



WSCAD 2022

Florianópolis / SC

XXIII Simpósio em
Sistemas Computacionais
de Alto Desempenho

Processamento e Análise de *Big Data* com HPCC Systems

Objetivo: Ao longo desse minicurso serão apresentados aos participantes os conceitos essenciais de processamento e análise de dados em quantidades massivas (*Big Data*) fazendo uso da plataforma “*open source*” HPCC Systems, possibilitando a aplicação dos conhecimentos adquiridos por meio de um ambiente de treinamento disponibilizado durante o evento.

Informações Técnicas: Curso de **nível básico**. O curso requer apenas um computador com acesso à Internet e uma conta no [Github](#).

Autores



- **Mauro D. Marques** – Engenheiro de SW na LexisNexis Risk Solutions
Engenheiro com pós-graduação nas áreas de TI e Educação
35 anos de atuação como engenheiro no setor automobilístico
11 anos de atuação como professor universitário nas áreas de Engenharia e Ciência da Computação
Mauro.Marques@lexisnexisrisk.com



- **Alysson R. Oliveira** – Engenheiro de SW na LexisNexis Risk Solutions
Engenheiro de Computação
03 anos como engenheiro de software e mentor de projetos acadêmicos
05 anos como instrutor de cursos técnicos na área de Engenharia de Computação
Alysson.Oliveira@lexisnexisrisk.com

Minicurso: Agenda

➤ **LexisNexis Risk Solutions: A Empresa**

- Quem somos nós?
- A nossa tecnologia: A evolução da plataforma HPCC Systems...

➤ **HPCC Systems: Visão Geral**

- Apresentação de conceitos;
- Aplicação de conhecimentos.

➤ **Próximos passos**

- Cursos online;
- Projetos de pesquisa;
- Oportunidades profissionais.

➤ **Considerações Finais**

Minicurso: Agenda

➤ **LexisNexis Risk Solutions: A Empresa**

- Quem somos nós?
- A nossa tecnologia: A evolução da plataforma HPCC Systems...

➤ **HPCC Systems: Visão Geral**

- Apresentação de conceitos;
- Aplicação de conhecimentos.

➤ **Próximos passos**

- Cursos online;
- Projetos de pesquisa;
- Oportunidades profissionais.

➤ **Considerações Finais**

LexisNexis Risk Solutions: A Empresa

- **Quem somos nós?**

A **LexisNexis Risk Solutions** é líder no fornecimento de informações essenciais que ajudam clientes de diversos setores e governos na avaliação, prevenção e gestão de riscos.

Fazemos parte do **LexisNexis Risk Solutions Group**, um portfólio de marcas que abrange vários setores que fornecem aos clientes tecnologias inovadoras, análises baseadas em informações e ferramentas de decisão e serviços de dados.

LexisNexis Risk Solutions Group



Saiba mais em: <https://risk.lexisnexis.com/group/our-brands>



RELX é um provedor global de análises baseadas em informações e ferramentas de decisão para clientes profissionais e empresariais. O Grupo atende clientes em mais de 180 países e possui escritórios em cerca de 40 países.

Saiba mais em www.relx.com

Científico



Eventos



Análise de risco



Legal



Minicurso: Agenda

➤ **LexisNexis Risk Solutions: A Empresa**

- ✓ Quem somos nós?

- A nossa tecnologia: A evolução da plataforma HPCC Systems...

➤ **HPCC Systems: Visão Geral**

- Apresentação de conceitos;
- Aplicação de conhecimentos.

➤ **Próximos passos**

- Cursos online;
- Projetos de pesquisa;
- Oportunidades profissionais.

➤ **Considerações Finais**

LexisNexis Risk Solutions: A Empresa

▪ A nossa tecnologia: A evolução da plataforma HPCC Systems...

2001



Primeira versão
da plataforma
HPCC é lançada

2011



Código aberto (licença
Apache e código no
GitHub)

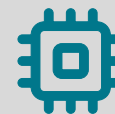
2012 – 16



Melhorias contínuas
com **FOCO NA
QUALIDADE**

Suporte e treinamento
aprimorado

2017-Presente



Aprimoramentos de
Arquitetura (*Cloud*)

Desenvolvimentos em
Machine Learning

Minicurso: Agenda

✓ **LexisNexis Risk Solutions: A Empresa**

✓ Quem somos nós?

✓ A nossa tecnologia: A evolução da plataforma HPCC Systems...

➤ **HPCC Systems: Visão Geral**

- Apresentação de conceitos;
- Aplicação de conhecimentos.

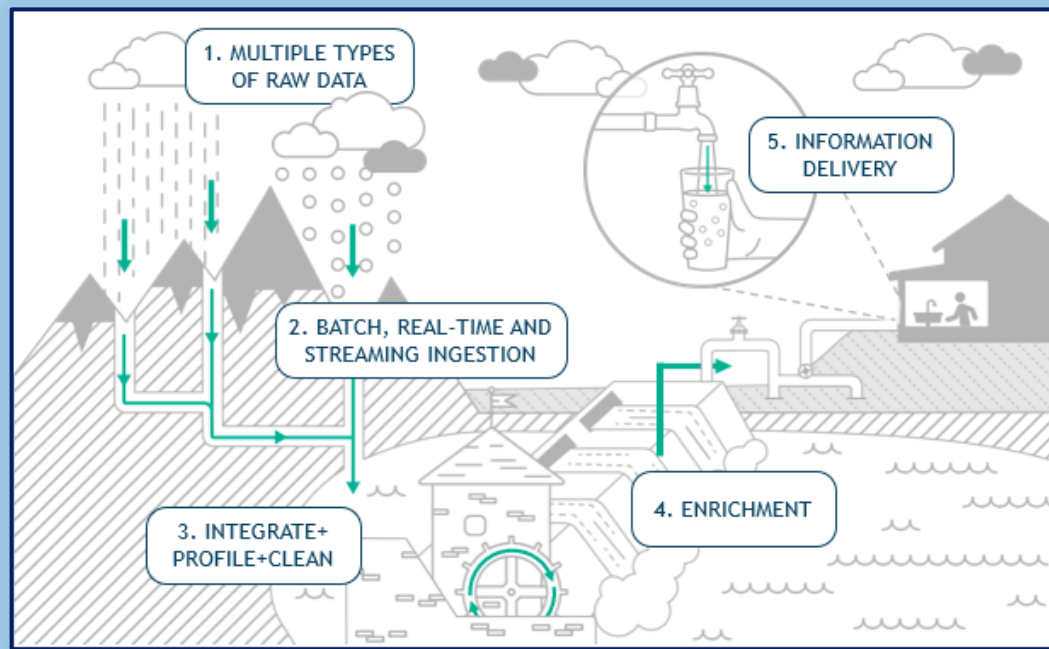
➤ **Próximos passos**

- Cursos online;
- Projetos de pesquisa;
- Oportunidades profissionais.

