

DATARMOR : la mer a son supercalculateur !

Mercredi 28 juin, l’Ifremer inaugure le supercalculateur DATARMOR, nouvelle infrastructure de calcul/traitement de données marines. Au premier semestre 2017, DATARMOR figure à la 446^{ème} place du top 500 mondial des supercalculateurs.

Localisée au Centre Ifremer Bretagne, DATARMOR est une infrastructure informatique au service des sciences marines qui remplace le précédent supercalculateur CAPARMOR. DATARMOR renforce les capacités de stockage, en passant de 0,5 pétaoctet à 7 pétaoctets, et offre une puissance de traitement des données marines 15 fois supérieure. Ainsi, de 27 téraflops avec CAPARMOR, on passe à une puissance de calcul de 426 téraflops.

Ces capacités accrues vont permettre de relever les défis du Big Data qui résultent de l’accroissement du volume et de l’hétérogénéité des données. Elles ouvrent de nouvelles perspectives d’investigation pour les chercheurs : une meilleure synergie entre le traitement et l’interprétation des données va permettre d’extraire des informations nouvelles, et par exemple d’améliorer les modèles de prévision océanique.



Supercalculateur DATARMOR © Ifremer/Olivier Dugornay



Les besoins des sciences marines en calcul et traitement de données

S'inscrivant dans la feuille de route nationale des Infrastructures de Recherche sous "pôles de données et de services du système terre", le projet DATARMOR vise à conforter et améliorer le positionnement national, européen et international du pôle breton en sciences marines en :

- Renforçant les capacités de stockage et de traitement des données marines
- Faisant évoluer les moyens de calculs dédiés à la modélisation et aux traitements
- Ouvrant de nouvelles perspectives d'investigation de données pour les chercheurs, grâce à l'apport des architectures parallèles de type « big data » (autrement appelé « mégadonnées » ou « données massives »).



« DATARMOR vise à relever les défis que représentent aujourd'hui les traitements et interprétations des données de gros volume », explique Pierre Cotty, responsable du Département Infrastructures Marines et Numériques de l'Ifremer. « La synergie entre ces deux volets permettra d'extraire des informations nouvelles à partir des données marines, d'améliorer les modèles de prévision océanique, et de développer des méthodes innovantes à chaque étape du cycle des données marines (collecte, bancarisation, traitement/interprétation, modélisation) en levant les verrous techniques résultant de l'accroissement du volume et de l'hétérogénéité des données ».

Pour en savoir plus sur les données, visitez le portail des données marines de l'Ifremer : <https://data.ifremer.fr>

Le milieu marin, un espace partagé

Les sciences marines ont leur centre de calcul depuis plus de 25 ans à Brest. Depuis 2001, le cofinancement des collectivités territoriales a grandement contribué à élever le niveau technique du centre, devenu « Pôle de Calcul Intensif pour la Mer » (PCIM) dans le cadre d'un partenariat élargi. Il est l'un des vecteurs de la coopération entre tous les chercheurs et ingénieurs de cette thématique.

On compte aujourd'hui plus de 600 utilisateurs issus de plus de 20 organismes et entreprises (PMEs et startups ciblant le domaine marin), avec un niveau d'usage de plus de 80 % pour la machine précédente CAPARMOR (arrêtée le 12 mai dernier), dont 20 % du temps de calcul attribué aux entreprises. DATARMOR va conforter les partenariats existants et ouvrir de nouvelles opportunités pour les données marines. L'implication du Pôle de compétitivité Mer Bretagne-Loire va permettre de mieux organiser les usages auprès des entreprises. L'élargissement de l'infrastructure DATARMOR aux données de gros volumes englobera de nouveaux utilisateurs en attente de moyens techniques à la hauteur de leurs objectifs de traitement de données et/ou de modélisation.

Ce projet réunit plusieurs partenaires au sein du Pôle de Calcul Intensif pour la Mer (PCIM) : l'Ifremer, le SHOM, l'IUEM/UBO, le CNRS, l'ENSTA Bretagne, l'École Navale, l'IRD et le Pôle Mer Bretagne Atlantique.

Le financement global (investissement) du projet CPER DATARMOR (2016-2020) est de 6,9 M€. Les financeurs sont l'Union européenne (fonds FEDER) (à hauteur de 2,2 M€), le ministère chargé de la Recherche (2,1 M€), la région Bretagne (0,95M€), le département du Finistère (0,475M€), Brest Métropole (0,475M€), le SHOM (0,5M€) et l'Ifremer (0,2M€). L'Ifremer assurera l'hébergement technique de la machine et les coûts de fonctionnement inhérents (administration technique, maintenance, énergie...)

