

Entrega Final – Desafio XPE (Pipeline Postgres → Kafka → S3 → Spark)

Informações:

Repositorio: <https://github.com/fdelfino93/desafioxpe>

Bucket AWS: `ferdelfino-xpe` (regiao `us-east-2`)

Arquivos de evidencias (prints): `docs/img`

Visao geral

Pipeline ponta-a-ponta:

- Fonte: PostgreSQL (tabelas `public.dadostesouroipca` e `public.dadostesouopre`).
- Streaming: Kafka (Confluent Platform) com topics `postgres-dadostesouroipca` e `postgres-dadostesouopre`.
- Ingestao Data Lake: Kafka Connect S3 Sink (Avro converter + Schema Registry) gravando camada Bronze em S3.
- Processamento: Spark (master local[*] via notebook) lendo Bronze e gravando Silver (Parquet) e Gold (Parquet agregacoes).

Camadas:

- Bronze: `s3a://ferdelfino-xpe/bronze/ipca/` e `s3a://ferdelfino-xpe/bronze/pre/` (JSON).
- Silver: `s3a://ferdelfino-xpe/processed-data/ipca/silver/` e `.../pre/silver/` (Parquet).
- Gold: `s3a://ferdelfino-xpe/analytics/ipca/gold/` e `.../pre/gold/` (Parquet).

Para eu subir o ambiente (Docker)

```
```pwsh
cd "D:\Documents\Projeto\Prova\postgres"
docker-compose up -d
```

```
cd "D:\Documents\Projeto\Prova"
docker-compose up -d
docker ps --format "table {{.Names}}\t{{.Status}}\t{{.Ports}}"
````
```

Esperado: broker, zookeeper, schema-registry, connect, postgres em estado Up.

Criar tópicos Kafka

```
```pwsh
docker exec broker kafka-topics --create --bootstrap-server localhost:9092 --partitions 1
--replication-factor 1 --topic postgres-dadostesouroipca
docker exec broker kafka-topics --create --bootstrap-server localhost:9092 --partitions 1
--replication-factor 1 --topic postgres-dadostesouopre
````
```

Registrar conectores

Source (Postgres → Kafka):

```
```pwsh
cd "D:\Documents\Projeto\Prova\connectors\source"
curl -X POST -H "Content-Type: application/json" --data
@connect_jdbc_postgres_ipca.config http://localhost:8083/connectors
curl -X POST -H "Content-Type: application/json" --data @connect_jdbc_postgres_pre.config
http://localhost:8083/connectors
````
```

Sink (Kafka → S3 Bronze) – ja configurados para bucket `ferdelfino-xpe` , regiao `us-east-2` , Avro + Schema Registry:

Evidencias (prints)

Conectores Kafka Connect RUNNING

Out[7]:

| Tipo | Titulo | Data Vencimento | Data Base | Taxa Compra Manha | Taxa Venda Manha | PU Compra Manha | PU Venda Manha | PU Base Manha | Data_Vencimento | Data_Base |
|------------------------------------|-------------------|-----------------|------------|-------------------|------------------|-----------------|----------------|---------------|-----------------|------------|
| Tesouro IPCA+ com Juros Semestrais | Tesouro Prefixado | 01/01/2006 | 06/12/2005 | 18.26 | 18.29 | 988.09 | 988.07 | 987.41 | 2006-01-01 | 2005-12-06 |
| | 15/08/2006 | 06/12/2005 | | 12.88 | 12.90 | 1525.14 | 1524.96 | 1523.96 | 2006-08-15 | 2005-12-06 |
| | 15/05/2045 | 06/12/2005 | | 9.07 | 9.17 | 1067.68 | 1056.28 | 1055.73 | 2045-05-15 | 2005-12-06 |
| | 15/08/2024 | 06/12/2005 | | 9.00 | 9.08 | 1183.68 | 1175.01 | 1174.39 | 2024-08-15 | 2005-12-06 |
| | 15/05/2015 | 06/12/2005 | | 8.98 | 9.04 | 1286.78 | 1281.80 | 1281.13 | 2015-05-15 | 2005-12-06 |
| | 15/05/2009 | 06/12/2005 | | 11.20 | 11.24 | 1351.99 | 1350.48 | 1349.67 | 2009-05-15 | 2005-12-06 |
| | 15/05/2007 | 06/12/2005 | | 11.90 | 11.92 | 1456.56 | 1456.20 | 1455.29 | 2007-05-15 | 2005-12-06 |
| | 15/08/2008 | 06/12/2005 | | 11.43 | 11.47 | 1408.42 | 1407.18 | 1406.32 | 2008-08-15 | 2005-12-06 |
| | Tesouro IPCA+ | 15/05/2015 | 06/12/2005 | 8.68 | 8.74 | 714.11 | 710.42 | 710.05 | 2015-05-15 | 2005-12-06 |
| | | 15/08/2024 | 06/12/2005 | 8.82 | 8.90 | 323.79 | 319.39 | 319.22 | 2024-08-15 | 2005-12-06 |
| ... | | | | | | | | | | |
| Tesouro Selic | 07/03/2015 | 02/08/2011 | | 0.00 | 0.04 | 4799.08 | 4792.19 | 4789.95 | 2015-03-07 | 2011-08-02 |
| | 07/03/2014 | 02/08/2011 | | 0.00 | 0.03 | 4799.08 | 4795.36 | 4793.11 | 2014-03-07 | 2011-08-02 |
| | 07/03/2017 | 02/08/2011 | | 0.00 | 0.04 | 4799.08 | 4788.39 | 4786.14 | 2017-03-07 | 2011-08-02 |
| | 07/03/2013 | 02/08/2011 | | 0.00 | 0.02 | 4799.08 | 4797.56 | 4795.32 | 2013-03-07 | 2011-08-02 |
| | 07/03/2012 | 02/08/2011 | | 0.00 | 0.01 | 4799.08 | 4798.80 | 4796.55 | 2012-03-07 | 2011-08-02 |
| Tesouro Prefixado | 01/01/2014 | 02/08/2011 | | 12.82 | 12.88 | 747.49 | 746.53 | 746.17 | 2014-01-01 | 2011-08-02 |
| | 01/01/2013 | 02/08/2011 | | 12.69 | 12.73 | 845.09 | 844.67 | 844.27 | 2013-01-01 | 2011-08-02 |
| | 01/01/2015 | 02/08/2011 | | 12.85 | 12.91 | 661.63 | 660.43 | 660.11 | 2015-01-01 | 2011-08-02 |
| | 01/01/2012 | 02/08/2011 | | 12.50 | 12.53 | 952.55 | 952.44 | 952.00 | 2012-01-01 | 2011-08-02 |
| Tesouro IPCA+ com Juros Semestrais | 15/08/2024 | 02/08/2011 | | 6.29 | 6.37 | 2067.28 | 2053.44 | 2052.81 | 2024-08-15 | 2011-08-02 |

