

概要

- STScI主導で提供されるRoman公式のクラウド型解析環境(JupyterLab on AWS)
Roman用解析ソフトウェア・データ・計算資源を単一プラットフォームに統合
- ミッション運用期も同一環境を継続提供（位置付けとしてRoman解析の主戦場となる）
- RRNは全研究者にオープン（使用自体は採択/非採択を問わないが、計算資源が変わる）

主な機能

✓ Roman解析環境(公式パイプライン等)

校正・処理・解析・シミュレーション一式（ローカル導入不要、最新版で検証済）

✓ スケーラブルな計算資源

CPU/メモリを用途に応じて選択・拡張、データは永続保存

✓ 共同研究支援システム

共有チームスペース、共通ソフト環境、リアルタイム共同セッション、Tutorial、ドキュメント

**Romanの科学解析は基本的に「各自のローカル環境」ではなく、
公式クラウドであるRRN上で、共通ソフト・共通データ・共有計算資源を用いて行う思想
(それ以外は自前のHPCなど膨大なRomanデータを扱える環境が必須になる)**

- ① RRNは全研究者に開かれている（採択者限定ではない）
- ② ProposerはRRN上で必要な計算資源を見積り、credit binを申告する必要がある
- ③ 採択プログラムにはRRN creditsが付与される

Nexus Resource Bin	Credits Needed
Bin 1	Credits \leq 4,000
Bin 2	4,000 < credits \leq 7,000
Bin 3	7,000 < credits \leq 10,000
Bin 4	Credits > 10,000

10k credits超えの解析はRRNではやらない前提で計画しなければならない。

10kを超える場合は解析環境を他で準備する。
その場合、例えば、

- Bin 1-3: RRN中心 + 必要なら外部併用
- Bin 4: RRNは入口・検証、主計算はHPC活用などの計画をproposalに明記する必要がある

想定されている解析規模

Server Type Offered in RRN	Specific AWS Server	vCPU	Memory*	Credits spent on 24 hrs of usage	Credits required for 1000 hrs of compute**
Small	r5.large	2	16 GiB	3	63
Medium	r5.2xlarge	8	64 GiB	12	63
Large - CPU Optimized	c6i.8xlarge	32	64 GiB	33	43
Large - Memory Optimized	r5.4xlarge	16	128 GiB	24	63

* 1 GiB = 1.07374 GB

** Assumes that computations will be spread out through the virtual CPUs (vCPU)

creditsの主な内訳: compute/storage/egress

- ingest(RRNへの持ち込み)は無料
- egress(外部持ち出し)は 84 credits / TB

必要リソースの具体化が求められている

1. Compute

- 使うインスタンス種別 (Small/Medium/…)
- 実行回数と1回あたりの実行時間
- 並列度 (vCPU利用の前提)

2. Storage

- 解析の作業領域に置くデータ量 (TB)
- 保持期間、EFSかS3か(頻度)

3. Egress

- RRN外に持ち出すデータ(基本は最小化)
- 84 credits / TB

これら合計でBinを決定する。

Storage Type	Credits required to store 1 TB of data/month
EFS - Frequent access	279
EFS - Infrequent access*	15
S3	10

* Inactive data accessed only a few times each quarter

想定されている解析規模

Server Type Offered in RRN	Specific AWS Server	vCPU	Memory*	Credits spent on 24 hrs of usage	Credits required for 1000 hrs of compute**
Small	r5.large	2	16 GiB	3	63
Medium	r5.2xlarge	8	64 GiB	12	63
Large - CPU Optimized	c6i.8xlarge	32	64 GiB	33	43
Large - Memory Optimized	r5.4xlarge	16	128 GiB	24	63

* 1 GiB = 1.07374 GB

** Assumes that computations will be spread out through the virtual CPUs (vCPU)

Storage Type	Credits required to store 1 TB of data/month
EFS - Frequent access	279
EFS - Infrequent access*	15
S3	10

* Inactive data accessed only a few times each quarter

Credits Bin見積もり例

- ① Medium(8CPU)を合計10日間使う(インスタンス)
計算時間は1000CPU時間

$$\text{compute: } 12 \text{ credits}/24\text{hrs} \times 10 \\ + 63 \text{ credits}/1000\text{CPUhr} = \underline{\underline{183 \text{ credits}}}$$

