Essa conversa ela tem uma finalidade bem interessante que é basicamente expandir o seu conhecimento acerca de leitura e escrita de dados é um pouco mais além daquele print do **input**

Que Muito provavelmente você já trabalhou em Python e acompanhante no slide por favor:

**I/O: input/output; entrada/saída**

* Arquivos em vários formatos

CSV, XML, JSON, parquet

* Encoding
* SQLite
* Persistência   
  Local Nuvem

Você pode ver aqui que temos na verdade de diferentes formatos para trabalharmos com entrada e saída de dados

Então não é somente mostrar uma mensagem na tela né somente a receber não ter em um algoritmo que receba o seu nome e com isso mostre aquele nome na tela

Ou então construir uma calculadora em que você soma aritmética dos números diferentes operações

E também mostrando tela existem diferentes formas de trabalharmos com informações tanto com o armazenamento e manipulação de arquivos

Então os formatos que são mais empregados aí em conexões em algoritmos no mercado de trabalho incluindo por exemplo **CSV,** **XML,** **Jason**, **Parquet**

É uma preocupação muito importante de processarmos essas informações e lermos corretamente acentuações afins então para isso conversaremos também sobre **Encodin**   
  
Além desses arquivos existe ou é *SQL light* ou **SQLite** existem diferentes pronúncias nesse tipo de aplicação que chega nem um pouco mais próximo aí de uma estrutura de banco de dados

E finalmente a persistência, ou seja, armazenamento desses arquivos seja de uma forma local ou seja no seu próprio computador como também na nuvem e interagir com diferentes serviços diferentes servidores vamos lá.

**XML e JSON**

Então para usarmos com exemplo essa questão de armazenamento de arquivos de manipulação de arquivos vamos pensar em um caso em que estamos tentando armazenar diferentes notas de diferentes estudantes

Então imagina um cadastro é o registro simples e eu tenho por exemplo códigos de estudantes eu tenho as notas deles e às vezes uma data sejam de cadastro hoje atualização só para vermos como que está a arquivos funcionam na prática Beleza

Então por favor me acompanhe nos slides:

<?xml version='1.0' encoding='utf‐8'?>

<estudantes>

<estudante>

<codigoestudante>10000144</codigoestudante>

<disciplina>25000</disciplina>

<dataatualizacao>2021‐12‐01 00:00:00</dataatualizacao>

<nota>5.4</nota>

</estudante>

<estudante>

<codigoestudante>10000145</codigoestudante>

<disciplina>54000</disciplina>

<dataatualizacao>2021‐06‐20 00:00:00</dataatualizacao>

<nota>5.7</nota>

</estudante>

<estudante>

<codigoestudante>10000146</codigoestudante>

<disciplina>54050</disciplina>

<dataatualizacao>2021‐06‐25 00:00:00</dataatualizacao>

<nota>7.9</nota>

</estudante>

</estudantes>

**JSON**

<?xml version='1.0' encoding='utf‐8'?>

<estudantes>

<estudante>

<codigoestudante>10000144</codigoestudante>

<disciplina>25000</disciplina>

<dataatualizacao>2021‐12‐01 00:00:00</dataatualizacao>

<nota>5.4</nota>

</estudante>

<estudante>

<codigoestudante>10000145</codigoestudante>

<disciplina>54000</disciplina>

<dataatualizacao>2021‐06‐20 00:00:00</dataatualizacao>

<nota>5.7</nota>

</estudante>

<estudante>

<codigoestudante>10000146</codigoestudante>

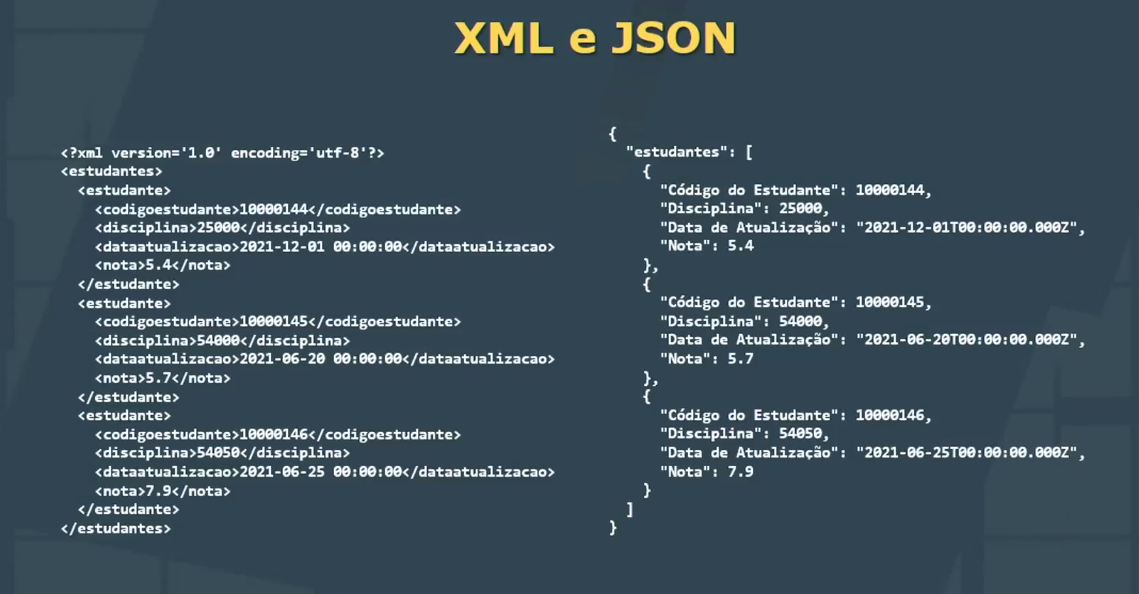
<disciplina>54050</disciplina>

<dataatualizacao>2021‐06‐25 00:00:00</dataatualizacao>

<nota>7.9</nota>

</estudante>

</estudantes>



aqui eu trago dois exemplos para você

Então eu gostaria que você enxergasse isso como se fosse um “De Para”, uma comparação do que em uma direita e uma comparação que tem na esquerda e vice-versa

O que eu tenho aqui no meu lado esquerdo é um exemplo de um arquivo **XML** enquanto isso no meu lado direito eu tenho exatamente a mesma informação porém em um formato **JSON** um arquivo **JSON**

Então é que eu vou começar a desenhar alguns exemplos eu espero que que isso fique mais legível para você

Note que primeiramente nós temos três estudantes então eu vejo aqui o código

Até atentar realçar eu tenho que ir por exemplo esse código 10000144 então é um exemplo de um estudante que tem esse código que eu tenho um estudante que tenho outro código né que seria 10000145 e outro estudante que eu tenho 10000146 esses seriam aí os códigos desses estudantes

Então esse exemplo não seria tão necessário aí muito com texto você consegue enxergar que eu tenho basicamente informações de três estudantes diferentes nós sabemos disso porque vemos três códigos diferentes de estudantes

Se você olhar na sua direita nós temos aqui também a mesma informação porém com uma sintase, uma forma de escrita um pouco diferente

Então tem o mesmo código é o 10000144 depois de 1000045 e o 10000146, então quando estamos falando aí de armazenagem de informações basicamente eu estou seguindo a mesma ordem aqui

Além disso para cada um desses estudantes pode perceber que temos que temos aí algumas informações adicionais a primeira informação adicional e disciplina então nós podemos sair de um código específico de uma disciplina dos Estudantes esse mesmo código pode ser códigos diferentes também atrelados aos estudantes diferentes então o primeiro estudante tem o código 25000 ou segundo ele tem a nota dele no caso referente a disciplina de código 54000 e por último no terceiro e estudante com sua nota atrelada a disciplina de código 54050

E mais uma vez exatamente a mesma coisa aqui na direita

E a mesma lógica também funciona para data de atualização né você pode ver que eu tenho sempre no formato ano mês dia hora Minuto segundo

Então aqui é a mesma coisa aqui embaixo a mesma coisa e não preciso né explicar muito direito na direita exatamente a mesma lógica se mantem, exatamente os mesmos valores se mantem

Finalmente aqui na esquerda temos a nota né 5.4, 5.1 7, 7.9 E que a direita também a mesma coisa de 5.4, 5.7 e 7.9

Agora porque que nós temos esses diferentes formatos é que semelhanças e diferenças nos temos entre eles?