164293 rows x 7 columns

S3/Gold IPCA (analytics)

Amazon S3 > Buckets > ferdefino-xpe > analytics/ > ipca/ > gold/

[Copiar URI do S3](#)

gold/

[Objetos](#) [Propriedades](#)

Objetos (2)

Os objetos são as entidades fundamentais armazenadas no Amazon S3. Você pode usar o [Inventário do Amazon S3](#) para obter uma lista de todos os objetos em seu bucket. Para outras pessoas acessarem seus objetos, você precisará conceder permissões explicitamente a eles. [Saiba mais](#)

| <input type="checkbox"/> | Nome | Tipo | Última modificação | Tamanho | Classe de armazenamento |
|--------------------------|--------------------------------------------------------------------|---------|-----------------------------|---------|-------------------------|
| <input type="checkbox"/> | _SUCCESS | - | 23 Nov 2025 09:24:45 AM -03 | 0 B | Padrão |
| <input type="checkbox"/> | part-00000-4d21a762-7972-4a90-a0eb-1e78e9fa221e-c000.snapy.parquet | parquet | 23 Nov 2025 09:24:42 AM -03 | 1.3 KB | Padrão |

S3/Gold PRE (analytics)

Amazon S3 > Buckets > ferdelfino-xpe > analytics > pre/ > gold/

gold/

Objetos (2)

Os objetos são as entidades fundamentais armazenadas no Amazon S3. Você pode usar o [inventário do Amazon S3](#) para obter uma lista de todos os objetos em seu bucket. Para outras pessoas acessarem seus objetos, você precisará conceder permissões explicitamente a elas. [Saiba mais](#)

| Nome | Tipo | Última modificação | Tamanho | Classe de armazenamento |
|---------------------------------------------------------------------|---------|-----------------------------|---------|-------------------------|
| _SUCCESS | - | 23 Nov 2025 09:25:12 AM -03 | 0 B | Padrão |
| part-00000-788a7d6b-abf5-4e48-b65d-e5b1b04b7800-c000.snappy.parquet | parquet | 23 Nov 2025 09:25:09 AM -03 | 1.3 KB | Padrão |

Consulta Postgres dadostesouroipca

DBeaver 25.2.5 - dadostesouroipca

Arquivo Editar Navegar Procurar Editor SQL Banco de dados Janela Ajuda

Navegador de banco de ... Projetos postgres public dadostesouroipca dadostesouropre

Filtre conexões por nome

postres (localhost:5432)

Bancos de dados

postres

- Esquemas
- Triggers de evento
- Extensões
- Armazenamento
- Informações do sistema
- Roles
- Administrador
- Informações do sistema

Propriedades Dados Diagrama

Insera uma expressão SQL para filtrar os resultados (use Ctrl+Espaço)

