Navigating the World of Data *from an enterprise perspective*

This two-page document has clickable links to posts at begincodingnow.com. I present these links from an enterprise-wide perspective and then from a data analytics perspective.
Page 1 - Enterprise Architecture – the four parts are: Business, Data, Technical & Application Architecture with undercurrents of Data Availability and Security
Page 2 – Data Analytics, Data Engineering and Data Science
Sources: Fundamentals of Data Engineering by Joe Reis & Matt Housley; Data Science and Big Data Analytics by EMC Education Services

Environment Analysis	External Stakeholders:	Business Architecture						External Stakeholders & Analysis:
Analyze to better understand the enterprise and how it relates to its environment. SLEPT Analysis: Social Legal Economic Political Technological SWOT Analysis: Strengths Weaknesses Opportunities Threats	Product Suppliers, Advertisers, Distributors Legal, Accounting	CEO Chief Executive Officer	CHRM Chief Human Resources Manager	CFO Chief Financial Officer	CTO/CIO Chief Technology Officer Chief Information Officer	COO Chief Operations Officer	CMO Chief Marketing Officer	Customers, Market Research, Surveys, A/B Testing
		Reports to Board, complies with Government Laws & Regulations, Triple Bottom Line	Hiring, Retention, Learning and Growth	Accounting, Finance, Budgeting, Project Management	Hardware, Software, Network,Cloud, Data, Reports, Dashboards, Monitoring	Manufacturing, Procurement, Production, Supply Chain, Technology, Robotics, Shipping	Sales, CRM, Market Research, Marketing Strategy, Target Marketing	
		Types of Data Analytics & Use Cases:	Culture, Values, Education, Authority	Business Intelligence - sales, profits,		Operational Data Analytics – inventory, real-time dashboards, application health,	Embedded Data Analytics – (customer-facing) sales, sales by segment, demographics, customer churn, customer satisfaction surveys	
	The Organization's Environment							

Data Architecture – The design of systems to support the changing data needs of an organization using the principles of good value, trade-offs and flexibility.
The Data Science Hierarchy of Needs (basic to advanced) are: Gather, Move and Store, Explore and Transform, Aggregate and Label, Learn and Optimize, AI and Deep Learning
Organizational Data Maturity - The progression toward higher data utilization, capabilities and integration. Below are the **three stages** of data maturity.

1. Starting with Data	2. Scaling with Data	3. Leading with Data
Data policies and goals need further refinement, data architectures are unfinished, most reports are ad hoc,	Formal data practices, few ad hoc data requests, creating a scaleable data architecture, planning for future,	Data-driven, automated pipelines, self-serve data analytics,

Data Analysis – A Few Methodologies and Frameworks – a methodology is a step-by-step systematic approach, whereas a framework is ...						
Plan Do Check Act	Plan	Do	Check	Act		
Project Management	Initiation	Planning	Starting	Performing	Closure	
EMC Education Services – The Data Analytics Life Cycle	Discovery Do we have enough information to draft a plan?	Data Prep Is there good enough quality data?	Model Plan Do we have a good idea about the model to try?	Model Build Is the model robust enough? Rebuild it?	Communication Are we successfully communicating results to technical and non-technical people?	Operations Take it out of the sandbox and into production.
Google's PACE Framework (Agile approach)	Plan & Analyze			Construct		Execute

The World of Data Analytics – Links to Blog Posts at BeginCodingNow.com

Types of Data	Sources of Data	Data Engineering Life Cycle Book: Fundamentals of Data Engineering by Joe Reis & Matt Housley					Data Analytics Life Cycle (Google's PACE from Coursera Course on Advanced Data Analytics) Data Analytics Types: Business Intelligence, Operational Embedded SEE BELOW for more information...				Data Science ML, AI, Deep Learning
Structured Unstructured Quantitative Qualitative Categorical Relationships -Nominal -Ordinal -Interval -Hierarchical	Internal External See EMC…	**Generation**	**Ingestion** ETL, Push Pull	**Transformation** After storage, make the data useful. Format, Clean, Structure,	**Serving** data to Data Analysis and Data Science (ML)	Data Analytics & Machine Learning (ML) – the data engineering life cycle serves data to Data Analysts and Data Scientists. Data engineering sits upstream from Data Analytics and Data Science	**Plan** start with business objectives and the users	**Analyze** EDA is: discovering, structuring, cleaning, joining, validating & presenting. **Google Course...**	**Construct** models, data visualization	**Execute** Visualization Presentation	Follows PACE ML Types: Supervised Unsupervised Reinforcement, Natural Language Processing (NLP), Clustering,
			Data Storage On-premise, Cloud, Batch, Streaming								
		Security A top priority	**Data Management** Data Governance	**DataOps** DataOps is similar to DevOps except for data	**Data Architecture** Data warehouse, data marts	**Orchestration** **DataEng. p. 66**	**Software Engineering (Software Dev.)** SDLC DevOps, Agile	Gartner Analytic Continuum,	Statistics, Sampling, Central Tendency Dispersion		
Roles RACI Matrix	Business Analyst	Database Administrator (DBA); Cloud Administrator, Data Engineer, Network Engineer, Cloud Engineer					Data Analyst Skills - Understanding context (this map will help!), curiosity, analytical, team player, communication, technical skills.				Data Scientist
Tools and Technologies	Databases, Documents, Excel, csv, tsv, JSON, XML, HTML)	Databases SQL, Hadoop,	Apache Spark	SQL, Feature Engineering,		Apache Airflow	Databases, Files (Excel, csv, tsv, JSON)	SQL, T-SQL, Python, R, pandas, Tableau, Excel, Power Query, Power BI,			Python, SciKit Learn,

Data Analytics & Machine Learning (ML) – Google's PACE Framework – Questions, Methodologies, Activities, and Models

Plan	Analyze	Construct		Execute
Who? What do they need to know and do? What data is required?	Import Data Exploratory Data Analysis (EDA) Discovering, Structuring, Cleaning, Joining, Validating & Presenting. Data Visualization (as part of EDA -see Anscombe's Quartet) **...Details here…** Outliers Duplicates Errors in data Feature Engineering	**Data Analytics:** Descriptive Statistics (mean, median, mode, SD, …) Probability Inferential Statistics Linear Regression Logistic Regression Multiple Linear Regression	**Machine Learning:** K-Means Clustering Decision Tree	Data Visualization Steps: -Understand the Context -Choose a Display -Eliminate Clutter -Focus Attention -Think like a Designer -Tell a Story Presentations