➤ **Considerações Finais**

HPCC Systems: Visão Geral

■ Apresentação de conceitos

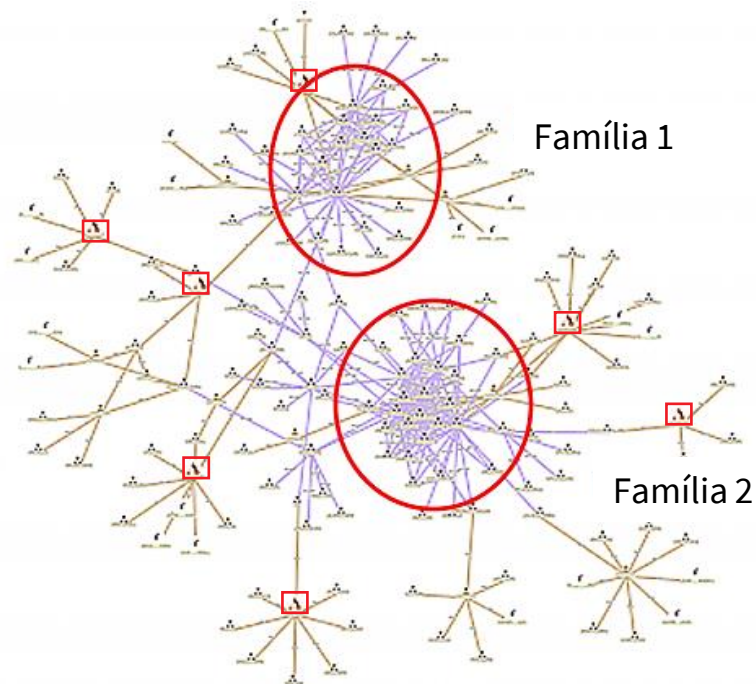
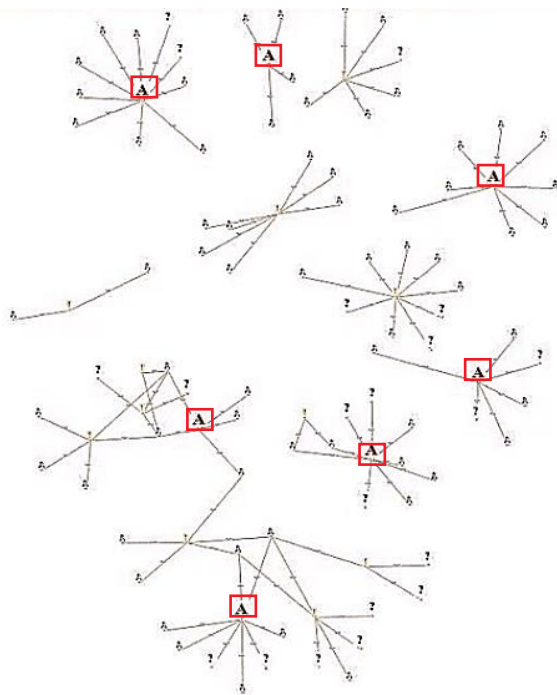




HPCC Systems

Processamento e Análise de *Big Data*

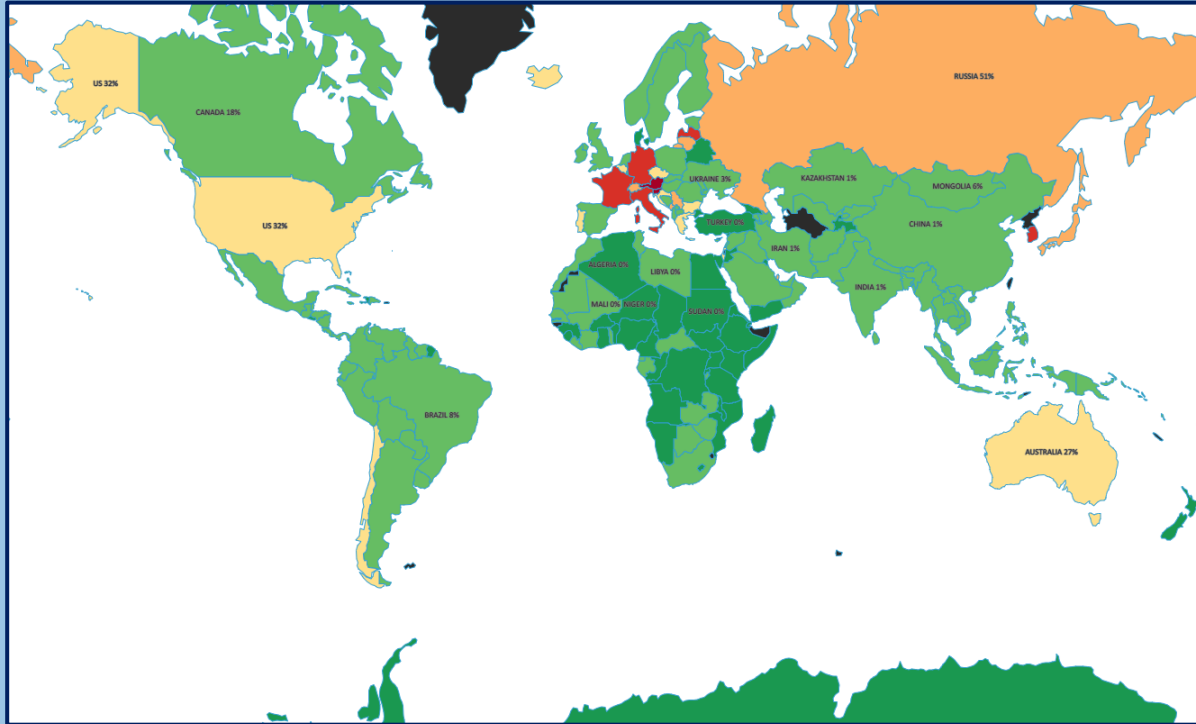


Exemplo de uso: “Detecção de Fraude”



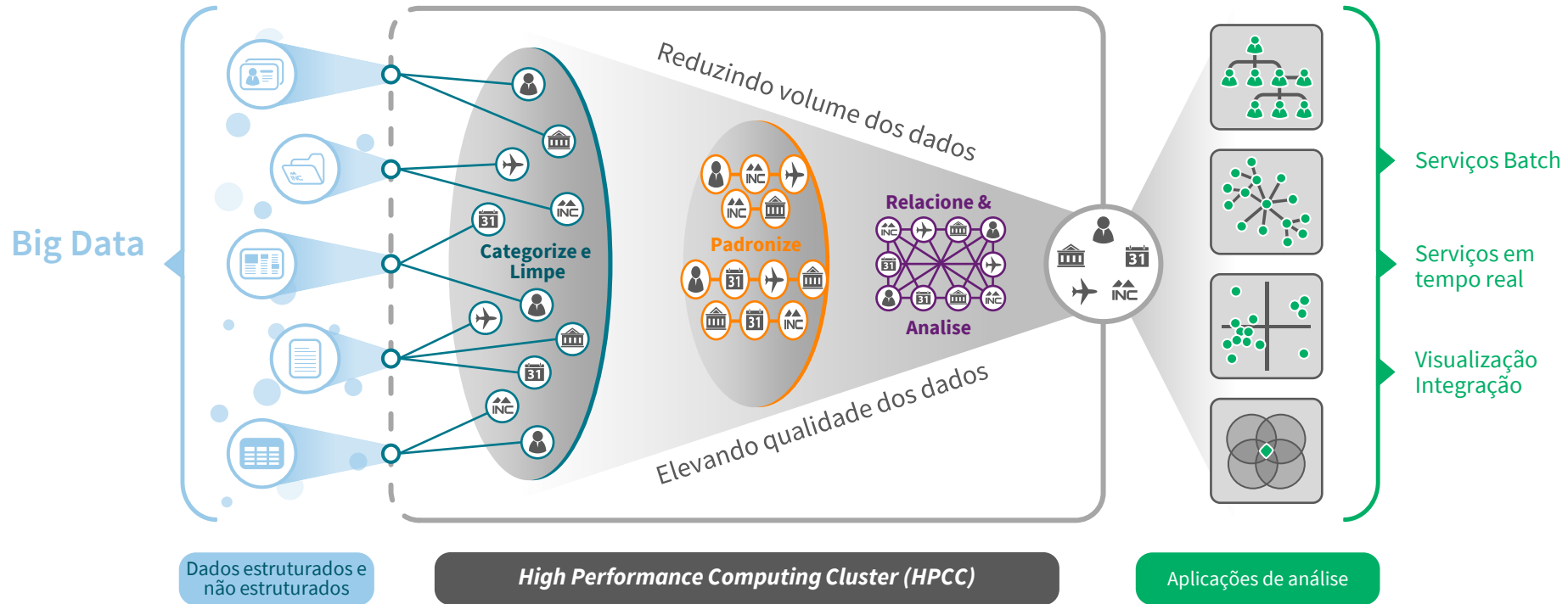
-  Acidentes com suspeitas de fraude
-  Pessoas associadas aos acidentes

Exemplo de uso: Uma situação mais atual...