Le changement climatique, l'acidification de l'océan, les migrations des espèces, les profonds bouleversements des écosystèmes, l'accroissement de la pression anthropique en zone côtière,... sont autant de défis qui demandent une approche scientifique intégrant à la fois les observations les plus nombreuses et précises possible que des modèles numériques performants.

L'évolution de la modélisation de l'océan

« La modélisation de l'océan évolue dans cinq directions. Elle accompagne à la fois les progrès faits en calcul numérique, en augmentation de la puissance de calcul, en capacité d'observation et en concepts théoriques », explique Pierre Garreau, président du Comité de Pilotage du projet DATARMOR et responsable du laboratoire Océan Côtier.

1) Vers une modélisation des conséquences du changement climatique

Le changement climatique s'impose maintenant comme une évidence. Il reste à en évaluer les conséquences sur les mers régionales et côtières. Cela nécessite la compréhension des variations inter-annuelles de la température, de la salinité, des courants marins. Ce travail est mené par exemple sur le golfe de Gascogne en comparant des simulations longues (au moins 10 ans) avec les observations.

2) Vers une modélisation haute résolution

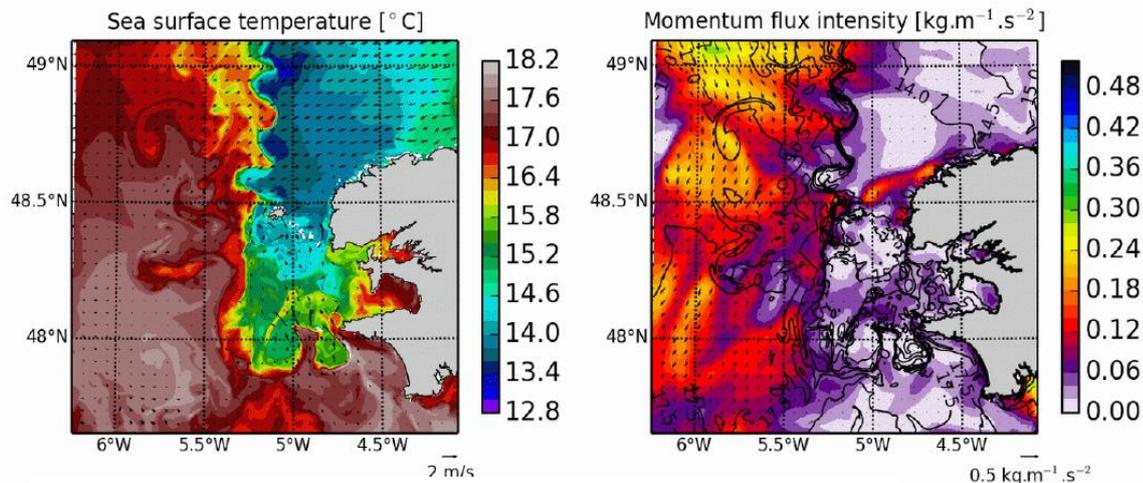
Il y a dans l'océan une gamme de phénomènes (tourbillons, filaments, fronts) dont la taille varie de quelques centaines de mètres à quelques dizaines de kilomètres. Ils sont maintenant peu à peu accessibles à la modélisation et à l'observation, et apparaissent de plus en plus comme une composante essentielle et non négligeable de la dynamique, à côté des courants généraux ou de la marée. En Méditerranée Occidentale durant les événements de mistral hivernaux, l'eau se refroidit et s'évapore tellement en surface qu'elle devient très dense et coule parfois jusqu'au fond. Ce phénomène de convection profonde nécessite de reproduire des processus de l'ordre du kilomètre.

3) Vers des études de processus

La simulation réaliste d'une zone particulière est parfois très complexe et le modélisateur peut être amené à reproduire « *en laboratoire numérique* », dans un océan parfait, à fond plat, sans limites latérales, un seul processus pour mieux le comprendre et le comparer à la théorie. Dans ce cas, on s'intéresse au rôle des tout petits tourbillons dans les mélanges verticaux et horizontaux des masses d'eau.

