



**GOVERNO DO
ESTADO DO CEARÁ**
Secretaria da Educação

Consultoria para consolidação e construção de banco de dados do programa de educação profissional

Coordenadoria de Avaliação e Acompanhamento da Educação – COAVE

Consultores:

Jussiê Ramos

Fernando Alcântara

Acompanhamento:

Eixo de Indicadores Educacionais

Fortaleza, 22 de novembro de 2018

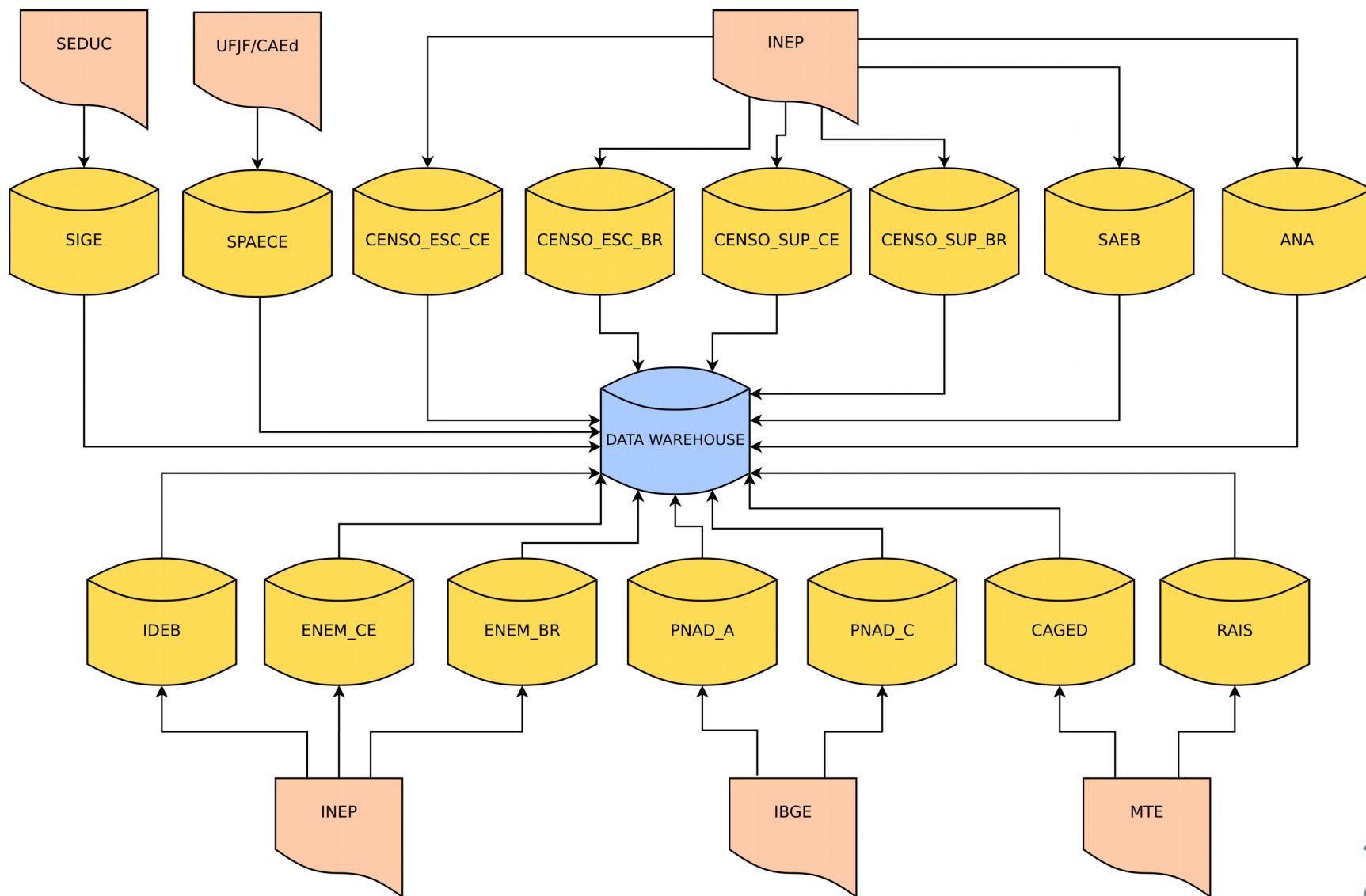


Objetivos Específicos

- Importar, transformar e estruturar no modelo de um único banco de dados relacional, as bases de dados produzidas e utilizadas pela Secretaria da Educação;
- Criar um mecanismo de mineração de dados (Datamining) a fim de promover uma normalização dos dados entre as adversidades das bases de dados;
- Projetar ambiente de Data Warehouse como parte de uma solução de BI (Business Intelligence) de acordo com a necessidade da Secretaria da Educação;
- Desenvolver documentação do banco de dados único bem como seu DER (Diagrama de Entidade Relacional);
- Criar métodos para consultas e produção de relatórios Ad-Hoc que possibilite análises e visões em várias dimensões;
- Elaborar oficinas de treinamento e capacitação dos técnicos da COAVE responsáveis pelas avaliações e indicadores a fim de que os mesmos possam utilizar e dar continuidade no processo supracitado.