Saiba mais em: <https://covid19.hpccsystems.com/>

Fluxo de Dados no HPCC Systems: 'Funil' de Dados



“Stack” tecnológico da plataforma HPCC Systems



Cluster Thor

ETL: Extração, Transformação e Carregamento de dados



Cluster ROXIE

Entrega online de consultas em *Big Data*



Ferramentas para manipulação de dados

Perfilamento, limpeza, consolidação de dados



Bibliotecas de *Machine Learning*

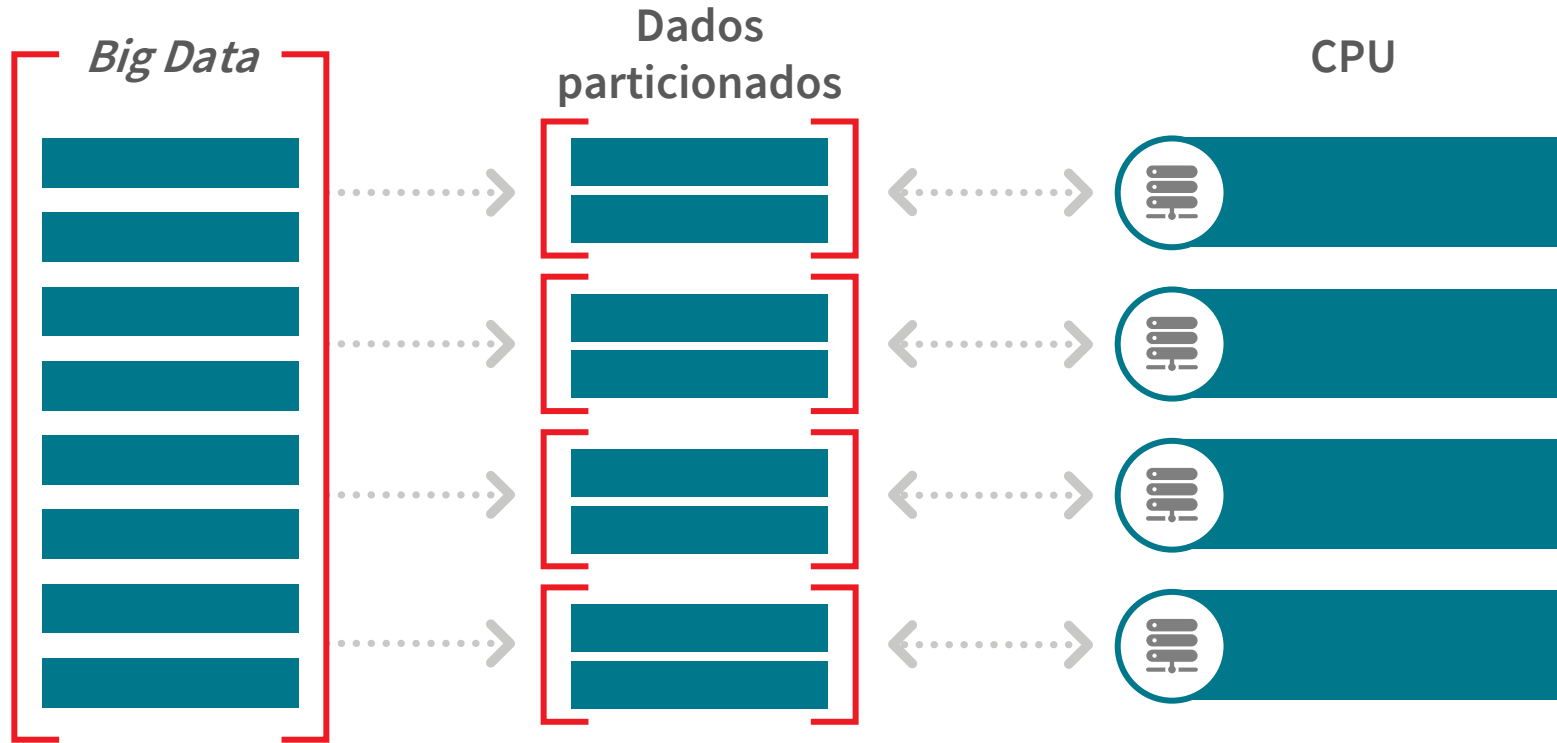
Supervisionado, não-supervisionado, aprendizagem profunda



