



MINISTÉRIO DA CIÊNCIA, TECNOLOGIA, INOVAÇÕES E COMUNICAÇÕES
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

ARQUITETURA DE CUBO DE DADOS PARA REPRESENTAÇÃO DE TELEMETRIA DE SATÉLITES

Yuri Matheus Dias Pereira
Mauricio Vieira Ferreira Gonçalves
Rodrigo Rocha Silva

Relatório Técnico resultado do
Exame de Proposta de Disserta-
ção do Curso de Pós-Graduação
em Engenharia e Gerenciamento de
Sistemas Espaciais.

URL do documento original:
<<http://urlib.net/>>

INPE
São José dos Campos
2019

PUBLICADO POR:

Instituto Nacional de Pesquisas Espaciais - INPE

Gabinete do Diretor (GB)

Serviço de Informação e Documentação (SID)

Caixa Postal 515 - CEP 12.245-970

São José dos Campos - SP - Brasil

Tel.:(012) 3945-6923/6921

Fax: (012) 3945-6919

E-mail: pubtc@sid.inpe.br

**COMISSÃO DO CONSELHO DE EDITORAÇÃO E PRESERVAÇÃO
DA PRODUÇÃO INTELECTUAL DO INPE (DE/DIR-544):****Presidente:**

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

Membros:

Dr. Gerald Jean Francis Banon - Coordenação Observação da Terra (OBT)

Dr. Amauri Silva Montes - Coordenação Engenharia e Tecnologia Espaciais (ETE)

Dr. André de Castro Milone - Coordenação Ciências Espaciais e Atmosféricas
(CEA)

Dr. Joaquim José Barroso de Castro - Centro de Tecnologias Espaciais (CTE)

Dr. Manoel Alonso Gan - Centro de Previsão de Tempo e Estudos Climáticos
(CPT)

Dr^a Maria do Carmo de Andrade Nono - Conselho de Pós-Graduação

Dr. Plínio Carlos Alvalá - Centro de Ciência do Sistema Terrestre (CST)

BIBLIOTECA DIGITAL:

Dr. Gerald Jean Francis Banon - Coordenação de Observação da Terra (OBT)

Clayton Martins Pereira - Serviço de Informação e Documentação (SID)

REVISÃO E NORMALIZAÇÃO DOCUMENTÁRIA:

Simone Angélica Del Ducca Barbedo - Serviço de Informação e Documentação
(SID)

Yolanda Ribeiro da Silva Souza - Serviço de Informação e Documentação (SID)

EDITORAÇÃO ELETRÔNICA:

Marcelo de Castro Pazos - Serviço de Informação e Documentação (SID)

André Luis Dias Fernandes - Serviço de Informação e Documentação (SID)



MINISTÉRIO DA CIÊNCIA, TECNOLOGIA, INOVAÇÕES E COMUNICAÇÕES
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

ARQUITETURA DE CUBO DE DADOS PARA REPRESENTAÇÃO DE TELEMETRIA DE SATÉLITES

Yuri Matheus Dias Pereira
Mauricio Vieira Ferreira Gonçalves
Rodrigo Rocha Silva

Relatório Técnico resultado do
Exame de Proposta de Disserta-
ção do Curso de Pós-Graduação
em Engenharia e Gerenciamento de
Sistemas Espaciais.

URL do documento original:
<<http://urlib.net/>>

INPE
São José dos Campos
2019



Esta obra foi licenciada sob uma [Licença Creative Commons Atribuição-NãoComercial 3.0 Não Adaptada](#).

This work is licensed under a [Creative Commons Attribution-NonCommercial 3.0 Unported License](#).

Informar aqui sobre marca registrada (a modificação desta linha deve ser feita no arquivo publicacao.tex).

“But I try not to think with my gut. If I’m serious about understanding the world, thinking with anything besides my brain, as tempting as that might be, is likely to get me into trouble. It’s OK to reserve judgment until the evidence is in.”

CARL SAGAN E ANN DRUYAN
em “O Mundo Assombrado pelos Demônios:
A Ciência Vista Como Uma Vela no Escuro”, 1995

RESUMO

Satélites são monitorados pelas equipes de solo via pacotes de telemetria, que informam o estado atual dos equipamentos e permitem avaliar a capacidade do satélite de continuar a sua missão. Esses pacotes de telemetria constituem um corpo de dados de tamanho e complexidade significativa, sendo que satélites que funcionam por vários anos geram dados históricos de grande volume, ainda úteis para a operação. Neste artigo apresentamos uma arquitetura baseada em conceitos de Big Data e Business Intelligence para criar uma representação de dados de telemetria pronta para a análise por operadores e engenheiros de satélite no Instituto Nacional de Pesquisas Espaciais (INPE), bem como apresentamos o fluxo de dados utilizado pelos dados históricos de telemetria de um dos satélites operados pelo INPE.

Palavras-chave: Cubo de Dados. Big Data. Operação. Satélite. Data Warehouse.

LISTA DE FIGURAS

	<u>Pág.</u>
4.1 Fluxo de dados em uma arquitetura de Big Data	10
4.2 Arquitetura de um cubo de dados	12
5.1 Aparência do SCD-Dashboard	16
5.2 Carregando dados de telemetria no RFragCubing	17
5.3 Curva de de agregação gerada pela SCD-Dashboard	17
5.4 Resultados preliminares da Medição de Relacionamento	18

LISTA DE TABELAS

	<u>Pág.</u>
3.1 Operadores e Arquiteturas de Big Data	6
4.1 Dados de Operação	9
A.1 Cronograma de atividades	27
A.2 Publicações planejadas	27

LISTA DE ABREVIATURAS E SIGLAS

WETAMC	–	Campanha de Mesoescala Atmosférica na Estação Úmida
IBGE	–	Instituto Brasileiro de Geografia e Estatística
MC	–	Método das Covariâncias
EDO	–	Equações Diferenciais Ordinárias
EDP	–	Equações Diferenciais Parciais
ECT	–	Energia Cinética Turbulenta
FDP	–	Função de Distribuição de Probabilidade
PR	–	Plot de Recorrência
FFT	–	Fast Fourier Transform
tS1200	–	Temperatura medida no nível superior às 12 horas
tS2300	–	Temperatura medida no nível superior às 23 horas
tM1200	–	Temperatura medida no nível médio às 12 horas
tM2300	–	Temperatura medida no nível médio às 23 horas
tI1200	–	Temperatura medida no nível inferior às 12 horas
tI2300	–	Temperatura medida no nível inferior às 23 horas
wS1200	–	Velocidade vertical do vento medida no nível superior às 12 horas

