অসীম সম্ভাবনা নিয়ে তথ্যবিজ্ঞানীদের ভবিষ্যৎ বিগডাটা

মর্তুজা আশীষ আহমেদ ও সৈয়দ হাসান মাহমুদ

তথ্যপ্রযুক্তি বলতে আমরা আসলে কী বুঝি। এক কথায় বলতে গেলে যে প্রযুক্তির সাহায্যে সঠিক ও কার্যকর উপায়ে তথ্য ও সামগ্রিক তথ্যব্যবস্থার ব্যবহার নিশ্চিত করা যায়, সেটাই তথ্যপ্রযুক্তি। কমপিউটার প্রকৌশল শুধুই যে হার্ডওয়্যার সম্পর্কিত প্রকৌশল, তা নয়। হার্ডওয়্যারের পাশাপাশি তথ্যের ব্যবস্থাপনাও এর বড় একটি অংশ। এজন্যই অন্যান্য প্রকৌশল বা বিজ্ঞানের চেয়ে কমপিউটার প্রকৌশল বা কমপিউটার বিজ্ঞান একটু আলাদা ধরনের বিষয়। এই তথ্যের আকার যখন নির্দিষ্ট সীমার মধ্যে থাকে, তখন এর ব্যবস্থাপনা নিয়ে তেমন বেগ পেতে হয় না। কিন্তু এই তথ্যের ব্যবস্থাপনা যখন বিশাল বা অসীম হয় তখন তথ্য ব্যবস্থাপনার আসল মাহাত্ম্য বেরিয়ে আসে। এই বিশাল তথ্যের ব্যবস্থাপনা থেকেই উৎপত্তি বিগডাটার।

তথ্য ও উপাত্ত

ইনফরমেশন অর্থ তথ্য এবং ডাটা অর্থ উপাত্ত। দুটোই নির্দিষ্ট কোনো বিষয়ের ওপর জ্ঞান হতে পারে। তবে এই তথ্য ও উপাত্তের মধ্যে আভিধানিক তেমন একটা পার্থক্য না থাকলেও এর ব্যাখ্যাতে বিশাল পার্থক্য রয়েছে। উপাত্ত বা ডাটা যেকোনো কিছুর সঙ্কলন বা সন্নিবেশন হতে পারে। তবে এই সঙ্কলন যখন কোনো নির্দিষ্ট উপায়ে বা যেকোনো ধারাবাহিকতায় সাজানো হবে এবং যা অর্থ বহন করবে সেটিই হচ্ছে তথ্য বা ইনফরমেশন। তথ্যপ্রযুক্তির প্রফেশনালদের যাবতীয় কাজ এই তথ্য ও উপাত্তকে ঘিরেই। আর উপাত্তকে প্রসেস করে যে তথ্য উৎপন্ন হয় তার সঠিক সন্নিবেশন হচ্ছে ডাটাবেজ। একটা ডাটাবেজ হচ্ছে অনেক তথ্যের সমাহার। এই তথ্য ও উপাত্ত নিয়ে তৈরি করা ডাটাবেজ যখন বিশাল বা অসীম ডাটা নিয়ে কাজ করে সেটাই বিগডাটা।

কী এই বিগডাটা?

বিগডাটা হচ্ছে তথ্যের বিশাল সন্নিবেশন ব্যবস্থাপনার ধরন। বর্তমানে বিগডাটা ম্যানেজ করছে এমন অনেকগুলো কোম্পানি আছে। এসব কোম্পানির কাজ হচ্ছে বিশাল আকারের তথ্যের ব্যবস্থাপনা ও তার ব্যবহারের যথোপযোগিতা ও তার পাশাপাশি বাহুল্যতা কমিয়ে আনা। যখন উপাত্ত সাধারণ আকারের হয় তখন তা থেকে সঠিক ব্যবস্থাপনার মাধ্যমে তাকে তথ্যে পরিণত করা হয়ে থাকে। কিন্তু যখন উপাত্ত অনেক বড় বা অনেক বিশাল হয়, তখন তাকে তথ্যে পরিণত করা কঠিন। উপাত্ত এত বিশাল হতে পারে যে সাধারণ ডাটাবেজের পক্ষে বিশ্লেষণ করা সম্ভব হবে না। এই তথ্যের ট্রাফিক এতই বিশাল ও চলাচল এতই দ্রুত হবে যে, ডাটাবেজ একে প্রসেস করতে পারবে না। এটাই বিগডাটা। বর্তমান বিশ্বে এই বিগডাটা নিয়ে বেশ আলোচনা হচ্ছে। ধারণা করা হচ্ছে যেহারে উপাত্তের আকার বাড়ছে তাতে এর ব্যবস্থাপনা বেশ দুরূহ হয়ে পড়ছে বলে নিকট ভবিষ্যতে বিগডাটা ব্যবস্থাপনার লোক পাওয়া যাবে না। ফলে বিগডাটা ম্যানেজ করার এক সঙ্কট তৈরি হবে। তাছাড়াও দিন-দিন বিগডাটা ম্যানেজারের চাহিদা দ্রুতহারে বাড়ছে বলে বিগডাটা নিয়ে আলোচনা চলছে সর্বত্র।