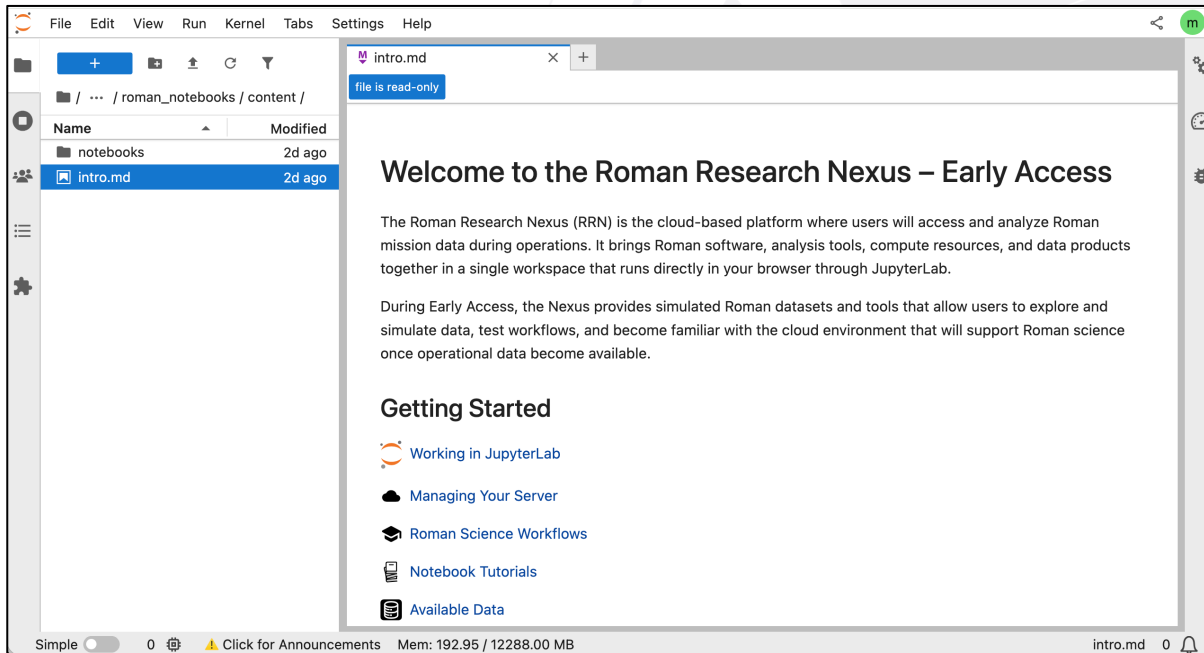
- ② 作業データ 5TB をS3に2ヶ月置く

$$\text{storage: } 10 \text{ credits}/\text{TB}/\text{month} \times 5\text{TB} \\ \times 2 \text{ months} = \underline{\underline{100 \text{ credits}}}$$

- ③ RRN外へ 2TB 持ち出す

$$\text{egress : } 84 \text{ credits}/\text{TB} \times 2 \text{ TB} = \underline{\underline{168 \text{ credits}}}$$

合計 : 100 + 168 + 183 = 451credits → Bin 1



STScIアカウントを作成すれば誰でも利用できる

Early Accessフェーズとして、

- WFIシミュレーションデータ(TVAC含む)
 - 公式解析ソフト、JupyterLab計算環境
 - チュートリアル(JupyterNotebook)
- が統合された形で提供されている。

Romanデータ解析の流れを実運用と同じ環境で事前に習得・検証できるプラットフォームとなっている。

The Nexus:
Documents:

<https://roman.science.stsci.edu/hub/>

<https://roman-docs.stsci.edu/data-handbook/roman-research-nexus>

- RomanのデータはMASTアーカイブで即時公開だが、それではダメか？
不可能ではないが、一般的には現実的ではない。RomanはTB～PB級データを想定しており、一部の機関を除けば、MASTダウンロード前提の解析は現実的でない。多くのユーザーにとってはRRN前提が合理的。
- RRNではどのような運用を想定すればいい？
RRNは「本番コードの実行環境」であり、試行錯誤の場ではない。
ジョブ起動=Creditsが即時不可逆消費される。設計・検証はRRN外(ローカル/国内HPCなど)で完結。
RRNでは再現可能なスクリプトを一回実行させる想定。
- RRNのCreditsは追加申請可能か？
基本的に申請・審査による割当になる(追加の枠組みが施行される可能性はある)。
 - ・個人アカウント：デフォルト30 credits、1 credit/dayで上限30まで補充 (Early Accessの方針)
 - ・チームアカウント：初期500 credits、追加は上限2000 creditsまで申請可能 (しかし、運用状況依存)
 - ・TAC承認プログラム：運用開始期にはチームが自動作成され、年上限などの枠内で運用される設計
計算量が多い場合は、外部計算資源との併用が前提になる。
- 個人アカウント同士でデータや解析環境を共有できる？
できない。個人アカウントのホーム領域は完全に分離、個人アカウント間では共有ストレージが使えない
チームに所属していれば、個人アカウントで起動してもチーム共有ディレクトリは見える
インフラに過度の負担をかけない等の条件が規約にあり、クレジット回避を意図した運用は停止される。