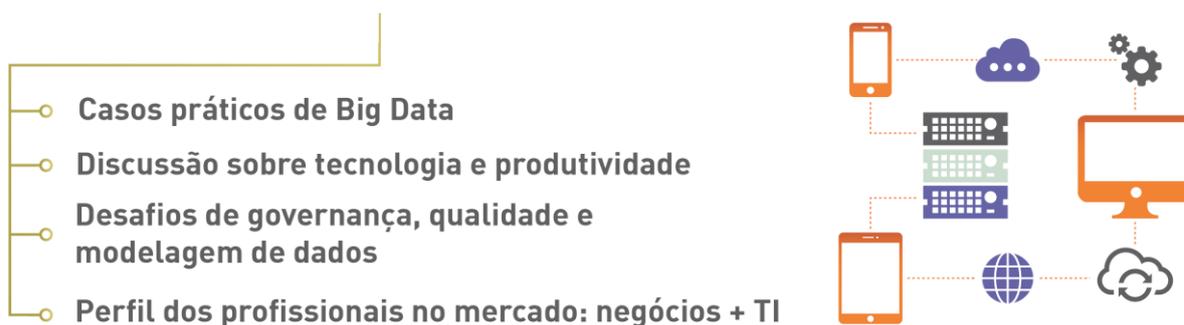


Resumo 11º Encontro de Usuários de BI.

Chegamos ao nosso 11º Encontro de Usuários de BI e o debate teve como tema “Como fazer com que Big Data tenha aplicações práticas de sucesso e efetivas nas empresas”.

Nesta edição, contamos com a presença de mais de 60 representantes de diversas empresas, todos em diferentes ramos, tais como, varejo, indústria e serviços; todas com iniciativas na área de Big Data e suas tecnologias.

Diferentemente das outras edições, o 11º Encontro seguiu um formato de debate mais aberto, com perguntas e respostas entre os participantes, com apoio do mediador. Os questionamentos, para efeito deste resumo, foram agrupados nos seguintes tópicos:



Casos Práticos de Big Data

O primeiro ponto abordado foram os casos práticos do uso de Big Data em empresas. Alguns participantes destacaram exemplos em que essas tecnologias ajudam seus negócios de forma efetiva. Profissionais de TI e de negócios trouxeram para o debate modelos e experiências em suas operações.

O primeiro exemplo veio de um dos participantes que atua como gestor de uma área de especialistas em captura/armazenamento de dados e um grupo de analistas de dados, os quais estão atualmente na área de TI, mas que tem um foco exclusivo de atendimento ao time de negócios da corporação. Essa equipe atua em diversas áreas atendendo a todo o grupo, o qual é composto por diversas empresas que têm atuação desde a parte industrial, até o ponto de vendas final. O gestor ilustrou aos participantes como a utilização e análise das informações do canal de vendas diretas ao consumidor (B2C), através de sua rede de farmácias, permite mostrar o impacto que essas vendas trazem na produção da indústria química do grupo, responsável pela produção de componentes químicos que formam o grupo de produtos de HPC (Higiene Pessoal e Cosméticos). Assim, essa indústria química pode realizar previsões de produção de ingredientes com maior precisão e assertividade; além de toda a informação para a área de negócios da área de farmácias, permitindo ações de *cross-sell* e *up-sell*.

A seguir, um dos representantes de uma grande rede de varejo de moda compartilhou com os participantes uma das utilizações de “Big Data” na empresa. A área de tecnologia investiu em um ambiente de processamento/armazenamento com foco na utilização de informações não estruturadas; tais como logs de internet. E foi nesse tipo de informações que o caso de uso desta empresa se fundamentou. A equipe de

negócios passou a realizar análises de quais produtos tinham maiores visitas no site da empresa e de posse dessas informações passou a realizar campanhas mais assertivas, enviando ofertas mais direcionadas a seus clientes. Com este tipo de análise, em princípio bastante simples, a empresa passou a ter uma taxa de conversão bastante elevada em suas ofertas, atingindo em alguns casos 30% de taxa.