| ComprasManha | VendaManha | PUCompraManha | PUVentaManha | Data_Vencimento | Data_Base | Tipo | dt_update |
|--------------|------------|---------------|--------------|-----------------|----------------------------|------|-------------------------|
| 6,64 | 6,72 | 593,43 | 586,17 | 985,93 | 2024-08-15 00:00:00.000000 | IPCA | 2025-11-23 09:56:35,647 |
| 2 | 7,11 | 7,17 | 1,039,66 | 1,035,47 | 2015-05-15 00:00:00.000000 | IPCA | 2025-11-23 09:56:35,648 |
| 3 | 6,67 | 6,75 | 590,48 | 583,25 | 2024-08-15 00:00:00.000000 | IPCA | 2025-11-23 09:56:35,649 |
| 7 | 7,17 | 7,23 | 1,036,06 | 1,030,9 | 2015-05-15 00:00:00.000000 | IPCA | 2025-11-23 09:56:35,650 |
| 5 | 6,66 | 6,74 | 591,13 | 583,9 | 2024-08-15 00:00:00.000000 | IPCA | 2025-11-23 09:56:35,651 |
| 6 | 7,15 | 7,21 | 1,036,99 | 1,031,82 | 2015-05-15 00:00:00.000000 | IPCA | 2025-11-23 09:56:35,652 |
| 7 | 7,23 | 7,29 | 1,029,98 | 1,025,83 | 2015-05-15 00:00:00.000000 | IPCA | 2025-11-23 09:56:35,653 |
| 8 | 6,67 | 6,75 | 589,97 | 582,75 | 2024-08-15 00:00:00.000000 | IPCA | 2025-11-23 09:56:35,654 |
| 9 | 6,7 | 6,78 | 586,81 | 579,63 | 2024-08-15 00:00:00.000000 | IPCA | 2025-11-23 09:56:35,655 |
| 10 | 7,32 | 7,38 | 1,022,96 | 1,018,85 | 2015-05-15 00:00:00.000000 | IPCA | 2025-11-23 09:56:35,656 |
| 11 | 7,37 | 7,43 | 1,019,07 | 1,014,97 | 2015-05-15 00:00:00.000000 | IPCA | 2025-11-23 09:56:35,657 |
| 12 | 6,7 | 6,78 | 586,56 | 579,13 | 2024-08-15 00:00:00.000000 | IPCA | 2025-11-23 09:56:35,658 |
| 13 | 6,68 | 6,76 | 580,12 | 580,91 | 2015-05-15 00:00:00.000000 | IPCA | 2025-11-23 09:56:35,659 |
| 14 | 7,37 | 7,43 | 1,018,61 | 1,014,91 | 2015-05-15 00:00:00.000000 | IPCA | 2025-11-23 09:56:35,660 |
| 15 | 6,72 | 6,8 | 584,25 | 577,1 | 2024-08-15 00:00:00.000000 | IPCA | 2025-11-23 09:56:35,661 |
| 16 | 7,44 | 7,5 | 1,018,37 | 1,009,29 | 2015-05-15 00:00:00.000000 | IPCA | 2025-11-23 09:56:35,662 |
| 17 | 7 | 7,08 | 564,37 | 559,29 | 2024-08-15 00:00:00.000000 | IPCA | 2025-11-23 09:56:35,663 |
| 18 | 7,81 | 7,87 | 1,007,18 | 997,21 | 2015-05-15 00:00:00.000000 | IPCA | 2025-11-23 09:56:35,664 |
| 19 | 7,02 | 7,1 | 564,44 | 557,38 | 2024-08-15 00:00:00.000000 | IPCA | 2025-11-23 09:56:35,665 |
| 20 | 7,78 | 7,84 | 1,002,77 | 998,4 | 2015-05-15 00:00:00.000000 | IPCA | 2025-11-23 09:56:35,666 |
| 21 | 6,99 | 7,07 | 568,82 | 559,94 | 2024-08-15 00:00:00.000000 | IPCA | 2025-11-23 09:56:35,667 |
| 22 | 7,75 | 7,81 | 1,004,37 | 1,000,39 | 2015-05-15 00:00:00.000000 | IPCA | 2025-11-23 09:56:35,668 |
| 23 | 6,94 | 7,02 | 570,8 | 563,86 | 2024-08-15 00:00:00.000000 | IPCA | 2025-11-23 09:56:35,669 |
| 24 | 7,64 | 7,7 | 1,011,03 | 1,007,01 | 2015-05-15 00:00:00.000000 | IPCA | 2025-11-23 09:56:35,670 |
| 25 | 7,77 | 7,83 | 1,001,96 | 997,98 | 2015-05-15 00:00:00.000000 | IPCA | 2025-11-23 09:56:35,671 |
| 26 | 7,06 | 7,14 | 560,22 | 553,42 | 2024-08-15 00:00:00.000000 | IPCA | 2025-11-23 09:56:35,672 |
| 27 | 7,69 | 7,75 | 1,006,89 | 1,002,89 | 2015-05-15 00:00:00.000000 | IPCA | 2025-11-23 09:56:35,673 |
| 28 | 6,97 | 7,05 | 567,77 | 560,87 | 2024-08-15 00:00:00.000000 | IPCA | 2025-11-23 09:56:35,674 |
| 29 | 6,85 | 6,93 | 578,09 | 571,05 | 2024-08-15 00:00:00.000000 | IPCA | 2025-11-23 09:56:35,675 |
| 30 | 7,57 | 7,63 | 1,014,56 | 1,010,52 | 2015-05-15 00:00:00.000000 | IPCA | 2025-11-23 09:56:35,676 |
| 31 | 6,9 | 6,98 | 573,32 | 566,34 | 2024-08-15 00:00:00.000000 | IPCA | 2025-11-23 09:56:35,677 |
| 32 | 7,49 | 7,55 | 1,019,31 | 1,015,25 | 2015-05-15 00:00:00.000000 | IPCA | 2025-11-23 09:56:35,678 |
| 33 | 6,76 | 6,84 | 585,5 | 578,36 | 2024-08-15 00:00:00.000000 | IPCA | 2025-11-23 09:56:35,679 |
| 34 | 7,34 | 7,4 | 1,028,08 | 1,024,97 | 2015-05-15 00:00:00.000000 | IPCA | 2025-11-23 09:56:35,680 |
| 35 | 6,83 | 6,91 | 579 | 571,94 | 2024-08-15 00:00:00.000000 | IPCA | 2025-11-23 09:56:35,681 |
| 36 | 7,44 | 7,5 | 1,021,78 | 1,017,7 | 2015-05-15 00:00:00.000000 | IPCA | 2025-11-23 09:56:35,682 |
| 37 | 6,77 | 6,85 | 584,05 | 576,93 | 2024-08-15 00:00:00.000000 | IPCA | 2025-11-23 09:56:35,683 |
| 38 | 7,33 | 7,39 | 1,028,75 | 1,024,64 | 2015-05-15 00:00:00.000000 | IPCA | 2025-11-23 09:56:35,684 |
| 39 | 6,73 | 6,81 | 587,4 | 580,23 | 2024-08-15 00:00:00.000000 | IPCA | 2025-11-23 09:56:35,685 |
| 40 | 7,19 | 7,25 | 1,037,96 | 1,033,81 | 2015-05-15 00:00:00.000000 | IPCA | 2025-11-23 09:56:35,686 |
| 41 | 6,6 | 6,68 | 598,8 | 591,48 | 2024-08-15 00:00:00.000000 | IPCA | 2025-11-23 09:56:35,687 |

Files - General Registro

Atualizar Salvar Cancelar Exportar dados 200 200+ 200 linhas recuperada(s) - 0,0s (0 recuperado) em 2025-11-23 às 10:08:36

Consulta Postgres dadostesouopre

DBeaver 25.2.5 - dadostesouopre

Arquivo Editar Navegar Procurar Editor SQL Banco de dados Janela Ajuda

Navegador de banco de... Projetos

postres public dadostesouopre dadostesouopre

Filtre conexões por nome

Bancos de dados postres

Esquemas Triggers de evento Extensões Armazenamento Informações do sistema Roles Administrar Informações do sistema