Fontes, Bases de Dados e Integração



Principais ações

- Carga dos dados / Processos de ETL;
- Descrição das Bases;
- Produção do Data Warehouse;
- Integração dos dados de sujeitos, instituições e processos;
- Apropriação e uso de ferramentas de visualização de dados;
- Formação de técnicos e disseminação das bases;
- Produção da base de dados das EEEP.



Carga dos dados / Processos de ETL

- Diversidade de fontes, meios e estruturas;
- Volume de dados e capacidade de armazenamento;
- Recursos: Pentaho Data Integration (PDI) e PostgreSQL;
- Técnicas: Input Text (.csv) e Output Table.



Carga dos dados / Processos de ETL

Diversidade de fontes e meios

INEP/Censo Escolar

- Inputs Oracle (.dat)

INEP/Sige

- Cópia do Banco

INEP/Saeb, INEP/Enem e UFJF/CAEd/Spaece

- Inputs de texto puro (.csv)

IBGE/Pnad

- Inputs SPSS e SAS (.txt)

INEP/Ideb

- Planilhas Eletrônicas (.xls, ods, xlsx etc.)



Carga dos dados / Processos de ETL

Tabelas

SCHEMA	TABELA	REGISTROS	TAMANHO PURO (BYTES)	TAMANHO FORMATADO
ana	tb_aluno_2014	2.708.375	1.042.440.192	994 MB
ana	tb_aluno_2016	2.707.309	905.207.808	863 MB
censo_esc_ce	tb_curso	2.618	344.064	336 kB
censo_esc_ce	tb_curso_educ_profissional	2.064	204.800	200 kB
censo_esc_ce	tb_docente	4.670.174	2.044.395.520	1.950 MB
censo_esc_ce	tb_escola	171.108	78.487.552	75 MB
censo_esc_ce	tb_etapa_ensino	495	122.880	120 kB
censo_esc_ce	tb_ies	52.752	4.562.944	4.456 kB
censo_esc_ce	tb_lingua_indigena	2.811	335.872	328 kB
censo_esc_ce	tb_matricula	28.691.672	9.025.036.288	8.607 MB
censo_esc_ce	tb_orgao_regional	6.960	524.288	512 kB
censo_esc_ce	tb_pais_origem	2.389	147.456	144 kB
censo_esc_ce	tb_pessoa_fisica	26.418.192	6.127.181.824	5.843 MB
censo_esc_ce	tb_situacao	26.409.492	5.090.385.920	4.855 MB
censo_esc_ce	tb_tipo_atividade	636	114.688	112 kB
censo_esc_ce	tb_turma	1.253.670	397.680.640	379 MB
censo_esc_ce	tb_uf_mun_dist	10.307	1.466.368	1.432 kB



Carga dos dados / Processos de ETL

Tabelas

SCHEMA	TABELA	REGISTROS	TAMANHO PURO (BYTES)	TAMANHO FORMATADO
ideb	tb_ideb_af_bra	0	8.192	8.192 bytes
ideb	tb_ideb_af_esc	43.607	21.135.360	20 MB
ideb	tb_ideb_af_mun	14.356	8.069.120	7.880 kB
ideb	tb_ideb_af_reg_uf	128	81.920	80 kB
ideb	tb_ideb_ai_bra	5	8.192	8.192 bytes
ideb	tb_ideb_ai_esc	59.922	28.278.784	27 MB
ideb	tb_ideb_ai_mun	14.436	8.642.560	8.440 kB
ideb	tb_ideb_ai_reg_uf	128	90.112	88 kB
ideb	tb_ideb_em_bra	0	8.192	8.192 bytes
ideb	tb_ideb_em_esc	19.624	3.416.064	3.336 kB
ideb	tb_ideb_em_reg_uf	96	65.536	64 kB
pessoa	tb_pessoa	5.642.591	249.864.192	238 MB
pessoa	tb_pessoa_censo	5.642.591	249.864.192	238 MB
pessoa	tb_pessoa_censo_info	5.642.591	757.522.432	722 MB
pessoa	tb_pessoa_info	5.493.460	737.935.360	704 MB
pessoa	tb_pessoa_spaece	2.457.524	128.229.376	122 MB