বিগডাটার ইতিহাস

ডাটা সায়েন্টিস্ট বা তথ্যবিজ্ঞানী (আভিধানিক অর্থে এটি উপাত্ত বিজ্ঞানী কিন্তু এখানে তথ্যবিজ্ঞানী উল্লেখ করা হলো) একটি নতুন শব্দ। কিছুদিন আগেও এই শব্দের কোনো অস্তিত্ব ছিল না। গুগলেও ইদানিং এই শব্দ খোঁজার পরিমাণ বাড়ছে। তথ্যবিজ্ঞানীরা পে-পাল, অ্যামাজন, ই-বে, এইচপি, আইবিএমের মতো বড় বড় কোম্পানিতে এখন কাজ করছেন। তথ্যবিজ্ঞানীদের কাজ হচ্ছে বিশাল আকারের ডাটাবেজ প্রসেস ও রক্ষণাবেক্ষণ করা। এর মধ্যে অন্যতম হচ্ছে বিগডাটা।

তথ্যবিজ্ঞানী শব্দটির উৎপত্তি হয় প্রথম ১৯৬৮ সালে। সে বছর আইএফআইপি কংগ্রেসে তথ্য-বিজ্ঞান শব্দের সৃষ্টি। মূলত তথ্য ব্যবস্থাপনা থেকেই ডাটা সায়েন্স বা তথ্য-বিজ্ঞান শব্দ যুগল আসে। ১৯৯৭ সালে উপাত্ত সংগ্রহ, সংরক্ষণ ও পুনরায় ব্যবহারোপযোগিতা আরো আধুনিক হয়। এ বছরে Knowledge Discovery and Data Mining বিষয়ে প্রকাশনা বের হয়। যদিও ডাটা মাইনিং ও বিগডাটা এক জিনিস নয়। তবে একটি আরেকটির পরিপূরক অবশ্যই। কাউকে ছেড়ে কাউকে বাদ দেবার কোনো উপায় নেই। ডাটা মাইনিং ও তথ্যবিজ্ঞানে এটি একটি গুরুত্বপূর্ণ অধ্যায়। ২০০১ সালে বিখ্যাত বেল ল্যাবরেটরি থেকে প্রকাশিত হয় Data Science: An Action Plan for Expanding the Technical Areas of the Field of Statistics। তখন থেকেই তথ্য বিজ্ঞান মানুষের নজরে আসে। এভাবেই তথ্যবিজ্ঞান ক্রমশ উন্নত হচ্ছে এবং এর চরম নিদর্শন হচ্ছে ▸

বিগডাটা।

বিগডাটা কেনো এত গুরুত্বপূর্ণ

বিগডাটা অনেকভাবে এবং অনেক কারণে গুরুত্বপূর্ণ। প্রথমেই বুঝতে হবে কেনো ডাটাকে বিগডাটা বলা হচ্ছে। যদি বলা হয় আগামী এক ঘণ্টায় পুরো পৃথিবীতে শুধু ইয়াহুতে কত ই-মেইল বিনিময় করা হবে, তার সঠিক পরিসংখ্যান বের করতে হবে। এখন একটু চিন্তা করলেই বোঝা যাবে, এক ঘণ্টায় কী বিশাল সংখ্যার ই-মেইল বিনিময় হতে পারে। এ সংখ্যা বিলিয়ন ছাড়িয়ে ট্রিলিয়নও হতে পারে। এধরনের বিশাল পরিমাণের ডাটা বা উপাত্ত প্রসেস করার জন্য বিগডাটা জানতে হবে।

প্রতিনিয়ত যেকোনো ধরনের ডাটার আকার বাড়ছেই। যেমন আজ থেকে দশ বছর আগে পিসির হার্ডডিস্ক ২০ গিগাবাইট মানেই বিশাল কিছু। অথচ আজকের বাস্তবতা হচ্ছে ২ টেরাবাইট ক্ষমতার হার্ডডিস্ক মানে মোটামুটি চলে। অর্থাৎ আমাদের ব্যবহার করা ডাটার আকার বেড়েছে। সেই সাথে প্রযুক্তির উৎকর্ষও বেড়েছে। শুধু অনলাইনে বা ক্লাউড কমপিউটিংয়ে এই এখন প্রতি অ্যাকাউন্টে ৫০ গিগাবাইটের সমান ডাটা বিনা খরচে স্টোর করা যায়।

বাণিজ্য ও বিগডাটা

যেকোনো কিছুর অর্থনৈতিক উৎকর্ষ না থাকলে তা নিয়ে মাতামাতি করার কিছু নেই। যেহেতু বিগডাটা নিয়ে এখন বেশ মাতামাতি হচ্ছে, তাই নিঃসন্দেহে বলে দেয়া যায়, বিগডাটার অর্থনৈতিক ও বাণিজ্যিক দিক অবশ্যই চমকপ্রদ। আগেই বলে হয়েছে, বিগডাটা প্রসেস ও রক্ষণাবেক্ষণ করা বেশ কষ্টসাধ্য। কারণ এতে বিপুল সংখ্যার উপাত্ত থাকে। তাই ডাটা ম্যানেজ ও প্রসেস করার জন্য আলাদা আলাদা প্রতিষ্ঠান গড়ে উঠছে। এর মধ্যে অনেক নামকরা প্রতিষ্ঠানও আছে, যারা এই বিগডাটার বাণিজ্যে আসছে। ধীরে ধীরে জমে উঠছে বিগডাটার বাণিজ্য। তৈরি হচ্ছে নিত্য-নতুন প্রতিষ্ঠান। আর তথ্যপ্রযুক্তিভিত্তিক বাণিজ্যিক প্রতিষ্ঠানগুলো তো আছেই।