Propriedades Dados Diagrama

Inserir uma expressão SQL para filtrar os resultados (use Ctrl+Espaço)

| Grade | 1 | 11,55 | 11,59 | 937,01 | 936,81 | 936,4 | 2008-10-01 00:00:00.000 | 2008-02-27 00:00:00.000 | PRE-FIXADOS | 2025-11-23 09:56:25,950 |
|-------|-------|-------|--------|--------|--------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
| 2 | 11,19 | 11,22 | 999,78 | 990,75 | 990,34 | 2008-04-01 00:00:00.000 | 2008-02-27 00:00:00.000 | PRE-FIXADOS | 2025-11-23 09:56:25,950 | |
| 3 | 12,21 | 12,25 | 857,22 | 856,81 | 856,41 | 2009-07-01 00:00:00.000 | 2008-02-27 00:00:00.000 | PRE-FIXADOS | 2025-11-23 09:56:25,900 | |
| 4 | 12,38 | 12,43 | 830,11 | 829,52 | 829,14 | 2009-10-01 00:00:00.000 | 2008-02-27 00:00:00.000 | PRE-FIXADOS | 2025-11-23 09:56:25,961 | |
| 5 | 12,46 | 12,51 | 805,18 | 804,52 | 804,14 | 2010-01-01 00:00:00.000 | 2008-02-27 00:00:00.000 | PRE-FIXADOS | 2025-11-23 09:56:25,962 | |
| 6 | 11,78 | 11,82 | 909,36 | 909,08 | 908,68 | 2009-01-01 00:00:00.000 | 2008-02-27 00:00:00.000 | PRE-FIXADOS | 2025-11-23 09:56:25,963 | |
| 7 | 11,33 | 11,36 | 964,85 | 964,76 | 964,35 | 2008-07-01 00:00:00.000 | 2008-02-27 00:00:00.000 | PRE-FIXADOS | 2025-11-23 09:56:25,964 | |
| 8 | 11,18 | 11,21 | 990,37 | 990,34 | 989,93 | 2008-04-01 00:00:00.000 | 2008-02-26 00:00:00.000 | PRE-FIXADOS | 2025-11-23 09:56:25,965 | |
| 9 | 12,39 | 12,44 | 829,61 | 829,02 | 828,63 | 2009-10-01 00:00:00.000 | 2008-02-26 00:00:00.000 | PRE-FIXADOS | 2025-11-23 09:56:25,966 | |
| 10 | 12,49 | 12,54 | 804,41 | 803,75 | 803,37 | 2010-01-01 00:00:00.000 | 2008-02-26 00:00:00.000 | PRE-FIXADOS | 2025-11-23 09:56:25,967 | |
| 11 | 11,33 | 11,36 | 964,44 | 964,35 | 963,94 | 2008-07-01 00:00:00.000 | 2008-02-26 00:00:00.000 | PRE-FIXADOS | 2025-11-23 09:56:25,968 | |
| 12 | 11,57 | 11,61 | 936,5 | 936,3 | 935,89 | 2008-10-01 00:00:00.000 | 2008-02-26 00:00:00.000 | PRE-FIXADOS | 2025-11-23 09:56:25,969 | |
| 13 | 11,81 | 11,85 | 908,75 | 908,47 | 908,06 | 2009-01-01 00:00:00.000 | 2008-02-26 00:00:00.000 | PRE-FIXADOS | 2025-11-23 09:56:25,970 | |
| 14 | 12,24 | 12,28 | 856,52 | 856,11 | 855,71 | 2009-07-01 00:00:00.000 | 2008-02-26 00:00:00.000 | PRE-FIXADOS | 2025-11-23 09:56:25,971 | |
| 15 | 11,17 | 11,2 | 989,95 | 989,94 | 989,52 | 2008-04-01 00:00:00.000 | 2008-02-25 00:00:00.000 | PRE-FIXADOS | 2025-11-23 09:56:25,972 | |
| 16 | 12,4 | 12,45 | 805,23 | 804,56 | 804,19 | 2010-01-01 00:00:00.000 | 2008-02-25 00:00:00.000 | PRE-FIXADOS | 2025-11-23 09:56:25,973 | |
| 17 | 11,32 | 11,35 | 964,05 | 963,97 | 963,56 | 2008-07-01 00:00:00.000 | 2008-02-25 00:00:00.000 | PRE-FIXADOS | 2025-11-23 09:56:25,974 | |
| 18 | 12,33 | 12,38 | 829,93 | 829,34 | 828,96 | 2009-01-01 00:00:00.000 | 2008-02-25 00:00:00.000 | PRE-FIXADOS | 2025-11-23 09:56:25,975 | |
| 19 | 12,17 | 12,21 | 856,84 | 856,43 | 856,04 | 2009-07-01 00:00:00.000 | 2008-02-25 00:00:00.000 | PRE-FIXADOS | 2025-11-23 09:56:25,976 | |
| 20 | 11,54 | 11,58 | 936,24 | 936,04 | 935,63 | 2008-10-01 00:00:00.000 | 2008-02-25 00:00:00.000 | PRE-FIXADOS | 2025-11-23 09:56:25,977 | |
| 21 | 11,76 | 11,8 | 908,69 | 908,41 | 908,01 | 2009-01-01 00:00:00.000 | 2008-02-25 00:00:00.000 | PRE-FIXADOS | 2025-11-23 09:56:25,978 | |
| 22 | 11,16 | 11,19 | 989,55 | 989,53 | 989,11 | 2008-04-01 00:00:00.000 | 2008-02-22 00:00:00.000 | PRE-FIXADOS | 2025-11-23 09:56:25,979 | |
| 23 | 11,29 | 11,32 | 963,74 | 963,65 | 963,24 | 2008-07-01 00:00:00.000 | 2008-02-22 00:00:00.000 | PRE-FIXADOS | 2025-11-23 09:56:25,980 | |
| 24 | 11,51 | 11,55 | 935,99 | 935,79 | 935,38 | 2008-10-01 00:00:00.000 | 2008-02-22 00:00:00.000 | PRE-FIXADOS | 2025-11-23 09:56:25,981 | |
| 25 | 12,4 | 12,45 | 804,85 | 804,19 | 803,81 | 2010-01-01 00:00:00.000 | 2008-02-22 00:00:00.000 | PRE-FIXADOS | 2025-11-23 09:56:25,982 | |
| 26 | 12,33 | 12,38 | 829,55 | 828,96 | 828,57 | 2009-10-01 00:00:00.000 | 2008-02-22 00:00:00.000 | PRE-FIXADOS | 2025-11-23 09:56:25,983 | |
| 27 | 12,14 | 12,18 | 856,76 | 856,35 | 855,96 | 2009-07-01 00:00:00.000 | 2008-02-22 00:00:00.000 | PRE-FIXADOS | 2025-11-23 09:56:25,984 | |
| 28 | 11,73 | 11,77 | 908,5 | 908,22 | 907,82 | 2009-01-01 00:00:00.000 | 2008-02-22 00:00:00.000 | PRE-FIXADOS | 2025-11-23 09:56:25,985 | |
| 29 | 11,17 | 11,2 | 989,18 | 989,1 | 988,69 | 2008-04-01 00:00:00.000 | 2008-02-21 00:00:00.000 | PRE-FIXADOS | 2025-11-23 09:56:25,986 | |
| 30 | 11,29 | 11,32 | 963,33 | 963,24 | 962,83 | 2008-07-01 00:00:00.000 | 2008-02-21 00:00:00.000 | PRE-FIXADOS | 2025-11-23 09:56:25,987 | |
| 31 | 11,52 | 11,56 | 935,53 | 935,33 | 934,92 | 2008-10-01 00:00:00.000 | 2008-02-21 00:00:00.000 | PRE-FIXADOS | 2025-11-23 09:56:25,988 | |
| 32 | 11,75 | 11,79 | 907,96 | 907,68 | 907,28 | 2009-01-01 00:00:00.000 | 2008-02-21 00:00:00.000 | PRE-FIXADOS | 2025-11-23 09:56:25,989 | |
| 33 | 12,17 | 12,21 | 856,06 | 855,65 | 855,26 | 2009-07-01 00:00:00.000 | 2008-02-21 00:00:00.000 | PRE-FIXADOS | 2025-11-23 09:56:25,990 | |
| 34 | 12,34 | 12,39 | 829,05 | 828,46 | 828,07 | 2009-10-01 00:00:00.000 | 2008-02-21 00:00:00.000 | PRE-FIXADOS | 2025-11-23 09:56:25,991 | |
| 35 | 12,42 | 12,47 | 804,21 | 803,55 | 803,17 | 2010-01-01 00:00:00.000 | 2008-02-21 00:00:00.000 | PRE-FIXADOS | 2025-11-23 09:56:25,992 | |
| 36 | 11,17 | 11,2 | 988,71 | 988,69 | 988,27 | 2008-04-01 00:00:00.000 | 2008-02-20 00:00:00.000 | PRE-FIXADOS | 2025-11-23 09:56:25,993 | |
| 37 | 11,32 | 11,35 | 962,83 | 962,74 | 962,33 | 2008-07-01 00:00:00.000 | 2008-02-20 00:00:00.000 | PRE-FIXADOS | 2025-11-23 09:56:25,994 | |
| 38 | 12,44 | 12,49 | 803,57 | 802,91 | 802,53 | 2010-01-01 00:00:00.000 | 2008-02-20 00:00:00.000 | PRE-FIXADOS | 2025-11-23 09:56:25,995 | |
| 39 | 12,38 | 12,43 | 828,19 | 827,6 | 827,21 | 2009-10-01 00:00:00.000 | 2008-02-20 00:00:00.000 | PRE-FIXADOS | 2025-11-23 09:56:25,996 | |
| 40 | 12,19 | 12,23 | 855,47 | 855,05 | 854,66 | 2009-07-01 00:00:00.000 | 2008-02-20 00:00:00.000 | PRE-FIXADOS | 2025-11-23 09:56:25,997 | |
| 41 | 11,79 | 11,83 | 907,28 | 907 | 906,59 | 2009-01-01 00:00:00.000 | 2008-02-20 00:00:00.000 | PRE-FIXADOS | 2025-11-23 09:56:25,998 | |

