



BAB 1

PENGENALAN DATA RAYA

Data is the new science.
Big data holds the answers.

PAT GELSINGER
(CEO of VMware)

1.1 LATAR BELAKANG

Manusia dan data tidak boleh dipisahkan kerana data menghasilkan ilmu pengetahuan yang boleh memajukan tamadun. Evolusi dan globalisasi teknologi telah memberi impak kepada kepentingan data.



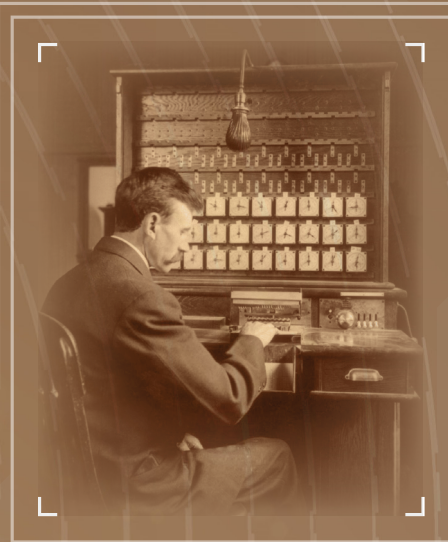
18,000 SM

Rekod jual beli di “Ishago Bone” adalah bukti manusia sudah mula menggunakan data.¹



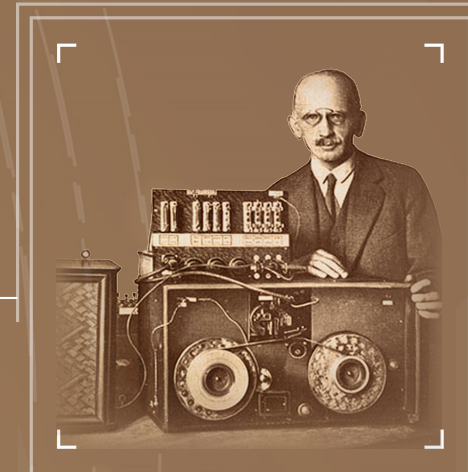
1880

Biro Bancian Amerika Syarikat berdepan masalah keupayaan pemprosesan data yang memerlukan masa selama lapan tahun untuk memproses data tersebut. Data banci 1890 dijangka ambil masa 10 tahun untuk analisis.²



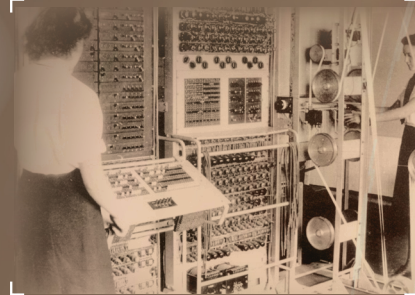
1881

Seorang pekerja biro bernama Herman Hollerith mencipta "Hollerith Tabulating Machine" yang dapat mengurangkan tempoh pemprosesan data.³



1927

Jurutera Jerman-Austria bernama Fritz Pfleumer mencipta kaedah penyimpanan data pita bermagnet yang dipatenkan pada 1928.⁴



1943-1945

Tentera British mencipta “Collosus” untuk memintas rangkaian komunikasi tentera Nazi Jerman.

Mesin boleh mengimbas 5,000 karakter sesaat dan mengurangkan beban kerja mingguan kepada beberapa jam sahaja.⁵



1945

ENIAC (Electronic Numerical Integrator and Computer) dibina untuk menyelesaikan pelbagai masalah pemprosesan data dengan hanya menggunakan satu mesin sahaja.⁶



1977

Komputer peribadi mula memasuki pasaran.⁷



1990

Tim Berners-Lee
daripada European
Organization for Nuclear
Research (CERN)
mencipta *World Wide
Web* (WWW) dengan
tiga pengaturcaraan
asas iaitu HTML, URL
dan HTTP.⁸



1997

Penyelidik NASA,
Michael Cox dan David
Ellsworth menggunakan
istilah “Big Data” untuk
menggambarkan
fenomena super komputer
menjana jumlah data
dengan sangat banyak
yang tidak boleh diproses
dan divisual.⁹



1999

Istilah *Internet of
Things* (IoT) digunakan
pertama kali oleh
Kevin Ashton dalam
pembentangan
di P&G.¹⁰

⁹ dezyre.com

¹⁰ foreignpolicy.com

1.2 ERA DATA RAYA

Ledakan teknologi maklumat dan perkembangan teknologi masa kini telah mencetuskan fenomena data raya yang membuka ruang baharu kepada proses membuat keputusan.