Carga dos dados / Processos de ETL

Tabelas

SCHEMA	TABELA	REGISTROS	TAMANHO PURO (BYTES)	TAMANHO FORMATADO
pnad	tb_pnad_2008	391.868	1.070.063.616	1.020 MB
pnad	tb_pnad_2009	399.387	545.300.480	520 MB
pnad	tb_pnad_2011	358.919	420.044.800	401 MB
pnad	tb_pnad_2012	201.614	424.173.568	405 MB
pnad	tb_pnad_2013	362.556	495.009.792	472 MB
pnad	tb_pnad_2014	362.627	594.132.992	567 MB
pnad	tb_pnad_2015	170.160	584.753.152	558 MB
pnad	tb_pnad_2016_educ	570.653	519.421.952	495 MB
pnad	tb_pnad_2017_educ	568.313	517.292.032	493 MB
enem	tb_enem_restrito_2014	114.412	20.537.344	20 MB
enem	tb_enem_restrito_2015	105.340	19.046.400	18 MB
enem	tb_enem_restrito_2016	116.278	20.914.176	20 MB
enem	tb_enem_2009	4.148.720	4.855.193.600	4.630 MB
enem	tb_enem_2010	4.626.092	5.413.855.232	5.163 MB
enem	tb_enem_2011	5.380.855	7.346.667.520	7.006 MB
enem	tb_enem_2012	5.791.674	3.189.202.944	3.041 MB
enem	tb_enem_2013	7.173.558	5.333.630.976	5.087 MB
enem	tb_enem_2014	8.722.502	6.402.088.960	6.106 MB
enem	tb_enem_2015	7.745.556	5.804.138.496	5.535 MB
enem	tb_enem_2016	8.626.360	6.203.572.224	5.916 MB



Carga dos dados / Processos de ETL

Tabelas

SCHEMA	TABELA	REGISTROS	TAMANHO PURO (BYTES)	TAMANHO FORMATADO
saeb	tb_aluno_2007	4.109.283	909.819.904	868 MB
saeb	tb_aluno_2009	5.931.406	1.001.529.344	955 MB
saeb	tb_aluno_2011_saeb	5.398.536	1.002.020.864	956 MB
saeb	tb_aluno_2015_3em	114.225	42.450.944	40 MB
saeb	tb_aluno_2015_5ef	2.497.696	1.030.332.416	983 MB
saeb	tb_aluno_2015_9ef	2.418.620	992.477.184	947 MB
saeb	tb_aluno_2015_escola	57.744	22.953.984	22 MB
saeb	tb_aluno_mat_2005_3ano	22.254	8.683.520	8.480 kB
saeb	tb_aluno_mat_2005_4s	41.782	16.302.080	16 MB
saeb	tb_aluno_mat_2005_8s	33.188	12.951.552	12 MB
saeb	tb_aluno_port_2005_3ano	22.284	8.699.904	8.496 kB
saeb	tb_aluno_port_2005_4s	42.145	16.441.344	16 MB
saeb	tb_aluno_port_2005_8s	33.163	12.943.360	12 MB
saeb	tb_escola_2005	4.850	851.968	832 kB
saeb	tb_escola_2007	64.755	4.423.680	4.320 kB
saeb	tb_escola_2009	75.603	5.251.072	5.128 kB
saeb	tb_escola_2011_saeb	72.808	7.684.096	7.504 kB
saeb	tb_municipio_2007	27.022	1.851.392	1.808 kB
saeb	tb_municipio_2009	37.078	2.547.712	2.488 kB
saeb	tb_municipio_2011_saeb	60.608	7.225.344	7.056 kB



Carga dos dados / Processos de ETL

Tabelas

SCHEMA	TABELA	REGISTROS	TAMANHO PURO (BYTES)	TAMANHO FORMATADO
spaece	tb_spaece_2007_lp_alfa	118.587	25.673.728	24 MB
spaece	tb_spaece_2009_lp_alfa_ef	130.836	32.407.552	31 MB
spaece	tb_spaece_2009_lp_em	416.115	96.346.112	92 MB
spaece	tb_spaece_2009_mt_em	416.115	96.354.304	92 MB
spaece	tb_spaece_2010_lp_ef	9.477	4.800.512	4.688 kB
spaece	tb_spaece_2010_lp_em	814.955	584.507.392	557 MB
spaece	tb_spaece_2010_mt_em	644.716	479.764.480	458 MB
spaece	tb_spaece_2011_lp_ef	385.633	315.432.960	301 MB
spaece	tb_spaece_2011_lp_em	351.942	315.670.528	301 MB
spaece	tb_spaece_2011_mt_ef	266.882	237.682.688	227 MB
spaece	tb_spaece_2011_mt_em	351.893	313.917.440	299 MB
spaece	tb_spaece_2012_ch_em	106.483	112.631.808	107 MB
spaece	tb_spaece_2012_cn_em	106.483	112.959.488	108 MB
spaece	tb_spaece_2012_lp_alfa	111.990	88.612.864	85 MB
spaece	tb_spaece_2012_lp_ef	261.211	257.835.008	246 MB
spaece	tb_spaece_2012_lp_em	374.234	343.498.752	328 MB
spaece	tb_spaece_2012_mt_ef	261.239	257.613.824	246 MB
spaece	tb_spaece_2012_mt_em	374.266	367.460.352	350 MB
spaece	tb_spaece_2013_lp_2f	100.993	103.415.808	99 MB
spaece	tb_spaece_2013_lp_5f	123.941	144.736.256	138 MB



