Tech
Mahindra

EVENT LOG: ENABLING AUTOMATED PROCESS MINING AND DISCOVERY

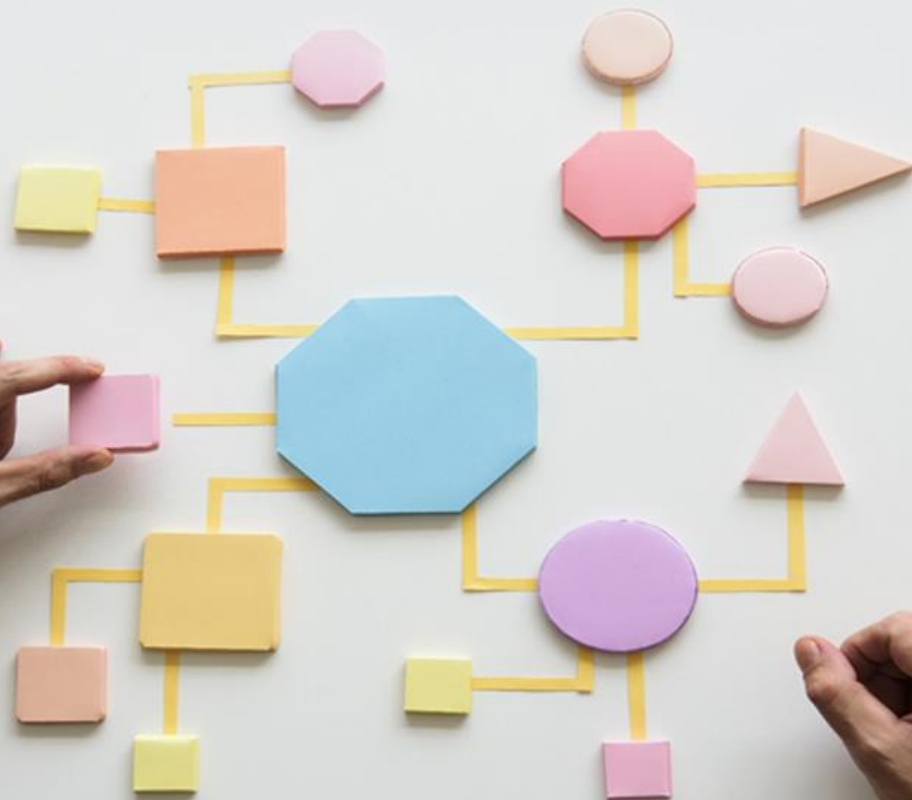## Sri Priya Vardhini N

Table of Contents

Abstract

Process mining is the bridge between data sciences and business process analysis. It is one of the greatest techniques leveraged across industries for process improvement and digital transformation. The input data for any process modelling is an event log. The data from system processes of any organization is extracted and converted into an event log which is to be used as an input for process mining.

This paper discusses the important aspects to consider when defining a conversion of systems data to an event log. Data from IT systems is converted to an XES (eXtensible event stream) format which is used as an input for process mining. There are different decisions that are defined and executed for conversion and these play a major role in the creation of the event log for process mining.

This paper also covers the framework to store aspects of such a conversion. This includes the extraction and definition of traces, their events and their attributes. Traces are smaller units of an event log. A collection of traces forms a log.

