

PDU LIneA rascunho (27/01/2015)

Realizações marcantes/impactantes (2014)

Científicas

- Renovação da participação do *Brazilian Participation Group* (BPG) envolvendo membros da comunidade astronômica brasileira no projeto SDSS-IV.
- Publicação de 12 artigos em revistas internacionais.
- Participação em reuniões das colaborações internacionais DES, SDSS-IV, DESI e LSST.
- Organização de 30 webinars internacionais trazendo palestrantes de todo o mundo incluindo PIs de grandes projetos internacionais como DES, SDSS-IV, DESI, LSST, GAIA entre outros mantendo a comunidade brasileira informada das tendências e resultados internacionais.
- Indicação de mais dois membros do LIneA como builders do DES reconhecendo a contribuição dada por estes a infraestrutura do projeto e autorizando a participação em todos os artigos da colaboração, levando ao total a 7 builders brasileiros.
- Organização em colaboração com Universidade de Stanford/SLAC da reunião XLDB Workshop 2014 para tratar da exploração científica de grandes volumes de dados como esperado na era do LSST.

Tecnológicas

- Revisão por comissão internacional do portal científico do LIneA no Fermilab.
- Revisão por comissão internacional do portal científico do LIneA no National Center for Supercomputing Applications (NCSA).
- Preparação de acordo de colaboração técnica com a divisão responsável pelo gerenciamento de dados do LSST.
- Preparação de acordo de colaboração técnica no âmbito dos projetos DES e LSST com o NCSA

- Instalação do portal do LIneA no Fermilab para uso da colaboração DES com quase 1,500 acessos por parte de 117 usuários 33 brasileiros e 84 internacionais.
- Instalação do portal científico no NCSA para servir como interface de usuário ao banco de dados Oracle para mineração de dados do projeto DES.
- Início de operação do LIneA cluster com 1000 cores para atender as demandas dos pesquisadores envolvidos em projetos apoiados pelo laboratório.

Propostas

- Concedido apoio da FINEP ao projeto APLIneA.
- Concedido recursos da FAPERJ para a compra de equipamentos.
- Submissão de proposta de formação de um INCT seguindo o modelo de cooperação do LIneA para a preparação e inserção de jovens pesquisadores em grandes projetos de colaboração internacional como DES, SDSS-IV, DESI e LSST que forma uma sequencia de projetos de complexidade cada vez maior procurando determinar a natureza da energia escura, ao mesmo tempo explorando volumes cada vez maiores do universo, o que beneficia diferentes áreas da Astronomia.
- Participação em proposta a ser submetida ao NSF para financiamento do esforço de redução e validação de dados do DES para os próximos 5 anos.

Necessidades para o exercício de 2015

- Transferência e adaptação dos equipamentos do LIneA atualmente localizados no PoP-RJ para o recém renovado CPD do LNCC em Petrópolis
- Instalação do sistema DMZ para a transferência de dados do NCSA, LBL, e Fermilab para o LNCC.
- Compra de máquina para o espelho de dados do projeto DES de forma a melhorar o desempenho do mesmo e se preparar para a distribuição pública de todos os dados acumulados pelos 4 levantamentos do projeto SDSS-III e para a distribuição de dados programas para 2016 e anos subsequentes.
- Iniciar compra de sistema de armazenamento para substituir/expandir o sistema já existente de 500 TB fora da

garantia, para o armazenamento dos vários projetos sendo apoiados pelo LIneA do qual atualmente dependem da ordem de mais de 30 pesquisadores espalhados em várias instituições nacionais (ON, UFRJ, UFF, CEFET, USP, UNESP, CEFET, UFRGS, UNSM) e que no futuro será usado para a distribuição de dados para a comunidade. Esta previsto que dados de valor científico agregado ficarão disponíveis em 2017 e 2021. Apenas o DES exigirá da ordem de 5 PB de armazenamento até o seu final previsto em 2021. No mesmo período a demanda dos outros projetos devera aumentar consideravelmente em particular do LSST que seria usado para testar seus várias componentes utilizando simulações numéricas. Esta compra deve dar início a compra de um sistema de armazenamento composto de discos SAS em unidades de 1 Pb (2 x 480 TB) por ano a um custo de US\$ 520 k de maneira a atingir 5 PB em 2021, quando um novo dimensionamento deverá ser feito para atender as necessidades do LSST e do centro regional de acesso a dados associado.

- Iniciar a compra de um novo cluster de processamento utilizando conexões infiniband para melhor desempenho dos workflows científicos como já ficou demonstrado ao analisar o primeiro ano de dados do DES onde para maior eficiência é necessária a otimização da execução de mais de 3600 processos simultâneos. Utilizando recursos da FAPERJ pretende-se comprar inicialmente 4 nós (64 cores) com um investimento da ordem de US \$ 60k. Para mais tarde ser expandido para 960 cores exigindo um investimento de US\$ 490k em 2016 e 2019.