A área de finanças trouxe alguns exemplos de utilização nesse setor. Um deles busca formas de identificação dos investimentos que os clientes realizam fora daquela instituição financeira, para assim, tentar mudar a estratégia de abordagem do banco com esses clientes e melhorar o seu *share wallet*. Um outro exemplo é a utilização de imagens, tais como *street view*, que permite a identificação de possíveis casos de fraudes, evitando assim esforços da equipe comercial em visitas não produtivas. Mais um exemplo nessa indústria foi a captura e análise de informações de navegação em redes sociais e nos sites de internet banking com foco no entendimento do comportamento dos clientes, permitindo assim, análises e insights que são direcionados e utilizados pelas equipes de CRM do banco para melhorar a eficiência e resultados de suas comunicações com os clientes/prospects.

Outro exemplo veio da área hospitalar. O participante compartilhou com o grupo a iniciativa de um projeto nesse grande hospital que irá permitir a otimização dos leitos por meio de análise das informações dos pacientes / prontuários dos mesmos, entendendo e prevendo todo o ciclo do mesmo desde sua chegada ao hospital.

Vimos assim que diferentes métodos exemplificam como a aplicação de Big Data pode se tornar útil e efetiva nas empresas. Mas, quais são as ferramentas disponíveis no mercado para auxiliar um projeto de Big Data? A resposta será detalhada no próximo tópico.



Tecnologias e produtividade

Inicialmente alguns dos participantes fizeram citações de diversas ferramentas que são utilizadas no ambiente de Big Data, tais como Cloudera, Horton Works, Hive, Impala, Sqoop, Flume, Cassandra, entre outras.

O grupo teve algumas interações sobre a questão das tecnologias utilizadas e entende que existem diversas ferramentas e opções para os projetos. Trata-se de um mercado oriundo do software livre, e, portanto, o conceito de utilizar um código único e dar suporte ao mesmo é que permite a criação de diversos *flavors*. As comunidades de desenvolvimento geram vários produtos, os quais acabam sendo concorrentes entre si. O grupo então teve debates sobre maturidade das ferramentas e concluiu que justamente por ainda estar em evolução, algumas ferramentas que hoje são utilizadas podem desaparecer daqui a 1 ou 2 anos, e em contrapartida, algumas podem virar “estrelas de mercado”. Portanto, é importante estar atento e preparado para essas mudanças.

Um dos blocos de ferramentas, talvez o principal neste ecossistema de Big Data, é o da estrutura de dados. Falamos aqui de Cloudera e de Horton Works, que são os principais em nosso mercado. O mediador fez, então, um desafio ao grupo para pensar sobre um dos exemplos citados como caso de Big Data e que não utiliza nenhum desses, mas sim, o tradicional banco de dados, neste caso, o Oracle. Como fica então essa questão de utilização e eventual migração entre tecnologias?

Segundo a opinião de alguns participantes, o tempo de migração depende do investimento que a empresa faz na ferramenta/tecnologia. Quanto mais utilizada e desenvolvida, mais dificuldade a sua empresa vai ter em sair dela. Outro ponto discutido foi a quantidade ideal de ferramentas para usar no negócio, os presentes concluíram que “menos é mais”.

Com relação à produtividade que se pode obter com as tecnologias/ferramentas, o grupo entende que toda adoção de tecnologia tem seu tempo de maturidade e as tecnologias são adotadas/escolhidas de acordo com os desafios de negócios apresentados. A adoção passa pelo conhecimento das tecnologias disponíveis e o que a empresa tem como linha diretiva – às vezes, queremos adotar uma nova ferramenta/tecnologia, mas o processo na empresa não nos permite ir em uma direção, portanto, temos que utilizar o que temos/conhecemos.