Carga dos dados / Processos de ETL

Tabelas

SCHEMA	TABELA	REGISTROS	TAMANHO PURO (BYTES)	TAMANHO FORMATADO
spaece	tb_spaece_2013_lp_9f	11.017	12.902.400	12 MB
spaece	tb_spaece_2013_lp_eja	9.495	11.116.544	11 MB
spaece	tb_spaece_2013_lp_em	175.375	205.266.944	196 MB
spaece	tb_spaece_2013_mt_5f	123.961	143.917.056	137 MB
spaece	tb_spaece_2013_mt_9f	11.017	12.894.208	12 MB
spaece	tb_spaece_2013_mt_eja	9.486	11.108.352	11 MB
spaece	tb_spaece_2013_mt_em	175.359	205.225.984	196 MB
spaece	tb_spaece_2014_mt_ef	234.647	368.140.288	351 MB
spaece	tb_spaece_2014_mt_em	226.448	314.761.216	300 MB
spaece	tb_spaece_2014_pt_ef	337.209	487.464.960	465 MB
spaece	tb_spaece_2014_pt_em	226.420	313.802.752	299 MB
spaece	tb_spaece_2015_alfa	106.512	143.499.264	137 MB
spaece	tb_spaece_2015_lp_5f	114.232	183.623.680	175 MB
spaece	tb_spaece_2015_lp_9f	109.099	171.180.032	163 MB
spaece	tb_spaece_2015_lp_eja	9.816	11.051.008	11 MB
spaece	tb_spaece_2015_lp_em	154.658	220.930.048	211 MB
spaece	tb_spaece_2015_mt_5f	114.240	183.656.448	175 MB
spaece	tb_spaece_2015_mt_9f	109.109	171.737.088	164 MB
spaece	tb_spaece_2015_mt_eja	9.815	11.223.040	11 MB
spaece	tb_spaece_2015_mt_em	154.661	220.553.216	210 MB