Orçamento Previsto

Em mil dólares

Equipamentos	2015	2016	2017	2018	2019
Espelho SDSS	60	60	60		
Cluster	65	490			490
Armazenamento	65	520	520	520	520
Estação de trabalho	6		6		6
Total em US\$	196	1070	586	520	1016
Total em R\$ (\$=2.7)	529	2889	1582	1404	2743

Em mil dólares

Participação Projetos	2015	2016	2017	2018	2019
SDSS	130	130	130	130	
DESI	80	80	80	80	80
Total em US\$	210	210	210	210	80
Total em R\$ (\$=2,7)	567	567	567	567	216

Em mil reais

Custeio	2015	2016	2017	2018	2019
Pessoal	3500	3700	3900	4100	4300
Passagens/diarias	50	50	50	50	50
Total	3550	3750	3950	4150	4350

Em mil reais

	2015*	2016	2017	2018	2019
Total	4630	7206	6083	6121	7293

Em 2015 o custo de pessoal está coberto pelo projeto FINEP. Portanto para este ano só são necessários R\$ 1130 k

Os investimentos em hardware a partir de 2019 devem estar alinhados com as necessidades para instalação de um Centro de Acesso a Dados do LSST em parceria com o LNCC, RNP e FAPERJ

Onde a Unidade precisa agir e soluções pendentes

- Assinatura do memorando de entendimento entre a RNP, LIneA, LNA e ANSP e o LSST para a entrada de uma equipe que será responsável pela a preparação de uma futura entrada de toda a comunidade brasileira sem custo.
- Firmar acordo com o SDSS-IV que envolve o pagamento de 4 parcelas anuais de US\$ 130 k, a operação adequada do espelho do SDSS para distribuição dos dados nacionalmente e como

alternativa ao site primário localizado na Universidade de Johns Hopkins e participação ainda a ser definida com o time responsável pelo gerenciamento de dados do projeto SDSS

- Firmar acordo com o projeto DESI com recursos do LIneA e possivelmente da FAPESP
- Assinar acordo de cooperação com o NCSA.
- Liberação da segunda parcela do projeto APLIneA da FINEP
- Identificar recursos e atividades junto ao projeto DESI que permita a participação de pelo menos 10 pesquisadores a um custo de apenas US\$ 200k por pesquisador sênior (contratado) mais 4 associados (pos-doc e estudantes) a ser pago em 4 ou 5 parcelas anuais. Este projeto é uma extensão do SDSS-IV da mesma forma que o LSST é em relação ao DES.
- Institucionalização do LIneA e criação de um orçamento próprio para garantir a continuidade de seu trabalho pelo menos ao longo do tempo de vida dos projetos DES e SDSS-IV (2021).

Desafios para o exercício de 2015

- A convivência do LIneA dentro da estrutura administrativa existente no ON e o modo de operação de diversos setores num período que se espera seja de transição
- A necessidade de um destaque orçamentário que garanta o bom funcionamento deste laboratório multiusuário com obrigações contratuais junto aos projetos internacionais DES e SDSS e agencias como DOE e NSF e usuários espalhados pelo mundo inteiro.

Unidades de Pesquisa/Organizações Sociais

O LIneA preenche uma enorme lacuna criada pela revolução na forma de trabalho científico causada pelo grande volume dados sendo gerados por experimentos modernos como o DES e SDSS. Esta tendência só aumentará com os futuros projetos como LSST e DESI, dos quais DES e SDSS são, respectivamente, precursores.

O trabalho do LIneA das seguintes vertentes:

- 1) A primeira vertente é científica e a atuação do LIneA serve para estimular a agregação de competência científica espalhada em diferentes instituições brasileiras para o melhor aproveitamento dos dados para diferentes áreas de conhecimento, a interação entre pesquisadores, a preparação de jovens pesquisadores para trabalhar e se inserir no ambiente típico de grandes colaborações, o fornecimento e manutenção de uma infraestrutura para o trabalho colaborativo, e a interface administrativa com as grandes colaborações internacionais geralmente representadas por organizações não governamentais criadas sem fins lucrativos. Em resumo, o LIneA seria o equivalente brasileiro a Fermi Research Alliance (FRA), Astrophysical Research Consortium (ARC), LSST Incorporation (LSSTC).
- 2) A segunda vertente é oferecer um ambiente computacional adequado as necessidades específicas do trabalho voltado para os casos de uso da Astronomia que inclui processamento altamente paralelizado, transporte de dados, mineração de grande volume de dados, sistemas eficientes e robustos de armazenamento, e a manutenção do acervo de dados a longo prazo.
- 3) A terceira vertente é a operação do centro de dados focado nas necessidades da área que inclui: a criação e manutenção de um repositório de códigos e de bibliotecas para o processamento científico, a criação de pacotes de software para distribuição e um serviço de helpdesk para apoio aos pesquisadores, monitoramento de diferentes sites (CTIO, Fermilab, NCSA, LNCC, P&D), monitoramento da transferência de dados entre os diferentes centros, monitoramento do serviço de distribuição de dados ao grande público, manutenção de serviço de educação e informação ao grande público
- 4) A quarta vertente é o desenvolvimento, manutenção e contínua integração de novos algoritmos ao portal científico acessado pela web. O portal científico desenvolvido pelo LIneA para o DES representa um investimento de mais 600 pessoas-meses ou da ordem de R\$ 25 a 30 milhões em valores de 2015 somente em pessoal. Para o desenvolvimento do portal emprega da ordem de 12 pessoas entre operadores, programadores, estudantes de informática, analistas de sistema e astrônomos formados, o que torna possível o desenvolvimento de um complexo sistema de gerenciamento, validação, análise e distribuição de dados voltado