SUMÁRIO

	<u>Pág.</u>
1 INTRODUÇÃO	1
2 FUNDAMENTAÇÃO	3
2.1 Operação	3
2.2 Big Data	3
2.3 Cubo de Dados	3
3 TRABALHOS CORRELATOS	5
3.1 Computação do Cubo de Dados	5
3.1.1 FragCubing	5
3.1.2 HIC	5
3.1.3 QCubing	5
3.1.4 HFrag	5
3.2 Outros Operadores	5
3.3 No INPE	7
4 PROPOSTA	9
4.1 Dados	9
4.2 Fluxo dos Dados de Operação	10
4.3 Arquitetura de um Cubo de Dados	11
4.3.1 Algoritmos de construção do cubo	13
4.4 Discussão	13
5 IMPLEMENTAÇÃO E RESULTADOS	15
5.1 SCD-Dashboard	15
5.2 RFragCubing	16
5.3 Medida de Similaridade	17
6 CONCLUSÃO E TRABALHOS FUTUROS	21
6.1 Planejamento	21
6.1.1 Trabalhos futuros	22
REFERÊNCIAS BIBLIOGRÁFICAS	23
ANEXO A - CRONOGRAMA E PUBLICAÇÕES	27

1 INTRODUÇÃO

[Resto da introdução]

Os capítulos restantes desta dissertação estão organizados da seguinte maneira:

- Capítulo 2: Este capítulo apresenta os conceitos e fundamentos teóricos, como apresentando os conceitos do Cubo de Dados, as definições utilizadas de *Big Data*, a definição do problema para a operação.
- Capítulo 3: Neste capítulo os trabalhos correlatos de Cubo de Dados são apresentados, bem como outros operadores de satélite estão resolvendo os problemas identificados.
- Capítulo 4: Neste capítulo a arquitetura proposta é apresentada e seus conceitos principais explicados, bem como o fluxo de dados atual do CCS e como a nova arquitetura vai melhorá-lo.
- Capítulo 5: Esse capítulo apresenta alguns resultados já alcançados, demonstrando os softwares que já foram escritos e as análises que já foram executadas.
- Capítulo 6: Com base na arquitetura proposta e nos resultados intermediários alcançados, esse capítulo apresentará as conclusões obtidas bem como as direções de implementação para o resto do trabalho.

2 FUNDAMENTAÇÃO

Neste capítulo, vamos apresentar os conceitos relevantes para este documento, bem como apresentação uma revisão da literatura na área de *Big Data*, terminando com uma visão geral de como o conceito e as tecnologias recentes estão sendo utilizadas por variadas agências espaciais.

2.1 Operação

[Dataflow? Explicar a operação de satélites bem por cima]

2.2 Big Data

[5 Vs]

[Data Science?]

2.3 Cubo de Dados

[Copiar a estrutura da tese do Rodrigo?]

3 TRABALHOS CORRELATOS

Nessa seção, vamos apresentar os trabalhos correlatos a essa proposta. Eles podem ser divididos em duas seções: uma focada nos algoritmos de construção do cubo de dados e outra em como outros operadores estão realizando tarefas semelhantes.

3.1 Computação do Cubo de Dados

O trabalho de (SILVA, 2015) mostra algumas das abordagens de construção do cubo de dados que são relevantes, bem como apresenta uma comparação de resultados de algumas das que vamos apresentar aqui.

Acho que isso pode ser feito igual na tese do rodrigo, ao invés dessas seções específicas. Não é para detalhar cada algoritmo, de qualquer forma...

3.1.1 FragCubing

3.1.2 HIC

3.1.3 QCubing

3.1.4 HFrag

3.2 Outros Operadores

A tabela 3.2 mostra uma revisão feita em artigos recentes sobre os operadores de satélite e quais tecnologias eles estão utilizando para atingir objetivos semelhantes, principalmente com o uso de *Big Data*, como demonstrado pelos artigos publicados.

Os objetivos em comum desses trabalhos são facilitar as atividades dos operadores por meio de algoritmos de detecção de anomalias e de verificação dos limites nos valores das telemetrias. Alguns dos operadores dessa lista estão responsáveis pela operação de constelações de satélites complexos, como constelações de sensoriamento remoto, que faz necessário um certo nível de automação ou a operação contínua teria um custo inviável.

É bom frisar que o uso dessas tecnologias é apenas para os operadores, pois em nenhum desses trabalhos eles estão na mesma estrutura de ingestão dos dados da carga útil, mesmo utilizando as mesmas tecnologias, como demonstrado em (MATEIK et al., 2017) e (ADAMSKI, 2016).

Alguns desses trabalhos não utilizam de estruturas completas que seguem um fluxo

Tabela 3.1 - Operadores e Arquiteturas de Big Data

Referência	Operador	Ferramenta	Tecnologias
(ADAMSKI, 2016)	L3 (EUA)	InControl	Hadoop, Spark, HBase, MongoDB, Cassandra, Amazon AWS
(BOUSSOUF et al., 2018)	Airbus	Dynaworks	Hadoop, Spark, HDFS, HBase, PARQUET, HIVE
(SCHULSTER et al., 2018)	EUMETSAT	CHART	MATLAB, MySQL, Oracle
(ZHANG et al., 2017)	SISSET (China)	-	Hadoop, HDFS, PostgreSQL, MongoDB, Logstash, Kibana, ElasticSearch, Kafka, MapReduce
(YVERNES, 2018)	Telespazio France	PDGS	OLAP (DataCube), Saiku, Pentaho, Jaspersoft OLAP
(DISCHNER et al., 2016)	SwRI + NOAA	CYGNSS MOC	SFTP, -
(EDWARDS, 2018)	EUMETSAT	MASIF	FTP, RESTful service, JMS Message Queue, PostgreSQL
(EVANS et al., 2016)	S.A.T.E + ESA/ESOC	-	Java, CSV
(FEN et al., 2016)	CSMT (China)	-	não menciona as tecnologias
(TROLLOPE et al., 2018)	EUMETSAT	CHART	algoritmos ad-hoc, estudo de caso
(GILLES, 2016)	L-3	InControl	Amazon EC2, LXC, Nagios
(HENNION, 2018)	Thales Alenia	AGYR	Logstash, Kafka, InfluxDB, ElasticSearch, Kibana, Grafana
(MATEIK et al., 2017)	Stinger, NASA	-	Logstash, ElasticSearch, Kibana, HDF5, CSV, R, Python, AWS, Excel
(FERNÁNDEZ et al., 2017)	NASA	MARTE	R, CSV, ad-hoc

Fonte: Produção do autor

de dados, como é o caso de (FERNÁNDEZ et al., 2017) e (TROLLOPE et al., 2018) que utilizam de scripts feitos conforme foram necessários, não mostrando uma visão da arquitetura completa do fluxo de dados e apenas na ferramenta utilizada para análise pontual, ao contrário de (YVERNES, 2018).

O trabalho de (YVERNES, 2018) é mais interessante que utilizam de estratégias OLAP e do Cubo de Dados, tendo utilizado uma modelagem dimensional para a operação de uma constelação, porém esse trabalho menciona apenas em alto nível a modelagem utilizada, e menciona que o trabalho foi somente na parte da modelagem dimensional e integração dos dados.