Introduction

Process Mining is the method of discovering, monitoring and improving real processes by extracting data from available event logs in the information systems of an organization.
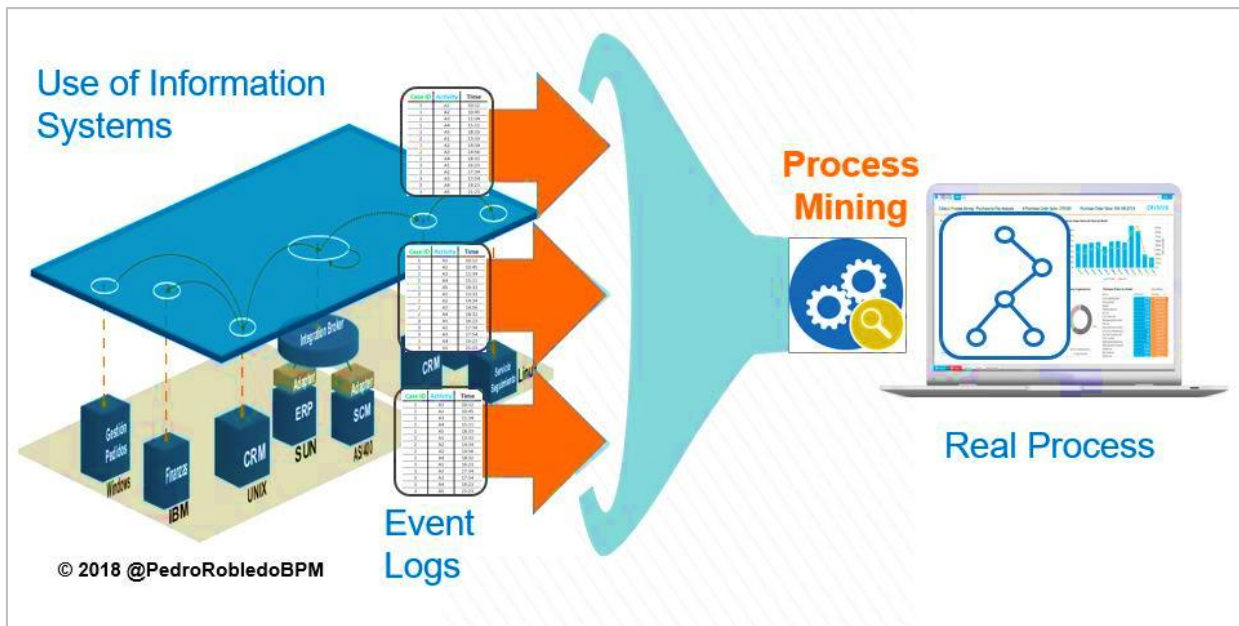

Fig 1

Process mining checks the observed behavior, represented by event data vis-à-vis the automatically discovered process model. An event log is represented by a case ID, activity name, timestamp and other data attributes. It can be extracted from the a database system, a CSV file, a transaction log, a business suite/ ERP system, a message log, an open API etc.

Consider the example of an event log given below:

Order No.	Activity	Timestamp	User	Product	Quantity
100	Register order	22-10-2020@09.15	ABC	iPhoneX	1
102	Register order	22-10-2020@09.18	ABC	iPhoneXl	2
103	Register Order	22-10-2020@09.25	ABC	iPhoneX	1

Table 1.1

The conversion of data to an event log consists of two steps:

1. Definition of conversion- How concepts of data source are mapped into event logs.
2. Execution of the conversion- Converting the data source to event log based on mapping.
A single data source can be subject to multiple event log extractions.

Things to Consider before Process Mining

Process mining can be executed to make an effect on the turnaround time, processing time, process efficiency and so on. Hence, it is required that a clear goal is kept in mind before initiating the activity. From the set goal. The project scope can be derived. This is to select the set of activities whose event logs need to be extracted, which will act as our input data for process mining. To drill down even further, focus of the activity includes extra details to be included in the event log, which needs to be considered before process mining.

