Sémantique de C2QL

Santiago Bautista

Juin 2017

Le but de ce document est de donner une définition formelle des fonctions dont est composé le langage C2QL.

Préambule

Définition 1 *On appelle* nom d'attribut *toute chaîne de caractères.*

Ici, pour simplifier, on appelle chaîne de caractères tout mot sur l'alphabet

$$\Sigma = \{a, \ldots, z\} \cup \{A, \ldots, Z\} \cup \{0, \ldots, 9\}$$

Vu que le nom d'attribut « id » joue un rôle particulier, on appelle, par opposition, nom d'attribut régulier *tout nom d'attribut autre que « id ».*

Définition 2 *On appelle* schéma relationnel *tout ensemble de noms d'attributs réguliers.*

Définitions générales

Définition 3 *On appellera* valeur *tout élément d'un certain ensemble $\mathcal{V}$, que l'on suppose non-vide, infini, dénombrable, et stable par formation de n-uplets (i.e. $\forall k \in \mathbf{N}, \mathcal{V}^k \subset \mathcal{V}$).*

Définition 4 *On appelle* relation *de schéma relationnel Δ tout ensemble de fonctions de $\Delta \cup \{id\}$ dans $\mathcal{V}$.*

Chacun des éléments de la relation (chacune de ces fonctions) est appelé(e) ligne.

Pour chaque ligne l de la relation et chaque α de Δ, $l(\alpha)$ est appelé attribut *de* nom α *pour la ligne l.*

L'image de id est appelée identifiant de la ligne*, et elle est, au sein de chaque relation, unique pour chaque ligne.*

Définition 5 *On appelle* S *l'ensemble des schémas relationnels possibles. Autrement dit, on pose* $\mathrm{S} = \mathcal{P}\left(\Sigma^* \setminus \{id\}\right)$.

On appelle T *l'ensemble des relations possibles,*

et on introduit la fonction sch *de* T *dans* S *qui à une relation associe son schéma relationnel.*

Projections et sélections

Définition 6 *Pour tout ensemble δ de noms d'attributs réguliers, on appelle* projection sur les attributs δ *la fonction suivante :*

$$\begin{array}{cccc} \pi_\delta : & \mathrm{T} & \to & \mathrm{T} \\ & r & \mapsto & \{l|_{(\delta \cap \mathrm{sch}(r)) \cup \{id\}} / l \in r\} \end{array}$$

Définition 7 *On appelle* L *l'ensemble de toutes les lignes possibles.*

On appelle prédicat *toute fonction de* L *dans* $\{true, false\}$.

On appelle domaine *d'un prédicat p le plus petit ensemble D tel que :*

$$\forall (l, l') \in \mathrm{L}^2, (l|_D = l'|_D \Rightarrow p(l) = p(l'))$$

et on le note $\mathrm{dom}(p)$.

Définition 8 *On appelle* sélection de prédicat p, *pour tout prédicat p, la fonction :*

$$\begin{array}{rccl} \sigma_p : & \mathrm{T} & \to & \mathrm{T} \\ & r & \mapsto & r \cap p^{-1}(\{true\}) \end{array}$$

Jointure naturelle

Définition 9 *Si r et r' sont deux relations et que l est un élément de r. On appelle* correspondants de l dans r' *l'ensemble des lignes l' telles que*

$$\forall \alpha \in \mathrm{sch}(r) \cap \mathrm{sch}(r'), l'(\alpha) = l(\alpha)$$

On note $\mathrm{cor}_{r,r'}(l)$ *l'ensemble de ces lignes-là.*

Définition 10 *Si l et l' sont deux lignes correspondantes, on appelle* concaténation de l et de l', *notée $l.l'$ la fonction de* $\mathrm{sch}(l) \cup \mathrm{sch}(l') \cup \{id\}$ *définie par :*

$$\begin{cases} l.l'(\alpha) = l(\alpha) & si\ \alpha \in \mathrm{sch}(r) \setminus \mathrm{sch}(r') \\ l.l'(\alpha) = l'(\alpha) & si\ \alpha \in \mathrm{sch}(r') \setminus \mathrm{sch}(r) \\ l.l'(\alpha) = l(\alpha) = l'(\alpha) & si\ \alpha \in \mathrm{sch}(r) \cap \mathrm{sch}(r') \\ l.l'(id) = \gamma & \gamma\ \textit{étant un identifiant frais} \end{cases}$$

où on appelle identifiant frais *une valeur qui ne soit l'identifiant d'aucune autre ligne dans le système.*