3.3 No INPE

O INPE já realiza análise de dados em outros setores, inclusive sobre as telemetrias de satélite. Os operadores devem monitorar os valores das telemetrias e informar a engenharia caso algum problema que não pôde ser corrigido aparece (TOMINAGA et al., 2017). Um exemplo está no trabalho (AES, 2012) feito sobre uma falha no satélite CBERS-2, onde o modelo proposto visa melhorar o conhecimento sobre avalanche térmica nas baterias para impedir que isso aconteça novamente em outros satélites. A motivação principal dos trabalhos da tabela 3.2 era a detecção de anomalias, que teve alguns algoritmos estudados em (AZEVEDO et al., 2011).

Para os outros setores, isso comumente se dá na análise de dados vindos da carga útil do satélite ou de agentes externos ao INPE, como dados de sensoriamento remoto, cuja análise não é trivial e estão classificados como Big Data. (MONTEIRO, 2017) utilizam de conceitos de Big Data para análise de trajetórias de objetos; (RAMOS et al., 2016) demonstram o uso de softwares como o Hadoop para a análise de dados do clima espacial, com uma arquitetura relacionada as arquiteturas revisadas na seção anterior; e (OES et al., 2018) mostra uma arquitetura que utiliza de Cubo de Dados para a análise de séries temporais.

4 PROPOSTA

Nesta seção apresentamos os dados que serão utilizados e são relevantes para a operação, bem como apresentamos o fluxo de dados utilizado nos trabalhos correlatos e a arquitetura proposta, por fim explicando o cubo de dados que será implementado.

4.1 Dados

A tabela 4.1 mostra os tipos de dados relevantes para a operação, a sua origem e o seu formato esperado, ignorando os dados provenientes da carga útil.

Tabela 4.1 - Dados de Operação

Tipo de Dado	Origem	Formato
Sensores de bordo	Equipamentos no satélite	Tabelas, CSV
Registros do Computador	Computador de Bordo	Texto (<i>Logs</i>)
Multimídia	Câmeras	MP4, JPG, RAW
Parâmetros orbitais	Operação, Rastreo	TLE, texto, tabelas
Documentação associada	Operadores, engenharia	Texto (Word, Excel)
Clima Espacial	Sensores no solo ou espaço	Texto, tabelas, avisos
<i>Situational Awareness</i>	Radares, US-STRACOM, etc	Texto, tabelas, avisos

Fonte: Adaptado de (ZHANG et al., 2017)

Para este trabalho, apenas os dados vindos de sensores de bordo serão considerados. Os outros dados nesta tabela poderiam ser considerados para uma *Data Warehouse* mais completa, pois um cubo de dados pode ser formado sobre quaisquer um desses dados.

Por exemplo, um cubo de dados textual poderia ser feito sobre os documentos associados a operação, como o CONOPS, tabelas de telecomandos e documentação de engenharia de sistemas para facilitar a análise da documentação sendo gerada pelo satélite. Um cubo multimídia poderia ser gerado sobre os dados multimídia tirados pelas câmeras do satélite para correlacionar com os dados gerados pelos sensores, e assim em diante. Alguns exemplos de cubos possíveis de serem feitos estão em (SILVA, 2015).

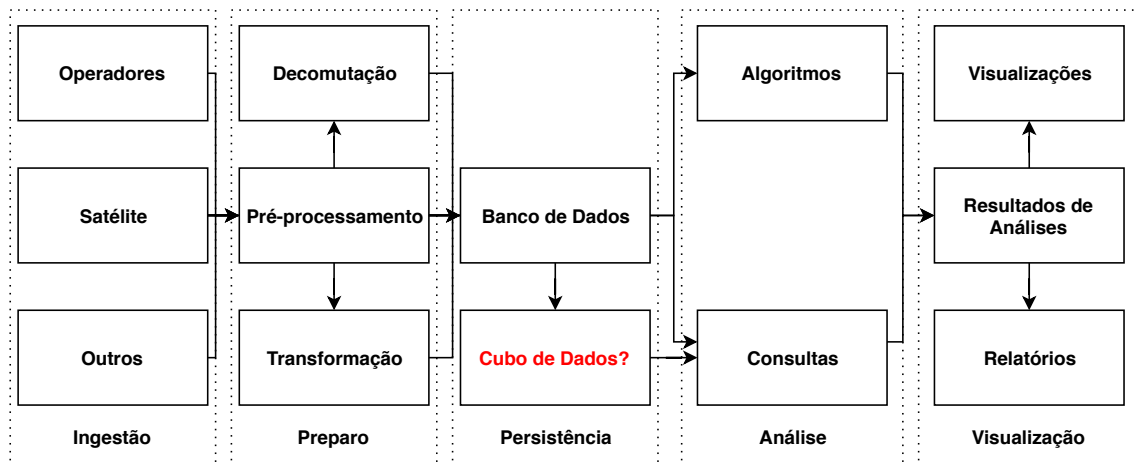
Esta lista não é exaustiva, e pode incluir dados da carga útil caso sejam relevantes para a análise em questão, como ajudar na georeferênciação de imagens tiradas pelo

satélite, bem como outros dados que os operadores acharem necessários e tiverem acesso a eles. Isso vai depender diretamente do fluxo de dados que é adotado, pois esses dados precisam ser coletados e preparados para serem utilizados.

4.2 Fluxo dos Dados de Operação

Baseado nos trabalhos correlatos e nos dados levantados na seção 4.1, a figura 4.1 demonstra o fluxo de dados esperado de uma arquitetura de *Big Data* para a operação de satélites.

Figura 4.1 - Fluxo de dados em uma arquitetura de Big Data



Fonte: Adaptado de (ZHANG et al., 2017)

Este fluxo está separado em cinco etapas que vão desde a origem dos dados até o seu resultado de análise, e este trabalho visa apenas mapear qual seria esse fluxo baseado nos trabalhos correlatos. Cada uma das etapas está detalhada a seguir:

- **Ingestão:** onde os dados serão coletados na sua fonte (satélites, sensores no solo, outras fontes, etc). Essa etapa se importa em **onde** estão os dados e **como** coletá-los, bem como **quais** são os dados importantes de serem coletados. A "fonte" aqui pode ser um serviço de terceiros, dentro da própria instituição ou disponível de outra forma.
- **Preparo:** os dados relevantes são selecionados, e transformações são reali-

zadas para inserir os mesmos na base de dados. Essa etapa se importa no formato específico dos dados, realizando operações de limpeza, verificação da qualidade e da relevância para a análise, entre outras. O seu objetivo é garantir que os dados tem qualidade, relevância, e estão no formato adequado para a base de dados.