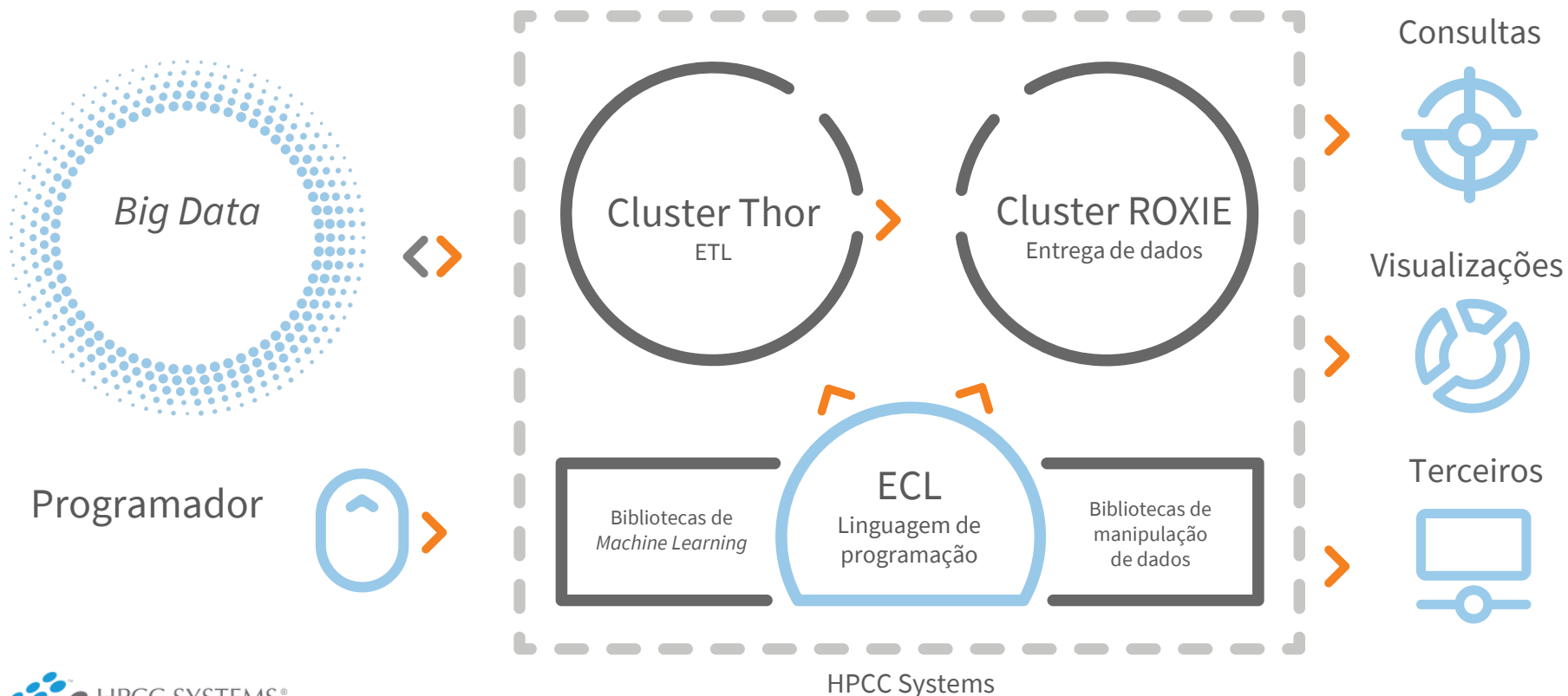
Conectividade

Plugins de
integração com
outros sistemas

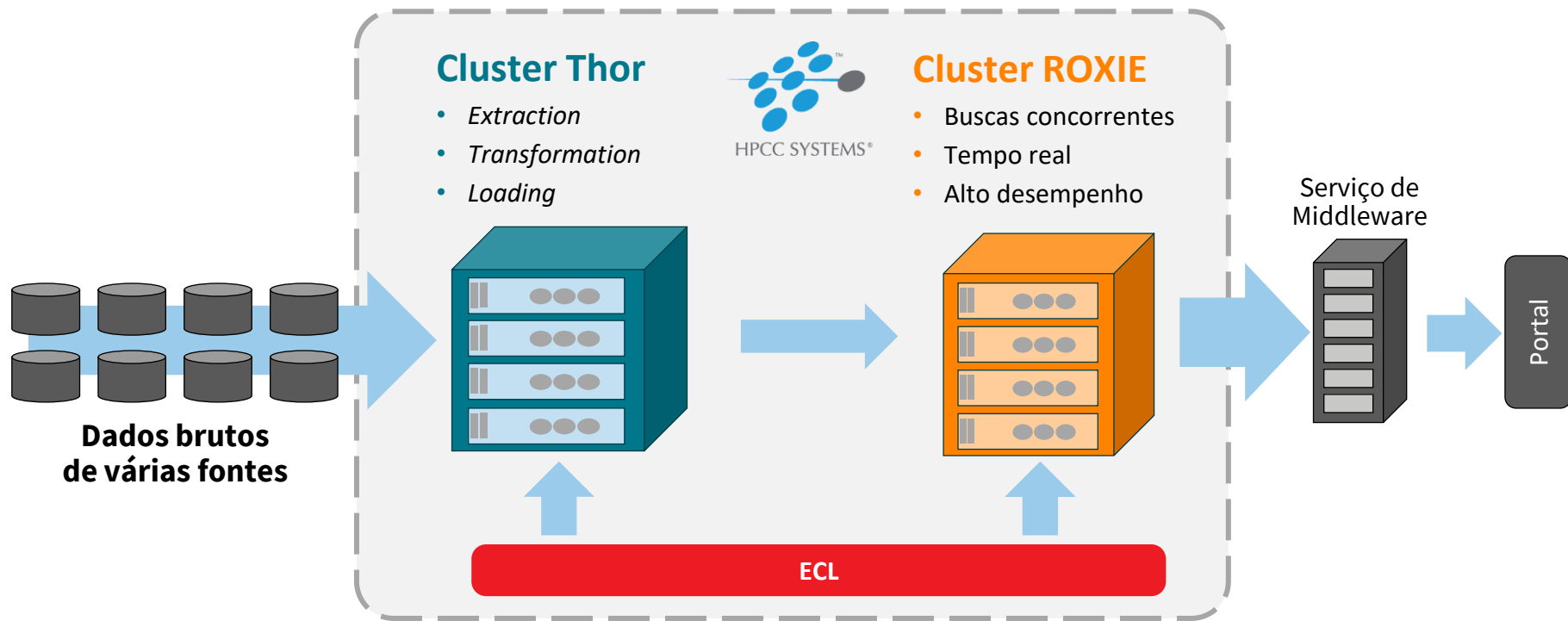
Dados Distribuídos e Processamento Paralelo



Arquitetura da plataforma HPCC Systems



O Power Trio da plataforma: Thor, ROXIE e ECL



Enterprise Control Language (ECL)

Linguagem de programação centrada em dados (*Data Flow*)

- Declarativa e não-procedural
- Códigos menores e reutilizáveis
- Biblioteca para manipulação de dados

Compilador

- Gera código otimizado (C++)
- Lógica para processamento paralelo e distribuído

Como fazer



vs.



O que fazer

HPCC Systems:

Ativos e Clientes

- 12 petabytes de dados públicos e privados
- 270 milhões de transações por hora
- Clientes em mais de **100** países
- **76%** de todas as empresas Fortune 500
- **7** dos 10 maiores bancos do mundo
- **100%** dos 50 maiores bancos americanos
- **95 das 100** maiores seguradoras
- **Mais de 7.500** órgãos governamentais: locais, estaduais e federais

Unidade	Símbolo	Número de Bytes
Kilobyte	KB	$2^{10} = 1024$ bytes
Megabyte	MB	$2^{20} = 1,048,576$ bytes
Gigabyte	GB	$2^{30} = 1,073,741,824$ bytes
Terabyte	TB	$2^{40} = 1,099,511,627,776$ bytes
Petabyte	PB	$2^{50} = 1,125,899,906,842,624$ bytes
Exabyte	EB	$2^{60} = 1,152,921,504,606,846,976$ bytes
Zettabyte	ZB	$2^{70} = 1,180,591,620,717,411,303,424$ bytes
Yottabyte	YB	$2^{80} = 1,208,925,819,614,629,174,706,176$ bytes

Minicurso: Agenda

✓ **LexisNexis Risk Solutions: A Empresa**

- ✓ Quem somos nós?
- ✓ A nossa tecnologia: A evolução da plataforma HPCC Systems...

➤ **HPCC Systems: Visão Geral**

- ✓ Apresentação de conceitos;
 - Aplicação de conhecimentos.

➤ **Próximos passos**

- Cursos online;
- Projetos de pesquisa;
- Oportunidades profissionais.

➤ **Considerações Finais**

HPCC Systems: Visão Geral

■ Aplicação de conhecimentos

The screenshot displays the ECL Watch application interface. The main window is titled "hthor" and shows a "Produto: Serviço de Consulta de Dados Pessoais" (Product: Personal Data Query Service). The interface includes a "fetch_persons_mdm" function with a "Dynamic Form" dropdown. Below this, there is a "FETCH_PERSONS_MDMREQUEST" section with input fields for "firstname_value", "lastname_value", and "state_value". At the bottom, there are buttons for "Output Tables", "FORM POST", "Submit", and "Clear All". The left sidebar shows a list of tasks under "Tarefas" and "Área de Recreio", including a list of WUIDs. The bottom status bar indicates "1 - 50 de 1252 resultados" and shows various error and warning counts.

1. Configuração do Ambiente: *Github / Gitpod* sem “Fork”

1 Sign in to GitHub

Username or email address

Password [Forgot password?](#)

Sign in

New to GitHub? [Create an account.](#)

2 mauromarx

Wikis **3** 0

Users **1**

[Advanced search](#) [Cheat sheet](#)

4 1 user

Mauro D. Marques mauromarx

Software Engineer

São Paulo - Brazil

5 Popular repositories

WSCAD_2022

Public

ECL 1

6 During the workshop GitPod will be used as main environment:

1. By using your GitHub credentials, just click on the following link for instantiate a environment via GitPod:
<https://gitpod.io/#https://github.com/mauromarx/WSCAD>

7 Gitpod

Open a cloud-based developer environment for the repository mauromarx/WSCAD

Continue with GitHub

8 Authorize Gitpod

Gitpod by **Gitpod** wants to access your mauromarx account

Personal user data
Email addresses (read-only)

Cancel **Authorize gitpod-io**

Authorizing will redirect to <https://gitpod.io>

9 Select Editor

Choose the editor for opening workspaces. You can always change later the editor in [user preferences](#).