Définition 11 *Pour r et r' deux relations, on appelle* jointure naturelle *de r et r' la relation*

$$r \bowtie r' = \{l.l' / l \in r, l' \in \mathrm{cor}_{r,r'}(l)\}$$

On utilisera aussi la notation préfixe. En effet, on vient de définir la fonction

$$\begin{array}{rccl} \bowtie : & \mathrm{T}^2 & \to & \mathrm{T} \\ & (r, r') & \mapsto & r \bowtie r' \end{array}$$

Fragmentation et défragmentation

La défragmentation est presque un cas particulier de jointure naturelle, où l'identifiant serait considéré comme un attribut en commun pour les deux tables et il serait le seul.

Définition 12 *Deux relations r et r' sont dites* unifiables *si :*

$$\mathrm{sch}(r) \cap \mathrm{sch}(r') = \emptyset$$

On remarquera que deux relations unifiables non vides sont également joignables.

On note Tu *l'ensemble des paires de relations unifiables, qui est un sous-ensemble de* T^2.

Définition 13 *Pour tout ensemble de noms d'attributs réguliers δ on appelle* fragmentation de fragment gauche δ *l'application suivante :*

$$\begin{array}{lccl} \mathrm{frag}_\delta & \mathrm{T} & \to & \mathrm{Tu} \\ & r & \mapsto & (\{l|_{(\mathrm{sch}(r)\cap\delta)\cup\{id\}}/l \in r\}, \{l_{(\mathrm{sch}(r)\setminus\delta)\cup\{id\}}/l \in r\}) \end{array}$$

Définition 14 *On dit que deux lignes l et l' sont* unifiables *si elles partagent le même identifiant et que les relations correspondantes sont unifiables.*

On définit alors leur unification $\mathrm{Unif}(l, l')$ *comme la fonction définie sur* $\mathrm{sch}(l) \cup \mathrm{sch}(l') \cup \{id\}$ *par*

$$\left\{\begin{array}{ll} \mathrm{Unif}(l, l')(\alpha) = l(\alpha) & si\ \alpha \in \mathrm{sch}(l) \\ \mathrm{Unif}(l, l')(\alpha) = l'(\alpha) & si\ \alpha \in \mathrm{sch}(l') \\ \mathrm{Unif}(l, l')(id) = l(id) = l'(id) & \end{array}\right.$$

Définition 15 *On appelle* défragmentation *la fonction de* Tu *à valeur dans* T *définie par :*

$$\begin{array}{lccl} \mathrm{defrag}: & \mathrm{Tu} & \to & \mathrm{T} \\ & (r, r') & \mapsto & \{\mathrm{Unif}(l, l')/l(id) = l'(id), l \in r, l' \in r'\} \end{array}$$

Chiffrement et déchiffrement

Vu que pour l'instant on s'intéresse uniquement aux contenus des tables pour démontrer la correction sémantique des lois de composition, on ne parlera pas pour l'instant des éventuelles clefs de chiffrement et déchiffrement.

Définition 16 *On appelle* chiffrement *tout couple* c *de fonctions* $(\mathrm{Enc}, \mathrm{Dec})$ *de* $\mathcal{V}$ *dans* $\mathcal{V}$ *vérifiant* $\mathrm{Dec} \circ \mathrm{Enc} = \mathrm{id}$.

Pour toute valeur v de $\mathcal{V}$ on note $\mathsf{c}(v) = \mathrm{Enc}(v)$ *et* $\mathsf{c}^{-1}(v) = \mathrm{Dec}(v)$

Définition 17 *Pour une ligne l définie sur Δ, pour α un attribut, et pour c un chiffrement, on appelle* version de l chiffrée pour α avec le chiffrement c *la ligne notée* $\mathsf{c}(l)_\alpha$ *définie par :*

$$\left\{\begin{array}{ll} \forall\beta \in \Delta \setminus \{\alpha\} & \mathsf{c}(l)_\alpha(\beta) = l(\beta) \\ & \mathsf{c}(l)_\alpha(\alpha) = \mathsf{c}(l(\alpha)) \quad si\ \alpha \in \Delta \end{array}\right.$$