- **Persistência:** após o devido processamento, os dados de alta qualidade são guardados em uma base de dados, de onde ficarão disponíveis para a análise. Nessa etapa um banco de dados é utilizado, se importando apenas em como esses dados estão guardados e como eles serão disponibilizados para as consultas e execução de algoritmos.
- **Análise:** nesta etapa são executadas as consultas e os algoritmos de interesse para a análise. Podem ser desde consultas simples ("qual era o valor da telemetria X durante a passagem Y?"), a execução de algoritmos complexos ("preveja os valores da telemetria X para a próxima passagem").
- **Visualização:** os resultados das consultas e algoritmos são visualizados. Podem conter desde gráficos simples, como um histograma de uma telemetria, a relatórios complexos de um subsistema/satélite, bem como resultados de algoritmos.

Esse fluxo é geralmente seguido pelos trabalhos correlatos na seção 3.2, sendo que os trabalhos de (ZHANG et al., 2017), (MATEIK et al., 2017) e (BOUSSOUF et al., 2018) definem esse processo mais claramente.

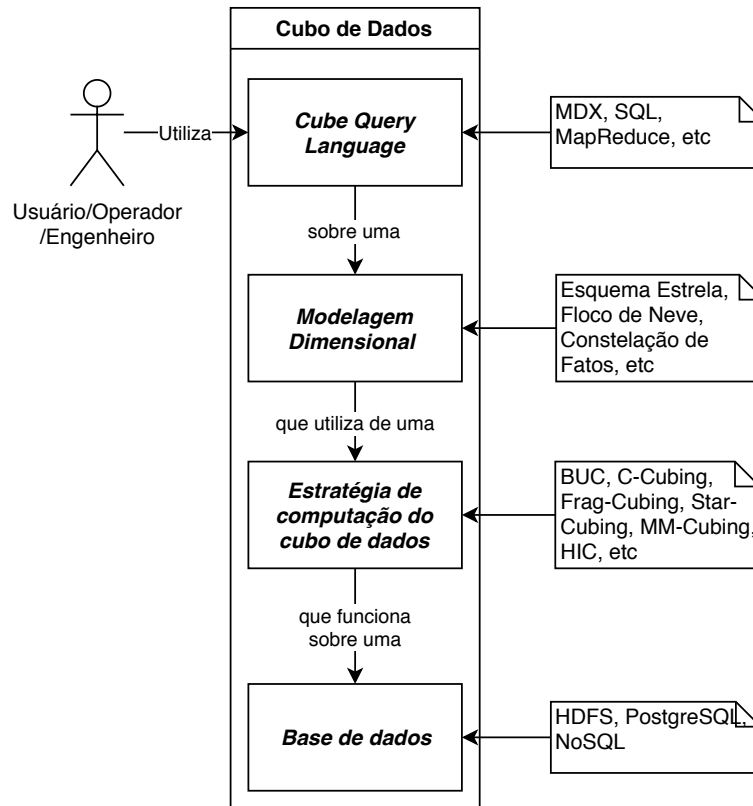
4.3 Arquitetura de um Cubo de Dados

A figura 4.2 demonstra a divisão em 4 camadas de uma estrutura de Cubo de Dados. Essas camadas demonstram tudo o que é necessário para a implementação de um Cubo de Dados, não sendo necessário que uma camada esteja fortemente atrelada a outra.

Para esta proposta, vamos nos concentrar apenas na proposição de um algoritmo de computação do cubo de dados mais apropriado, utilizando das outras seções quando elas vão se tornando necessárias. Os detalhes, algoritmos e conceitos listados na figura estão majoritariamente descritos na seção 2.3.

Uma informação interessante é que esta estrutura mostra o uso de pelo menos duas linguagens de computação sobre os dados: uma é a Cube Query Language que será

Figura 4.2 - Arquitetura de um cubo de dados



Fonte: Autor (2019)

utilizada pelo usuário para realizar as operações sobre o cubo (Drill-Down, Roll-up, etc), e outra é a linguagem que será utilizada pelo cubo para realizar essas operações, e elas podem ser independentes, por exemplo, pode-se utilizar SQL estendida com vocabulário de OLAP, porém o algoritmo de cubo de dados internamente pode consultar uma estrutura feita com MapReduce para o cálculo das medidas e das agregações.

Porém, utilizar duas linguagens muito diferentes nesse ponto pode não ser uma boa ideia, pois adicionaria um nível de diferença entre o usuário e os dados. Caso seja necessário realizar uma consulta OLTP normal, sem o uso do cubo de dados, por exemplo, essa diferença ficaria mais óbvia, por exemplo traduzir uma consulta de SQL para MapReduce não seria muito fácil simplesmente por ter que entender de ambas as linguagens bem para conseguir fazer isso. Deste modo, é interessante manter a mesma linguagem ao longo da estrutura, apenas alterando nas operações

relevantes para o cubo de dados.

Com isso se torna necessário ressaltar o último nível, a Base de Dados: a escolha de banco de dados vai impactar como o algoritmo funciona, visto que existem diferentes sistemas de arquivos e como eles são atingidos, bem como o estilo do banco vai mudar como o algoritmo deve gerar o cubo, pois a base pode utilizar diferentes paradigmas de banco de dados (CUZZOCREA et al., 2013).

4.3.1 Algoritmos de construção do cubo

Uma das necessidades de usar algoritmos diferentes de cubo de dados está no número de dimensões que um certo cubo consegue realizar pesquisas: consultas com mais que 15 dimensões não são comumente(ou praticamente) executadas em alguns algoritmos, como o trabalho de (SILVA, 2015) demonstra.

Como estabelecido na seção 4.1, os dados de telemetria de interesse possuem muito mais do que o limite de consultas em até 60 dimensões: com mais de 130 telemetrias para os satélites da família da SCD, e milhares para satélites maiores, a execução de consultas complexas seria normalmente inviável nos algoritmos de construção do cubo. Esse problema é geralmente resolvido pela modelagem dimensional, como em (AZEVEDO; AMBRÓSIO, 2010), porém isso transforma os dados de um formato "largo" para um formato "longo", aumentando o número de tuplas.

Deste modo é necessário investigar as abordagens de construção de cubo que funcionem com muitas dimensões, e que permitam o cálculo das medidas necessárias para a operação. Neste trabalho, iremos investigar algumas dessas abordagens e como elas se comparam para a análise.

4.4 Discussão

TODO: Passos do que seria feito? Estou achando melhor deixar isso no final, nas conclusões...

5 IMPLEMENTAÇÃO E RESULTADOS

Para a implementação da arquitetura, é necessário conhecimento sobre o domínio dos dados e como eles estão organizados. Para esse estágio da pesquisa, foram utilizados dados do SCD-2 fornecidos pelo CCS, porém alguns softwares foram implementados como parte da pesquisa sobre o cubo de dados e como resultados da análise dos dados.