para a Astronomia. Embora o foco atual seja o desenvolvimento do portal científico para atender as necessidades do DES e que deverá ser mantido em funcionamento até 2021 (vide acordo DES-Brazil/DES e LIneA/NCSA), a tecnologia desenvolvida pode ser muito útil para atender outros projetos como são os casos da Rede Brasileira de Ocultação e a reorganização dos dados acumulados por pesquisadores no SOAR como solicitado pelo LNA. Desde 2010 o portal foi objeto de avaliação por parte de uma comissão internacional formada por cientistas e técnicos que após 2 dias de apresentação produziram relatórios incluindo comentários e recomendações que nortearam o seu desenvolvimento. Instâncias do portal estão atualmente disponíveis nos seguintes locais:

- a. Cerro-Tololo Interamerican Observatory (CTIO) do National Optical Astronomy Observatory (NOAO) onde as imagens obtidas pela câmera de 500 pixels são avaliadas em tempo real.
- b. Fermilab onde o portal é usado para validação dos dados reduzidos pelo NCSA através da inspeção das imagens e dos catálogos produzidos, e para mineração dos dados.
- c. NCSA onde será usado como interface ao banco de dados da Oracle evitando desta forma a transferência de grandes volumes.
- d. LIneA onde são mantidos um ambiente de produção com 1000 cores e um sistema lustre de compartilhamento de arquivos destinado a execução de workflows científicos disponível para os participantes do DES-Brazil e BPG, além de um ambiente de P&D.

É a experiência de coordenar este amplo escopo de competências para atender a grande variedade de demandas que distingue o LIneA de outras iniciativas e sua atuação estratégica colaborando com os principais laboratórios nacionais nas áreas de física de altas-energias, astronomia e informática americanos (LBL, SLAC, Fermilab, NOAO, NCSA) na construção da infraestrutura do LSST e no desenvolvimento de ferramentas para o gerenciamento de dados.

A experiência do LIneA em bigdata, de interesse de diversas áreas de conhecimento, o reconhecimento internacional de seus produtos e seu envolvimento em projetos internacionais que se estenderão até pelo menos 2032 requerem a institucionalização deste laboratório. Em

particular, tendo em vista a natureza dos experimentos que exigem respostas rápidas, a volatilidade de pessoal típica da área de TI, e a diversidade de expertise envolvida a melhor solução para ser a de adotar o modelo de uma organização social regida por um termo de compromisso especificando entregáveis e sujeita a revisões periódicas dos stake-holders. Acreditamos que só assim o LIneA terá a agilidade necessária para manter qualidade de serviço exigida de um laboratório multiusuário atendendo a comunidade nacional e internacional. Uma OS oferece ainda os mecanismos para atender novos projetos que podem ter seu mérito científico julgado pelo Conselho Científico da OS e recursos aprovados pelo Conselho de Administração, além de um comitê de usuários para avaliar regularmente a qualidade do serviço prestado. Pode-se imaginar uma OS sozinha ou como parte de uma organização que forneça os serviços básicos de operação (e.g. Centro de e-ciência) mas que mantenha sua individualidade e destaque orçamentário e da qual seja um cliente, principalmente ao se imaginar um centro regional de acesso aos dados do LSST que terá sozinho mais demanda que atualmente previsto em termos de conectividade, armazenamento e processamento

Isto deve ser complementado por um acordo envolvendo outros parceiros do MCTI interessados ou que possam contribuir como LNA, LNCC, ON e RNP, criando assim a possibilidade de cessão de pessoal em tempo parcial ou integral e de mecanismos para transferência bidirecional de conhecimento que possa beneficiar outras aplicações ou mesmo outras áreas sem prejuízo do foco principal do LIneA.

Em resumo, atualmente nenhuma unidade de pesquisa do MCTI tem o foco, agilidade, infraestrutura e isoladamente a competência exigida para atender esta demanda. Além disso, não basta uma infraestrutura técnica que pode ser atendida em parte pelo LNCC, RNP ou um Centro de e-Ciência sendo discutido em alguns forums. É a habilidade de coordenar e adaptar tudo isto as necessidades de uma área específica com demandas já existentes e sua capacidade de desenvolvimento de software aplicado que o LIneA faz é que poderá servir de modelo para outras áreas de conhecimento.