A primeira coisa que vale a pena não é considerar que o XML é um formato um pouco mais antigo né você pode ver que ele é um pouco mais verbos ou seja tem muita informação né do que começa, do que termina

Se você já trabalhou por exemplo HTML ou será que tem na sintase bem parecida né E tem aquelas chavezinhas, do menor que maior que eu defino onde que o comércio onde que eu termino um determinado elemento ou como a gente chama HTML em uma tag então é que tem uma mesma lógica dessas tags porém no final isso pode ser muito, como eu falei muito verboso

Eu posso ter um arquivo gigantesco onde a informação em si é basicamente metade ou até menos da metade do arquivo e mais da metade desses arquivos são somente definições

Olhe de novo aqui no slide você pode ver, até eu vou apagar aqui para não ficar muito complexo da gente enxergar, mas você pode vir aqui pegar uma parte da que é somente da sintase

Até vou escolher aqui uma nota uma cor um pouco diferente para mapearmos, talvez um verde, tudo isso aqui que eu estou desenhando, melhor né que estou pintando aqui numa cor diferente são somente as tags

Então você pode ver que grande parte da informação, a melhor grande parte do que temos no nosso arquivo é somente informações sobre o que eu tenho no arquivo e não exatamente o conteúdo do arquivo

Imagine pensando em analogias recebe um presente esse presente uma caixa gigantesca só que dentro dela eu tenho tanto papelão papel bolha e o presente em si, que é uma coisa tão pequena às vezes não precisaria de uma embalagem tão grande

Então essa analogia que gostaria de trazer para você né em relação ao arquivo XML Mas isso não significa que é de todo ruim, ele nos trás uma

boa definição de hoje realmente começa e termina as informações então olhando até aqui nessa linha específica em específico no estudante 10.000 144 se sabe exatamente onde que é inflamação começa e onde ela termina

E exatamente a mesma coisa aí para as demais informações um outro uma outra informação que ela bem importantes consegue só de bater o olho saber exatamente ter acesso na estrutura de árvore ou seja, você tem la elementos Pais e elementos filhos e aí com isso seja a estrutura que é próximas uma estrutura de árvore

Então você sabe que por exemplo, que vai ficar bem colorida,todas essas informações é do código estudante disciplina, data de atualização e nota elas se referem a exatamente este estudante

então não tem como essa informação aqui da disciplina tem alguma relação com esses outros estudantes e baixo porque só de bater o olho você sabe exatamente aonde que pertence aquela informação então por isso não é essas tags por isso que a gente sabe onde que um caso é um elemento começa e outro elemento termina

e também uma outra informação bem importante aqui no começo é o ENCODING eu já vou falar um pouco mais sobre ENCODING mas só dando em breve spoiler basicamente uma informação que se refere é como é que nós podemos processar aquela informação do ponto de vista de acentuações o ponto de vista de outros caracteres especiais então isso ajuda a evitar erros na nossa conversão de dados e

finalmente a direita os consegue ver que temos uma estrutura muito mais limpa né quando eu até eu estava pensando aquela estrutura que eu estava pensando em você pode ver que enquanto ele sabe explicar nem terminei de pintar porque tem tanta né informação aqui se nós olhamos exatamente a mesma coisa no JSON temos uma estrutura muito mais leve né muito mais enxuta

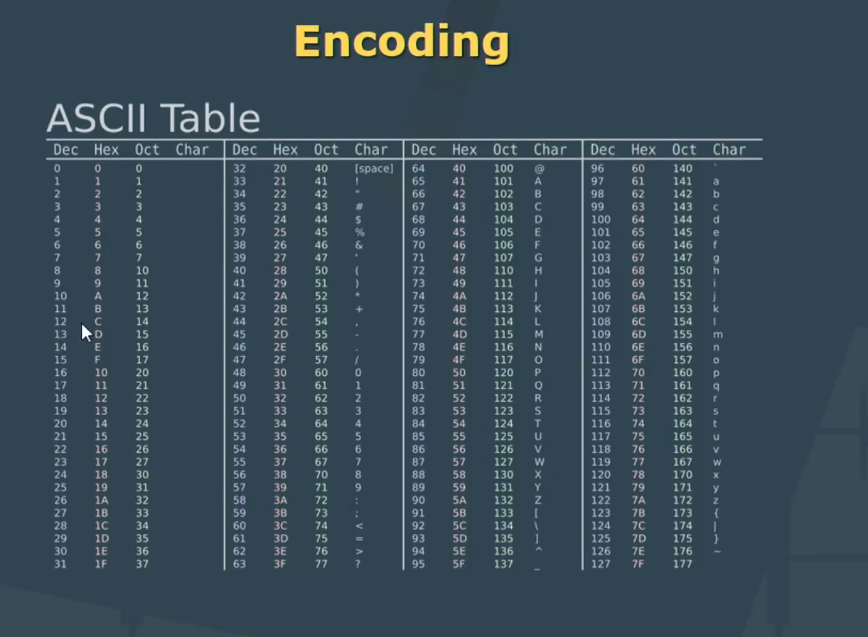
Então se a gente for comparar né o que eu tenho de verde você pode compara aqui comigo que eu tenho um caso Bem mais leve né assim realmente muito menos tags eu tenho uma informação que ela mais limpa de nós visualizarmos e também processarmos e finalmente você pode também perceber que a gente consegue identificar nesses elementos esses essas diferentes entre aspas Tags utilizando os espaços utilizando as acentuações o que não era possível um exemplo ai da esquerda então por exemplo a data de atualização né pode ver que aqui no nosso caso da direita temos data Espaço De Espaço atualização tenho maiúscula em minúscula tem espaço tem cedilha tem um tio agora aqui no nosso exemplo a esquerda não é exatamente o caso tem que juntar tudo aquilo né como se fosse realmente código ou algum outro elemento