5.1 SCD-Dashboard

Os dados vindos do SCD2 possuem 135 dimensões e estão distribuídos em um período de 4 anos, tornando a sua análise não trivial. Durante atividades de *Data Science* era necessário criar muitos gráficos e visualizações com períodos e dimensões variadas, e com tantas dimensões isso levava vários pedaços de código copiados para criar relatórios sobre os dados e dimensões.

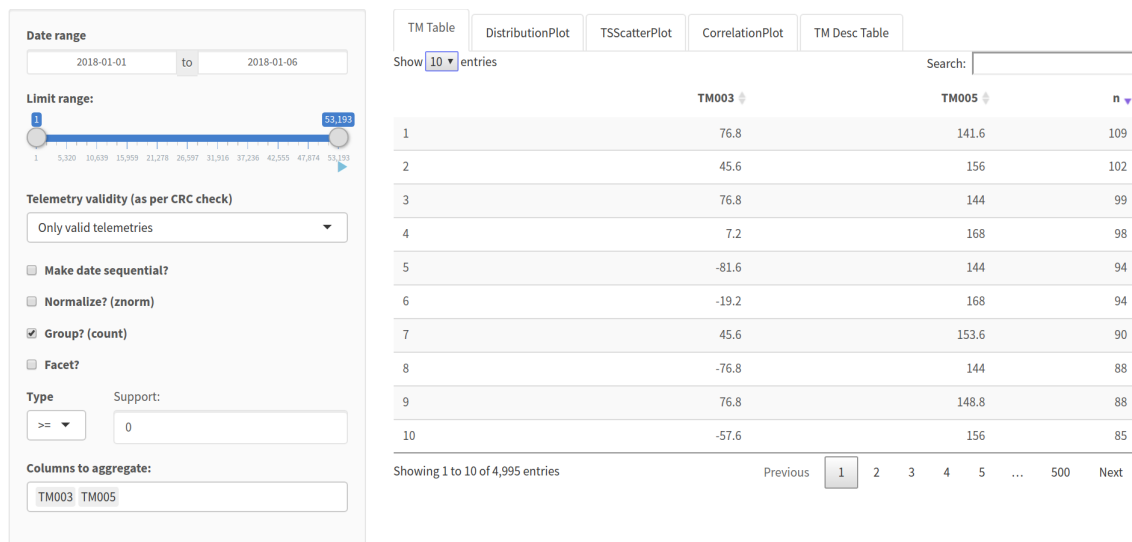
Para facilitar essa análise e automatizar esse problema, foi criado um software chamado **SCD-dashboard** para facilitar na visualização dos dados de telemetria. Ele foi implementado utilizando o pacote Shiny ([CHANG et al., 2019](#)) da linguagem R ([R Core Team, 2018](#)), que já estava sendo utilizada para criar as visualizações, a interface permite apenas criar as visualizações e aplicar os outros algoritmos relacionados de uma interface amigável e sem ter a necessidade de escrever código.

Ele funciona sobre um banco de dados PostgreSQL que possui o histórico das telemetrias já importadas e propriamente transformadas por outro script feito só para os dados do SCD2, porém devem funcionar para qualquer dado exportado via CSV pelo SatCS.

Essa ferramenta também pode ser vista como um piloto para implementar algo semelhante as *dashboards* criadas por outras agências, sendo que tem um precedente na ferramenta MARTE utilizada pela NASA ([FERNÁNDEZ et al., 2017](#)), que utiliza das mesmas tecnologias e conceitos, porém focada no algoritmo de detecção de anomalias. Ferramentas mais parecidas, e mais completas, estão no CHART e nas dashboard criadas usando o Kibana em ([MATEIK et al., 2017](#)) e ([ZHANG et al., 2017](#)).

Essa ferramenta foi utilizada extensamente para visualização dos dados e para análise exploratória, sendo que foi melhorada durante as disciplinas de *Data Science* e Algoritmos de *Data Mining*, com apresentações da mesma feitas utilizando ela.

Figura 5.1 - Aparência do SCD-Dashboard



Fonte: Produção dos autores

5.2 RFRagCubing

O algoritmo FragCubing, apresentado na seção 3.1.1, foi disponibilizado em forma compilada via código de C++, porém, ele foi feito com uma interface complicada de ser utilizada e automatizada, e não permitia a importação dos dados de telemetria sem um trabalho de pré-processamento não trivial antes. Como essa implementação foi a utilizada no trabalho de (SILVA, 2015), os resultados possuem histórico para comparação, portanto o seu uso contínuo seria interessante.

Para isso, foi criado um pacote na linguagem R chamado de **RFRagCubing**, que permite a integração com o algoritmo implementado em C++. Esse pacote faz a interface com o código do FragCubing, permitindo importar os dados, executar as queries e retornar os resultados das mesmas, bem como algumas adições de medição de memória e tempo de processamento. Uma das vantagens do pacote está na distribuição, com uma interface que permite a execução da mesma consulta em outros algoritmos de construção do cubo, uma das razões para fazer a implementação, para facilitar a comparação entre os algoritmos.

Figura 5.2 - Carregando dados de telemetria no RFragCubing

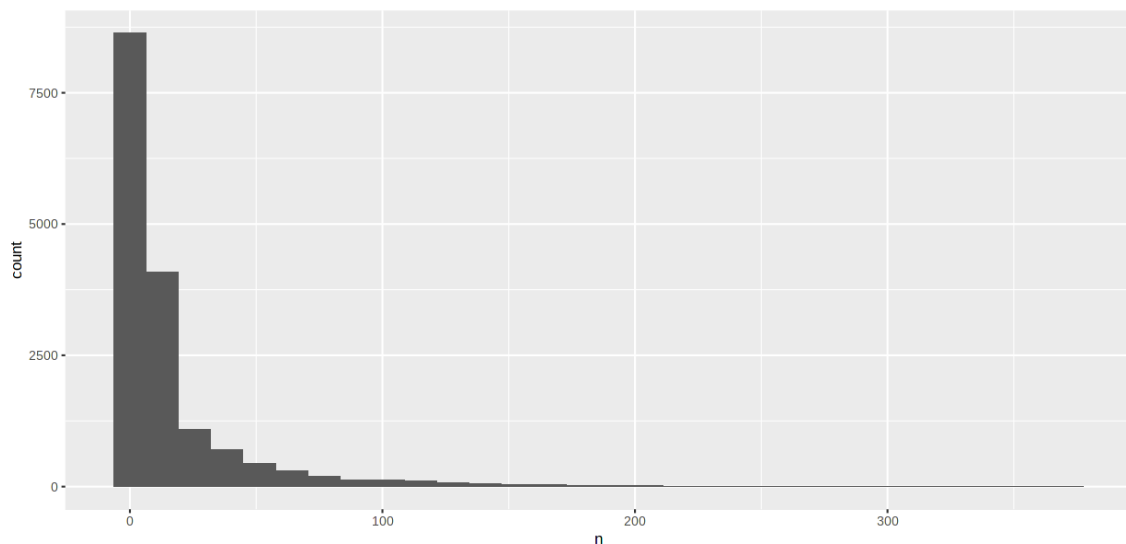
```
Initializing ... used time: 26 ms.  
Reading input file ... used time: 1654 ms.  
Computing shell fragments ... used time: 1681 ms.  
  
135 dimensional data loaded.  
296909 tuples read.  
135 shell fragments of size 1 constructed.
```

Fonte: Produção dos autores

5.3 Medida de Similaridade

Utilizando a ferramenta criada em 5.1 para uma atividade de exploração de dados que envolvia realizar consultas multidimensionais baseadas no conceito do cubo de dados, um padrão foi notado entre as telemetrias: quando uma medida de agregação por contagem era executada sobre telemetrias que se sabia ter algum tipo de relacionamento entre elas, elas possuíam uma curva característica, como da figura 5.3.