O grupo também entendeu que a tecnologia em si talvez não seja o mais importante, pois a mesma vai mudando/evoluindo, e, portanto, a utilização adequada da mesma seja fundamental, sabendo reconhecer o momento de mudanças/migrações tecnológicas. Foi citado o exemplo de utilização de tecnologia em startups, as quais adotam determinadas tecnologias em inícios de operações e vão evoluindo conforme enfrentam seus desafios de crescimento.

A seguir discutiu-se sobre utilização de soluções em nuvem para os projetos de Big Data e para os projetos de TI de forma geral. Levantaram-se questões principalmente com a segurança das informações. Entretanto, de forma geral, o grupo entende que adotar a nuvem em si não é o grande problema, pois se você estiver em um data center seguro, como, por exemplo, Amazon ou Google, o mesmo é possivelmente mais seguro que o data center das empresas, mesmo as maiores. A preocupação que o grupo coloca como sendo de maior importância é a utilização das informações que estão nesses data centers por parte dos provedores. Foi citado o exemplo de como o Google tem a noção exata de onde cada um de nós está utilizando a geolocalização e de como obtém as informações através de APIs dos aplicativos que utilizamos em nosso dia a dia – consentimos a coleta desses dados, mesmo as vezes não notando isso nos temos de uso.

“Saber que estamos sendo vigiados é mais eficiente do que colocar a chave”.

A frase acima citada por um dos presentes representa exatamente a visão de algo extremamente importante no quesito segurança. Devemos atender o usuário, orientar ele na medida do possível e estar preparado para ajudar, porque o que faz a equipe de negócios não seguir as melhores práticas de governança e utilização é justamente não atendê-lo, fazendo com que o mesmo busque caminhos alternativos para suas necessidades.

É neste ponto do debate que entra a questão sobre os desafios da governança que veremos no tópico a seguir.



Desafios de governança, qualidade e modelagem de dados

Um dos participantes que trabalha na área de finanças levantou um ponto que foi debatido intensamente, não somente neste encontro, mas em outros – como estão as iniciativas de governança da informação nas empresas, considerando em especial o

Big Data, que tem potencial imenso de armazenamento? Como tem sido e quais os desafios?

O debate iniciou com a pontuação de um dos presentes de que esse desafio já existe nos ambientes tradicionais, e que cada vez mais teremos esse desafio presente com a passagem para a mão dos usuários de negócio da utilização e manipulação dos dados e informações. A sensação desse participante é de que teríamos “voltado no tempo”, pois anteriormente a ideia do ambiente analítico era que tivéssemos a “única versão da verdade”, e hoje, enfrentamos o risco de termos várias pessoas com diferentes visões dos dados.

O desafio de governança em ambientes tradicionais é importante e nos ambientes de Big Data/Hadoop se torna ainda maior. Neste momento houve a pontuação sobre terminologia com relação ao termo Big Data, o qual se entende que por si só não representa algo relevante ao dizer apenas “dado grande”, mas sim a relevância e a importância da terminologia Business Intelligence, que de fato apresenta o conceito verdadeiro dos ambientes analíticos que trata da “Inteligência de Negócios” – para as corporações não importa o fato de utilizar Oracle, MySQL, Cloudera, o que de fato importa são os conceitos de negócio.

O grupo então debateu sobre a pontuação de “visão única da verdade” e entende que essa é uma discussão importante e que a maioria tem encontrado nas corporações que isso não é algo tão relevante hoje em dia como foi há algum tempo. Se entende que o foco não deve estar em termos únicos e sim em termos bem definidos. Ao invés de buscar quais são os principais indicadores da empresa, os quais normalmente têm revisões e discussões mensais sobre quais devemos utilizar, a relevância está em definir muito bem o que se está medindo. Desta forma, se as equipes investirem seu tempo na definição do que elas estão falando/medindo, as discussões se tornam mais produtivas. Nessa discussão, a participação de TI é fundamental para auxiliar os usuários de negócios na preparação e definição de termos.