2001



Penganalisis Gartner, Doug Laney memperkenalkan konsep 3V (*high-volume, high-velocity and high-variety*) data raya.¹¹

Roger Mougals, pengarah kajian pasaran syarikat latihan dan teknologi Amerika Syarikat, O'Reiley Media menggunakan secara meluas istilah “Big Data”.¹²

2005



2005

A circular diagram with a blue elephant logo in the center. The word "HADOOP" is written below the elephant. The diagram includes various icons representing data storage and processing, such as folders, documents, and a network of nodes connected by lines. The entire diagram is set against a dark blue background with a subtle pattern of concentric circles and dots.

Yahoo! mencipta perisian Hadoop berdasarkan kerangka perisian sumber terbuka Nutch dan digabungkan bersama model pengaturcaraan Google, MapReduce.

Hadoop mengindekskan keseluruhan WWW dan kini merupakan sumber terbuka yang diguna pakai oleh pelbagai organisasi seluruh dunia untuk pemprosesan data.¹³

¹¹ forbes.com

¹² researchtrends.com

¹³ wired.com

2008



Google memproses
20 PETABYTE
data setiap hari.¹⁴

Ketua Pegawai Eksekutif Google ketika itu, Eric Smith berkata sejak bermulanya tamadun manusia sehingga 2003, **5 EXABYTE** data telah dihasilkan oleh manusia. Kini, 5 exabyte data dicipta manusia hanya dalam tempoh dua hari.¹⁵

2010



2020



Dalam tempoh hanya seminit, 204 juta e-mel dihantar, 47,000 aplikasi dimuat turun, US\$83,000 jualan di Amazon, 20 juta tontonan gambar di Flickr, 320 akaun Twitter baharu dicipta dengan 100,000 twit dikeluarkan, 2 juta carian di Google dan 1.3 juta tontonan video di YouTube.¹⁶

datacenterknowledge.com 14

techcrunch.com 15

Malaysia Digital Economy Corporation 16

1.3

DEFINISI DATA RAYA

McKinsey
& Company

Kuantiti set data yang saiznya melampaui keupayaan perisian pangkalan data biasa bagi merakam, menyimpan, menguruskan dan menganalisisnya.

(Sumber: mckinsey.com)

Gartner®

Data raya didefinisikan sebagai 3V iaitu *High-Volume*, *High-Velocity* dan *High-Variety*.

(Sumber: gartner.com)



VOLUME

Saiz atau kuantiti data yang besar sehingga tidak dapat ditampung oleh storan biasa.

Dianggarkan 2.5 quintillion bytes data dicipta setiap hari dan hasilnya 40 zettabytes data dicipta menjelang 2020, kenaikan 300 kali ganda berbanding data dicipta pada 2005.



VELOCITY

Kadar kepantasan aliran data yang tinggi.

Data perlu diproses pada masa yang tepat untuk menentukan potensi kegunaan data tersebut.



VARIETY

Kepelbagaian format data yang tercipta.

Format data sama ada dalam bentuk berstruktur (pangkalan data konvensional), semi-struktur dan tidak berstruktur (audio, video, e-mel)



1.4 ANALITIS DATA RAYA

Analitis data raya ialah proses menganalisis data yang berskala besar untuk merungkai corak dan perkaitan tersembunyi.