Traces

Different activities are executed in a process. Every process has several instances of its execution. These instances can also be called cases. Traces are the activities that are recorded in every instance of a process. Information relating to every activity in a trace can be recorded as an attribute. These are business objects that are handled or used by the business. A trace should contain events relating to a single business object only. Eg: patients, machines, washing machines, orders or items handled etc. Business objects are stored in the master database where information about business objects are recorded and attributes are added to the traces. Based on the scope of the project, the trace is selected and the scope in turn is determined by selecting the business object.
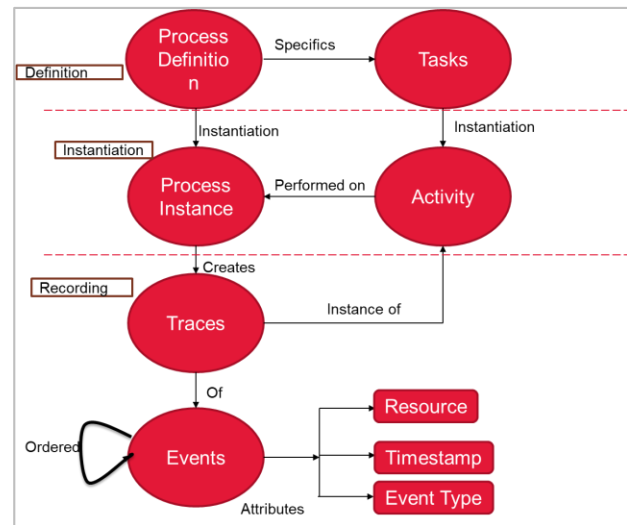

Fig 2. General Event Log Structure

Event Selection

Once the process instance is selected, the events constituting this instance can be selected and specified. Usually all the events are treated equally. The events can be grouped together to form a higher level activity for an instance.

The transactional model of event types are start and complete to indicate the state of a certain activity. The complete type is commonly used. Including the start type is mandatory for certain type of analysis, e.g. for measuring actual processing times, waiting time and other performance measurement analysis.

A single event can relate to multiple process instances and vice versa. A same event can happen multiple times in a process instance and this is called divergence. When a single activity is executed in multiple process instances then it is called convergence.
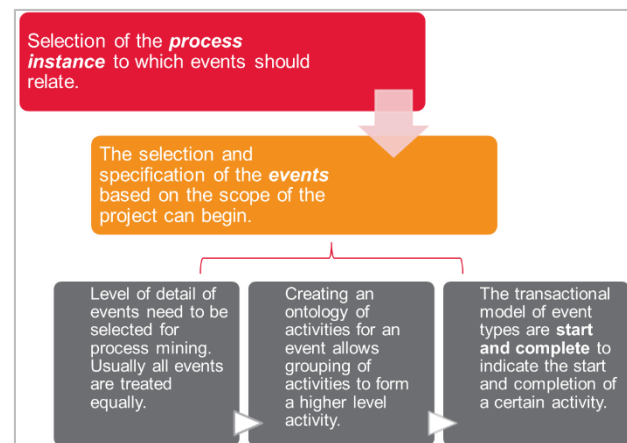

Fig 3. Event Selection

Convergence and divergence can pose issues while working with event logs for process mining. To combat this issue, the representation of the process instance is made same as the identifier of the activity instance, making it to be recognized and treated differently.

One can select and filter an event log by selecting either the traces to be included or the events to be included in an event log. The time period for the trace selection should be five times as long as the duration of an average case. This helps in including complete traces instead of incomplete traces. One can also extract traces based on specifying the start and end of the activity of the trace. This way one can focus on traces that have been executed completely. The subset of traces can be used to discover a process model. For eg. The order process has activities such as create, receive goods, pay invoice. Since these activities follow each other sequentially, they form blocks of a process model.
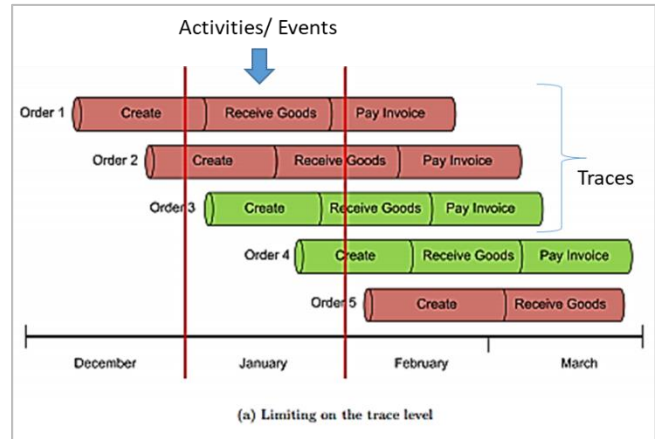
Fig 4. Order 1, 2, 3, 4, 5 are traces of the instance "Order", Create, Receive Goods, Pay Invoice form activities or events of the individual trace