VS Code **VS Code** **IntelliJ IDEA** **GoLand**

BROWSER

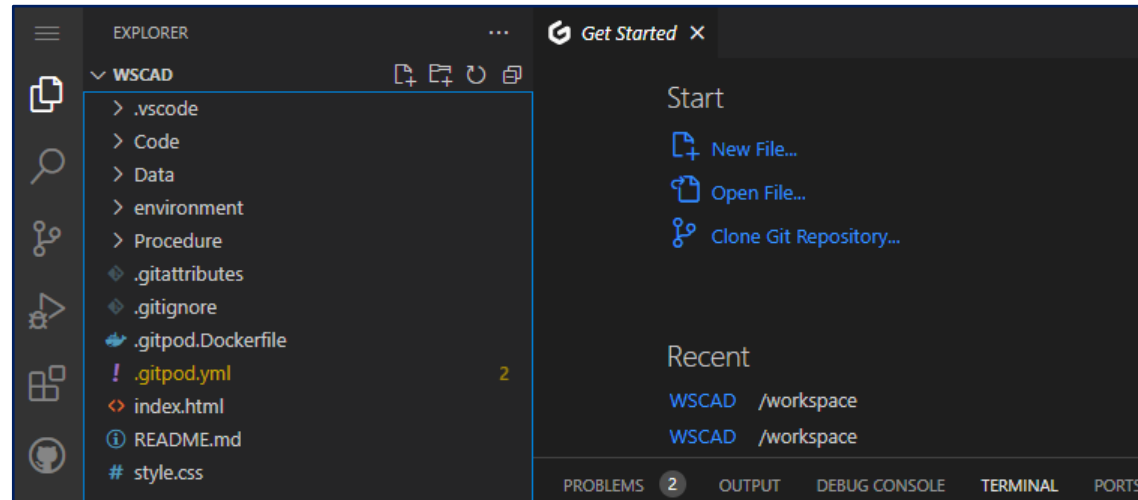
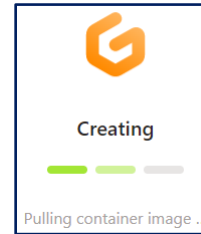
10 User Validation Required

Enter a mobile phone number you would like to use to verify your account. Having trouble? [Contact support](#)

Mobile Phone Number

Send Code via SMS

1. Configuração do Ambiente: *Github* / *Gitpod* sem “Fork”



1. Configuração do Ambiente: *Github* / *Gitpod* com “Fork”

The image is a collage of 11 numbered screenshots illustrating the process of setting up a development environment using GitHub and Gitpod.

- 1**: GitHub Sign in page. Fields for Username or email address and Password are visible, along with a 'Sign in' button.
- 2**: GitHub profile page for 'mauromarx'.
- 3**: GitHub navigation menu showing Wikis (0) and Users (1).
- 4**: GitHub user profile for 'Mauro D. Marques' (Software Engineer, São Paulo - Brazil).
- 5**: 'Popular repositories' section showing 'WSCAD_2022' by 'ECL'.
- 6**: 'Fork' button and 'Your existing forks' section (You don't have any forks of this repository).
- 7**: 'Create a new fork' dialog. Repository name is 'WSCAD'. A note states: 'By default, forks are named the same as their parent repository. You can customize the name to distinguish it further.'
- 8a**: 'Recent Repositories' list showing 'maurodmarques/WSCAD'.
- 8b**: 'Code' button and 'Sync fork' option.
- 9**: Gitpod.io URL: `gitpod.io/#https://github.com/maurodmarques/WSCAD`.
- 10**: Gitpod 'Creating' workspace screen with a progress bar and the text 'Pulling container image ...'.
- 11**: VS Code Explorer showing the file structure of the 'WSCAD' repository. The 'Start' panel on the right shows options: 'New File...', 'Open File...', and 'Clone Git Repository...'. The 'Recent' panel shows the workspace path: `/workspace`.

2. Enterprise Control Language (ECL)

■ Conceitos básicos de ECL:

- Paradigma declarativo (não-procedural)
- ECL não é sensível a caixa alta/baixa
- Espaço em branco é ignorado para melhor leitura
- Comentários em linha (//) e em bloco (/* e */)
- ECL utiliza sintaxe **objeto.propriedade**

Dataset.Campo

// referencia um campo em um dataset

NomedoDiretorio.Definicao

// referencia uma definição em outro módulo

2. Enterprise Control Language (ECL)

- **Conceitos básicos de ECL:**

- O código ECL é constituído de:

- ❖ **Definições** estabelecem o que as coisas são (arquivos de definição ECL)

mydef := 'People'; // não inicia uma **WU**

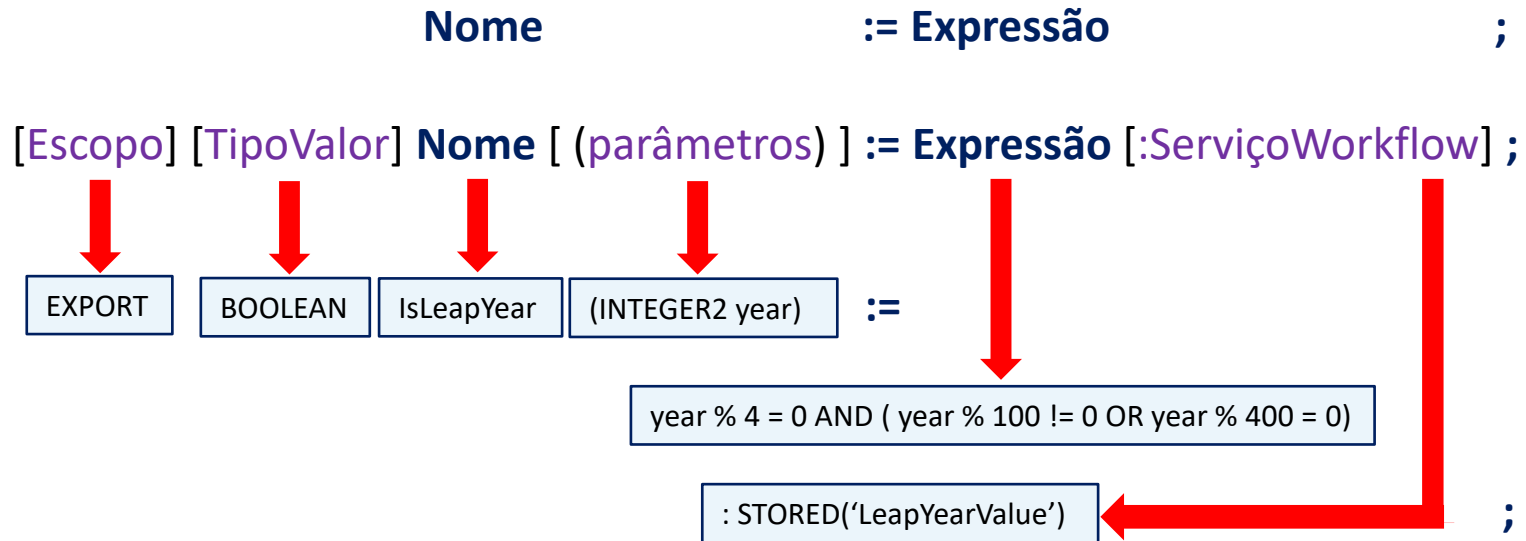
- ❖ **Ações** resultam em compilação e execução (arquivos **BWR** – *Builder Window Runnable*)