Atualizar Salvar Cancelar Exportar dados 200 200+ ... 200 linha(s) recuperada(s) - 0,0s (0,0 recuperado), em 2025-11-23 às 10:09:27

S3 Bronze - IPCA

partition=0/

Objetos (280)

Copiar URI do S3 Copiar URL Fazer download Abrir Excluir Ações Criar pasta Carregar

Localizar objetos por prefixo

| Nome | Tipo | Última modificação | Tamanho | Classe de armazenamento |
|------------------------------------------|------|-----------------------------|---------|-------------------------|
| postgres-dadostesouopca+0-000000000.json | json | 23 Nov 2025 09:15:01 AM -03 | 405.0 B | Padrão |
| postgres-dadostesouopca+0-000000002.json | json | 23 Nov 2025 09:15:02 AM -03 | 405.0 B | Padrão |
| postgres-dadostesouopca+0-000000004.json | json | 23 Nov 2025 09:15:04 AM -03 | 405.0 B | Padrão |
| postgres-dadostesouopca+0-000000006.json | json | 23 Nov 2025 09:15:05 AM -03 | 406.0 B | Padrão |
| postgres-dadostesouopca+0-000000008.json | json | 23 Nov 2025 09:15:06 AM -03 | 406.0 B | Padrão |
| postgres-dadostesouopca+0-000000010.json | json | 23 Nov 2025 09:15:08 AM -03 | 406.0 B | Padrão |
| postgres-dadostesouopca+0-000000012.json | json | 23 Nov 2025 09:15:09 AM -03 | 405.0 B | Padrão |
| postgres-dadostesouopca+0-000000014.json | json | 23 Nov 2025 09:15:10 AM -03 | 405.0 B | Padrão |
| postgres-dadostesouopca+0-000000016.json | json | 23 Nov 2025 09:15:12 AM -03 | 405.0 B | Padrão |
| postgres-dadostesouopca+0-000000018.json | json | 23 Nov 2025 09:15:13 AM -03 | 405.0 B | Padrão |
| postgres-dadostesouopca+0-000000020.json | json | 23 Nov 2025 09:15:14 AM -03 | 405.0 B | Padrão |
| postgres-dadostesouopca+0-000000022.json | json | 23 Nov 2025 09:15:15 AM -03 | 406.0 B | Padrão |
| postgres-dadostesouopca+0-000000024.json | json | 23 Nov 2025 09:15:17 AM -03 | 405.0 B | Padrão |
| postgres-dadostesouopca+0-000000026.json | json | 23 Nov 2025 09:15:18 AM -03 | 406.0 B | Padrão |
| postgres-dadostesouopca+0-000000028.json | json | 23 Nov 2025 09:15:20 AM -03 | 406.0 B | Padrão |