Attribute Selection

Information recorded in event logs is called attribute. Attributes themselves can contain information to store more detailed information.

Generic attributes are those that can be used to specify both for traces and events in general. These are defined by the five standard extensions of XES.

- Concept extension – The name for the log, trace and event elements is specified by the concept extension. Names of events, eg. Invoice Generation should provide the name of the executed activity, eg. Generation of invoice during order processing, represented by the event itself.
- Time extension – The timestamp attribute for the event log containing events is given by the time extension.
- Lifecycle extension – Event types, namely start andcomplete are provided by this extension.
- Organizational extension - This extension defines three different attributes for events – resource, role and group. The distribution of work over different resources can be visualized. The actor of the executing activity needs to be stored as an attribute within each event.
- Semantic extension - This extension adds a modelReference attribute to all the elements in the event log. Grouping of events based on similarity can be performed using this extension.
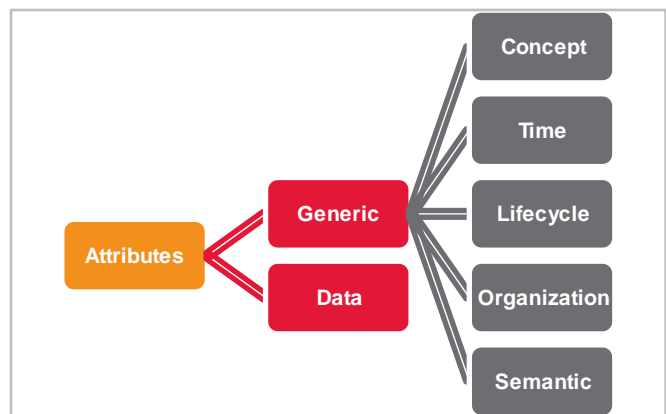
Fig 5. Attribute Classification

Data Attributes - These store the properties of the process instance/ activity executed. Data attributes can form business rules.

Solution Approach

The application prototype to convert s systems data to an XES format must first include the conversion definition from a source data and the conversion execution based on the logic defined in the former.

Requirements of application:

Data source connectivity- Sources of data include relational databases, CSV files, XML files.

Conversion – Conversion definition can start either at the data source or the target event log format. It depends on which area the user has a better understanding.

User friendliness of the application- Structured Query Language (SQL) is one of the easiest and most convenient applications to easily convert the conversion definition to a query that can be run on the data source.

Domain Model:

The domain model stores the entirety of the information needed for the conversion of systems data to an XES event log. This means that for each element that will be present in the created event log, the conversion definition will provide the details on the value and location that can be found in the data source. It consists of the data source, the mapping element, the connection class, the log class, the classifier class, the trace definition, the event class, the General Mapping Item Class, Extension Class, Property Class and Link Class.
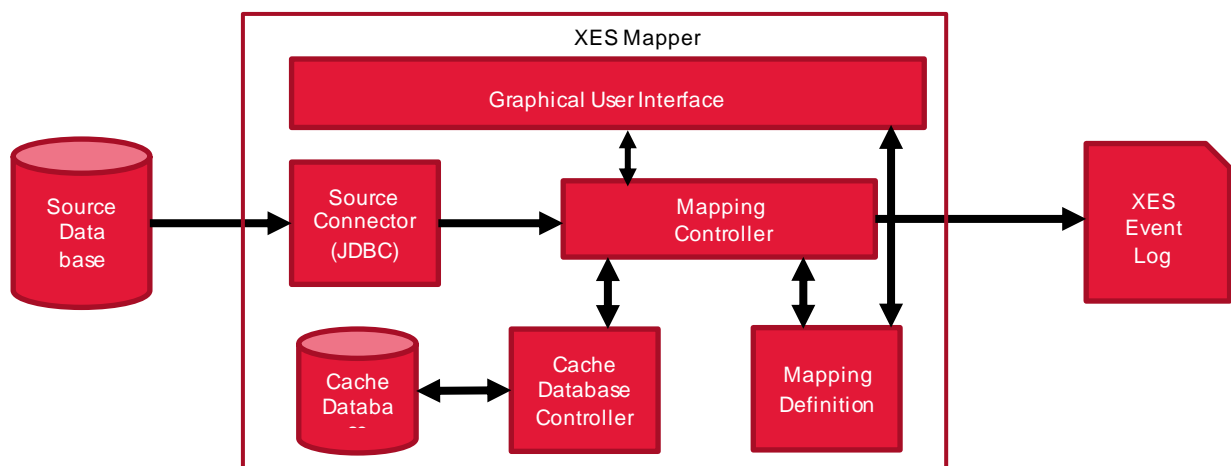
Implementation

Fig 7. The Environment of the Application

Conversion Execution: The data format used for process mining

The XES Mapper or 'XESMa' is an application that guides the definition of a conversion. There is no need to program. This application can also perform the data source to event log conversion. Hence a process analyst can define and execute the conversion on their own using the mapper.

SQL Creation:
- The first step is to create an SQL query for each log, trace and event instance from their conversion definition.
- The total number of queries is therefore equal to the total number of event definitions plus 2 (one log and one trace definition).

Run Queries:
- The second step is to run each of these queries on the source system's database.
- The results of these queries need to be stored in an intermediate database.

Conversion:
- The third step is to convert this intermediate database to the XES event log.
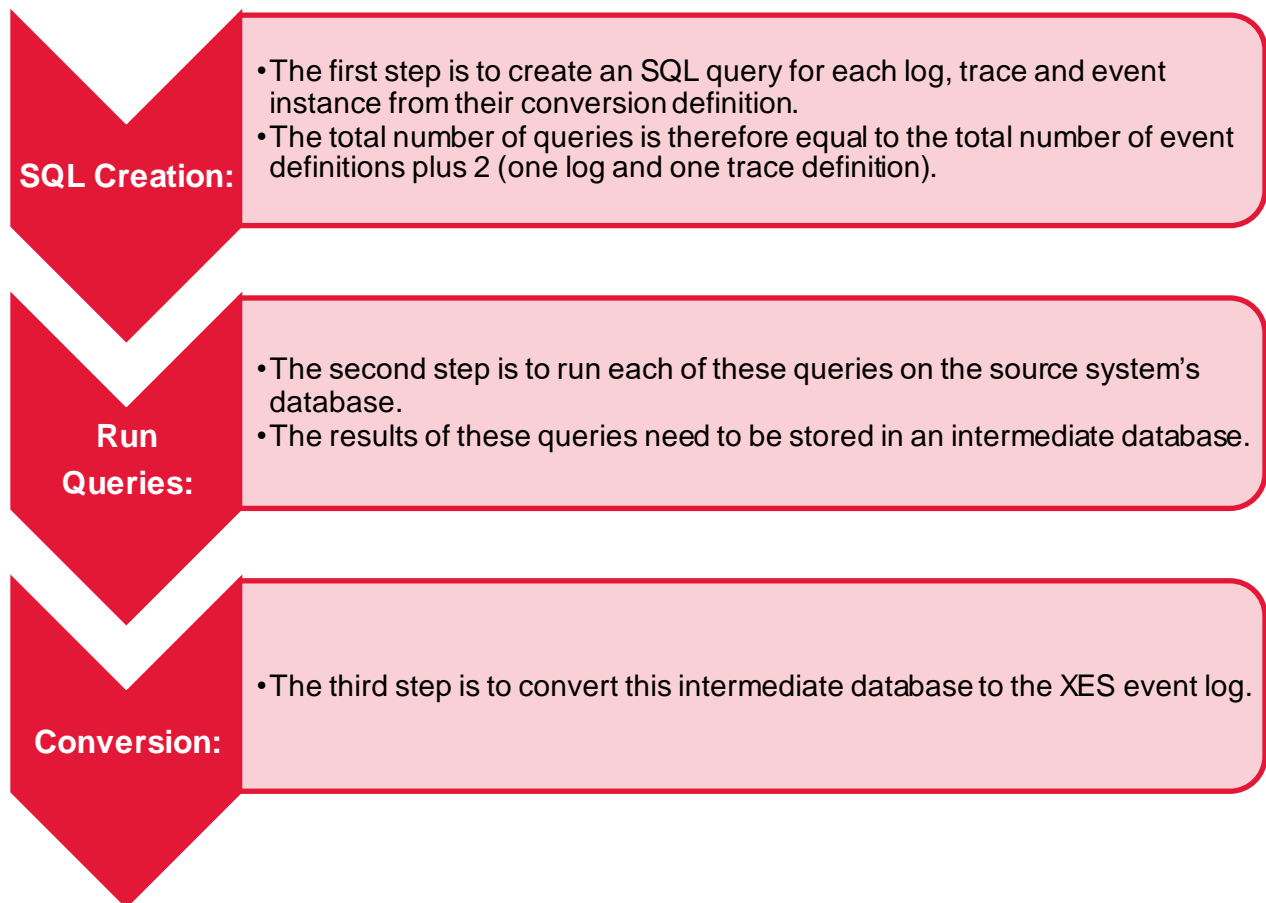
Fig 8. Conversion Execution