শীর্ষের দিকে থাকা কিছু বিগডাটা কোম্পানি

অনেক বড় আকারের ডাটা বা অনেক বেশি ডাটা এক সাথে নড়াচড়া করা কোনো সহজ কাজ নয়। আমরা এখন বাজারে হার্ডডিস্ক দেখছি টেরাবাইট আকারের। চিন্তা করে দেখুন আমাদের সাধারণ ব্যবহারের জন্য টেরাবাইট হার্ডডিস্ক ব্যবহার করছি। এ বড় আকারের হার্ডডিস্ক ব্যবহারের সময় কোনো কিছু সার্চ করতে দিলে বা পুরো ডিস্ক ভাইরাস স্ক্যান দিলে কতটা সময় লাগে। আমাদের হার্ডডিস্ক নিয়েই আমরা কত হিমশিম খাচ্ছি। এখন ভাবুন, সার্ভারের জন্য যেসব হার্ডডিস্ক বা ডাটা সংরক্ষণ করার ব্যবস্থা গ্রহণ করা হয় তা কত বড় হতে পারে। সেগুলো কি আর টেরাবাইটের হলে হবে? কারণ সেখানে কোটি কোটি ডাটা সংরক্ষিত থাকে। যেমন ফেসবুক বা টুইটারের কথাই ধরা যাক, তাদের ডাটাবেজ কত বড় হতে পারে তা কি একবার ভেবে দেখেছেন? ডাটাবেজের কথা মনে এলেই মনে আসে ওরাকল, এসকিউএল, মাইএসকিউএল, এক্সেস ইত্যাদির কথা। এসব নামকরা ডাটাবেজের ডাটা প্রসেস করার ক্ষমতা সীমিত। বিশাল বড় আকারের ডাটা এগুলো সামলাতে পারে না। টেরাবাইট, পেটাবাইট ইত্যাদি আকারের ডাটা প্রসেস করার জন্য রয়েছে বিশেষ প্রযুক্তি। এ প্রযুক্তি নতুনই বলা চলে। এ প্রযুক্তির চর্চা এখনো অনেক উচ্চ পর্যায়ে রয়েছে। এটি সাধারণ জনগণের হাতে এখনো পৌঁছায়নি। এ প্রযুক্তি এতটাই নতুন যে এখনো কোনো বিশ্ববিদ্যালয়ের সাবজেক্ট হিসেবে এটি স্থান করে নিতে পারেনি। তবে কয়েক বছরের মধ্যে এ নিয়ে বেশ তোলপাড় হবে তা বলার অপেক্ষা রাখে না। তাই আগেভাগে এ প্রযুক্তি নিয়ে কিছু জেনে রাখা ভালো যাতে সময় হলে তা কাজে লাগানো সম্ভব হয়।

বিগডাটা হ্যান্ডেল করার জন্য এখনকার বাজারে সবচেয়ে কার্যকর ও ভালো প্রযুক্তিটি হচ্ছে হাডুপ (Hadoop)। সংক্ষেপে হাডুপ হচ্ছে বিশাল ডাটা প্রসেসিংয়ের একটি অন্যতম প্রযুক্তি, যা তৈরি হয় ওপেন সোর্স পণ্য হিসেবে। এটি অ্যাপাচির ওপেন সোর্স প্লাটফর্মের সাথে যুক্ত।

বিগডাটা নিয়ে কাজ করছে অনেকগুলো কোম্পানি। কিন্তু অনেকেই এ খবর জানেন না। তাই দেখে নেয়া যাক, কারা এ নতুন প্রযুক্তি নিয়ে কাজ করছে।

ইএমসি

ইএমসি (EMC) বিগডাটা ক্ষেত্রের শীর্ষ-সারির প্রতিষ্ঠান। ২০১০ সালে তারা গ্রিনপ্লাম (Greenplum) নামের একটি কোম্পানিকে কিনে নেয়, যারা বিশাল আকারের ডাটা প্রসেসিং নিয়ে কাজ করতো। সেই কোম্পানির প্রযুক্তিকে আরো শক্তিশালী করে তারা একটি প্লাটফর্ম তৈরি করে যার নাম ইএমসি গ্রিনপাম ডাটা কমপিউটিং অ্যাপ্লায়েন্স। ইএমসি মূলত বড় আকারের ডাটা স্টোরেজ কোম্পানি। তাদের সেই স্টোরেজ সলিউশনের সাথে খুব ভালোভাবে মিশে গেছে এই ডাটা প্রসেসিং প্রযুক্তি। ইএমসি হাডুপ (Hadoop) প্রযুক্তি নিয়ে কাজ করে।

ক্লাউডএরা

ক্লাউডএরা (Cloudera) মূলত হাডুপ এবং ম্যাপরিডিউস নিয়ে কাজ করে। এ প্রযুক্তি দুটি বানানোর সময় যারা এ প্রজেক্টে কাজ করেছেন, তাদের অনেকেই কাজ করেন ক্লাউডএরাতে। তাই অন্যান্য কোম্পানির চেয়ে তাদের হাডুপ ও ম্যাপরিডিউস নিয়ে কাজের দক্ষতা কিছুটা বেশি। যদিও ইয়াহু প্রথম হাডুপ নিয়ে কাজ শুরু করে, কিন্তু বর্তমানে এরাই মূলত হাডুপ এবং ম্যাপরিডিউসের প্রথম সারির প্রতিষ্ঠান হিসেবে পরিণত হয়েছে। এই প্রযুক্তির মাধ্যমে তারা কয়েক পেটাবাইট ডাটা পর্যন্ত অ্যানালাইসিস করতে পারে। ইন্টারনেটের বড় বড় কোম্পানি, যেমন– অ্যাডকোনিয়ন, অ্যাডগোরো, অ্যাপ্রেগেট নলেজ, এওএল অ্যাডভারটাইজিং, অ্যাপোলো গ্রুপ, নাওটেক, নকিয়া, ট্রেন্ড মাইক্রো,গ্রুপন, স্যামসাং, ট্রুলিয়া, স্কাইবক্স ইমেজিং, টাইন্ট, ইত্যাদি প্রতিষ্ঠানগুলো এদের পণ্য এবং সেবা নিয়ে থাকে।