Carga dos dados / Processos de ETL

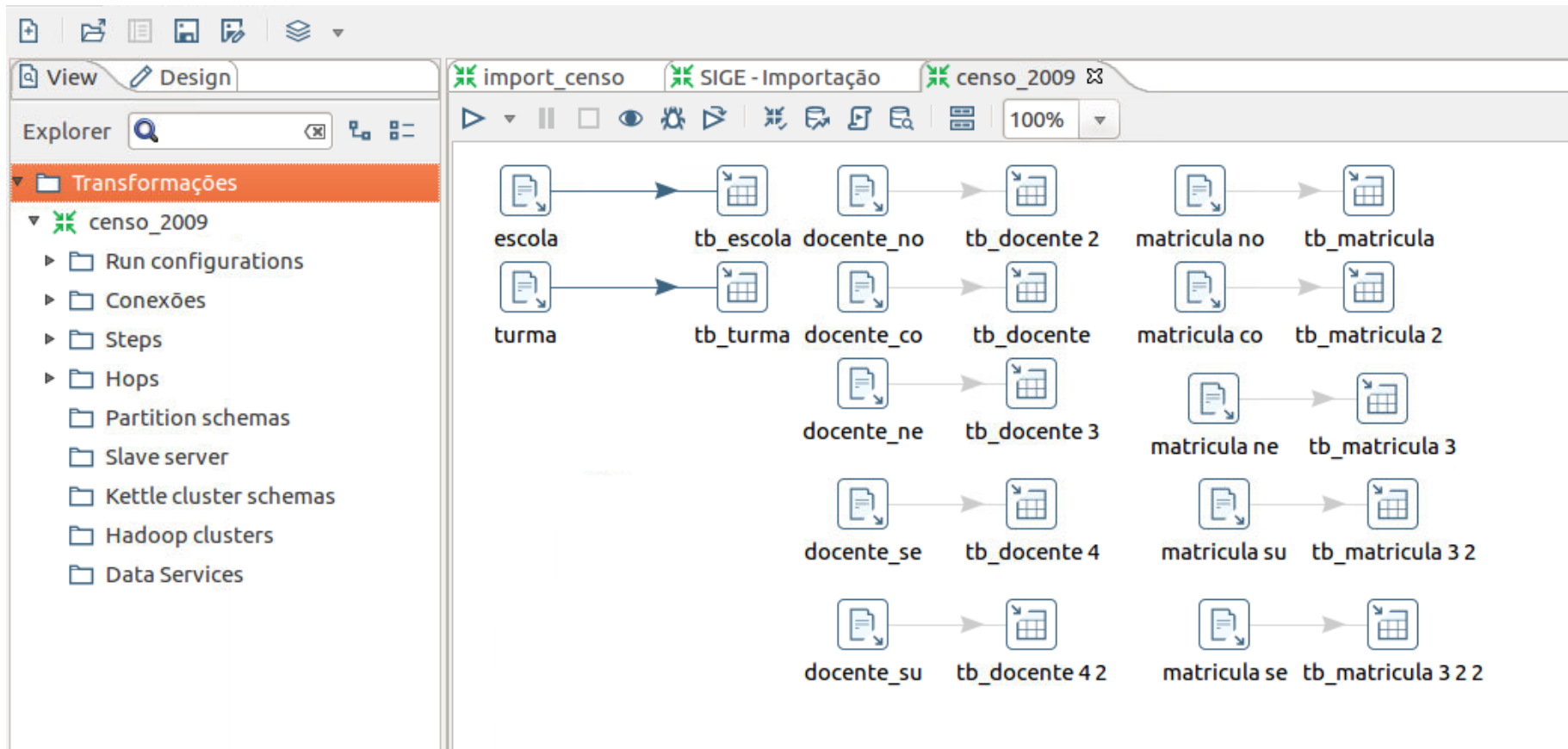
Tabelas

SCHEMA	TABELA	REGISTROS	TAMANHO PURO (BYTES)	TAMANHO FORMATADO
spaece	tb_spaece_2016_lp_2ef	98.723	158.187.520	151 MB
spaece	tb_spaece_2016_lp_5ef	111.340	180.101.120	172 MB
spaece	tb_spaece_2016_lp_9ef	111.566	176.136.192	168 MB
spaece	tb_spaece_2016_lp_eja	6.610	8.552.448	8.352 kB
spaece	tb_spaece_2016_lp_em	95.802	139.018.240	133 MB
spaece	tb_spaece_2016_mt_5ef	111.340	180.117.504	172 MB
spaece	tb_spaece_2016_mt_9ef	111.566	176.185.344	168 MB
spaece	tb_spaece_2016_mt_eja	6.610	8.601.600	8.400 kB
spaece	tb_spaece_2016_mt_em	95.804	139.575.296	133 MB
spaece	tb_spaece_2017_ef_lp	311.659	395.165.696	377 MB
spaece	tb_spaece_2017_ef_mt	217.025	285.720.576	272 MB
spaece	tb_spaece_2017_em_lp	114.937	157.794.304	150 MB
spaece	tb_spaece_2017_em_mt	114.937	157.728.768	150 MB



Carga dos dados / Processos de ETL

Pentaho Data Integration (PDI)



Produção do Data Warehouse

- Apropriação dos conceitos de Business Intelligence (BI) como “dimensão”, “fato”, “granulidade” e “cubo de dados”;
- Estudo das bases, suas variáveis e possibilidades de relacionamento;
- Aprimoramento da linguagem SQL;
- Recursos: PostgreSQL;
- Técnicas: Criação de tabelas, chaves e relacionamentos.



Integração dos dados de sujeitos, instituições e processos

- Integração dos dados de sujeitos (Tabela de Pessoas);
- Criação das dimensões de municípios, escolas, turmas, etapas, cursos e disciplinas;
- Recursos: PostgreSQL e Linguagem R;
- Técnicas: Pareamento determinístico de registros e deduplicação.



Apropriação e uso de ferramentas para visualização de dados

- Pentaho Saiku;
- Power BI;
- Google DataStudio.



Formação de técnicos e disseminação das bases

- Materiais descritivos e formativos;
- Oficina de utilização e produção de dados;
- Recursos: Google Classroom e Google Docs;
- Técnicas: Reuniões e oficinas.



Produção da Base de Dados das EEEP

- Correções de dados no Censo Escolar acerca das escolas, turmas e cursos nos primeiros anos de implantação;
- Produção de um Data Warehouse das EEEP;
- Integração dos dados de identificação de todos os alunos que passaram pelas EEEP;
- Identificação dos alunos que participaram do projeto nas bases do Censo Escolar, SIGE, SPAECE e ENEM.



Data Warehouse das EEEP

Dimensoes

- Séries
- Cursos
- Áreas
- Turmas
- Escolas
- Localização
- Crede/Sefor
- Município

Fatos e Medidas

- Matrículas
- Contagem de matrículas
- Resultados em avaliações
- Médias de proficiência
- Notas escolares
- Médias das notas
- Frequência escolar
- Percentual de faltas



bi_ce.tb_censo_dm_escola		
nr_ano_censo	integer	
id_escola	integer	
nm_escola	text	
nm_escola_recente	text	
cd_dependencia	integer	
ds_dependencia	text	
id_distrito	integer	
id_municipio	integer	
cd_uf	integer	
cd_localizacao	integer	
ds_localizacao	text	
cd_localizacao_diferenciada	integer	
nr_latitude	text	
nr_longitude	text	
ds_localizacao_diferenciada	text	
ds_orgao_regional	text	
id_crede_sefor	integer	
nm_crede_sefor	text	
fl_indigena	integer	
ds_indigena	text	
cd_categoria_escola_privada	integer	
ds_categoria_escola_privada	text	
cd_categoria_escola_estadual	integer	
ds_categoria_escola_estadual	text	
cd_prioritaria_estadual	integer	
ds_prioritaria_estadual	text	