4) Vers un couplage atmosphère-océan-vagues

L'atmosphère et l'océan échangent sans arrêt à leur interface de la chaleur, de l'eau (évaporation et précipitation), du mouvement (le vent). Ces échanges dépendent fortement de l'état de mer (les vagues). Jusqu'à présent modélisés séparément, l'atmosphère, l'océan et les vagues, sont maintenant modélisés ensemble, chacun agissant sur les deux autres. Cette modélisation dite « couplée » appliquée à la mer d'Iroise par exemple montre l'importance de ces échanges.



Couplage atmosphère-océan-vagues : le front d'Ouessant

5) Vers une utilisation intensive conjointe modèles/mesures *in situ*/images satellites

S'il reste encore des efforts à faire en zone côtière, le développement rapide de moyens de mesure autonomes augmente considérablement la masse des données *in situ* disponibles pour valider les modèles numériques. Associée aux images satellites, cette information est maintenant indissociable d'une activité de modélisation.

L'utilisation de DATARMOR pour l'étude de la dynamique des écosystèmes côtiers

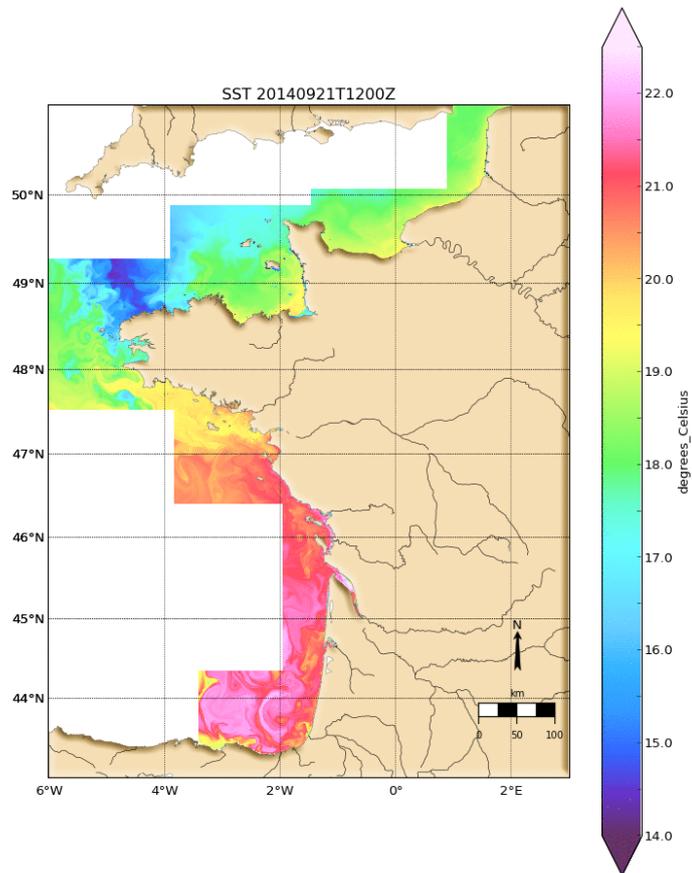
Les études en environnement côtier nécessitent des moyens croissants en puissance de calcul et en capacité de stockage pour :

1) Améliorer les modèles numériques

« *Il s'agit d'améliorer les modèles numériques utilisés maintenant depuis de nombreuses années et dont la complexité et la finesse (spatiale et temporelle) augmentent régulièrement* », explique Philippe Cugier, chercheur en modélisation des écosystèmes marins côtiers au Centre Ifremer Bretagne. A l'Ifremer, la modélisation océanographique côtière se fait principalement grâce au modèle communautaire MARS qui est un modèle pluridisciplinaire intégrant des modules permettant de simuler la dynamique sédimentaire, biogéochimique, des contaminants. Les modèles mis en œuvre considèrent la colonne d'eau et le sédiment, et couvrent différentes échelles spatiales depuis des zones très côtières (estuaires, baies) jusqu'aux façades maritimes (Manche, golfe de Gascogne,...) avec des résolutions pouvant aller de quelques dizaines de mètres à plusieurs kilomètres.

Les questions et la complexité des systèmes étudiés nécessitent parfois l'utilisation de plusieurs modules simultanément associés à un grand nombre de variables d'état afin de représenter les composantes essentielles de l'écosystème (sédiment, phytoplancton, faune du fond marin, etc.). Les échelles de temps abordées peuvent être de quelques heures/jours (marée, dispersion d'un contaminant, dynamique d'un bouchon vaseux,...), de quelques semaines (dispersion de larves), de l'ordre de la saison (dynamique phytoplanctonique) jusqu'aux simulations pluriannuelles/décennales (dynamique de population d'organismes du fond marin à long cycle de vie), voire centennales (morphodynamique côtière).