এইচপি ভার্টিকা

আমরা বেশিরভাগ মানুষই জানি এইচপি (HP) একটি কমপিউটার, কমপিউটার পণ্য ও ইলেকট্রনিক্স সামগ্রী প্রস্তুতকারী প্রতিষ্ঠান। কিন্তু এইচপি এই বছরের ফেব্রুয়ারি মাসে ভার্টিকা (Vartica) নামের একটি কোম্পানিকে কিনে নিয়ে এই বিগডাটা প্রসেসিং বাজারে ঢুকে গিয়েছে। এরা মূলত ই-কমার্স অ্যানালাইটিক্সের জন্য বিখ্যাত। ভার্টিকার বড় সুবিধা হলো, এটা চলার জন্য বিশেষ কোনও মেশিনের প্রয়োজন হয় না। বাজারে পাওয়া যায় সাধারণ মেশিনেই চলতে পারে। এর কম্প্রেশন প্রযুক্তি খুবই উন্নত। আবার কোনও অনুসন্ধান চালালে সেটা প্রচন্ড দ্রুততার সাথে নিয়ে আসতে পারে। আমেরিকা অনলাইন, টুইটার এবং গ্রুপন এই প্রযুক্তি ব্যবহার করছে।

আইবিএম

আইবিএম (IBM)-এর রয়েছে ডিবি-২ (DB-২) ভিত্তিক স্মার্ট অ্যানালাইটিক সিস্টেম। আইবিএম বিগডাটা মার্কেটে ভালোভাবে আসার জন্য কিনে নিয়েছে নেটেজ্জা অ্যাপ্লায়েন্স নামের প্লাটফর্ম। ডিবি-২ ব্যবহার হয় খুবই হাই-স্কেল এন্টারপ্রাইজ ডাটাওয়্যার হাউজের কাজে, যেমন– কলসেন্টার। আইবিএমের নেটেজ্জা সলিউশনটি ডিবি-২ এর চেয়েও আরো বড় আকারের ডাটা অ্যানালাইসিস করতে পারে। এটি মূলত তৈরি করা হয়েছে বড় বড় টেলিফোন কোম্পানি, ডিজিটাল মার্কেটিং প্রতিষ্ঠান বা এমন প্রতিষ্ঠান যারা টেরাবাইট কিংবা পেটাবাইট পরিমাণ ডাটা নিয়ে কাজ করে থাকে।

ইনফোব্রাইট

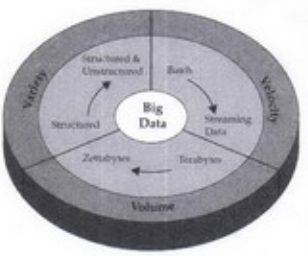
ইনফোব্রাইট (Infobright) ব্যবহার করে কলাম-স্টোর ডাটাবেজ প্রযুক্তি, যা প্রসেস করতে পারে কয়েকশ' গিগাবাইট থেকে শুরু করে কয়েক টেরাবাইট ডাটা। কোটি কোটি রেকর্ডের লগ ফাইল প্রসেসের কাজ করা এ প্রযুক্তির কাছে মামুলি ব্যাপার। ইনফোব্রাইটের মতে, তারা প্রতিষ্ঠানগুলোর ডাটাবেজ অ্যাডমিনিস্ট্রেটরদের (ডিবিএ) কাজের পরিমাণ শতকরা ৯০ ভাগ কমিয়ে ফেলেছে।

এছাড়াও বিগডাটা মার্কেটের দিকে এগিয়ে আসা আরো কয়েকটি কোম্পানির মধ্যে রয়েছে– Dataguise, DataSift, Lattice Engines, ▸