bi_ce.tb_censo_dm_turma		
nr_ano_censo	integer	
id_escola	integer	
id_turma	integer	
nm_turma	text	
tp_etapa_ensino_censo	integer	
no_etapa_ensino_censo	text	
cd_classe	integer	
ds_classe	text	
cd_mediacao	integer	
ds_mediacao	text	
cd_etapa_turma	integer	
ds_etapa_turma	text	
cd_etapa_fase_turma	integer	
ds_etapa_fase_turma	text	
cd_ano_serie_turma	integer	
ds_ano_serie_turma	text	
fl_regular	integer	
ds_regular	text	
fl_eja	integer	
ds_eja	text	
fl_profissional	integer	
ds_profissional	text	
cd_tipo_curso_profissional	integer	
ds_tipo_curso_profissional	text	
id_curso_profissional	integer	
nm_curso_profissional	text	
id_area_profissional	integer	
nm_area_profissional	text	
cd_unificacao	integer	
ds_unificacao	text	
cd_seriacao	integer	
ds_seriacao	text	
nr_duracao_turma	integer	
fl_integral	integer	
ds_integral	text	
cd_turno	integer	
ds_turno	text	
cd_tipo_turma	integer	
ds_tipo_turma	text	

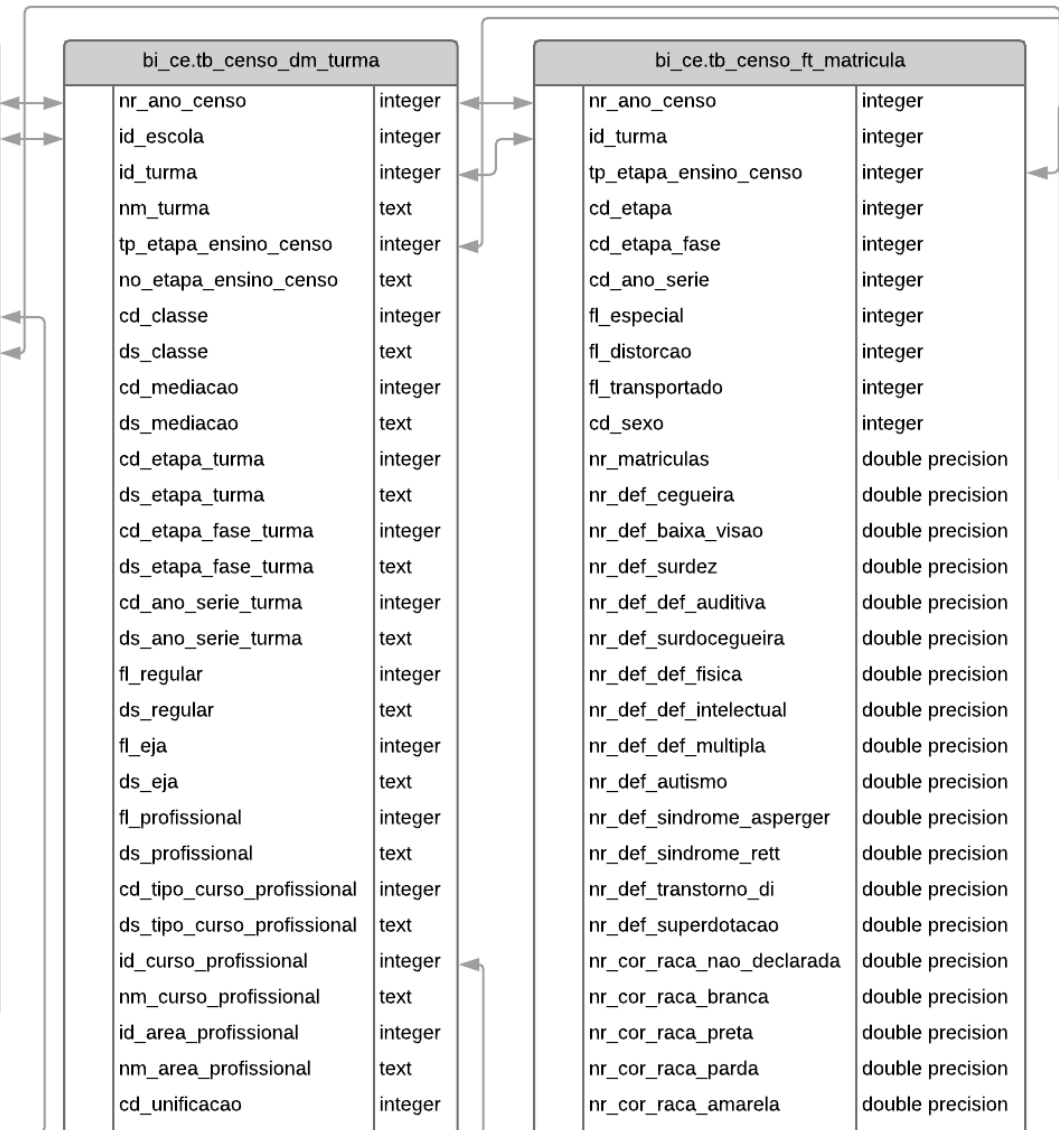
bi_ce.tb_censo_ft_matricula			
nr_ano_censo	integer		
id_turma	integer		
tp_etapa_ensino_censo	integer		
cd_etapa	integer		
cd_etapa_fase	integer		
cd_ano_serie	integer		
fl_especial	integer		
fl_distorcao	integer		
fl_transportado	integer		
cdsexo	integer		
nr_matriculas	double precision		
nr_def_cegueira	double precision		
nr_def_baixa_visao	double precision		
nr_def_surdez	double precision		
nr_def_def_auditiva	double precision		
nr_def_surdocegueira	double precision		
nr_def_def_fisica	double precision		
nr_def_def_intelectual	double precision		
nr_def_def_multipla	double precision		
nr_def_autismo	double precision		
nr_def_sindrome_asperger	double precision		
nr_def_sindrome_rett	double precision		
nr_def_transtorno_di	double precision		
nr_def_superdotacao	double precision		
nr_cor_raca_nao_declarada	double precision		
nr_cor_raca_branca	double precision		
nr_cor_raca_preta	double precision		
nr_cor_raca_parda	double precision		
nr_cor_raca_amarela	double precision		
nr_cor_raca_indigena	double precision		