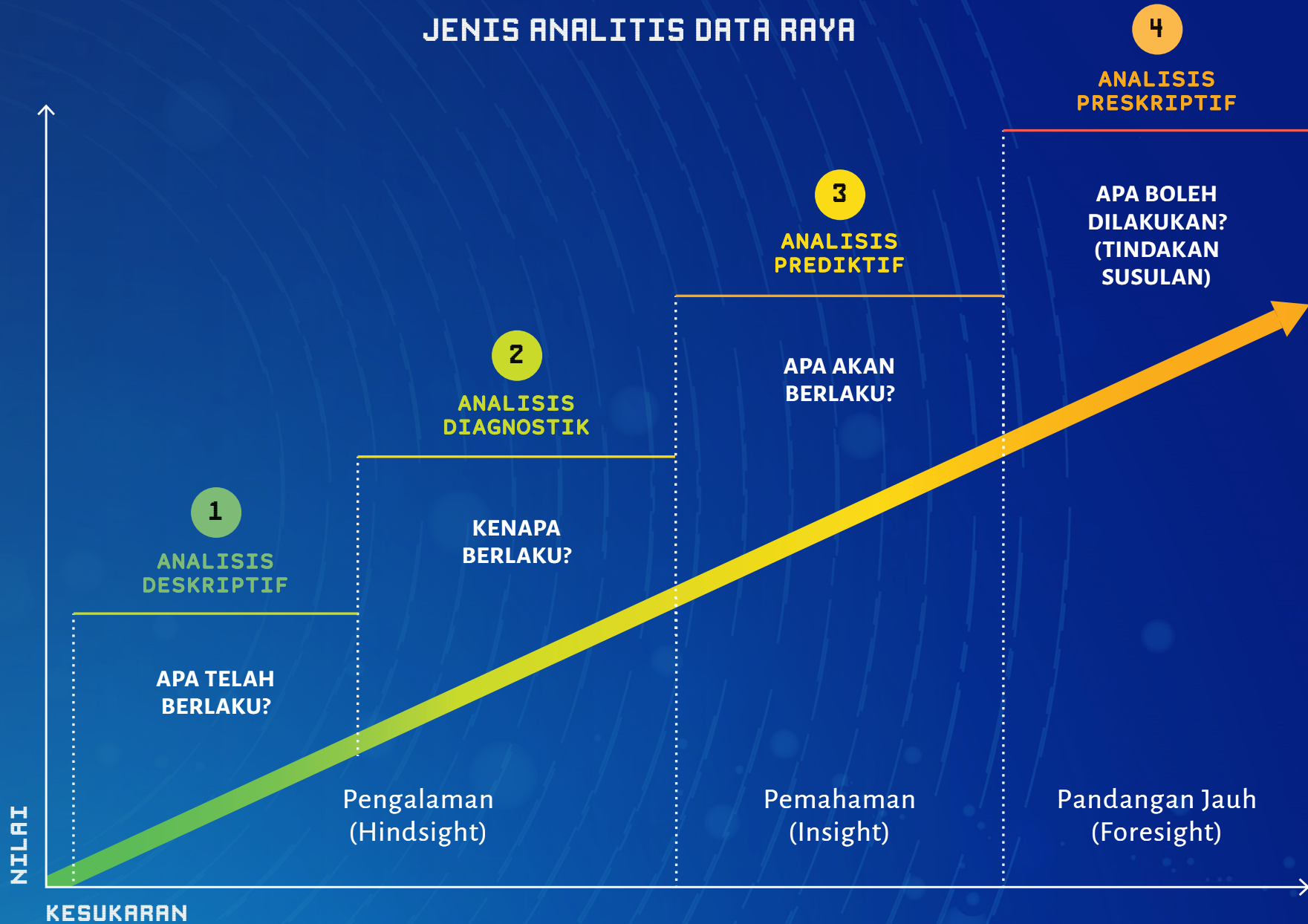
(Sumber: whalts.com)