De même, on définit la version de l déchiffrée pour α avec le chiffrement c, *notée* $\mathsf{c}^{-1}(l)_\alpha$, *par :*

$$\left\{\begin{array}{ll} \forall\beta \in \Delta \setminus \{\alpha\} & \mathsf{c}^{-1}(l)_\alpha(\beta) = l(\beta) \\ & \mathsf{c}^{-1}(l)_\alpha(\alpha) = \mathsf{c}^{-1}(l(\alpha)) \quad si\ \alpha \in \Delta \end{array}\right.$$

Définition 18 *Pour α un nom d'attribut et c un chiffrement, on appelle* fonction de chiffrement de α par c *la fonction*

$$\begin{array}{lccl} \mathrm{crypt}_{\alpha,\mathsf{c}}: & \mathrm{T} & \to & \mathrm{T} \\ & r & \mapsto & \{\mathsf{c}(l)_\alpha/l \in r\} \end{array}$$

De même, on appelle fonction de déchiffrement de α par c *la fonction*

$$\begin{array}{lccl} \mathrm{decrypt}_{\alpha,\mathsf{c}}: & \mathrm{T} & \to & \mathrm{T} \\ & r & \mapsto & \{\mathsf{c}^{-1}(l)_\alpha/l \in r\} \end{array}$$

Définition 19 *On dit d'un prédicat p et un chiffrement c sont compatibles pour l'attribut α s'il existe un autre prédicat $\mathsf{c}_\alpha \Rightarrow p$ ne dépendant que de p, c et du nom d'attribut α tel que*

$$\forall l \in L, p(l) = (\mathsf{c}_\alpha \Rightarrow p)(\mathsf{c}(l)_\alpha)$$

Agrégation

Définition 20 *Pour δ un ensemble de noms d'attributs réguliers, on appelle* nom de groupe pour δ *toute application n définie de δ à valeurs dans $\mathcal{V}$.*

On remarque que tout nom de groupe est une ligne.

δ est appelé domaine *du nom de groupe n, et noté* $\mathrm{dom}(n)$.

De plus, pour r une relation, on définit l'ensemble des noms de groupe de r pour δ *:*

$$r_\delta = \{l|_\delta / l \in r\}$$

Définition 21 *Pour r une relation et n un groupe, on appelle* groupe de r pour le nom n *l'ensemble des éléments de r coïncidant avec n sur* $\mathrm{sch}(r) \cap \mathrm{dom}(n)$*. On le note r_n.*

Autrement dit :

$$r_n = \{l \in r / l|_{\mathrm{sch}(r)\cap\mathrm{dom}(n)} = n|_{\mathrm{sch}(r)\cap\mathrm{dom}(n)}\}$$

De plus, on appelle identifiants du groupe r_n *l'ensemble des identifiants des lignes du groupe. On note* $\mathrm{IDs}(r_n)$ *cet ensemble.*

Autrement dit :

$$\mathrm{IDs}(r_n) = \{l(id)/l \in r_n\}$$

Définition 22 *On dira qu'une application f est* plus petite *qu'une application g si f est une restriction de g.*

On dira qu'un nom de groupe n_0 est minimal *pour une relation r donnée si c'est une plus petite application n pour laquelle le groupe de r pour n vaut r_{n_0}.*

Définition 23 *Pour r une relation, n un nom de groupe, et α un attribut de* $(\mathrm{sch}(r) \setminus \mathrm{dom}(r)) \cup \{id\}$*, on appelle* valeurs du groupe r_n pour l'attribut α *la fonction*

$$\begin{array}{rccc} r_n(\alpha) : & \mathrm{IDs}(r_n) & \rightarrow & \mathcal{V} \\ & l(id) & \mapsto & l(\alpha) \end{array}$$

Remarque : Souvent, on supposera que l'ensemble des identifiants possible est totalement ordonné et on s'en servira pour considérer des fonctions définies sur un ensemble d'identifiants (par exemple les $r_n(\alpha)$ définis ci-dessus) comme des listes.