bi_ce.tb_censo_dm_etapa		
tp_etapa_ensino_censo	integer	
no_etapa_ensino_censo	text	
cd_etapa	integer	
ds_etapa	text	
cd_etapa_fase	integer	
ds_etapa_fase	text	
cd_ano_serie	integer	
ds_ano_serie	text	

bi_ce.tb_censo_dm_municipio		
id_municipio	integer	
nm_municipio	text	
id_microrregiao	integer	
nm_microrregiao	text	
id_mesorregiao	integer	
nm_mesorregiao	text	
id_uf	integer	
nm_uf	text	
sg_uf	text	
id_regiao	integer	
nm_regiao	text	
fl_capital	integer	
ds_capital	text	

bi_ce.tb_censo_dm_curso_profissional		
nr_ano_censo	integer	
id_area_profissional	integer	
nm_area_profissional	text	
id_curso_profissional	integer	
nm_curso_profissional	text	

bi_ce.tb_censo_dm_distrito		
id_distrito	integer	
nm_distrito	text	
id_municipio	integer	



Data Warehouse das EEEP

Cubos

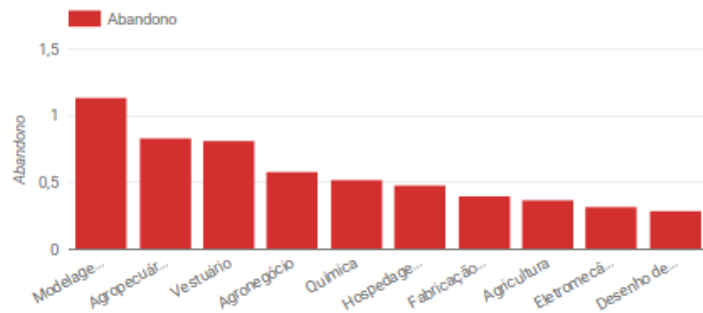
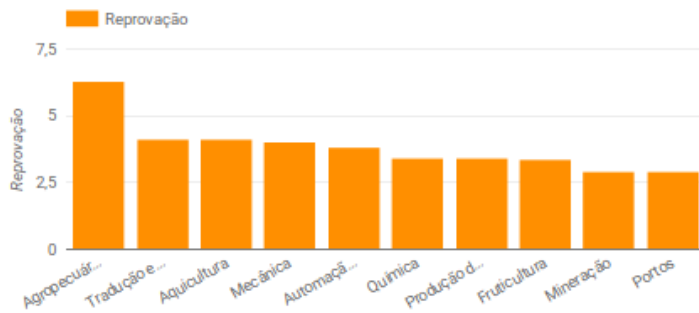
- Matrícula (Censo e Sige)
- Spaece
- Enem
- Notas escolares (Sige)
- Frequência escolar (Sige)



Data Warehouse das EEEP

Ano	Aprovados	Aprovação (%)	Reprovados	Reprovação (%)	Abandono	Abandono (%)
2008	3.876	97,07	80	2	37	0,93
2009	12.269	95,26	549	4,26	62	0,48
2010	16.725	96,33	565	3,25	73	0,42
2011	22.715	96,32	763	3,24	106	0,45
2012	28.417	97,25	671	2,3	134	0,46
2013	34.365	97,44	799	2,27	103	0,29
2014	39.130	97,63	813	2,03	138	0,34
2015	42.228	97,9	799	1,85	109	0,25
2017	48.229	98,62	599	1,22	75	0,15
2016	46.019	97,63	1.013	2,15	106	0,22