Figura 5.3 - Curva de de agregação gerada pela SCD-Dashboard

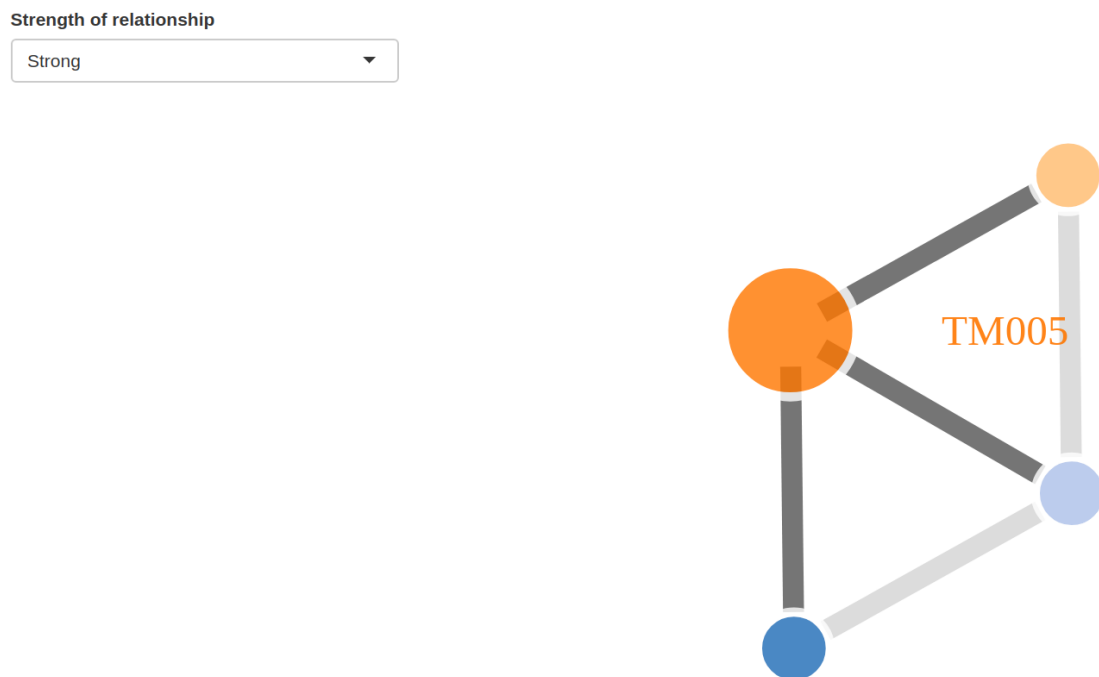


Agregação das telemetrias TM003 e TM004 ao longo de 01/2018

Fonte: Produção dos autores

Essa curva foi utilizada para desenvolver um algoritmo de classificação do relacionamento entre as telemetrias, pois possuía algumas vantagens: a medida de contagem dos valores que se repetem é independente do tipo dos valores em si, permitindo comparar um valor contínuo com um discreto bem como discretos com discretos e contínuos com contínuos, de equipamentos diferentes e com características diferentes; e é uma medida que funciona com qualquer número de dimensões, sendo uma operação $\Theta(D)$, com D sendo o número de dimensões, tornando a sua integração, e subsequente otimização, com um algoritmo de construção de cubo de dados simplificada.

Figura 5.4 - Resultados preliminares da Medição de Relacionamento



Um grupo de telemetrias com relacionamento forte

Fonte: Produção dos autores

O algoritmo está sujeito a Maldição de Dimensionalidade, pois ele verifica a relação entre um número n de telemetrias definida pelo usuário, tendo 2^D possíveis combinações para D dimensões, assim ele atualmente só foi executado para todas as combinações de 2 dimensões dos dados do SCD. Mesmo assim, e por ser uma operação demorada mesmo com apenas um ano de telemetrias, ainda são 8911 com-

binações. Esse número só cresce com o número de combinações, sendo que a execução de todas as combinações com mais de 4 dimensões é inviável.

O algoritmo está sendo revisado atualmente, mas o objetivo é implementar o resultado no algoritmo de construção do cubo, sendo que as dimensões que tem um relacionamento associado como forte ou média seriam as combinações com maior interesse pelos operadores, pois provavelmente seriam as operações mais comumente executadas por eles.

Isto também abre espaço para um algoritmo de detecção de anomalias: caso durante uma passagem um relacionamento que antes era tido como forte muda para um relacionamento fraco ou perde o relacionamento, isso pode ser um sinal de que alguma coisa está de errado com esse grupo de telemetrias. Isso pode ajudar no conhecimento dos operadores, pois os relacionamentos entre grupos de telemetrias são difíceis de serem visualizados e apenas os operadores com mais experiência em um dado satélite conseguem visualizar alguns desses relacionamentos. *Citação?*

Existem alguns trabalhos de visualização dos relacionamentos, porém estes geralmente utilizam algoritmos de indução de regras, cuja saída é de difícil interpretação, como demonstrado em (KANNAN; DEVI, 2016). A abordagem aqui proposta tem potencial de ter respostas mais relevantes para os operadores, porém precisa ser avaliada por um operador ainda, um dos próximos passos deste trabalho.

6 CONCLUSÃO E TRABALHOS FUTUROS

Este trabalho apresenta uma abordagem de cubo de dados para executar operações de análise nos dados de telemetrias de satélites. Essa abordagem utiliza de conceitos de *Big Data* para orientar a execução de consultas em dados com muitas dimensões e de alta complexidade. Uma revisão da literatura de arquiteturas de *Big Data* é apresentada, demonstrando que tipos de tecnologias e abordagens estão em uso por outros operadores de satélite, bem como uma revisão sobre os conceitos e abordagens de cubo de dados.