OUTPUT('People'); // inicia uma **WU**

OUTPUT(mydef); // inicia uma **WU**

2. Enterprise Control Language (ECL)

- Sintaxe completa de uma Definição ECL:



2. Enterprise Control Language (ECL)

- **“Inline” Datasets** utilizados durante a primeira parte da demonstração:

Firstname	Lastname	Gender	Age	Nationality	Occupation	Account	Balance	Income
Isaac	Newton	M	84	english	scientist	cc100	100	3500.00
Albert	Einstein	M	76	german	scientist	cc200	-100	4000.30
Marie	Curie	F	66	polish	scientist	cc300	200	3640.10
Victor	Hugo	M	83	french	writer	cc400	150	1900.00
Jane	Austen	F	41	english	writer	cc500	180	2000.00
Emily	Bronte	F	30	english	writer	cc600	120	1800.00
Jane	Doe		25	brazilian	unemployed	cc700	-500	0.00
John	Doe	U	65	american	retired	cc800	750	3211.11

Firstname	Lastname	Email	Phone
ISAAC	NEWTON	isaac.newton@cam.ac.uk	16431727
ALBERT	EINSTEIN	albert.einstein@princeton.edu	18791955
MARIE	CURIE	marie.curie@sorbonne.fr	18671934
VICTOR	HUGO	victor.hugo@lacroix.fr	18021885
JANE	AUSTEN	jane.austen@hampshire.uk	17751817
EMILY	BRONTE	emily.bronte@thornton.uk	18181848
JANE	DOE	jane.doe@hotmail.com	
JOHN	WAYNE	john.wayne@paramount.com	12345678

2. Enterprise Control Language (ECL)

The screenshot displays the HPCC Systems IDE interface. The Explorer panel on the left shows the project structure for WSCAD_2022, with the file BWR_Demo01.ecl selected. The main editor window shows the ECL code for BWR_Demo01.ecl, which defines a record structure and a table named RECORD. The bottom panel shows the results of a query, displaying a table with 8 rows of data.

```
1 //
2 // ***** Estruturas de dados basicas em ECL
3 //
4 // ***** SHOW 01
5 // ***** Estrutura RECORD
6 rec := RECORD
7     STRING10    Firstname;
8     STRING      Lastname;
9     STRING1     Gender;
10    UNSIGNED1    Age;
11    STRING       Nationality;
12    STRING       Occupation;
13    STRING       Account;
```

Result 1: [8 rows]

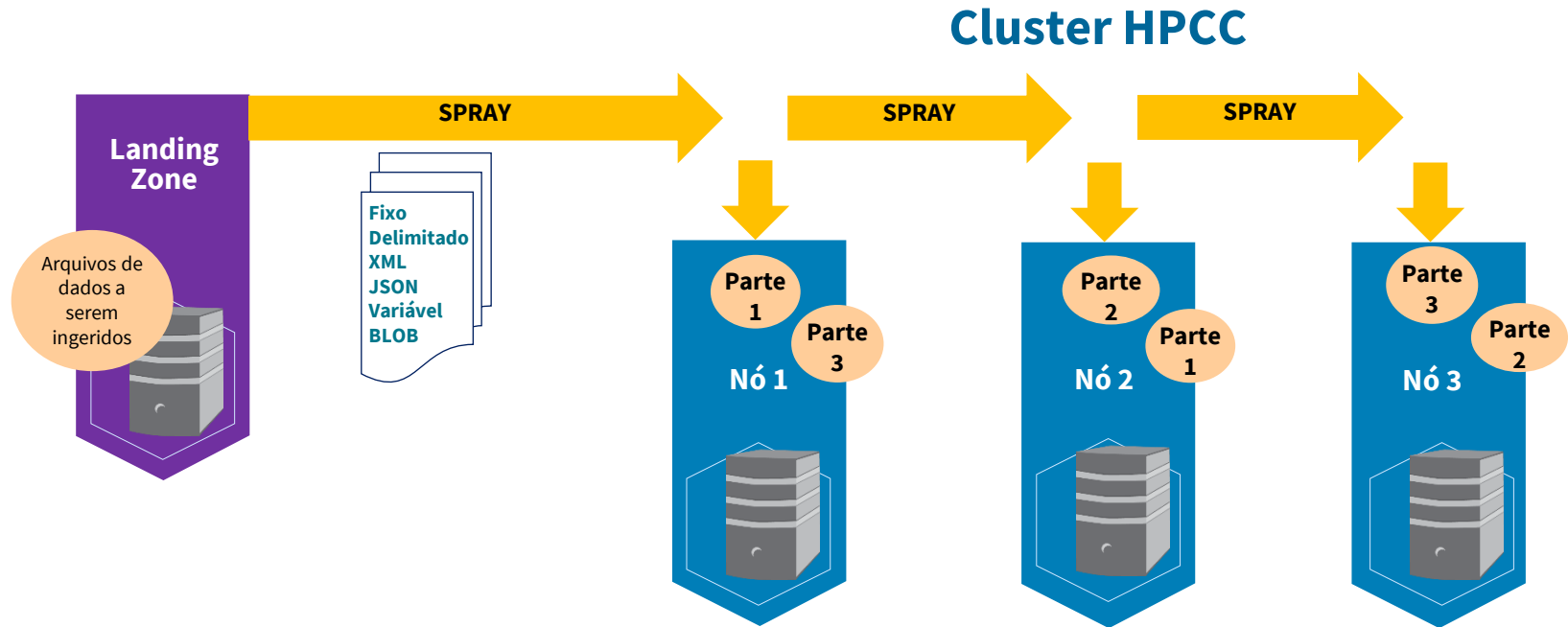
firstname	lastname	gender	age	nationality	occupation
Isaac	Newton	M	84	english	scientist
Albert	Einstein	M	76	german	scientist
Marie	Curie	F	66	polish	scientist
Victor	Hugo	M	83	french	writer
Jane	Austen	F	41	english	writer
Emily	Bronte	F	30	english	writer
Jane	Doe		25	brazilian	unemployed
John	Doe	U	65	american	retired

3. ETL

- Os 4 objetivos principais em ETL:



3. Extração dos dados



As partes do arquivo são referenciadas em ECL como um único arquivo lógico...

3. Extração dos dados

- **Escopo e Nomes de arquivos lógicos**

Os nomes de arquivos lógicos sempre começam com um escopo (estrutura de diretórios) e termina com o nome do arquivo.

O HPCC busca por **arquivos cujos nomes começam com um escopo padrão (THOR):**

'DIR1::DIR2::NomeArquivo' //dado isso, HPCC procura por:
'**THOR**::DIR1::DIR2::NomeArquivo' //esse arquivo

O sinal de “til” (~) indica a **supressão do escopo padrão:**

'~DIR1::DIR2::NomeArquivo' //dado isso, HPCC procura por:
'DIR1::DIR2:: NomeArquivo' //esse arquivo

3. Extração dos dados

- **Upload do “raw” Dataset para a ‘Zona de Entrada de Arquivos’:**
 - Importação dos Dados Brutos (Upload = *Landing Zone*)

Uploader(Enviador) de Arquivo [X]

Zona de Entrada de Arquivos:

Máquinas:

Pasta:

#	Type	File Name	Size
1	CSV	<u>DatasetInline01.csv</u>	495

Sobrescrever ☐ **Começar** Fechar

3. Extração dos dados

- **Spray para a distribuição do arquivo entre os 'nós' do *Cluster*:**
 - Spray do arquivo (Spray = *Logical Files*)