Um ponto comentado em conjunto com a governança foi a segurança da informação. O fato é que garantir a segurança é um tema extremamente complexo e difícil – o foco aqui tem que ser em lidar da melhor forma possível com o tema sem prejudicar a utilização da mesma. Devemos buscar a segurança sempre e estar preparado para lidar com possíveis brechas, pois em tempos de internet qualquer celular no bolso se torna uma forma de acesso. Na estratégia de segurança é melhor monitorar do que tentar controlar acessos. Obviamente controles básicos tem que ser implementados, tais como criptografias, mas amarrações em excesso levam os usuários de negócio a outras soluções tais como “shadow IT”, ou seja, seus próprios servidores de informações.

Neste ponto um dos participantes levanta o tema de qualidade de dados com foco nos dados do tipo “não estruturados”, que se encontram mais presentes em ambientes de Big Data.

É comentado que esse tipo de dados não necessariamente são dados “ruins ou sujos”, são apenas dados que não são estruturados (imagem, texto, vídeo, logs etc.). Assim como existem dados estruturados que não são bons, ou não estão tratados, existem dados não estruturados a serem tratados. Entretanto a questão levantada é um pouco mais abrangente, pois além da qualidade em si, existe a diversidade do dado. Então como se trata isso em ambientes Hadoop / Como se faz em Big Data? É diferente do BI tradicional?

Uma visão simplista de mercado traz uma abordagem de que em big data não se trata os dados, são apenas utilizados – essa é uma visão trazida principalmente pelos fornecedores de tecnologia. Isso é dito da mesma forma em que é dito que também não é necessária modelagem de dados em big data. Na visão do grupo e da prática de implementações de projetos, é de que não é exatamente uma visão correta, pois tanto em projetos tradicionais de BI, quanto de Big Data, são necessárias a estruturação dos dados e a qualidade de dados. Caso contrário existirá uma grande dificuldade de utilização por parte das equipes de negócio. E caso não sejam modelados e/ou tratados, essa missão acabará sendo transferida de TI para as equipes de negócio – trabalho de alta complexidade. Um dos grandes valores de TI, se não o maior, está justamente no tratamento e na preparação/modelagem desses dados para negócios utilizar da melhor forma, com tranquilidade, segurança, performance, governança, um dado mais limpo etc.

Com relação ao tipo de dados, a visão entendida pelo grupo após debates é de que de fato não tem como usar um dado não estruturado. Como exemplo, processamento de imagem nada mais é do que estruturar uma coisa que visivelmente parece não estruturada. Você utiliza um algoritmo, divide a imagem em pedacinhos, dá um significado para cada um e pronto, estruturou o dado. Portanto, o dado na origem pode não ser estruturado, mas de alguma maneira você vai dar uma estrutura ao mesmo. E a maioria dos dados não estruturados que a gente lida, não é que eles não são estruturados, eles são estruturados de uma maneira que um banco de dados relacional não consegue lidar muito bem. Um log é um dado estruturado. A navegação da internet é totalmente estruturada, só que ela é estruturada de uma maneira que para o banco de dados tradicional não funciona bem.

Portanto, se o dado não estiver estruturado, se não conseguir estruturar os mesmos, você não vai conseguir usar. Isso é diferente de dizer que é um dado de um banco de dados relacional ou um dado de um ambiente Hadoop – e modelar/organizar os dados é sim muito importante, bem como tratar da qualidade dos dados é muito importante. Como citado no encontro, existem diversos graus de qualidade, a gente trabalha desde o dado mais sujo até o mais limpo, só temos que saber qual nível de qualidade que você tem ali na sua utilização.

Ao se falar de utilização, o grupo iniciou conversas sobre perfil das equipes.