On définit la longueur de telles listes comme le cardinal de leur ensemble de départ. Par exemple, la longueur de $r_n(\alpha)$ est $|r_n(\alpha)| = |\mathrm{IDs}(r_n)|$

Définition 24 *Pour r une relation, et n un nom de groupe, on appelle* ligne de groupe de r pour n *la ligne notée* $\mathrm{lg}_{r,n}$ *définie sur* $\mathrm{sch}(r) \cup \{id\}$ *par :*

$$\begin{cases} \mathrm{lg}_{r,n}(\alpha) = n(\alpha) & \textit{si } \alpha \in \mathrm{sch}(r) \cap \mathrm{dom}(n) \\ \mathrm{lg}_{r,n}(\alpha) = r_n(\alpha) & \textit{si } \alpha \in (\mathrm{sch}(r) \setminus \mathrm{dom}(n)) \\ \mathrm{lg}_{r,n}(id) = \gamma & \textit{où } \gamma \textit{ est un identifiant frais} \end{cases}$$

Définition 25 *Pour δ un ensemble de noms d'attributs, on appelle* fonction d'agrégation pour les attributs δ *la fonction suivante :*

$$\begin{array}{rccc} \mathrm{group}_\delta : & \mathrm{T} & \rightarrow & \mathrm{T} \\ & r & \mapsto & \{\mathrm{lg}_{r,n} / n \in r_\delta\} \end{array}$$

Réduction

La plupart du temps, les agrégations sont faites pour pouvoir faire une réduction ensuite.

On suppose que les identifiants des lignes peuvent être totalement ordonnés et donc que les fonctions définies sur des ensembles d'identifiants peuvent être vues comme des listes.

Pour toute liste l on notera $\mathrm{hd}(l)$ le premier élément de la liste, et $\mathrm{tl}(l)$ le reste de la liste.

Dans les définitions qui suivent, f est une fonction de $\mathcal{V}^2$ dans $\mathcal{V}$ et z est un élément de $\mathcal{V}$.

Définition 26 *On appelle* réduction d'une liste t par la fonction f avec l'élément neutre z *la valeur* $\mathrm{red}_{f,z}(t)$ *définie par induction sur la liste par :*

$$\begin{cases} \mathrm{red}_{f,z}(\emptyset) = z \\ \mathrm{red}_{f,z}(t) = \mathrm{red}_{f,f(z,hd(t))}(\mathrm{tl}(t)) \end{cases}$$

Si une valeur v de $\mathcal{V}$ n'est pas une liste, on la considère alors comme une liste à un seul élément et on pose donc $\mathrm{red}_{f,z}(v) = f(z,v)$.

Définition 27 *Pour l une ligne définie sur δ, et α un nom d'attribut régulier, on appelle* réduction de l'attribut α dans la ligne l par la fonction f avec l'élément neutre z *la ligne* $\mathrm{red}_{\alpha,f,z,l}$ *définie sur δ par :*

$$\begin{cases} \mathrm{red}_{\alpha,f,z,l}(\alpha) = \mathrm{red}_{f,z}(l(\alpha)) & si\ \alpha \in \delta \\ \mathrm{red}_{\alpha,f,z,l}(\beta) = l(\beta) & si\ \beta \neq \alpha \end{cases}$$

Définition 28 *On appelle* fonction de réduction de l'attribut α par la fonction f avec l'élément neutre z *la fonction suivante :*

$$\begin{cases} \mathrm{fold}_{\alpha,f,z} : & \mathrm{T} & \rightarrow & \mathrm{T} \\ & r & \mapsto & \{\mathrm{red}_{\alpha,f,z,l}\,/l \in r\} \end{cases}$$

Opérations ensemblistes : union, différence, fragmentation horizontale

Définition 29 *On dit que deux tables r et r' sont* défragmentables horizontalement *si elles ont même schéma relationnel e leurs ensembles d'identifiants sont disjoints. Autrement dit,*

$$\begin{cases} \mathrm{sch}(r) = \mathrm{sch}(r') \\ \{l(id)/l \in r\} \cap \{l(id)/l \in r'\} = \emptyset \end{cases}$$

On appelle union *ou* défragmentation horizontale *de deux tables r et r' défragmentables horizontalement la table $r \cup r'$, aussi notée* $\mathrm{hdefrag}(r, r')$.

Définition 30 *On appelle* différence ensembliste *de deux tables r et r' ayant le même schéma relationnel la table $r \setminus r'$.*

Définition 31 *Pour p un prédicat, on appelle* fragmentation horizontale de critère p *la fonction*

$$\begin{array}{lll} \mathrm{hfrag} : & \mathrm{T} \rightarrow & \mathrm{T}^2 \\ & r & \mapsto \quad (\{l \in r/p(l)\}, \{l \in r, \neg p(l)\}) \end{array}$$

On remarquera que les deux tables du résultat sont horizontalement défragmentables.