São apresentados resultados intermediários de análises dos dados de telemetria, bem como os produtos de software utilizados para realizar as análises e algumas descobertas desse processo. Esses resultados mostram que a aplicação do cubo é interessante para os dados disponíveis, bem como que é possível implementar algumas etapas do fluxo de dados, propriamente adequadas para lidar com *Big Data*.

Como essa arquitetura é melhor(diferente?) da utilizada por outros operadores? O que o Cubo de Dados traz de diferente?

6.1 Planejamento

Para o trabalho da dissertação, os passos seguintes são:

- a) Formalizar quais são as consultas relevantes para os operadores de satélite, e quais são as atividades de análise que podem ser expressas como consultas;
- b) Criar uma representação dimensional do cubo de dados apropriada para as consultas identificadas, mapeando as medidas que são necessárias e quais os seus tipos;
- c) Implementar a representação com as medidas em vários algoritmos da literatura recente, mais notadamente os revisados no capítulo 3.1 e coletar os resultados da execução das consultas relevantes para os operadores;
- d) Avaliar os resultados da implementação dos algoritmos e mostrar qual das abordagens é mais apropriada para o cenário da operação.

O passo *c* está parcialmente implementado no pacote mostrado na seção 5.2, porém precisa de trabalho significativo de implementação para executar outros algoritmos e realizar os testes das consultas relevantes.

Como resultado esperados da dissertação, teríamos o mapeamento das consultas que são relevantes para os operadores de satélite, e a sua representação em um cubo de dados, que teria resultados de implementações diferentes para conseguir avaliar qual dos algoritmos disponíveis é o mais adequado para o cenário de operação.

6.1.1 Trabalhos futuros

Como trabalhos futuros, a arquitetura do fluxo de dados mostrada na seção 4.2 pode ser implementado nos moldes das outras agências a exemplo de 3.2, numa arquitetura que permita a inclusão de todos os tipos de dados elecandos neste trabalho na seção 4.1. Também seria interessante expandir os tipos de algoritmos que serão testados para esse trabalho.

Expandir o uso dos dados para a abordagem de cubo seria relevante, pois existem desafios diferentes quando se lida com um satélite de tamanho grande (ex. GEO) e uma constelação de satélites menores, porém com um volume de dados comparável. Alguma abordagem para lidar com dados de CubeSats seria relevante para o momento, principalmente se forem de diferentes cubesats e/ou de constelações.

REFERÊNCIAS BIBLIOGRÁFICAS

- ADAMSKI, G. Data Analytics for Large Constellations. In: **SpaceOps 2016 Conference**. [S.l.]: American Institute of Aeronautics and Astronautics, 2016. (SpaceOps Conferences). 00000. [5](#), [6](#)
- AES, R. O. de M. **Estudo de avalanche térmica em um sistema de carga e descarga de bateria em satélites artificiais**. 00000. Tese (Doutorado) — Instituto Nacional de Pesquisas Espaciais, São José dos Campos, fev. 2012. Acesso em: 01 ago. 2018. [7](#)
- AZEVEDO, D. N. R.; AMBRÓSIO, A. M. Dependability in Satellite Systems: An Architecture for Satellite Telemetry Analysis. In: Workshop em Engenharia e Tecnologia Espaciais, 1. (WETE)., 30 mar. - 1 abr. 2010, São José dos Campos. **Anais...** São José dos Campos: INPE, 2010. IWETE2010-1065, p. 6. ISSN 2177-3114. 00000. Acesso em: 30 jul. 2018. [13](#)
- AZEVEDO, D. N. R.; AMBRÓSIO, A. M.; VIEIRA, M. **Estudo sobre técnicas de detecção automática de anomalias em satélites**. Tese (Doutorado) — Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2011. Acesso em: 10 ago. 2018. [7](#)
- BOUSSOUF, L.; BERGELIN, B.; SCUDELER, D.; GRAYDON, H.; STAMMINGER, J.; ROSNET, P.; TAILLEFER, E.; BARREYRE, C. Big Data Based Operations for Space Systems. In: **2018 SpaceOps Conference**. [S.l.]: American Institute of Aeronautics and Astronautics, 2018. 00000. [6](#), [11](#)
- CHANG, W.; CHENG, J.; ALLAIRE, J. J.; XIE, Y.; MCPHERSON, J. **Shiny: Web Application Framework for R**. [S.l.: s.n.], 2019. [15](#)
- CUZZOCREA, A.; BELLATRECHE, L.; SONG, I.-Y. Data Warehousing and OLAP over Big Data: Current Challenges and Future Research Directions. In: **Proceedings of the Sixteenth International Workshop on Data Warehousing and OLAP**. New York, NY, USA: ACM, 2013. (DOLAP '13), p. 67–70. ISBN 978-1-4503-2412-0. 00000. [13](#)
- DISCHNER, Z.; REDFERN, J.; ROSE, D.; ROSE, R.; RUF, C.; VINCENT, M. CYGNSS MOC; Meeting the challenge of constellation operations in a cost-constrained world. In: **2016 IEEE Aerospace Conference**. [S.l.: s.n.], 2016. p. 1–8. 00000. [6](#)

EDWARDS, T. Dealing with the Big Data - The Challenges for Modern Mission Monitoring and Reporting. In: **15th International Conference on Space Operations**. Marseille, France: American Institute of Aeronautics and Astronautics, 2018. ISBN 978-1-62410-562-3. 00000. [6](#)

EVANS, D. J.; MARTINEZ, J.; Korte-Stapff, M.; VANDENBUSSCHE, B.; ROYER, P.; RIDDER, J. D. Data Mining to Drastically Improve Spacecraft Telemetry Checking: A Scientist's Approach. In: **SpaceOps 2016 Conference**. [S.l.]: American Institute of Aeronautics and Astronautics, 2016, (SpaceOps Conferences). 00000. [6](#)

FEN, Z.; YANQIN, Z.; CHONG, C.; LING, S. Management and Operation of Communication Equipment Based on Big Data. In: **2016 International Conference on Robots Intelligent System (ICRIS)**. [S.l.: s.n.], 2016. p. 246–248. 00000. [6](#)

FERNÁNDEZ, M. M.; YUE, Y.; WEBER, R. Telemetry Anomaly Detection System Using Machine Learning to Streamline Mission Operations. In: **2017 6th International Conference on Space Mission Challenges for Information Technology (SMC-IT)**. [S.l.: s.n.], 2017. p. 70–75. 00003. [6](#), [7](#), [15](#)

GILLES, K. Flying Large Constellations Using Automation and Big Data. In: **SpaceOps 2016 Conference**. [S.l.]: American Institute of Aeronautics and Astronautics, 2016, (SpaceOps Conferences). 00000. [6](#)

HENNION, N. Big-data for satellite yearly reports generation. In: **2018 SpaceOps Conference**. [S.l.]: American Institute of Aeronautics and Astronautics, 2018. 00000. [6](#)

