. Ce qui change au jour le jour, c'est les scripts d'analyse et les résultats en cours, et ça, c'est tout petit.

Analyste : Et les postes ? Vous m'avez parlé de Linux et de Mac.

Karim : Sept Linux, trois Mac, deux serveurs Linux. Les trois Mac, ce sont les chefs d'équipe, je ne vais pas leur demander de changer. Si votre truc ne marche que sur Linux, ça ne me sert qu'à moitié.

Analyste : Dernière question pour aujourd'hui : qu'est-ce qui vous empêche de dormir, dans cette histoire ?

Karim : Qu'on perde une thèse. Un doctorant, c'est trois ans de travail. S'il perd six mois de données et qu'on ne peut pas les remonter, ce n'est pas un incident informatique, c'est un drame humain. Le reste, franchement, je m'en arrangerai.

4. Ce que vous ferez de ce dossier

Ce dossier est votre **source initiale** sur le besoin du client. Pendant les quatre TD, un membre de votre groupe tiendra le rôle de *client* : sa mission sera de distinguer faits du dossier, hypothèses et compléments. Les décisions supplémentaires du client fictif sont explicitement identifiées dans les TD et ajoutées à vos sources.

TD1 vous en tirerez les problèmes métier, puis vos premières exigences

TD2 vous écrirez les *User Stories* correspondantes

TD3 vous chercherez tout ce qui peut mal se passer

TD4 vous en extrairez le vocabulaire et le modèle du métier

✗ À ne pas faire : L'erreur à ne pas commettre

Inventer les besoins de Karim. Vous serez tentés, en TD, d'écrire des exigences qui vous paraissent évidentes — « il voudra sûrement du chiffrement », « il faudra une interface web ». **Il ne l'a jamais dit.**

Une proposition qui ne s'appuie ni sur ce dossier ni sur un complément explicite reste une hypothèse à confirmer. Notez-la comme telle et préparez la question qui permettrait de la valider.