Simulation des conditions hydrologiques de la façade Manche Atlantique à l'aide du modèle MARS3D pour alimenter une base de données de paramètres physiques pour la communauté scientifique - exemple de la température de surface de la mer



Le besoin de moyens de calcul performants concerne aussi et de manière plus récente le traitement de volumes de données de plus en plus importants acquis par les moyens d'observation modernes (données satellites, données génomiques,...), voire générées par les modèles eux-mêmes. Ces traitements nécessitent des algorithmes performants, parallélisés et l'accès à une mémoire vive importante.

La puissance de calcul de DATARMOR (exécution, mémoire) associée à la parallélisation des codes de calcul et/ou des algorithmes permettent d'aborder les différentes échelles spatiales et temporelles pour l'étude de la dynamique des écosystèmes côtiers et le traitement de données volumineuses.

Quelques exemples non exhaustifs d'objets d'étude peuvent illustrer ces différents points :

- La simulation des conditions hydrologiques sur la façade Manche-Atlantique.
- La simulation fine des conditions hydrodynamiques en rade de Brest pour l'étude des connectivités larvaires.
- La dynamique du bouchon vaseux de l'estuaire de Seine.
- La modélisation de la biodiversité phytoplanctonique en rade de Brest.
- La dynamique invasive d'un mollusque gastéropode, la crépidule, en rade de Brest.
- L'analyse d'images obtenues par capteurs hyperspectraux (*in situ* ou aéroportés) pour l'extraction des données biologiques ou environnementales (ex. cartographie des habitats, identification des espèces et description de leur organisation spatiale).
- La bio-informatique : analyser la diversité génétique des micro-algues toxiques à l'échelle du génome.

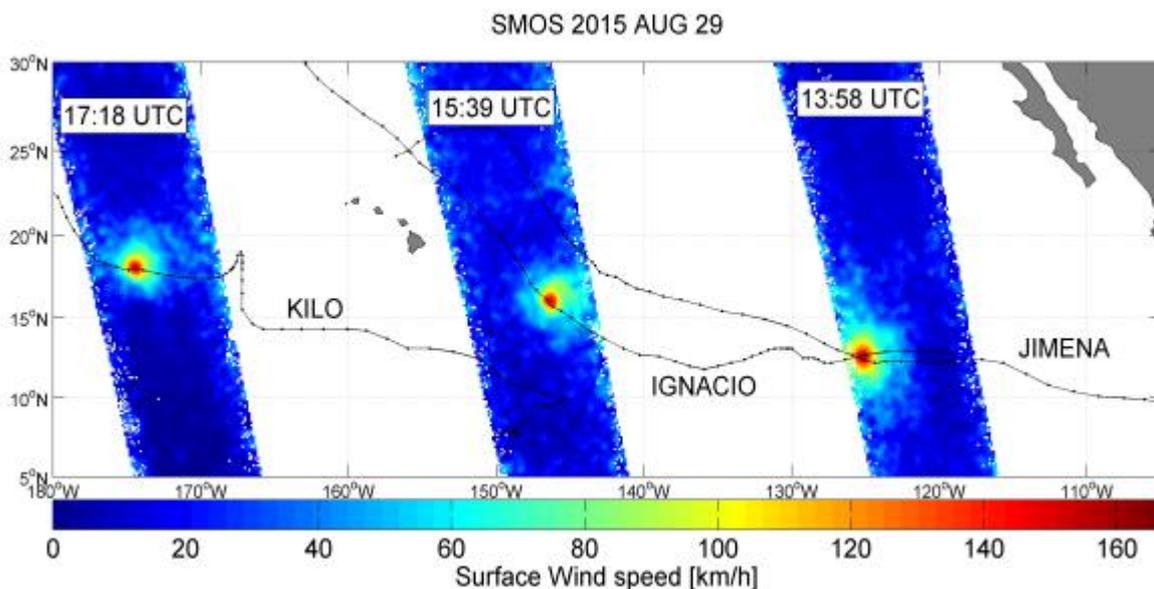
Enrichir le diagnostic du changement global grâce à DATARMOR

L'Ifremer participe aux travaux de recherche liés à la connaissance de la dynamique de l'océan pour enrichir le diagnostic du changement global à partir, entre autres, d'observations *in situ* et satellites. DATARMOR peut répondre aux besoins scientifiques liés à l'analyse des données océanographiques notamment issues des capteurs embarqués sur satellites.