Então até na parte de legibilidade do ponto de vista humano né Para nós humanos entendermos o que tem aqui em informação acaba sendo aí um pouco melhor de trabalharmos com Jason mas novamente quando entrarmos no ponto de vista do mercado de trabalho de integrações algoritmos nem tudo será JASON

Podemos ter casos que está em XML por N motivos que sejam a integração compatibilidade de sistemas e assim por diante e tem casos que realmente JASON será uma opção então por isso é que apresenta os dois para você para você entender para que servem esses diferentes formatos e não é que no mercado de trabalho somente existem JASON somente existe XML

Invariavelmente você terá ai contato com ambos Ok

Além disso né o que eu estava comentando sobre ENCODING é isso aqui é uma tabela ASCII Table

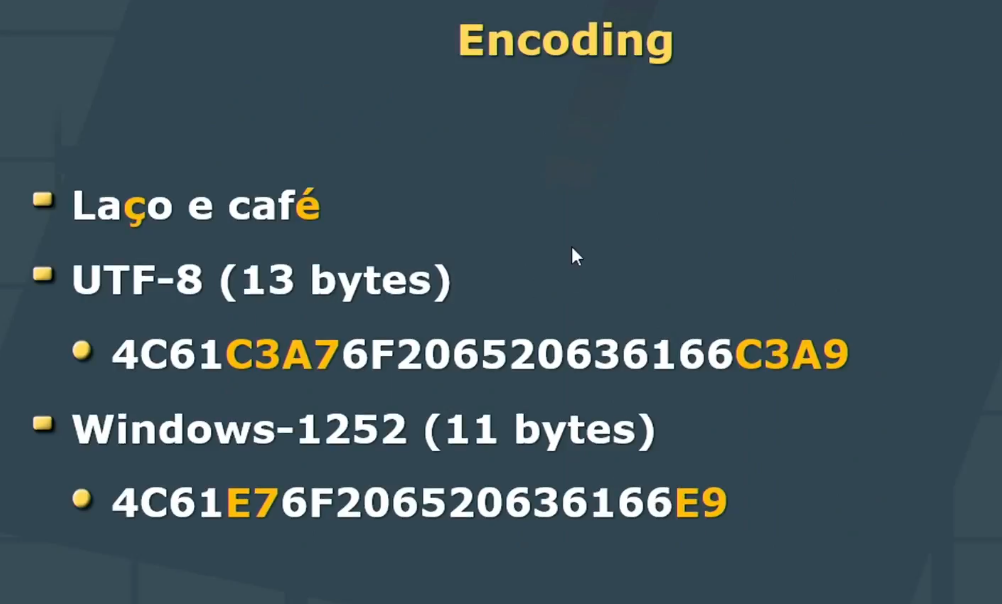


então é basicamente o formato né que nós trabalhamos identificamos os caracteres e traduzimos aquilo para que os computadores consigo trocar informações entre si então arquivo de texto, documentos e afins tudo isso de alguma forma é acaba sendo convertido para Bytes para Bits (O termo bit foi originado de Binary digit, ou seja, dígito binário. O bit corresponde aos valores de 0 e 1. Juntando 8 bits, temos 1 byte. Portanto, 1 byte é oito vezes maior que 1 bit.) como você bem sabe e essas informações aqui elas são guardados em números e valores então geralmente é o melhor aquela versão original