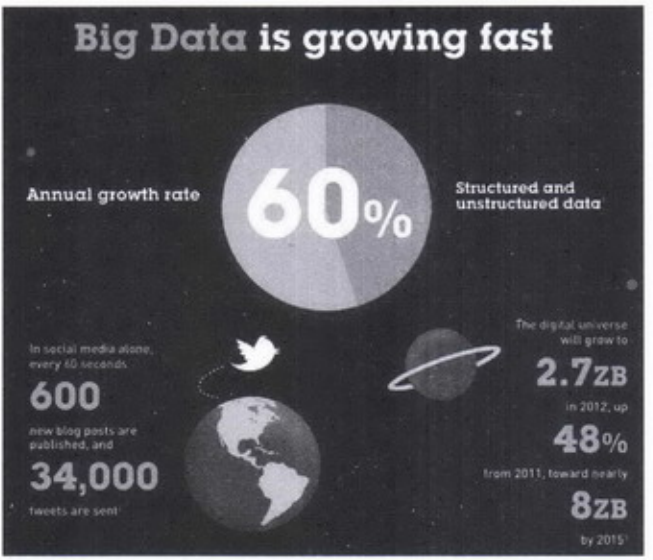
Palantir Technologies, Cloudera, SS Inc., i2, Centrifuge Systems, Wavii, Factual, Hyperpublic, Infochimps, Avanade, The Trade Desk, Recorded Future, Fluidinfo, Teradata, Aspera, Fusion IO, Par Accel, EMC, Data Stax, Pervasive, NetApp, Sybase, Datameer, dataspora, OpenHeatMap, nPario, Sociocast, Metamarkets, SaveWave, Kinetic Global Markets, Expan, Sulia, DataPop, NewsCred, Wonga, Klarna, ReadyForZero, Stormpulse, ClickFox, Kaggle, Acunu, Hadapt, Crowdflower, Mapr, General Sentiment ইত্যাদি। এতো গেলো কোম্পানির কথা, এবার আসা যাক ব্যক্তির কথায়। বিগ ডাটা বিশেষজ্ঞদের মধ্যে নামকরা কয়েকজনের মধ্যে রয়েছেন– আইবিএমের জেফ জোনাস, লিঙ্কডইনের পিট স্কোমোরোচ, কালার ল্যাবের ভিজে পাতিল, ক্লাউডেরা ও ফেসবুকের জেফ হ্যামারবেচার, অ্যামাজনের দীপক সিং প্রমুখ। বিগডাটা নিয়ে কাজ করা বড় কোম্পানিগুলোর ওয়েবসাইট ভিজিট করে তাদের সার্ভিস এবং সাপোর্টগুলো দেখলে বিগডাটা মার্কেটে কী ধরনের কাজ হয়, সে সম্পর্কে স্পষ্ট ধারণা পাওয়া যাবে।

বিগডাটার বৈশিষ্ট্য

আইবিএম বা ইন্টারন্যাশনাল বিজনেস মেশিনের মতে- বিগডাটার তিনটি বৈশিষ্ট্য রয়েছে। এগুলো হচ্ছে– ভলিউম, ভ্যারাইটি ও ভেলোসিটি যাকে সংক্ষেপে ভিকিউব (V^3) বলা হয়। সংক্ষেপে বিগডাটার বৈশিষ্ট্যের কথা বলতে গেলে বলা যায়, এটি বিশাল আয়তনের নানা ধরনের ডাটা নিয়ে অনেক দ্রুততার সাথে কাজ করতে সক্ষম। ২০০০ সালের হিসাব অনুযায়ী, পৃথিবীতে ৮ লাখ পেটাবাইট (১ পেটাবাইট = ১০ লাখ গিগাবাইট বা ১ হাজার টেরাবাইট)



ডাটা সংরক্ষণ করা হয়েছে। টুইটার প্রতিদিন ৭ টেরাবাইট, ফেসবুক ১০ টেরাবাইট এবং কিছু এন্টারপ্রাইজ প্রতি ঘণ্টায় টেরাবাইট পরিমাণ ডাটা জেনারেট করছে। এসব ডাটার মধ্যে রয়েছে ভিন্ন ধরনের ডাটা তার ইয়ত্তা নেই। এত ডাটা সামাল দেয়ার জন্যেই জন্ম নিয়েছে এ নতুন প্রযুক্তি। কারণ, আগের যেসব প্রযুক্তি ছিল তার সাহায্যে এত ডাটা প্রসেস করতে গেলে কয়েক বছর লেগে যাবে। নতুন প্রযুক্তির সাহায্যে তা কয়েক দিন বা কয়েক ঘণ্টার মধ্যেই করা সম্ভব। বড় কোম্পানিগুলো যাদের বিশাল আকারের ডাটা প্রসেস করার প্রয়োজন পড়ে তারা শরণাপন্ন হচ্ছেন বিগডাটা অ্যানালাইটিক্স কোম্পানিগুলোর কাছে। তারা নিমেষেই তাদের কাজ সমাধা করে দিচ্ছে। ধারণা করা হচ্ছে, ২০১৫ সাল নাগাদ এ ডাটার পরিমাণ দাঁড়াবে ৮ জেটাবাইট (১ জেটাবাইট = ১০ লাখ পেটাবাইট বা ১ হাজার ইয়োট্টাবাইট) যা ২০২০ সালে গিয়ে ৩৫ জেটাবাইটে পৌঁছতে পারে। যেভাবে ডাটার পরিমাণ বাড়ছে সেভাবে তা হ্যান্ডেল করার লোকের প্রয়োজনীয়তাও বাড়বে। তাই বলা যায়, ধীরে ধীরে একটি নতুন কর্মক্ষেত্র জন্ম নিচ্ছে।



বিগডাটার ভবিষ্যৎ

কমপিউটার বিজ্ঞান ও কমপিউটার প্রকৌশলে পড়াশোনা করে বিভিন্ন বিষয়ে বিশেষজ্ঞ হওয়া যায়। এ বিষয়ে পড়াশোনা করে যেকোনো একটি উইং ধরা যায়। এগুলোর মধ্যে মূল উইংগুলো হচ্ছে সফটওয়্যার ইঞ্জিনিয়ারিং, ডাটাবেজ অ্যাডমিনিস্ট্রেশন, ম্যানেজমেন্ট ইনফরমেশন সার্ভিসেস, টেলিকম ও নেটওয়ার্কিং, হার্ডওয়্যার ইত্যাদি। এগুলোর মধ্যে আগামী কয়েক বছরে ডাটাবেজ অ্যাডমিনিস্ট্রেশনের মূল জায়গা দখল করে নেবে বিগডাটা, এ ব্যাপারে কোনো সন্দেহ নেই। বিশেষজ্ঞদের মতে, আগামী কয়েক বছর পর বিগডাটা প্রফেশনালদের চরম অভাব বিরাজ করবে। তাই এখন থেকেই আমাদের এ বিষয়ে দক্ষ জনবল গড়ার ব্যাপারে মনোযোগী হতে হবে।