« Les satellites fournissent une quantité d'observations en croissance quasi exponentielle. Nous disposons à présent de plus de 25 ans de mesures continues pour la majorité des paramètres d'océan, produites par différents capteurs avec des technologies et des résolutions spatiales et temporelles hétérogènes », expose Fanny Arduin, responsable du laboratoire Satellites et Interfaces Air-Mer. « La production d'indicateurs long terme du changement climatique requiert un travail important à la fois de collecte des observations, de définition d'indicateurs pertinents, de connaissance de la physique de la mesure des capteurs, de validation des paramètres d'océan déduits, d'inter-calibration entre les différentes sources de données, etc. »

Le travail d'homogénéisation des données de différents capteurs utile à toute analyse de type "observations longue série" est nécessaire. Ces longues séries à haute valeur ajoutée sont construites, archivées, analysées à l'Ifremer puis diffusées vers la communauté nationale, européenne et internationale. Ce travail ne peut se faire sans avoir une capacité de stockage de données adéquate et la possibilité de traitement global sur un ensemble de données historiques (traitement de données volumineuses).

Ces données servent de base à des axes de recherche thématiques divers, par exemple la compréhension de la dynamique de l'interface air-mer (vagues, vent, échanges air-océan...), la quantification des glaces en mer (banquise et icebergs), l'évolution de la salinité de surface, du cycle de l'eau, l'analyse des événements extrêmes, etc.

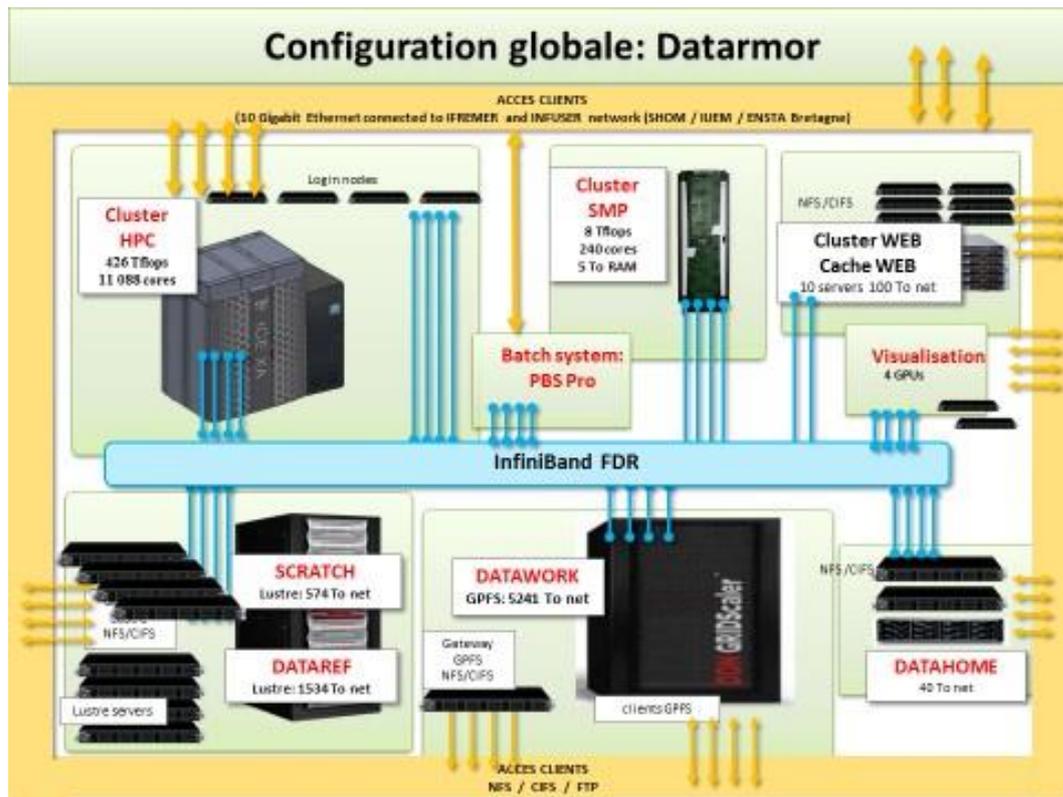


Estimation du vent pour les 3 tempêtes successives Kilo, Ignacio et Jimena en septembre 2015 à partir du capteur européen SMOS.