S3 Bronze - PRE

Amazon S3 > Buckets > ferdeflino-xpc > bronze/ > pre/ > postgres-dadosesouropre/ > partition=0/

partition=0/

Objetos | Propriedades

Objetos (256)

Copiar URI do S3 | Copiar URL | Fazer download | Abrir | Excluir | Ações | Criar pasta | Carregar

Localizar objetos por prefixo

| Nome | Tipo | Última modificação | Tamanho | Classe de armazenamento |
|-------------------------------------------|------|-----------------------------|---------|-------------------------|
| postgres-dadosesouropre+0+0000000000.json | json | 23 Nov 2025 09:14:46 AM -03 | 424.0 B | Padrão |
| postgres-dadosesouropre+0+0000000002.json | json | 23 Nov 2025 09:14:47 AM -03 | 424.0 B | Padrão |
| postgres-dadosesouropre+0+0000000004.json | json | 23 Nov 2025 09:14:49 AM -03 | 424.0 B | Padrão |
| postgres-dadosesouropre+0+0000000006.json | json | 23 Nov 2025 09:14:50 AM -03 | 422.0 B | Padrão |
| postgres-dadosesouropre+0+0000000008.json | json | 23 Nov 2025 09:14:51 AM -03 | 423.0 B | Padrão |
| postgres-dadosesouropre+0+0000000010.json | json | 23 Nov 2025 09:14:53 AM -03 | 424.0 B | Padrão |
| postgres-dadosesouropre+0+0000000012.json | json | 23 Nov 2025 09:36:09 AM -03 | 423.0 B | Padrão |
| postgres-dadosesouropre+0+0000000014.json | json | 23 Nov 2025 09:36:10 AM -03 | 424.0 B | Padrão |
| postgres-dadosesouropre+0+0000000016.json | json | 23 Nov 2025 09:36:11 AM -03 | 423.0 B | Padrão |
| postgres-dadosesouropre+0+0000000018.json | json | 23 Nov 2025 09:36:12 AM -03 | 424.0 B | Padrão |
| postgres-dadosesouropre+0+0000000020.json | json | 23 Nov 2025 09:36:14 AM -03 | 424.0 B | Padrão |
| postgres-dadosesouropre+0+0000000022.json | json | 23 Nov 2025 09:36:15 AM -03 | 421.0 B | Padrão |
| postgres-dadosesouropre+0+0000000024.json | json | 23 Nov 2025 09:36:16 AM -03 | 424.0 B | Padrão |
| postgres-dadosesouropre+0+0000000026.json | json | 23 Nov 2025 09:36:17 AM -03 | 423.0 B | Padrão |
| postgres-dadosesouropre+0+0000000028.json | json | 23 Nov 2025 09:36:18 AM -03 | 422.0 B | Padrão |
| postgres-dadosesouropre+0+0000000030.json | json | 23 Nov 2025 09:36:20 AM -03 | 422.0 B | Padrão |

Containers Confluent/Connect/DB (docker ps)

docker desktop

Containers | Give feedback

Container CPU usage: 7.04% / 800% (8 CPUs available) | Container memory usage: 3.24GB / 6.57GB | Show charts

Search | Only show running containers

| Name | Image | Status | Port(s) | CPU (%) | Last started | Actions |
|-----------------------------|---------------------------------------|---------------|-------------|---------|----------------|--------------------------|
| postgres | | Running (1/1) | | 0.3% | 44 minutes ago | [...] |
| postgres_12185cd10497 | postgres:13.2 | Running | 5432-5432 | 0.3% | 44 minutes ago | [...] |
| prova | | Running (7/7) | | 6.7% | 36 minutes ago | [...] |
| zookeeper_21f0146794 | confluentinc/cp-zookeeper:7.0.0 | Running | 2181-2181 | 0.12% | 36 minutes ago | [...] |
| broker_0add275e12e | confluentinc/cp-kafka:7.0.0 | Running | 29092-29092 | 1.99% | 36 minutes ago | [...] Show all ports (3) |
| connect_d3e1ea8223c | connect-custom:1.0.0 | Running | 8083-8083 | 3.47% | 36 minutes ago | [...] |
| schema-registry_d22dac2e2d7 | confluentinc/cp-schema-registry:7.0.0 | Running | 8081-8081 | 0.63% | 36 minutes ago | [...] |
| rest-proxy_0dd542ea77b | confluentinc/cp-kafka-rest:7.0.0 | Running | 8082-8082 | 0.1% | 36 minutes ago | [...] |
| ksqldb-cli_c4d9fb138d7 | confluentinc/cp-ksqldb-cli:7.0.0 | Running | | 0% | 36 minutes ago | [...] |
| ksqldb-server_cbf230008d96 | confluentinc/cp-ksqldb-server:7.0.0 | Running | 8088-8088 | 0.39% | 36 minutes ago | [...] |

Showing 10 items

Engine running | RAM 6.93 GB CPU 0.50% | Signed in | New version available 4.6

Containers Kafka/Spark/Connect (docker ps)

Docker desktop

Containers

Container CPU usage: 11.97% / 800% (8 CPUs available)