তথ্যবিজ্ঞানী হওয়ার জন্য ক্যারিয়ার যেভাবে ডেভেলপ করা যায়

যারা আইটিতে পড়াশোনা করছেন তাদের জন্য বিগডাটা একটি সম্ভাবনাময় খাত হিসেবে বিবেচিত হতে পারে। ডাটাবেজ অ্যাডমিনিস্ট্রেটর এমনিতেই খুব ভালো একটি পেশা। এবং সম্ভবত আইটি সেক্টরে এই পেশাতেই সবচেয়ে কম খাটুনি। ডাটাবেজ অ্যাডমিনিস্ট্রেটররা খুব সহজেই বিগডাটা প্রফেশনাল হতে পারেন। ডাটাবেজ অ্যাডমিনিস্ট্রেশন এবং বিগডাটা প্রসেসিং সমগোত্রীয় পেশা হলেও বিগডাটা ম্যানেজমেন্ট অনেক বেশি চ্যালেঞ্জিং এ ব্যাপারে কোনো সন্দেহ নেই। আমাদের দেশে বিগডাটা প্রফেশনালদের জন্য এখনও কোনো ইনস্টিটিউশন তৈরি হয়নি। তবে বিগডাটা প্রফেশনালদের চাহিদা যে হারে বাড়ছে তাতে এধরনের ইনস্টিটিউশন তৈরি হতে খুব বেশি সময় লাগবে না। আশা করা যায় খুব শিগগির আমাদের দেশেই বিগডাটা প্রফেশনালদের চাহিদা তৈরি হবে। দেশে বিগডাটা ম্যানেজ করার প্রয়োজন হতে পারে এমন প্রতিষ্ঠানের এখন অভাব নেই।

বিগডাটা নিয়ে কথা বলতে গিয়ে প্রায়ই তথ্যবিজ্ঞানী শব্দটি উঠে এসেছে এই প্রতিবেদনে। মূলত যারা বিগডাটা নিয়ে কাজ করেন তারাই তথ্যবিজ্ঞানী। তবে কোনো ইউনিভার্সিটি বা প্রতিষ্ঠানেই তথ্য-বিজ্ঞান বিষয়ে কোনো কারিকুলাম নেই। কারণ তথ্যবিজ্ঞানী টার্মটি থাকলেও তথ্য-বিজ্ঞান নামে পড়াশোনা বা উচ্চশিক্ষা বলতে কিছু নেই। তাই এ বিষয়ে সেভাবে প্রাতিষ্ঠানিক শিক্ষার ব্যবস্থা নেই। তাহলে যারা বিগডাটা প্রফেশনাল হতে চাচ্ছেন তারা এ বিষয়ে শিক্ষা নেবেন কিভাবে এটা একটা বড় প্রশ্ন।

স্নাতক পর্যায়ের অত্যাবশ্যকীয় কোর্স

বিগডাটা প্রফেশনাল হতে চাইলে অ্যাকাডেমিক ▸

লেভেলে কয়েকটি কোর্স রাখতে হবে। প্রথমেই ম্যাট্রিক্স ফ্যাকটোরাইজেশন সম্পর্কে ভালোভাবে পড়াশোনা করতে হবে। এজন্য গ্র্যাজুয়েশন লেভেলে বা পোস্ট গ্র্যাজুয়েশন লেভেলে লিনিয়ার অ্যালজেব্রা কোর্সটি রাখতে হবে। বিশ্ববিদ্যালয় বা প্রতিষ্ঠানভেদে একেক জায়গায় এই একই কোর্স একেক নামে পরিচিত। যেমন একে কখনও কখনও নিউমেরিক্যাল মেথডস, ম্যাট্রিক্স অ্যানালাইসিস বা ম্যাট্রিক্স কমপিউটেশনও বলা হয়। আমাদের দেশ থেকে যারা কমপিউটার প্রকৌশল বা কমপিউটার বিজ্ঞান নিয়ে পড়াশোনা করেছেন এবং বিগডাটা প্রফেশনাল হতে চান তারা আগেই হতাশ হবেন না। আমাদের দেশের প্রায় সব ইউনিভার্সিটি ও প্রতিষ্ঠানে এই কোর্সটি করানো হয়। ডাটা মাইনিং, নিউমেরিক অ্যালগরিদম ইত্যাদি নিয়ে যেনো এই কোর্সে অবশ্যই আলোচনা করা হয় সে ব্যাপারে লক্ষ রাখতে হবে। আর এই কোর্স করার সময় যদি ম্যাটল্যাব শেখানো হয় তবে তা হবে একটি প্লাস পয়েন্ট।

