Em linhas gerais, usar o **CNPJ** como chave‐estrangeira “principal” (em vez do código CVM) faz sentido principalmente porque:

1. **É o identificador comum a todos os arquivos da CVM**
   * Quase todos os relatórios (ITR, DFP, Formulário de Referência, eventuais, números de ações etc.) vêm referenciados pelo CNPJ.
   * Se você fizer do CNPJ a chave‐estrangeira principal, não precisará ficar fazendo “lookup” ou join extra para converter CNPJ → código CVM toda vez que for carregar um arquivo novo (já vem “pronto” para relacionar com a tabela empresas).
2. **É um identificador universal (natural key) mais aderente ao mundo externo**
   * Fora da base da CVM, bancos, sistemas de mercado e outras fontes (como Serasa, Receita Federal, B3 etc.) sempre usam o CNPJ para cruzar dados de uma empresa.
   * Manter o CNPJ como FK simplifica integrações futuras (por exemplo, misturar dados de balanço CVM com dados de mercado, ratings, ou mesmo consultar outras APIs que retornam CNPJ).
3. **Evita duplicidade semântica**
   * Se a tabela empresas já tiver CNPJ como PK, e você usar o código CVM apenas como outro campo, está criando redundância: toda vez que entrar num “parecer” você vai precisar converter CNPJ → código CVM e depois fazer join código CVM → Tabela empresas.
   * Ao usar diretamente CNPJ → empresas(cnpj\_companhia), você corta essa etapa intermediária.
4. **Maior estabilidade em certos contextos**
   * Na prática, o código CVM existe enquanto a companhia está registrada. Se uma empresa for baixada da CVM e depois reabrir, pode eventualmente receber outro código. Já o CNPJ, quando uma companhia encerra ou muda de razão social, costuma permanecer o mesmo (a menos de reorganizações societárias profundas).
   * Ou seja, usar CNPJ torna menos arriscada a possibilidade de “perder o vínculo” para um mesmo CNPJ que, por algum evento especial, mude de código na CVM.
5. **Legibilidade e manutenção**
   * Quem for dar manutenção no banco vai se “encontrar” mais rápido ao ver “empresas.cnpj\_companhia = '12.345.678/0001-90'” do que ter de rastrear “qual empresa é esse código CVM 12345?”.
   * Para auditoria interna, relatórios e até scripts simples, referenciar CNPJ deixa o modelo mais “autoexplicativo”.

**Quando, mesmo assim, vale usar o código CVM como FK principal?**

O código CVM pode ser vantajoso se:

* **Você só consome dados estritamente vindos da base CVM** (e não integra com outros sistemas que usam CNPJ).
* **Quiser evitar problemas de formatação** (pontos, barras e traços) na coluna CNPJ, que às vezes vêm em formatos diferentes:
  + Ex.: 12345678000190 vs. 12.345.678/0001-90.
* **Quiser garantir que cada registro de empresa tenha exatamente um identificador numérico curto** (o código CVM costuma ser menor que um CNPJ inteiro), o que pode reduzir ligeiramente o tamanho das FKs e acelerar joins em tabelas gigantes.

Mas, mesmo nesses casos, muitos grupos simplesmente armazenam o CNPJ sem pontuação (apenas dígitos) e definem um índice único; assim, resolvem a formatação sem abrir mão de usar o CNPJ como FK.

**Resumo: por que preferir o CNPJ como FK principal**

1. **Disponível em todos os arquivos**: não há casos de “parecer” ou “itr” que tragam só o código CVM sem antes informar CNPJ.
2. **Sem “lookup” extra**: cada linha já aponta direto para empresas(cnpj\_companhia), sem precisar consultar a própria tabela empresas para converter CNPJ → código CVM.
3. **Integração futura**: se mais adiante você quiser mesclar com outras bases (bancos, Receita, APIs de mercado), todas elas usarão CNPJ.
4. **Menos risco de inconsistência**: se uma empresa migrar de código CVM ou tiver alguma reclassificação, o CNPJ segue constante.

Em suma, **manter o CNPJ como chave‐estrangeira principal é a decisão que torna seu modelo mais direto, menos propenso a steps de conversão e mais fácil de integrar com dados externos**, pois o CNPJ é o campo natural que aparece em todos os relatórios da CVM e em praticamente qualquer outra fonte que trate de pessoas jurídicas.