Container memory usage: 4.89GB / 6.57GB

Search: Search

Only show running containers

| Name | Image | Status | Port(s) | CPU (%) | Last started | Actions |
|------------------|--------------------------------------|---------------|-------------|---------|--------------|--------------------|
| c4d3fb138d7 | ksqldb-server | Running | 8088:8088 | 0.26% | 8 hours ago | [...] |
| cbf230008d96 | confluentinc/cp-ksqldb-server:7.0.0 | Running | 8082:8082 | 0.09% | 8 hours ago | [...] |
| 0dd1542ea77b | rest-proxy | Running | 8081:8081 | 0.42% | 8 hours ago | [...] |
| d22dec2e22d7 | schema-registry | Running | 29092:29092 | 2.01% | 8 hours ago | Show all ports (3) |
| 0acd3d75e12e | broker | Running | 2181:2181 | 0.11% | 8 hours ago | [...] |
| 21f9014bd794 | zookeeper | Running | 8083:8083 | 2.86% | 7 hours ago | [...] |
| b13878729815 | connect | Running | 5432:5432 | 5.09% | 8 hours ago | [...] |
| postres | | Running (1/1) | | 5.09% | 8 hours ago | [...] |
| postres | postgres:13.2 | Running | 8888:8888 | 0.3% | 6 hours ago | [...] |
| spark | | Running (3/3) | | 0.48% | 6 hours ago | [...] |
| master | bde2020/spark-master:3.3.0-hadoop3.3 | Running | 7077:7077 | 0.08% | 6 hours ago | Show all ports (2) |
| jupyter-notebook | jupyter/nyspark-notebook:latest | Running | 8085:8085 | 0.1% | 6 hours ago | [...] |
| worker | bde2020/spark-worker:3.3.0-hadoop3.3 | Running | 8086:8086 | 0.1% | 6 hours ago | [...] |

Showing 14 items

Engine running RAM 6.91 GB CPU 6.12% Signed in

Restart to update

Bucket AWS ferdelfino-xpe

ferdelfino-xpe Informações

Objetos | Metadados | Propriedades | Permissões | Métricas | Gerenciamento | Pontos de acesso

Objetos (3)

Copiar URI do S3 | Copiar URL | Fazer download | Abrir | Excluir | Ações | Criar pasta | Carregar

Localizar objetos por prefixo

| Nome | Tipo | Última modificação | Tamanho | Classe de armazenamento |
|-----------------|-------|--------------------|---------|-------------------------|
| analytics/ | Pasta | - | - | - |
| bronze/ | Pasta | - | - | - |
| processed-data/ | Pasta | - | - | - |

Spark Gold IPCA/PRE (visao)

The screenshot shows the AWS S3 console with the path: Amazon S3 > Buckets > ferdelfino-xpe > analytics/ > ipca/ > gold/. The 'gold/' folder contains two objects: '_SUCCESS' (0 B) and a parquet file named 'part-00000-4d21a762-7972-4a90-a0eb-1e78e9fa221e-c000.snappy.parquet' (1.3 KB). The table has columns: Nome, Tipo, Última modificação, Tamanho, and Classe de armazenamento.

| Nome | Tipo | Última modificação | Tamanho | Classe de armazenamento |
|---------------------------------------------------------------------|---------|-----------------------------|---------|-------------------------|
| _SUCCESS | - | 23 Nov 2025 09:24:45 AM -03 | 0 B | Padrão |
| part-00000-4d21a762-7972-4a90-a0eb-1e78e9fa221e-c000.snappy.parquet | parquet | 23 Nov 2025 09:24:42 AM -03 | 1.3 KB | Padrão |

Spark Silver IPCA/PRE (visao)

The screenshot shows the AWS S3 console with the path: Amazon S3 > Buckets > ferdelfino-xpe > processed-data/ > pre/ > silver/. The 'silver/' folder contains two objects: '_SUCCESS' (0 B) and a parquet file named 'part-00000-3f45fda8-77c4-41a9-b300-b870ead6b68-c000.snappy.parquet' (14.4 KB). The table has columns: Nome, Tipo, Última modificação, Tamanho, and Classe de armazenamento.

| Nome | Tipo | Última modificação | Tamanho | Classe de armazenamento |
|--------------------------------------------------------------------|---------|-----------------------------|---------|-------------------------|
| _SUCCESS | - | 23 Nov 2025 09:24:09 AM -03 | 0 B | Padrão |
| part-00000-3f45fda8-77c4-41a9-b300-b870ead6b68-c000.snappy.parquet | parquet | 23 Nov 2025 09:24:06 AM -03 | 14.4 KB | Padrão |

Jupyter Notebook executado (Spark)

The screenshot shows a Jupyter Notebook cell with the code:

```
[21]: import os
from pyspark.sql import SparkSession
from pyspark.sql.functions import col, from_unixtime, avg, count
from dotenv import load_dotenv

load_dotenv() # carrega AWS_* do ambiente

spark = (
    SparkSession.builder
    .appName("S3 Access Example")
    .master("local[*]")
    .config("spark.jars", "/opt/spark/jars/hadoop-aws-3.4.4.jar,/opt/spark/jars/aws-java-sdk-bundle-1.12.262.jar")
    .config("spark.hadoop.fs.s3a.impl", "org.apache.hadoop.fs.s3a.S3AFFileSystem")
    .config("spark.hadoop.fs.s3a.aws.credentials.provider", "com.amazonaws.auth.EnvironmentVariableCredentialsProvider")
    .config("spark.hadoop.fs.s3a.path.style.access", "true")
    .config("spark.hadoop.fs.s3a.connection.ssl.enabled", "true")
    .getOrCreate()
)

aws_access_key = os.getenv("AWS_ACCESS_KEY_ID")
aws_secret_key = os.getenv("AWS_SECRET_ACCESS_KEY")
aws_region = os.getenv("AWS_REGION", "us-east-2")

spark._jsc.hadoopConfiguration().set("fs.s3a.access.key", aws_access_key)
spark._jsc.hadoopConfiguration().set("fs.s3a.secret.key", aws_secret_key)
spark._jsc.hadoopConfiguration().set("fs.s3a.endpoint", f"s3.{aws_region}.amazonaws.com")

print("Spark session ready")
```

Spark session ready

```
[22]: bronze_ipca = "s3a://ferdelfino-xpe/bronze/ipca/"
bronze_pre = "s3a://ferdelfino-xpe/bronze/pre/"

silver_ipca = "s3a://ferdelfino-xpe/processed-data/ipca/silver/"
silver_pre = "s3a://ferdelfino-xpe/processed-data/pre/silver/"

gold_ipca = "s3a://ferdelfino-xpe/analytics/ipca/gold/"
gold_pre = "s3a://ferdelfino-xpe/analytics/pre/gold/"

for p in [bronze_ipca, bronze_pre, silver_ipca, silver_pre, gold_ipca, gold_pre]:
    print(p)

s3a://ferdelfino-xpe/bronze/ipca/
s3a://ferdelfino-xpe/bronze/pre/
s3a://ferdelfino-xpe/processed-data/ipca/silver/
```

Jupyter Notebook - contagens/paths

The screenshot shows a Jupyter Notebook interface with a file browser on the left and a code editor on the right.