ডিস্ট্রিবিউটেড কমপিউটিং নিয়ে জ্ঞান

ডিস্ট্রিবিউটেড কমপিউটিং নিয়ে জানতে হবে। সাধারণত তেমন কোনো সরাসরি কোর্স নেই এ বিষয়ে গ্র্যাজুয়েশন লেভেলে। তবে কিছু এক্সট্রা কারিকুলার অ্যাড অন কোর্স আছে এ বিষয়ে পড়াশোনার জন্য। ডিস্ট্রিবিউটেড কমপিউটিং নিয়ে জানতে হলে প্রথমেই লিনআক্স নিয়ে বেশ ভালোভাবে জানতে হবে। বলে রাখা ভালো বিগডাটা যেসব অপারেটিং সিস্টেমে রান করানো হয় সেগুলো মূলত সার্ভার না হলে মিনি কমপিউটার। যেগুলোর অপারেটিং সিস্টেম থাকে লিনআক্স বা ইউনিক্স। এই কোর্সে জানতে হবে কীভাবে স্কেলেবল ডিস্ট্রিবিউটেড অ্যালগরিদম ডিজাইন করতে হয়।

স্ট্যাটিস্টিক্যাল অ্যানালাইসিস

যারা ছাত্রজীবনে পরিসংখ্যান নিয়ে পড়াশোনা করেছেন তারা স্ট্যাটিস্টিক্যাল অ্যানালাইসিসে এগিয়ে থাকবেন। স্ট্যাটিস্টিক্যাল অ্যানালাইসিস সম্পর্কে ভালোভাবে জানতে হবে। ডাটাবেজ কিন্তু পরিসংখ্যানগত তথ্য ছাড়া আর কিছুই নয়। তাই স্ট্যাটিস্টিক্যাল অ্যানালাইসিস অত্যাবশ্যকীয়।

মাস্টার অ্যালগরিদম ও ডাটা স্ট্রাকচার

মাস্টার অ্যালগরিদম ও ডাটা স্ট্রাকচার সম্পর্কে বেশ ভালোভাবে জানতে হবে। বিশেষ করে ডাটা স্ট্রাকচার। কারণ ডাটা স্ট্রাকচার না জানলে অ্যালগরিদম জানা সম্ভব নয়।

আলোচ্য এই বিষয়গুলোর বেশিরভাগই কমপিউটার বিজ্ঞান ও কমপিউটার প্রকৌশল গ্র্যাজুয়েশনে পড়ানো হয়। যদি কোনো কারণে এই বিষয় বাদ দেয়া হয় তাহলে এই বিষয়গুলো ভালোভাবে আয়ত্ত করে তথ্যবিজ্ঞানী হওয়ার প্রাথমিক ধাপ সম্পন্ন হবে।

ওরাকল কোর্স

ডাটাবেজ নিয়ে কাজ করতে গেলে ওরাকল নিয়ে কাজ করতেই হবে। ওরাকল এবং এসকিউএল জানতেই হবে এ ব্যাপারে কোনো সন্দেহের অবকাশ নেই। আমাদের দেশে ওরাকল কোর্স করায় এমন প্রচুর প্রতিষ্ঠান আছে। এখান থেকে ওরাকল শিখে ভেন্ডর সার্টিফিকেট নিতে হবে। তাহলে বিগডাটা নিয়ে কাজ করার পথ সুগম হবে।

শুধু যে ওরাকল শিখলেই হবে তা নয়। ওরাকল ছাড়াও আরো অনেক ভারি ডাটাবেজ প্রোগ্রাম আছে। এগুলোর মধ্যে এন্টারপ্রাইজ ডাটাবেজ আছে, যেমন আইবিএম DB/2 (UDB), Sybase ASE এবং মাইক্রোসফট এসকিউএল সার্ভার। ওরাকল হচ্ছে বেস্ট মার্কেটেড ডাটাবেজ সিস্টেম। এছাড়া সম্প্রতি সান মাইক্রোসিস্টেমকে কিনে নেয়ার পর ওরাকলের অধিপত্য বাড়ছে। তবে ওরাকল একটি বিশাল ডাটাবেজ সিস্টেম।

শুধুই ট্রেনিং সেন্টার বা ইনস্টিটিউশনগুলোর ওপরে ভরসা করলে কিন্তু চলবে না। নিজেকে অনেক সময় ব্যয় করতে হবে ওরাকল শিখার জন্য। এক্ষেত্রে এর ট্রেনিং ম্যানুয়াল যা ইন্টারনেটে পাওয়া যায় যেগুলো অনেক উপকারী। আর ডিবিএ হতে হলে শুধু ওরাকল RDBMS শিখলেই চলবে না, জাভা প্রোগ্রামিং ডেস্কটপ এবং ওয়েব ডেভেলপমেন্ট দুটোতেই, ইউনিক্স সিস্টেম অ্যাডমিনিস্ট্রেশন, নেটওয়ার্ক অ্যাডমিনিস্ট্রেশন ইত্যাদি সম্পর্কে জানতে হবে।

শেখার জন্য ওরাকল আর মাইএসকিউএল প্রায় কাছাকাছি। এসকিউএল সার্ভার এবং মাইএসকিউএল শিখতে হবে। এরপর ওরাকলে সুইচ করা যেতে পারে। ভালো হয় যদি ওরাকলও প্রথমে নিজে নিজে শিখা যায়। এরপর কোনো সমস্যা হলে বিশেষায়িত প্রতিষ্ঠানে ভর্তি হয়ে সেসব সমাধান করে নেয়া যায়। এক্ষেত্রে প্রতিষ্ঠান বড় ব্যাপার নয়, ব্যাপার হচ্ছে কে শেখাচ্ছে। কারণ ইনস্ট্রাক্টর যত ভালো হবে শেখাটা তত সহজ হয়ে যাবে। নিজে নিজে না শিখে যদি শুধুই প্রতিষ্ঠানের ভরসা করা হয় তাহলে একটা সার্টিফিকেট ছাড়া আর তেমন কিছু পাওয়া যাবে না এবং সেই সার্টিফিকেটের কোনো ভ্যালু থাকবে না।