JENIS ANALITIS DATA RAYA

ANALISIS DESKRIPTIF	ANALISIS DIAGNOSTIK	ANALISIS PREDIKTIF	ANALISIS PRESKRIPTIF
Gambaran atau rumusan data sedia ada untuk menerangkan apa yang sedang atau telah berlaku	Analisis lanjutan untuk mengenal pasti sebab sesuatu berlaku	Ramalan hasil akhir menggunakan model statistik dan teknik pembelajaran mesin	Cadangan satu atau lebih tindakan susulan atau penyelesaian selepas analisis data
CONTOH Mengetahui pasti bilangan murid yang tercalar daripada sistem pembelajaran di Malaysia	CONTOH Mengetahui pasti sebab-sebab berlakunya kecacatan	CONTOH Meramalkan implikasi kecacatan murid daripada aspek sosial dan ekonomi	CONTOH Mengetahui pasti tindakan penyelesaian bagi mengurangkan masalah kecacatan

Sumber:
www.analyticsinsight.net/four-types-of-business-analytics-to-know
www.bigdatatraining.com/blog/types-of-data-analytics

JENIS ANALITIS DATA RAYA



1.5

DATA RAYA SEBAGAI PEMBOLEH DAYA PEMBANGUNAN NEGARA

Agenda pembangunan negara memerlukan data raya sebagai salah satu pemboleh daya bagi memastikan kejayaan dan kelancaran pelaksanaan dasar. Inisiatif data raya dinyatakan dalam Rancangan Malaysia Lima Tahun (RMLT), Wawasan Kemakmuran Bersama 2030, Rangka Tindakan Ekonomi Digital Malaysia (MyDigital), Polisi Nasional Industri 4.0 (Industry 4WRD) dan Pelan Strategik Pendigitalan Sektor Awam 2021-2025.

RANCANGAN MALAYSIA KESEBELAS

2016-2020

BAB 9 BIDANG FOKUS A:

Mentransformasi perkhidmatan awam untuk produktiviti

STRATEGI A3:

Memanfaatkan data bagi meningkatkan *outcome* dan mengurangkan kos

RANCANGAN MALAYSIA KEDUABELAS

2021-2025

TONGGAK

- ◆ Reformasi Institusi
- ◆ Dimensi Pemerkasaan Ekonomi

PELAN STRATEGIK PENDIGITALAN SEKTOR AWAM 2021-2025

T1: Pemanfaatan
Kepintaran Data

S3: Pemerkasaan Penyampaian
Perkhidmatan Kerajaan
Berasaskan Kepintaran Data



■ VISI

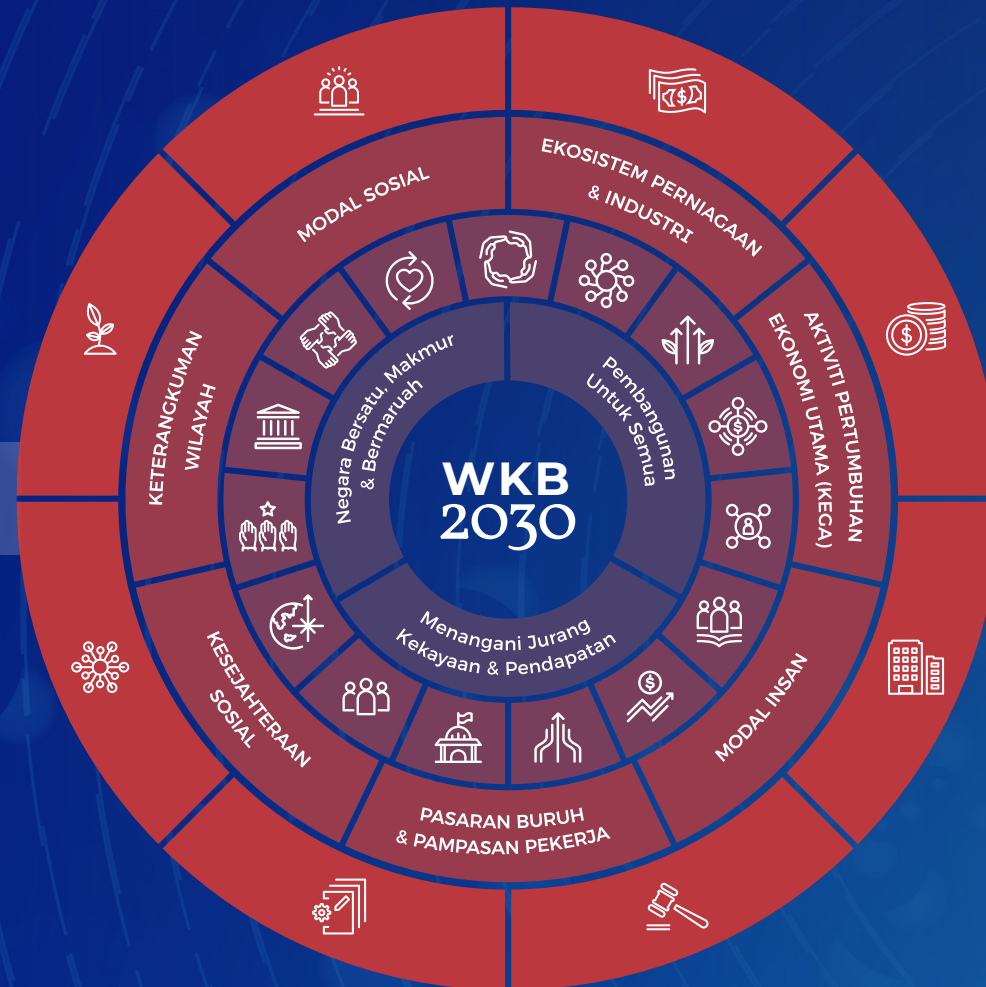
■ TERAS
STRATEGIK

■ PRINSIP
PANDUAN

■ EKOSISTEM
PEMBOLEH DAYA

WAWASAN KEMAKMURAN BERSAMA 2030

Pemboleh daya: Data raya



OBJEKTIF
 PRINSIP PANDUAN
 TERAS STRATEGIK
 PEMBOLEH DAYA



PENJIMATAN MASA

Memudahkan proses mengenal pasti sumber data untuk membantu analisis dilakukan serta merta dan keputusan pantas dibuat.



PENAWARAN “HINDSIGHT, INSIGHT DAN FORESIGHT”

Meningkatkan kemampuan organisasi untuk memenuhi permintaan pelanggan.



PENJIMATAN KOS

Mengurangkan kos pengurusan data dalam organisasi.



PEMACU INOVASI DAN PRODUKTIVITI

Membolehkan penciptaan idea/perkhidmatan/produk baru berdasarkan hasil analisis.

KEPENTINGAN DATA RAYA

Sumber:

www.rd-alliance.org/group/big-data-ig-data-development-ig/wiki/big-data-definition-importance-examples-tools

1.6

TAHAP KEMATANGAN BIG DATA ANALYTICS (BDA)