After the building and execution of the query, the results are stored in the intermediate database. Hereafter, the third step, the part where the XES event log is created from the data present in the intermediate database, is executed. The OpenXES Java library is used to create the event log table. The header of the table needs to be added before the contents under each of these headers is added. The extension URL, name and prefix are added. The traces and event globals (attributes for traces with value definition) are added. The event classifiers as defined in the conversion is added. The data stored in the intermediate database is used and added to the event log in the following sequence:

- Extraction of log attributes.

- Extraction of traces - The attributes and events are retrieved from the intermediate database per trace, in the order as specified by the conversion definition.
- Combination of the trace attributes and events with their attributes, in the same order as how the data occurs in the event log to maintain the sequence.

Event Log Anomalies

There are certain inconsistencies that are observed in the event log when compared to the process model:

Skipping the Standard Process Entry Point: There are cases where event logs contain data that do not start at the designated starting point. We need to be able to distinguish if this is how data is being recorded and stored in the systems or it is an actual depiction of the process in reality.

Skipping Steps: Process logs also often skip steps. The absence of a step in the event log does not mean that the step did not happen. If a dataset is subjected to process mining for the first time, there is a good chance that it contains such regions of missing data.

Rapidly Progressing through Steps: Related to the previous case are situations where a process quickly skips through a number of steps at a speed that is inconsistent with the expected pace. Some systems do not allow steps within a process to be skipped and thus forcing users to quickly cycle through multiple statues in quick succession. Such rapid progressing through steps is often legitimate, similar to a system completing a series of automation steps.

Going Forward

Given that there are different event log anomalies, for an organization to go ahead with process mining, there needs to be the following key considerations:
- Setting the goal of process mining
- Setting the scope of process mining
- Setting the focus of the project

There are other key aspects that should be looked at from a broader standpoint. The event log may not always provide the correct depiction of the process. Hence a process mining consultant must explore the possibilities of the steps/ events of every process instance. Celonis has introduced task mining as a technique wherein the data is collected from the actors of the business processes by automatically detecting the systems that the users are interacting with.