KANNAN, S. A.; DEVI, T. Mining satellite telemetry data: Comparison of rule-induction and association mining techniques. In: **2016 IEEE International Conference on Advances in Computer Applications (ICACA)**. [S.l.: s.n.], 2016. p. 259–264. 00000. [19](#)

MATEIK, D.; MITAL, R.; BUONAIUTO, N. L.; LOUIE, M.; KIEF, C.; AARESTAD, J. Using Big Data Technologies for Satellite Data Analytics. In: . [S.l.]: American Institute of Aeronautics and Astronautics, 2017. ISBN 978-1-62410-483-1. 00001. [5](#), [6](#), [11](#), [15](#)

MONTEIRO, D. V. **A FRAMEWORK FOR TRAJECTORY DATA MINING**. Tese (Doutorado), 2017. [7](#)

OES, R. E. d. O. S.; CAMARA, G.; QUEIROZ, G. R. de. Sits: Data analysis and machine learning using satellite image time series. In: Workshop de Computação Aplicada, 18. (WORCAP), 21-23 ago., São José dos Campos, SP. **Resumos...** [S.l.], 2018. p. 18. Acesso em: 02 maio 2019. [7](#)

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2018. 00000. [15](#)

RAMOS, M. P.; TASINAFFO, P. M.; de Almeida, E. S.; ACHITE, L. M.; da Cunha, A. M.; DIAS, L. A. V. Distributed Systems Performance for Big Data. In: LATIFI, S. (Ed.). **Information Technology: New Generations**. [S.l.]: Springer International Publishing, 2016, (Advances in Intelligent Systems and Computing). p. 733–744. ISBN 978-3-319-32467-8. [7](#)

SCHULSTER, J.; EVILL, R.; PHILLIPS, S.; FELDMANN, N.; ROGISSART, J.; DYER, R.; ARGEMANDY, A. CHARTing the Future – An offline data analysis and reporting toolkit to support automated decision-making in flight-operations. In: **15th International Conference on Space Operations**. Marseille, France: American Institute of Aeronautics and Astronautics, 2018. ISBN 978-1-62410-562-3. 00001. [6](#)

SILVA, R. R. **Abordagens para Cubo de Dados Massivos com Alta Dimensionalidade Baseadas em Memória Principal e Memória Externa: HIC e BCubing**. 00000. Tese (Doutorado) — Instituto Tecnológico de Aeronáutica, São José dos Campos, 2015. Acesso em: 01 ago. 2018. [5](#), [9](#), [13](#), [16](#)

TOMINAGA, J.; FERREIRA, M. G. V.; AMBRÓSIO, A. M. Comparing satellite telemetry against simulation parameters in a simulator model reconfiguration tool. In: CERQUEIRA, C. S.; BÜRGER, E. E.; YASSUDA, I. d. S.; RODRIGUES, I. P.; LIMA, J. S. d. S.; OLIVEIRA, M. E. R. de; TENÓRIO, P. I. G. (Ed.). **Anais...** São José dos Campos: Instituto Nacional de Pesquisas Espaciais (INPE), 2017. ISSN 2177-3114. 00000. Acesso em: 30 jul. 2018. [7](#)

TROLLOPE, E.; DYER, R.; FRANCISCO, T.; MILLER, J.; GRISO, M. P.; ARGEMANDY, A. Analysis of automated techniques for routine monitoring and contingency detection of in-flight LEO operations at EUMETSAT. In: **2018 SpaceOps Conference**. Marseille, France: American Institute of Aeronautics and Astronautics, 2018. ISBN 978-1-62410-562-3. 00001. [6](#), [7](#)

YVERNES, A. Copernicus Ground Segment as a Service: From Data Monitoring to Performance Analysis. In: **15th International Conference on Space**

Operations. Marseille, France: American Institute of Aeronautics and Astronautics, 2018. ISBN 978-1-62410-562-3. 00000. [6](#), [7](#)

ZHANG, X.; WU, P.; TAN, C. A big data framework for spacecraft prognostics and health monitoring. In: **2017 Prognostics and System Health Management Conference (PHM-Harbin)**. [S.l.: s.n.], 2017. p. 1–7. 00000. [6](#), [9](#), [10](#), [11](#), [15](#)

ANEXO A - CRONOGRAMA E PUBLICAÇÕES

A tabela mostra o cronograma esperado para as próximas atividades do mestrado.

Tabela A.1 - Cronograma de atividades

Atividade	maio	jun.	jul.	ago.	set.	out.	nov.	dec.	jan.	fev.
Exame de Proposta	X									
Submissão Artigo Periódico					X					
Apresentação Conferência						X				
Defesa final							X			X

A tabela mostra os veículos de publicação planejados e os já publicados/em processo de publicação.

Tabela A.2 - Publicações planejadas

Nome	Qualis	Prazo	Notas
WETE 2018	Conferência		Publicado
IAC 2019	Conferência		Artigo aceito, porém feito pela matéria do Prof. Geilson (não é nessa área), porém sou o 1 autor
BDCAT	Conferência	29/08/2019	Nova Zelândia...
WETE 2019	Conferência		?
IEEE América Latina	B2		?
International Journal of Data Warehousing and Mining	B1 (CC)		Não tem qualis para ENG-III, caro, JCR 0,66

PUBLICAÇÕES TÉCNICO-CIENTÍFICAS EDITADAS PELO INPE

Teses e Dissertações (TDI)

Teses e Dissertações apresentadas nos Cursos de Pós-Graduação do INPE.

Manuais Técnicos (MAN)

São publicações de caráter técnico que incluem normas, procedimentos, instruções e orientações.

Notas Técnico-Científicas (NTC)

Incluem resultados preliminares de pesquisa, descrição de equipamentos, descrição e ou documentação de programas de computador, descrição de sistemas e experimentos, apresentação de testes, dados, atlas, e documentação de projetos de engenharia.

Relatórios de Pesquisa (RPQ)

Reportam resultados ou progressos de pesquisas tanto de natureza técnica quanto científica, cujo nível seja compatível com o de uma publicação em periódico nacional ou internacional.

Propostas e Relatórios de Projetos (PRP)

São propostas de projetos técnico-científicos e relatórios de acompanhamento de projetos, atividades e convênios.

Publicações Didáticas (PUD)

Incluem apostilas, notas de aula e manuais didáticos.

Publicações Seriadas

São os seriados técnico-científicos: boletins, periódicos, anuários e anais de eventos (simpósios e congressos). Constam destas publicações o Internacional Standard Serial Number (ISSN), que é um código único e definitivo para identificação de títulos de seriados.

Programas de Computador (PDC)

São a seqüência de instruções ou códigos, expressos em uma linguagem de programação compilada ou interpretada, a ser executada por um computador para alcançar um determinado objetivo. Aceitam-se tanto programas fonte quanto os executáveis.

Pré-publicações (PRE)

Todos os artigos publicados em periódicos, anais e como capítulos de livros.