DEFINISI Mengukur tahap kesediaan dan pelaksanaan BDA



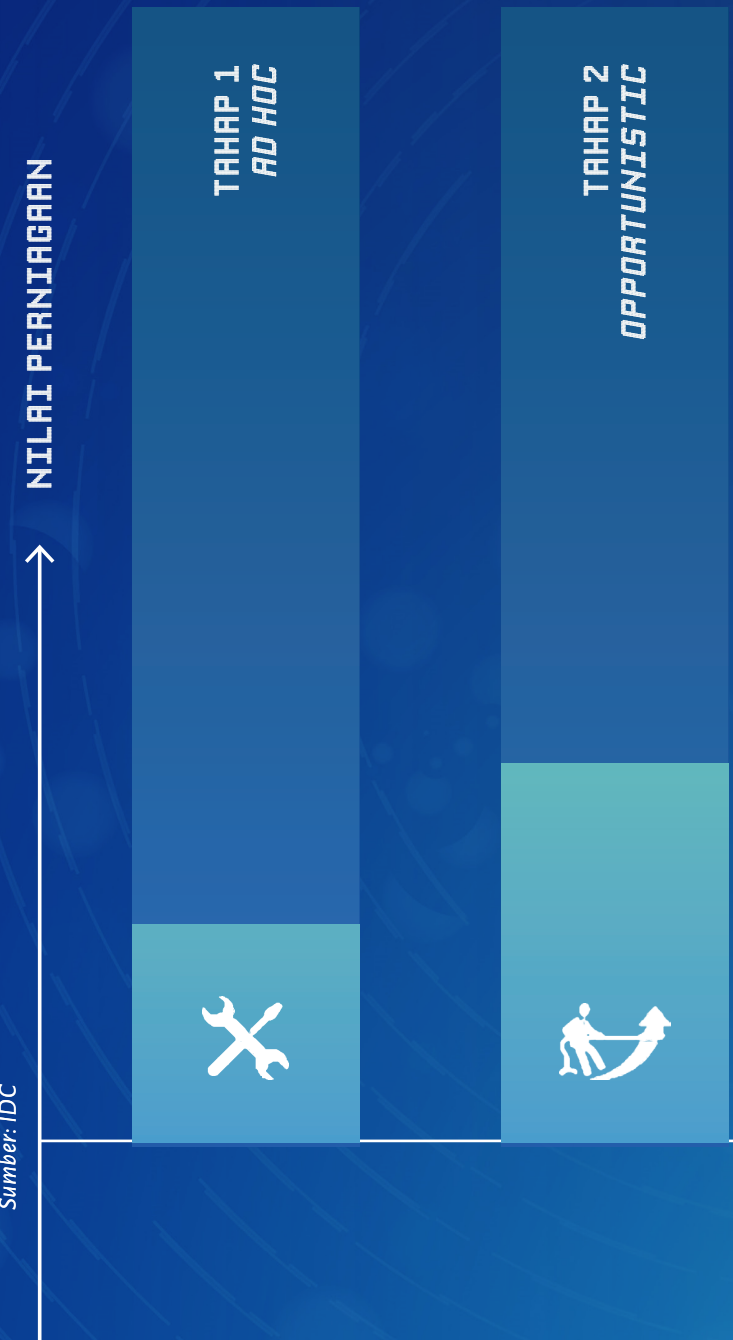
TAHAP AD HOC

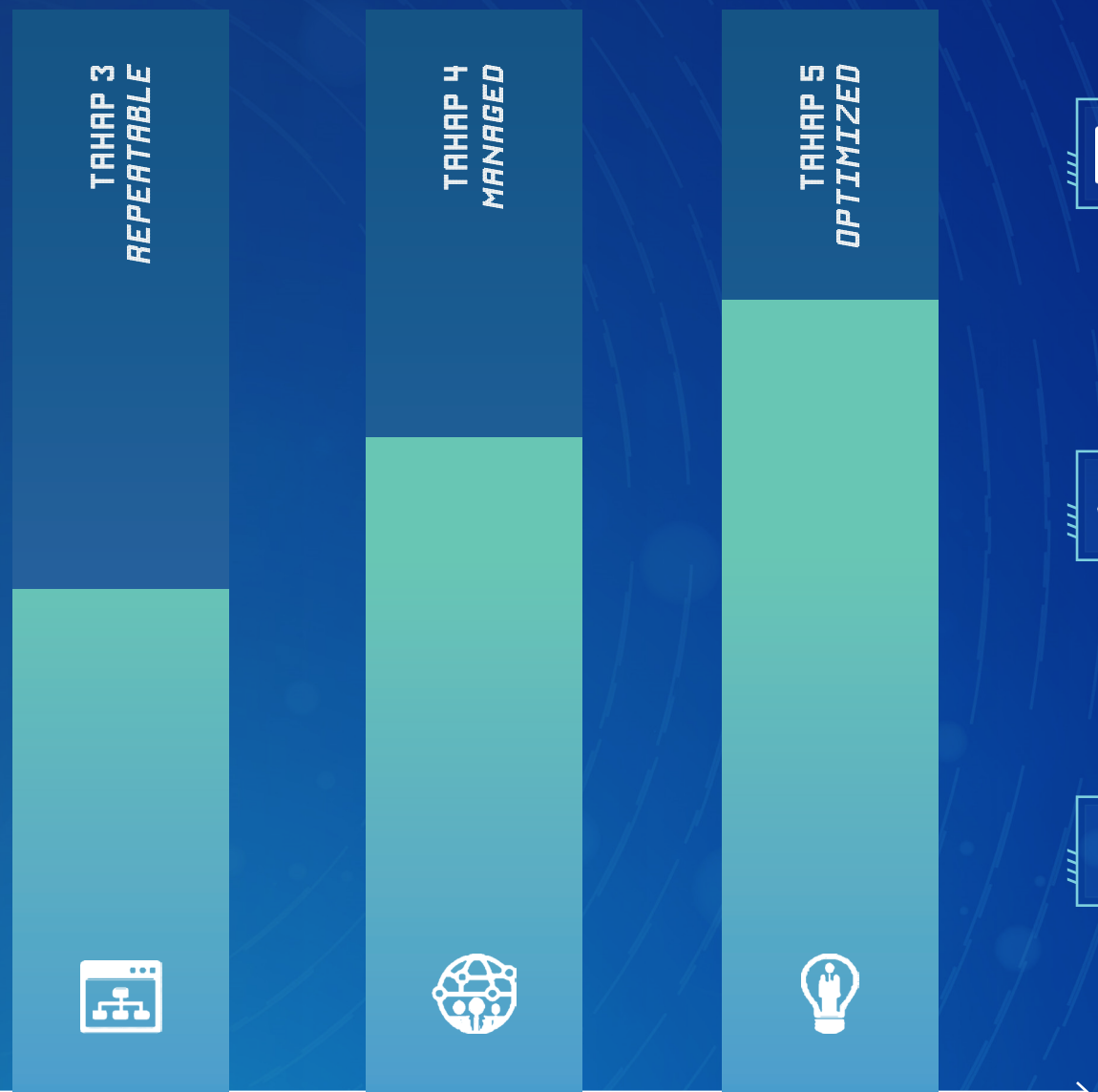
- ◆ Projek perintis atau 'proof-of-concept'
- ◆ Pengalaman pelaksanaan memberikan pengalaman baharu



TAHAP OPPORTUNISTIC

- ◆ Pengetahuan baharu bertambah
- ◆ Peluang perniagaan bernilai tinggi lebih jelas





TAHAP REPEATABLE

- ◆ Pengamalan berterusan
- ◆ Nilai kepada perniagaan sudah diketahui tetapi masih tertumpu kepada unit tertentu sahaja