1. Recording the user interactions – clicks, scrolls, actions with timestamps
2. Adding context – Adding attributes such as user name and other related information
3. Identifying and Clustering – Identifying the activities and grouping them into clusters based on their types. This can be considered as grouping of different events in a trace.
4. Case Matching – Matching the userID and case ID and assessing the impact of the user interactions on business processes and outcomes.

This way, task mining can track the effectiveness against the outcomes that really matter. It has impact on the growth metrics and innovation.

There are different process mining tools that ease the process of data extraction from the source systems. This is done by custom connectors that enable data extraction by means of applications that are

9

designed to be compatible with source systems such as SAP, Oracle, Coupa etc. They have pre-built extraction and transformation to accelerate the process mining journey. Several players in the market have collaborated with analysts and data engineers to build new tools for different data sources specific to a particular industry they operate in. The role of these connectors address the pain point of having to extract data from unknown source systems for process mining. With these connectors deployed, the event collection is done much faster, thereby reducing the overall time and effort of the process mining activity.

References

1. https://medium.com/@pedrorobledobpm/process-mining-plays-an-essential-role-in-digital-transformation-384839236bbe

2. http://www.processmining.org/_media/presentations/event_logs_the_input_for_process_mining.pdf

3. http://www.processmining.org/logs/start

4. http://www.processmining.org/_media/xesame/xesma_thesis_final.pdf

5. https://fluxicon.com/blog/2014/11/data-preparation-for-process-mining-part-i-human-vs-machine/

6. https://www.celonis.com/process-mining/what-is-task-mining/#how-task-mining-works

About the Author

Sri Priya Vardhini

Sri Priya is a business consultant with experience in creating digital transformation stories and strategy consulting for telecom and enterprise clients. She is a six sigma green belt certified professional and a trained Celonis data engineer for process mining. She has also handled CPG and SCM partnerships in the BPS space. She has attained her MBA from NITIE, Mumbai. Currently she is a part of C Consult. She can be contacted on nv00556385@techmahindra.com

About Tech Mahindra

Tech Mahindra represents the connected world, offering innovative and customer-centric information technology experiences, enabling Enterprises, Associates and the Society to Rise™. We are a USD 5.2 billion company with 123,400+ professionals across 90 countries, helping 981 global customers including Fortune 500 companies. Our convergent, digital, design experiences, innovation platforms and reusable assets connect across a number of technologies to deliver tangible business value and experiences to our stakeholders. Tech Mahindra has been recognized amongst India's 50 best companies to work for in 2020 by the Great Place to Work® Institute.

We are part of the USD 21 billion Mahindra Group that employs more than 240,000 people in over 100 countries. The Group operates in the key industries that drive economic growth, enjoying a leadership position in tractors, utility vehicles, after-market, information technology and vacation ownership.

About Tech Mahindra's Business Excellence Services

We are the Business Excellence team, Tech Mahindra's consulting unit. We help clients achieve business objectives in the digital era

- We work with clients to develop and implement digital transformation strategies that impact their products and business models
- We help our clients transform their operations and processes in line with this strategy
- We also help them build a key enabler for achieving these objectives: agility and automation in the technology function
- Our program and change management services ensure on-track implementation of the various transformation initiatives

Proven methodologies, frameworks and tools underpin all of these services. These are based on design thinking approaches that ensure stakeholder buy-in at each stage. Our clients find our global experience, collaborative approach, and the ownership we bring to ensure outcomes in every one of our engagements, as a key differentiator.

BE. Tech Mahindra Consulting
befutureready@techmahindra.com

Tech Mahindra

www.techmahindra.com

connect@techmahindra.com

www.youtube.com/user/techmahindra09

www.facebook.com/techmahindra

www.twitter.com/tech mahindra

www.linkedin.com/company/tech-mahindra