Perfil dos profissionais no mercado: negócios + TI

Existem alguns termos novos que chegaram com o termo Big Data, como, por exemplo, o termo cientista de dados – o grupo debateu um pouco em torno desse tema. Iniciou-se com uma questão sobre a diferença entre um estatístico e um cientista de dados – será que é possível diferenciar os dois? Hoje em dia, as empresas em sua maioria, enxergam o estatístico como alguém com o foco mais acadêmico. E esse foi um ponto importante no encontro: a discussão sobre pessoas acadêmicas e pessoas práticas.

O grupo comenta em torno da teoria de que o conhecimento seria um atributo muito valorizado e isso leva as empresas no sentido de cometerem erros. Enquanto que o mais importante nessa teoria seria a capacidade de aprender, e aprender de forma rápida.

Ao se contratar um especialista, com conhecimento determinado e específico, as chances de errar são grandes, pois o que está se contratando é o conhecimento, até

porque está se contratando um currículo, por vezes supervalorizado. Ao invés disso, se nos processos de entrevistas se buscar domínio de conceitos, capacidade de absorção do seu ambiente/negócio, capacidade de entendimento da conversa, a chance de sucesso é fortemente ampliada.

Em geral, houve consenso de que quem trabalha na área de informação, quem é técnico, tem que aprender de negócio e quem é de negócios tem que aprender tecnologia. Não tem saída, obviamente existem nuances, existirá o especialista técnico que entende pouco de negócio e existirá a pessoa que tem foco somente no negócio, mas também entende um pouco de tecnologia.

Portanto, o grupo percebeu que temos que mudar o perfil radicalmente – as pessoas de TI têm que ser pessoas focadas no negócio, e as pessoas de negócio tem que gostar de tecnologia. O mundo é tecnologia, a pessoa que não gosta de tecnologia hoje está fora do mercado. Hoje em dia, um profissional da área de negócios que não saiba mexer no excel, tem sobrevivido? Não tem. Daqui alguns anos, um profissional que não souber operar um tableau, entrar numa base de dados, fazer análise, criar um modelo estatístico básico, estará fora do mercado. E por outro lado o pessoal de TI tem que entender que o papel de TI vai mudar, vai mudando – antigamente dava-se suporte aos usuários de excel. A mesma coisa vai acontecer com essas ferramentas, mas é um engano achar que a complexidade está diminuindo, não está, ela está aumentando.

Em geral, as empresas líderes de seus segmentos já contam com equipes de negócios que entendem e dominam completamente tecnologia – esse é o caminho para se tornar empresa líder.

“A capacidade de aprender e se adaptar é mais importante do que conhecimentos específicos”



Conclusão

Projetos de Big Data têm alguns desafios: coletar os dados, organização e tratamento dos dados, e saber o que fazer com esse ambiente que você está criando.

Portanto temos que saber coletar o dado, tratar e modelar/organizar o dado com os devidos cuidados para não termos excessos, pois se tivermos muito cuidado no tratamento, os usuários podem se recusar a usar, e por outro lado, se tiver pouco, o usuário não consegue usar.

O ambiente tem que ter utilidade, objetivo, tem que ter gente querendo usar. Não é necessariamente um business case todo estruturado; tem que ter do lado de negócios gente com vontade de usar. Em todos esses anos de implementação e acompanhamento de utilização de ambientes analíticos, os que são descritos como de sucesso foi onde o usuário tinha um papel relevante e principalmente entendia o valor da informação, a qual é utilizada de maneira prática. Projetos que são muito focados em indicadores corporativos (KPIs), geralmente tem dificuldade de sobrevivência. Projetos que estão ligados a casos práticos, como melhorar a conversão no site, melhorar o relacionamento com o cliente, melhorar o problema de fraude, melhorar efetividade do uso de leito, melhorar a assertividade de lançamento de produtos, como vimos no encontro, esses são os que dão certo.

Questão/Mensagem final: o que fazemos para um projeto de big data ter sucesso e ser prático para a empresa? Os participantes responderam: atenda o negócio, equipe multidisciplinar e mista, mudar o perfil das nossas equipes, dar mais valor ao negócio do que à tecnologia.