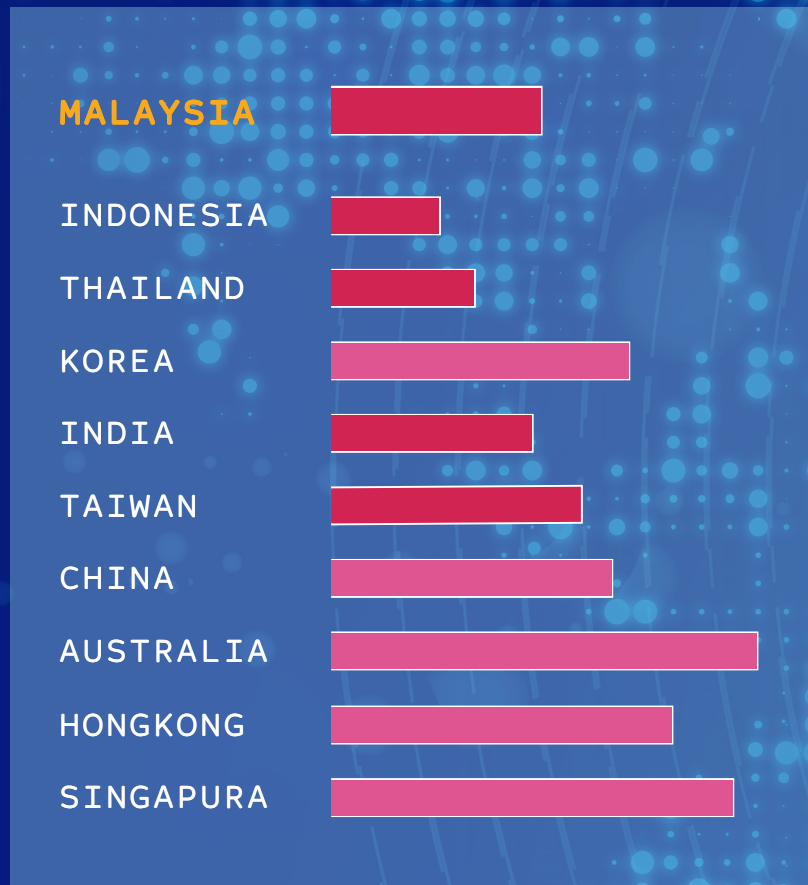
TAHAP MANAGED

- ◆ Pelaksanaan menyeluruh
- ◆ Produk dan peluang baharu perniagaan menjadi pelan perniagaan syarikat



TAHAP OPTIMIZED

- ◆ Pembelajaran dan penambahbaikan berterusan
- ◆ Peluang dan nilai baharu dalam perniagaan dihasilkan secara berterusan



Sumber: IDC

TAHAP KEMATANGAN BDA DI MALAYSIA

Malaysia berada pada Tahap *Ad Hoc* Kematangan BDA menurut kajian yang dibuat oleh MDEC dan IDC pada tahun 2015.

TAHAP

AD HOC

OPPORTUNISTIC

REPEATABLE

MANAGED

OPTIMIZED

1.7

PENCAPAIAN E-GOVERNMENT DEVELOPMENT INDEX (EGDI) MALAYSIA



E-Government Development Index (EGDI) adalah merupakan pembolehdya kepada pelaksanaan agenda Pembangunan Mampan Dunia 2030 Pertubuhan Bangsa-Bangsa Bersatu (UN 2030 SDG AGENDA).

Kedudukan EGD Malaysia meningkat saban tahun. Pada tahun 2016, negara berada pada tangga ke-60 dan melonjak kepada kedudukan ke-48 pada tahun 2018.

Pada tahun 2020, Malaysia berada pada kedudukan ke-47 daripada 193 buah negara anggota Pertubuhan Bangsa-Bangsa Bersatu (PBB). Laporan EGD 2020 meletakkan Malaysia antara tujuh negara baharu Asia yang memasuki kategori *Very-High EGD* (V1 EGD), salah satu daripada empat kategori V1 EGD, membuktikan wujud kemajuan dalam pelaksanaan inisiatif e-Kerajaan di Malaysia. Nilai OSI dan perkhidmatan dalam talian Malaysia juga dianggap PBB hampir setaraf dengan negara berpendapatan tinggi.¹⁹





Peningkatan kedudukan Malaysia saban tahun ini didorong oleh beberapa inisiatif data raya dalam agenda pendigital Negara yang dilancarkan sejak tahun 2012.

Pada tahun 2016, Suruhanjaya Komunikasi dan Multimedia Malaysia (SKMM) telah menjalankan kerjasama strategik dengan Digi Telecommunications dan Majlis Organisasi Wanita Kebangsaan (NCWO) dalam inisiatif Digi Wanita Era Digital yang menggunakan Pusat Internet 1Malaysia bagi latihan penggunaan peranti teknologi pintar dan internet.

Malaysia juga melancarkan inisiatif Asean Data Analytics Exchange (ADAX) pada tahun 2017 yang merupakan platform digital serantau bagi mengumpulkan bakat, model pembangunan dan teknologi analitik terkini.²⁰

Sejak dilancarkan, ADAX telah melatih lebih 1,800 orang daripada 298 buah syarikat dalam 19 industri.