Bem antigo que existem de Décadas atrás é essa tabela ASCII aqui que você consegue enxergar na sua tela essa tabela ASCII eu gostaria que você prestasse atenção especificamente nessa última coluna aqui, que é a Char achar ela basicamente se refere aos caracteres que nós estamos trabalhando então se você quer saber por exemplo o A minúsculo como que ele é salvo em um computador né como que a gente consegue salvar aquela informação sabendo que é tudo 0 e 1 é sabendo que tudo ação de quatro números e daí você sabe que aquele a minúsculo recebe um códigozinho com o número 97 em decimal se estivéssemos trabalhando em hexadecimal seria o 61 se fosse em Octa seria 141 e assim por diante

mas é o padrão aqui ou é a mensagem principal é você saber que cada carácter ele tem lá um códigozinho específico porém ao bater o olho aqui você pode sentir falta de algumas coisas por exemplo Estamos vendo que é tudo do minúsculas temos aqui alguns caracteres especiais por exemplo & Comercial o $ cifrão eu tenho ponto exclamação tem o espaço tem o números dígitos mas estão faltando algumas coisas por exemplo o cedilha na onde que está o c com a cedilha Ç, onde que esta a trema ¨, onde que está o a com acento agudo acento agudo com acento circunflexo e assim por diante

Então como você deve imaginar essa tabela ASCII, ela não comporta todos os caracteres do mundo imagina por exemplo alfabeto em cirílico né que é utilizado aí na Russia se alguns países do leste europeu ou alfabetos né que a seguem o alfabeto grego, japonês ou chinês assim por diante então pensando em todas essas combinações essa tabela ASCII ela não dá conta do recado então existem expansões a essa tabela aqui então é esse aqui o ponto a gente gostaria de trazer para você



* Laço e café
* UTF-8 (13 bytes)
* 4C61C3A76F206520636166C3A9
* Windows-1252 (11 bytes)
* 4C61E76F206520636166E9

porque pensando nessas tabelas dessa expansão da tabela ASCII para guardar por exemplo laço né e o café que como você está vendo na sua tela tem lá cedilha eu tenho neoway com o acento com acento em cima dessa letra então houveram ao longo do tempo e expansões para conseguir completar esse texto e aí que vem o canal do TF 8 e aí que vem por exemplo Windows 252 e vários várias outras formas de encoding como você viu comigo até anteriormente aqui lembre-se que eu mostrei para você de se incomodem utf-8 então vários arquivos ou praticamente várias informações que são salvas em texto esses dados eles tem o incomodem por traz Eles foram escritos com algum encolhimento se por exemplo a gente salva uma informação em um Windows 252 os pode ver que essa frase em cima laço e café a gente por baixo dos panos na ela recebe todos iguais porque veja que o cedilha e o e eles receberam Esses códigos aqui embaixo então o Seco a seguir ele recebeu o E7 e o e com acento recebeu e 9 Se tentarmos ler essa informação é que esses B está com e 7 e 9 e por algum outro software algum outro programa que está esperando em utf-8 vai aparecer em bom português ou na com os devidos aí anglicismos vai ficar tudo bugado esse cedilha aparece às vezes um ser com percentual uma vai aparecer eu mexi completamente a fora do esperado esse texto porque realmente não esperava de uma forma e o outro e o outro só que eu esperava com outro Encore então por isso que essa informação Ela é bem útil mesmo que no teste 8 por exemplo aquele mesmo se e o e a cedilha receberam outro código não foi mais o E7 por baixo dos panos foi o C3 Acerte a mesma coisa do É com acento e se meu C3 A9 Então como você deve imaginar se um não está esperando determinar em que ordem realmente aquele seu texto parecer completamente bugado Dulce Você já trabalhou com por exemplo dashboard a trabalho em ciências de dados trabalhos por exemplo uma inflação de textos de fato se já deve ter se deparado com algum caso sim se esperavam um caractere um determinado cento e não foi esse o caso Então realmente era um problema de encoding o tema de encoding ele é muito trabalhado do ponto de vista assim por várias por vários ambientes de estudo mas isso realmente montado com um importante que você conheça para saber da onde que vieram aqueles erros né porque que esses erros estão sendo apresentados então por isso que a gente apresenta este este tema para você até Finalmente né quando pensamos aqui no teste 8 existe um outro caso que esse o teste 8 é possibilita o seu emprego emojis então quando a gente pensa nos emojis seja WhatsApp ou qualquer outro aplicativo aqueles emojis na verdade que corresponde também a outro código aqui então é esse código que acaba sendo visto somente por alguns tipos de podem nem todos então por isso que se você vai carregar nenhum emoji até em outro software de outras soluções que não estão esperando aquele que podem aparecer um quadradinho sem nada dentro ou vai aparecer um triângulo com um ponto de interrogação vai aparecer qualquer coisa menos o Emoji algum erro na comunicação alguma coisa assim não exatamente em caso de beleza