1 - 10 / 10 < >

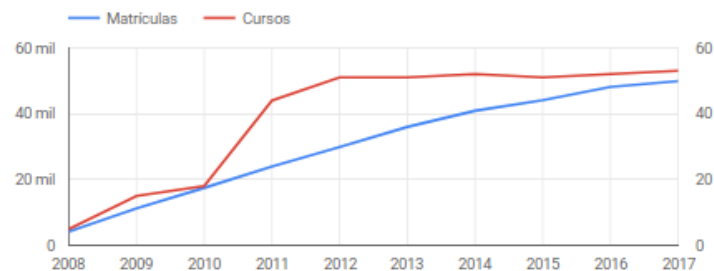
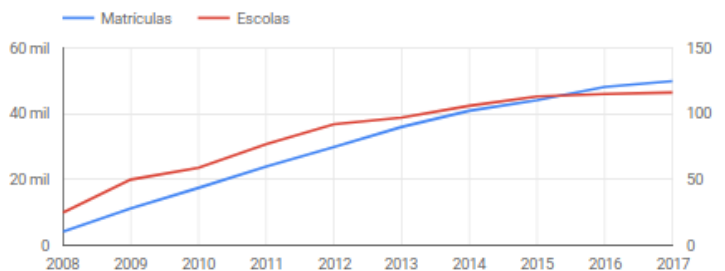
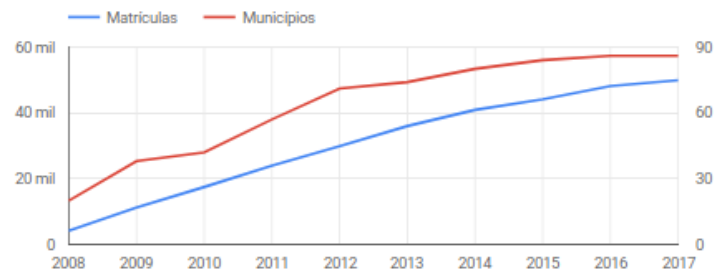
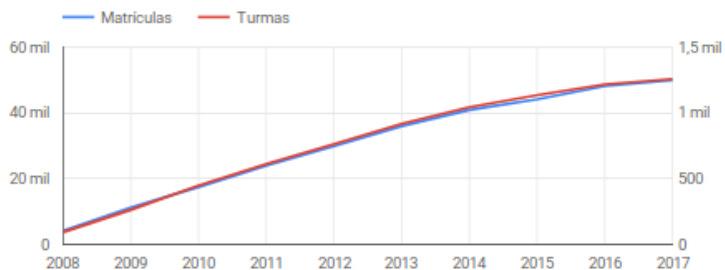
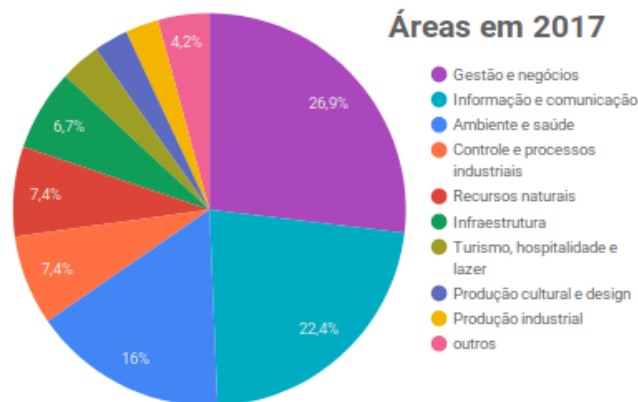


Data Warehouse das EEEP

CREDE/SEFOR ▾ Município ▾ Localização ▾ Série ▾ Área ▾ Curso ▾

Ano	Matrículas	Turmas	Cursos	Escolas	Municípios
2008	4.181	94	5	25	20
2009	11.229	264	15	50	38
2010	17.481	447	18	59	42
2011	23.916	612	44	77	57
2012	29.885	762	51	92	71
2013	35.981	918	51	97	74
2014	40.897	1.043	52	106	80
2015	44.069	1.134	51	113	84
2016	48.089	1.215	52	115	86
2017	49.894	1.257	53	116	86

1 - 10 / 10 < >



Data Warehouse das EEEP

Ano	Disciplina / Padrão / Escolas						Total geral
	Língua Portuguesa			Matemática			
	Intermediário	Crítico	Adequado	Intermediário	Crítico	Adequado	
2017	107	1	5	62	46	5	113
2016	97	6	3	53	50	3	106
2014	86	5	1	59	31	2	92
2013	70	6	1	53	19	5	77
2012	51	8	-	33	25	1	59



Coordenadoria de Avaliação e Acompanhamento da Educação - COAVE

Luciano Nery
Coordenador

Célula de Gestão de Dados e Avaliação – CEGDA

Anderson Araújo
Orientador

Eixo de Indicadores

George Gomes
Gerente

Francisco Silveira
Assistente Técnico

Philippe Azevedo
Técnico





**GOVERNO DO
ESTADO DO CEARÁ**
Secretaria da Educação