ANNEXE La configuration technique de DATARMOR

Le premier objectif du projet DATARMOR consistait à remplacer le cluster opérationnel CAPARMOR (qui date de la période 2007-2009) par un système technologiquement à jour, offrant une puissance de traitement bien supérieure. Le second objectif de DATARMOR est de développer des compétences et des techniques pour analyser/interpréter les très gros volumes de données marines hérités des systèmes d'observation en continu (satellite, ARGO, ...) ou de capteurs particuliers (vidéo des engins sous-marins, sondeurs multifaisceaux, sismique). Pour ce faire, des espaces de stockage volumineux étaient nécessaires pour rendre les données de référence disponibles avec de hautes performances, et mettre à disposition les espaces de travail en rapport avec le volume des données.

L'architecture de DATARMOR se compose de plusieurs ensembles (« grappes » ou « clusters ») de calcul/traitement ou de stockage de données interconnectés par un réseau ultra-rapide Infiniband. Cette architecture permet de répondre aux différents besoins exprimés par les utilisateurs scientifiques, tant en matière de capacité de traitement et/ou de stockage, qu'en matière de performances intégrées.



Grappes dédiées au traitement/calcul :

La grappe de calcul parallèle « ClusterHPC » permet de faire tourner des modèles ou des traitements de données « parallélisés », c'est-à-dire capables d'utiliser des processeurs et de la mémoire distribués sur des machines (« nœuds ») indépendantes. Cette grappe représente la plus grande part de la puissance de calcul de DATARMOR.

ClusterHPC se compose de 396 nœuds incluant chacun :

- 2 processeurs INTEL Broadwell E5-2680 v4, cadencé à 2,4GHz, comportant chacun 14 cœurs de calcul.
- 128 Go de mémoire

La grappe de calcul à mémoire partagée « ClusterSMP » permet de faire tourner des modèles ou des traitements qui ne peuvent pas être parallélisés, mais ont néanmoins besoin d'importantes ressources mémoire ; par exemple, les logiciels dont on ne maîtrise pas les sources peuvent faire partie de ces applications.

ClusterSMP inclut au total :

- 20 processeurs INTEL Xeon E5-4650v3, cadencés à 2.1GHz, comportant chacun 12 cœurs de calcul
- Un total de 5To de mémoire adressable par tous les 240 cœurs de calcul

La grappe de serveurs WEB « ClusterWEB » a pour fonction d'héberger les logiciels et applications ayant une interface WEB qui permet d'activer des traitements et/ou de visualiser des données. La visualisation de données sous forme de cartes géographiques ou la bio-informatique sont, par exemple, utilisatrices d'interfaces Web.

ClusterWEB est composé de :

- 10 nœuds de 2 processeurs 2620v3, soit 120 cœurs de calcul et 640 Go de mémoire
- 100 To de disque cache

Grappe « DataREF » : Les données résidant sur cet espace disque sont sélectionnées par un Groupe de Travail scientifique dépendant du Comité de Pilotage. Il s'agit de données « de référence », c'est-à-dire d'intérêt commun et bien cataloguées. La mise à jour de ces données est réalisée via des processus déterminés par le groupe de travail ; les utilisateurs de DATARMOR n'y ont accès qu'en lecture.

Grappe « DataHOME » : Il s'agit d'un espace disque traditionnellement destiné aux fichiers sources et autres données précieuses des utilisateurs. « DataHOME » est constitué de 40 To de disques sauvegardés.

Grappe « DataWORK » : Cet espace disque est le lieu de stockage pour les données temporaires des utilisateurs. Il privilégie la taille et les performances plutôt que la sauvegarde, qui n'y est disponible que sous forme de « snapshots » sous responsabilité des utilisateurs.

Logistique électrique / refroidissement

Alimentation électrique : Elle est basée sur un ensemble d'équipements alimentés en double accès physique ; un onduleur garantit la qualité du courant envoyé vers DATARMOR et l'autonomie est assurée par un groupe électrogène.

Refroidissement : Le refroidissement du « clusterHPC » se fait par un circuit d'eau fermé, rafraîchi par la température de l'air extérieur ; son intégration étroite à l'architecture machine permet de fonctionner avec très peu d'énergie. Le refroidissement des autres parties de DATARMOR est réalisé par une ventilation directement intégrée aux racks. La technologie de DATARMOR intègre donc complètement l'informatique et son refroidissement.