File Browser:

- File: etl-spark.ipynb
- Folder: / work /
- File: -py (5 hours ago)
- File: etl-spark.ipynb (40 seconds ago)
- File: etl-spark.ipynb (6 hours ago)

Code Editor (Cells):

[22]:

```
bronze_ipca = "s3a://ferdelfino-xpe/bronze/ipca/"
bronze_pre = "s3a://ferdelfino-xpe/bronze/pre/"

silver_ipca = "s3a://ferdelfino-xpe/processed-data/ipca/silver/"
silver_pre = "s3a://ferdelfino-xpe/processed-data/pre/silver/"

gold_ipca = "s3a://ferdelfino-xpe/analytics/ipca/gold/"
gold_pre = "s3a://ferdelfino-xpe/analytics/pre/gold/"

for p in [bronze_ipca, bronze_pre, silver_ipca, silver_pre, gold_ipca, gold_pre]:
    print(p)

s3a://ferdelfino-xpe/bronze/ipca/
s3a://ferdelfino-xpe/bronze/pre/
s3a://ferdelfino-xpe/processed-data/ipca/silver/
s3a://ferdelfino-xpe/processed-data/pre/silver/
s3a://ferdelfino-xpe/analytics/ipca/gold/
s3a://ferdelfino-xpe/analytics/pre/gold/
```

[23]: # Ler Bronze

```
df_bronze_ipca = spark.read.json(bronze_ipca)
df_bronze_pre = spark.read.json(bronze_pre)
print("Bronze IPCA count:", df_bronze_ipca.count())
print("Bronze PRE count:", df_bronze_pre.count())

Bronze IPCA count: 560
Bronze PRE count: 512
```

[24]: # Silver: deduplicar, converter timestamps, tratar nulos

```
def to_silver(df):
    df_s = df.dropDuplicates()
    df_s = df_s.withColumn("Data_Vencimento", from_unixtime(col("Data_Vencimento")/1000, "yyyy-MM-dd")) \
        .withColumn("Data_Base", from_unixtime(col("Data_Base")/1000, "yyyy-MM-dd")) \
        .withColumn("dt_update", from_unixtime(col("dt_update")/1000, "yyyy-MM-dd HH:mm:ss"))
    df_s = df_s.fillna({"PUCompraManha": 0, "PUVendaManha": 0, "PUBaseManha": 0})
    return df_s

silver_ipca_df = to_silver(df_bronze_ipca)
silver_pre_df = to_silver(df_bronze_pre)

silver_ipca_df.write.mode("overwrite").parquet(silver_ipca)
silver_pre_df.write.mode("overwrite").parquet(silver_pre)
print("Silver gravado")

Silver gravado
```

Jupyter Notebook - Silver/Gold

The screenshot shows a Jupyter Notebook interface running on localhost:8888. The left sidebar displays a file tree with a file named 'etl-spark.ipynb' selected. The main area contains Python code for ETL operations:

```
silver_ipca_df.write.mode("overwrite").parquet(silver_ipca)
silver_pre_df.write.mode("overwrite").parquet(silver_pre)
print("Silver gravado")

[25]: # Gold: agregações (médias e contagem por tipo)
gold_ipca_df = silver_ipca_df.groupBy("Tipo").agg(
    avg("PUCompraManha").alias("Media_PUCompraManha"),
    avg("PUVendaManha").alias("Media_PUVendaManha"),
    count("*").alias("Total_Registros")
)

gold_pre_df = silver_pre_df.groupBy("Tipo").agg(
    avg("PUCompraManha").alias("Media_PUCompraManha"),
    avg("PUVendaManha").alias("Media_PUVendaManha"),
    count("*").alias("Total_Registros")
)

gold_ipca_df.write.mode("overwrite").parquet(gold_ipca)
gold_pre_df.write.mode("overwrite").parquet(gold_pre)
print("Gold gravado")

[ ]: spark.stop()
```

S3/Silver IPCA (processed-data)

The screenshot shows the Amazon S3 console at the URL [Amazon S3 > Buckets > ferdeflino-xpe > processed-data/ > ipca/ > silver/](#). The 'silver/' folder contains two objects:

| Nome | Tipo | Última modificação | Tamanho | Classe de armazenamento |
|---------------------------------------------------------------------|---------|-----------------------------|---------|-------------------------|
| _SUCCESS | - | 23 Nov 2025 09:23:41 AM -03 | 0 B | Padrão |
| part-00000-67e5d22d-1f63-467e-8505-17c3cafb8e36-c000.snappy.parquet | parquet | 23 Nov 2025 09:23:37 AM -03 | 15.9 KB | Padrão |

S3/Silver PRE (processed-data)

Amazon S3 > Buckets > ferdelfino-xpe > processed-data/ > pre/ > silver/

silver/

Copiar URI do S3

Objetos Propriedades

Objetos (2)

Os objetos são as entidades fundamentais armazenadas no Amazon S3. Você pode usar o [inventário do Amazon S3](#) para obter uma lista de todos os objetos em seu bucket. Para outras pessoas acessarem seus objetos, você precisará conceder permissões explicitamente a eles. [Saiba mais](#)

Localizar objetos por prefixo

< 1 > |

| <input type="checkbox"/> | Nome | Tipo | Última modificação | Tamanho | Classe de armazenamento |
|--------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------|---------|-----------------------------|---------|-------------------------|
| <input type="checkbox"/> |  _SUCCESS | - | 23 Nov 2025 09:24:09 AM -03 | 0 B | Padrão |
| <input type="checkbox"/> |  part-00000-3f45fda8-77c4-41a9-b300-b4870ead6b68-c000.snappy.parquet | parquet | 23 Nov 2025 09:24:06 AM -03 | 14.4 KB | Padrão |