Spray (Distribuir aos Nós): Fixo ▾ Delimitado ▾ XML ▾ JSON ▾ Variável ▾ BLOB ▾

Destino

Grupo: mythor ▾

Fila: dfusever_queue ▾

Escopo de Destino: CLASS::MDM::DEMO::

Nome do Destino

Datasetline01

Opções

Formato: ASCII ▾

Maximo tamanho do registro: 8192

Seperadores: \

Omitir Separador: ☐

Escapar:

Terminador de linhas: \n\r\n

Aspas: "

Sobrescrever: ☐

Sem Separador: ☐

Compactar: ☐

Estrutura de registro disponivel: ☐

Expira em (dias):

Replicar: ☐

Incomum: ☒

Falha em caso de arquivo sem fonte: ☐

Terminador em Aspas: ☐

Replicação atrasada: ☒

Spray (Distribuir aos Nós)

3. Extração dos dados

- **“Raw” Dataset utilizado durante a segunda parte da demonstração:**
 - Importação dos Dados Brutos (Upload = *Landing Zone*)

Persons

Fixo: 151

963.512
registros

Zona de Entrada de Arquivos						
Largura:		151	EBCDIC:		<input type="checkbox"/>	
1	0000	AB••Cherianne	Khatchaturian	N	19990922•• M•	69 BOULDER RIDGE RD # 25A
2	0097	BB••Muyesser	Raplee	X	20001111a• F•	55 SWAMP RD
3	012e	CB••Roselin	Viceconte		19990325X• F•	19800113107 HILL TER
4	01c5	DB••Inda	Provines		20000909•• U•	290 W MOUNT PLEASANT AVE
5	025c	EB••Inderdeep	Laurence	D	20001228X• M•	44 PROSPECT PL
6	02f3	FB••Chrystine	Mangiapane		19990827;• F•	197803061806 1ST AVE APT 8F
7	038a	GB••Adelene	Stock	R	20000827U• M•	1117 FARM RD
8	0421	HB••Mendy	Rufenblanchette		20000903•• M•	3 W 83RD ST APT 4C
9	04b8	IB••Lannie	Amerantes	I	200012199• U•	200 W 20TH ST APT 909
10	054f	JB••Tare	Gonyeau	T	199308070• F•	197508016 CANDLE CT
11	05e6	KB••Finney	Aristilde	P	19900621X• M•	19560920222 1ST AVE APT 2B
12	067d	LB••Oreoluwa	Marthaler		19931006F• F•	19731201176 CLAREMONT GDNS
13	0714	MB••Surge	Abbottkrepp	D	20000308•• F•	22 LE PARC CT
14	07ab	NB••Dave	Mcjury		20001129i• U•	510 COOPER RD # 1
15	0842	OB••Ramsay	Ping		20001129i• M•	404 AVENUE L
16	08d9	PB••Lacides	Wisniveskydr	Q	20001227•• M•	23 JEFFERSON LN
17	0970	QB••Hazele	Scoggins	H	20001102•• M•	43 RENAISSANCE DR
18	0a07	RB••Laini	Mandrake	B	20001205•• M•	5 SHEVCHENKO AVE
19	0a9e	SB••Settnaphon	Zeulli	N	200009228• M•	196212011714 E WILLOW GROVE AVE
20	0b35	TB••Tomas	Drabick	U	19830116•• F•	19550901155 E 34TH ST # 20
21	0bcc	UB••Ginabelle	Munkel	Q	20000914•• M•	833 SUMMER ST STE 3B
22	0c63	VB••Ornah	Aschermann		200012079• M•	36 UNION ST
23	0cfa	WB••Moisey	Shupp	R	200012299• M•	12 STEEPLECHASE LN
24	0d91	XB••Jynevelyne	Hirschkind		19870618•• F•	19630324404 SUMMIT AVE
25	0e28	YB••Norli	Pisciotta	V	19870320•• F•	1933120123 SEVINOR RD # 0
26	0ebf	ZB••Toai	Ibric		20000706•• M•	17 TOLKIEN PSGE
27	0f56	[B••Murshell	Bykov		20001209a• F•	2142 MOUNTAIN VIEW AVE

3. Extração dos dados

- **Spray para a distribuição do arquivo entre os ‘nós’ do *Cluster*:**
 - Spray do arquivo (Spray = *Logical Files*)

Spray (Distribuir aos Nós): Fixo Delimitado XML JSON Variável BLOB

Destino

Grupo: mythor

Fila: dfuserver_queue

Escopo de Destino: CLASS::MDM::DEMO::

Nome do Destino	Tamanho do Registro
persons	151

Opções

Sobrescrever: ☐

Sem Separador: ☐

Compactar: ☐

Expira em (dias):

Replicar: ☐

Incomum: ☒

Falha em caso de arquivo sem fonte: ☐

Replicação atrasada: ☒

Spray (Distribuir aos Nós)

3. Extração dos dados

▪ Definição da Estrutura de Dados & Análise do Perfil (*Profiling*) dos Dados

```
EXPORT modPersons := MODULE
  EXPORT Layout := RECORD
    INTEGER4 RecID;
    STRING15 FirstName;
    STRING25 LastName;
    STRING15 MiddleName;
    STRING2 NameSuffix;
    STRING8 FileDate;
    UNSIGNED2 BureauCode;
    STRING1 MaritalStatus;
    STRING1 Gender;
    UNSIGNED1 DependentCount;
    STRING8 BirthDate;
    STRING42 StreetAddress;
    STRING20 City;
    STRING2 State;
    STRING5 ZipCode;
  END;
  EXPORT File := DATASET('~CLASS::MDM::DEMO::persons', Layout, THOR);
END;
```

#	recid	firstname	lastname	middlename	namesuffix	filedate	bureaucode	maritalstatus	gender	dependentcount	birthdate	streetaddress
1	1000001	Cherianne	Khatchatourian	N		19990922	24		M	4		69 BOULDER RIDGE RD #
2	1000002	Muyesser	Raplee	X		20001111	353		F	0		55 SWAMP RD
3	1000003	Roselin	Viceconte			19990325	344		F	4	19800113	107 HILL TER
4	1000004	Inda	Provines			20000909	13		U	3		290 W MOUNT PLEASANT #
5	1000005	Inderdeep	Laurence	D		20001228	344		M	3		44 PROSPECT PL
6	1000006	Chrystine	Mangiapane			19990827	315		F	0	19780306	1806 1ST AVE APT 8F
7	1000007	Adelene	Stock	R		20000827	252		M	0		1117 FARM RD
8	1000008	Mendy	Rufenblanchette			20000903	24		M	1		3 W 83RD ST APT 4C
9	1000009	Lannie	Amerantes	I		20001219	313		U	0		200 W 20TH ST APT 909
10	1000010	Tare	Gonyeau	T		19930807	48		F	0	19750801	6 CANDLE CT
11	1000011	Finney	Aristilde	P		19900621	344		M	2	19560920	222 1ST AVE APT 2B
12	1000012	Oreoluwa	Marthaler			19931006	358		F	5	19731201	176 CLAREMONT GDNS
13	1000013	Surge	Abbottkrepp	D		20000308	13		F	4		22 LE PARC CT
14	1000014	Dave	Mcjury			20001129	238		U	1		510 COOPER RD # 1

4. Preparação dos dados

“Limpeza, padronização e consolidação de registros ”

##	id	firstname	lastname	middlename	namesuffix	filedate	bureaucode	maritalstatus	gender	dependentcount	birthdate	streetaddress
1	1000001	Cherianne	Khatchatourian	N		19990922	24		M	4		69 BOULDER RIDGE RD #
2	1000002	Muyesser	Raplee	X		20001111	353		F	0		55 SWAMP RD
3	1000003	Roselin	Viceconte			19990325	344		F	4	19800113	107 HILL TER
4	1000004	Inda	Provines			20000909	13		U	3		290 W MOUNT PLEASANT P
5	1000005	Inderdeep	Laurence	D		20001228	344		M	3		44 PROSPECT PL
6	1000006	Chrystine	Mangiapane			19990827	315		F	0	19780306	1806 1ST AVE APT 8F
7	1000007	Adelene	Stock	R		20000827	252		M	0		1117 FARM RD
8	1000008	Mendy	Rufenblanchette			20000903	24		M	1		3 W 83RD ST APT 4C
9	1000009	Lannie	Amerantes	I		20001219	313		U	0		200 W 20TH ST APT 909
10	1000010	Tare	Gonyeau	T		19930807	48		F	0	19750801	6 CANDLE CT
11	1000011	Finney	Aristilde	P		19900621	344		M	2	19560920	222 1ST AVE APT 2B
12	1000012	Oreoluwa	Marthaler			19931006	358		F	5	19731201	176 CLAREMONT GDNS
13	1000013	Surge	Abbottkrepp	D		20000308	13		F	4		22 LE PARC CT

Minicurso: Agenda

✓ **LexisNexis Risk Solutions: A Empresa**

- ✓ Quem somos nós?
- ✓ A nossa tecnologia: A evolução da plataforma HPCC Systems...

✓ **HPCC Systems: Visão Geral**

- ✓ Apresentação de conceitos;
- ✓ Aplicação de conhecimentos.

➤ **Próximos passos**

- Cursos online;
- Projetos de pesquisa;
- Oportunidades profissionais.

➤ **Considerações Finais**

Próximos passos

▪ **Cursos online: +170 aulas** (<https://learn.lexisnexis.com/hpcc>)

Introdução ao ECL (parte 1)

- Conceitos e consultas

Introdução ao ECL (parte 2)

- ETL com ECL

ECL Avançado (parte 1)

- Dados relacionais

ECL Avançado (parte 2)

- Superarquivos, XML/JSON e PLN

ECL Aplicado

- Geração e automação de código ECL

ROXIE ECL (parte 1)

- Índices e consultas

ROXIE ECL (parte 2)

- Otimização de consultas

Machine Learning com HPCC Systems

- Fundamentos para uso dos *plugins*

Administração de Sistemas

- Conceitos e operação básica

HPCC para gestores

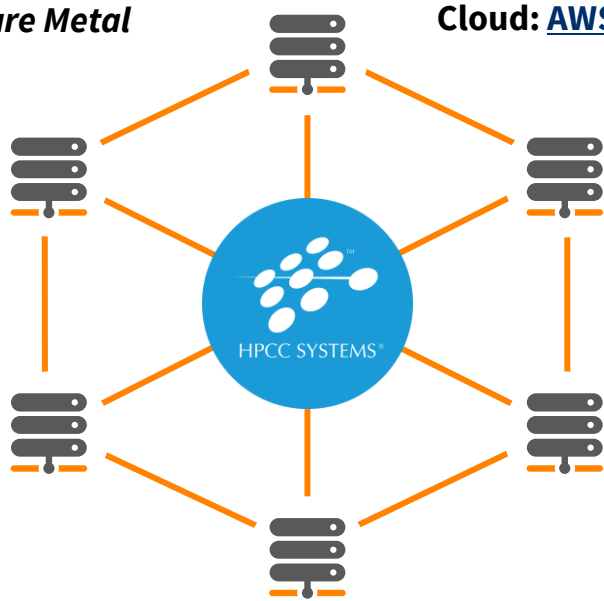
- Visão geral e aplicações da plataforma

Próximos passos

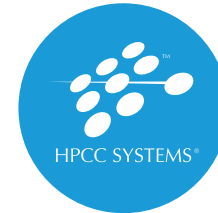
- **Opções de utilização: *Playground*** (<http://play.hpccsystems.com:8010/>)

Bare Metal

Cloud: AWS/Azure



Oracle Virtual Box
HyperV
Docker

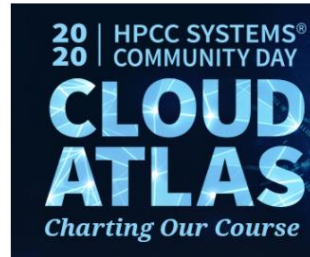


HPCC Máquina Virtual

Saiba mais em: <https://hpccsystems.com/getting-started/>

Próximos passos

- **Relacionamento com a Comunidade Acadêmica:**
<https://hpccsystems.com/community/academics>



Universidade de São Paulo
Brasil



Próximos passos

■ Universidades Brasileiras

Universidade de São Paulo
Brasil



- Disciplina Optativa na Poli/USP ([Link](#))
- Cursos de extensão ([Link](#))
- Co-Orientação de IC's (PIBIC [Link1](#) [Link2](#) [Link3](#))



UNIVERSIDADE FEDERAL
DE SANTA CATARINA

- Co-Orientação de TCC's ([Link1](#) [Link2](#))
- Co-Orientação de IC's ([Link](#))
- Co-Autoria de artigos científicos ([Link](#))
- Auxílio para aquisição de equipamentos

Minicurso: Agenda

✓ **LexisNexis Risk Solutions: A Empresa**

- ✓ Quem somos nós?
- ✓ A nossa tecnologia: A evolução da plataforma HPCC Systems...

✓ **HPCC Systems: Visão Geral**

- ✓ Apresentação de conceitos;
- ✓ Aplicação de conhecimentos.

➤ **Próximos passos**

- ✓ Cursos online;
 - Projetos de pesquisa;
 - Oportunidades profissionais.

➤ **Considerações Finais**

Próximos passos

- **Projetos de pesquisa**



Saiba mais em: <https://wiki.hpccsystems.com/display/hpcc/Available+Projects>

Minicurso: Agenda

✓ **LexisNexis Risk Solutions: A Empresa**

- ✓ Quem somos nós?
- ✓ A nossa tecnologia: A evolução da plataforma HPCC Systems...

✓ **HPCC Systems: Visão Geral**

- ✓ Apresentação de conceitos;
- ✓ Aplicação de conhecimentos.

➤ **Próximos passos**

- ✓ Cursos online;
- ✓ Projetos de pesquisa;
- Oportunidades profissionais.

➤ **Considerações Finais**

Próximos passos

■ Oportunidades profissionais

#ExploreMore


<https://risk.lexisnexis.com/about-us/careers>
<https://www.linkedin.com/company/lexisnexis-risk-solutions/>
<https://www.vagas.com.br/v2273659>

#Contato

Ana Cristina Vieira
Senior Talent Acquisition
LexisNexis Risk Solutions Group
☎ +55.11.97075.5659
ana.vieira@lexisnexisrisk.com

Links Úteis

- Site principal: hpccsystems.com
- Primeiros passos: hpccsystems.com/Why-HPCC-Systems
- Canal do youtube: youtube.com/user/HPCCSystems
- Fórum da Comunidade: hpccsystems.com/forums
- Poster Competition: [Link](#)



Faça parte da Comunidade

Registre-se em:
<https://hpccsystems.com/pt-br>

Minicurso: Agenda

✓ **LexisNexis Risk Solutions: A Empresa**

- ✓ Quem somos nós?
- ✓ A nossa tecnologia: A evolução da plataforma HPCC Systems...

✓ **HPCC Systems: Visão Geral**

- ✓ Apresentação de conceitos;
- ✓ Aplicação de conhecimentos.

✓ **Próximos passos**

- ✓ Cursos online;
- ✓ Projetos de pesquisa;
- ✓ Oportunidades profissionais.

➤ **Considerações Finais**



perguntas
& RESPOSTAS