ওরাকলে বর্তমানে বাংলাদেশ এবং বাইরে আসলেই অনেক জবের সম্ভবনা সৃষ্টি হচ্ছে। বর্তমানে ব্যাংক, বীমা, মোবাইল কোম্পানি সব কিছুতেই ওরাকল ডেভেলপারের প্রয়োজন পড়ছে। তবে প্রযুক্তি যেভাবে সামনে এগুচ্ছে সে অনুসারে আমাদের বাংলাদেশ এখনও অনেক পিছিয়ে আছে। বাংলাদেশে এখনও ওরাকল ১০জির ওপর ব্যবহার নেই। কিন্তু অনেক আগেই ওরাকল ১১জি রিলিজ হয়েছে। মোটকথা শুধু জানলেই হবে না। ডাটাবেজ ও লিনআক্স অপারেটিং সিস্টেম সম্পর্কে ভালোভাবে জানতে হবে তথ্যবিজ্ঞানী হওয়ার জন্য এবং বিগডাটা নিয়ে কাজ করার জন্য।

বিগডাটার ওপর অনলাইন কোর্স

পড়াশোনা এখন আর বইয়ের পৃষ্ঠা ও শিক্ষাপ্রতিষ্ঠানের গণ্ডির মধ্যে সীমাবদ্ধ নেই। প্রযুক্তির উন্নতির সাথে সাথে শিক্ষা-ব্যবস্থায় এসেছে ব্যাপক পরিবর্তন। জ্ঞান আহরণ করার জন্য এখন স্কুল, কলেজ বা বিশ্ববিদ্যালয়ের কড়া নাড়ার প্রয়োজন নেই। যদি ইচ্ছা এবং ধৈর্য থাকে তবে তা ঘরে বসেই লাভ করা সম্ভব। ইন্টারনেট হচ্ছে বিশাল এক জ্ঞানভান্ডার। শিক্ষার উপকরণ সংগ্রহে কিছু ক্ষেত্রে শিক্ষা-প্রতিষ্ঠান বা শিক্ষকের চেয়ে ভালো ভূমিকা পালন করে থাকে ইন্টারনেট। বেশ কিছু ইউনিভার্সিটি অনলাইনে কোর্স চালু করেছে বিভিন্ন বিষয়ে। সেসব কোর্সের কিছু বিনামূল্যে করা যায়, আবার কিছুর ক্ষেত্রে ব্যয় করতে হয় অর্থ, আবার কিছু আছে পুরো কোর্স বিনামূল্যে করার পর পরীক্ষা দিয়ে সার্টিফিকেট অর্জন করার সময় অর্থ খরচ করতে হয়। এসব অনলাইন কোর্সে লেকচার ভিডিও, লেকচার শিট, অডিও ফাইল, ইমেজ, রেফারেন্স গাইড ও অন্যান্য আনুষঙ্গিক বিষয়বস্তু ইন্টারনেট থেকে ডাউনলোড করে নেয়ার সুযোগ থাকে। কোর্স কো-অর্ডিনেটরের সাথে মেইল বা চ্যাট ম্যাসেঞ্জারের মাধ্যমে যোগাযোগ করার ব্যবস্থা থাকে। কোর্সে অংশ নেয়া অন্যান্য শিক্ষার্থীর সাথে যোগাযোগ করার মাধ্যম হিসেবে সেই কোর্সের জন্য বানানো থাকে ফোরাম যাতে নিজের আইডিয়া, সমস্যা ও যেকোনো প্রাসঙ্গিক প্রশ্ন করার ও অন্যের প্রশ্নের উত্তর দেয়ার সুযোগ থাকে। এসব অনলাইন কোর্সের সবচেয়ে মজার ব্যাপার হচ্ছে এতে ধরাবাধা কোনো সময় নেই ক্লাসের, যার যখন

(বাকি অংশ ৩০ পৃষ্ঠায়)

আসন্ন বাজেট নিয়ে প্রত্যাশার কথা জানতে চাইলে শুরুতেই বাজেটে আইসিটি নীতিমালার প্রতিফলন বাস্তবায়নের প্রতি জোর দেন বাংলাদেশ অ্যাসোসিয়েশন অব সফটওয়্যার অ্যান্ড ইনফরমেশন সার্ভিসেস তথা বেসিস সভাপতি মাহাবুব জামান। তিনি বলেন, ২০০৯ সালে আইসিটি নীতিমালা তৈরি হয়েছিল। নীতিমালায় আইসিটি শিল্প উন্নয়ন তহবিল হিসেবে ৭০০ কোটি টাকা গত বছরই বরাদ্দ থাকার কথা ছিল। কিন্তু এখনো তা হয়নি। গত বছরের মতো এ বছরও আমরা বাজেটে শিল্প উন্নয়ন তহবিল থেকে ১০ শতাংশের বরাদ্দ চাই।

দেশের আউটসোর্সিংয়ের অবকাঠামো তৈরিতে এই অর্থ ব্যবহারের প্রতি জোর দিয়ে মাহাবুব জামান বলেন, এতে আউটসোর্সিংয়ের জন্য প্রতিবছর কমপক্ষে দশ হাজার আইটি প্রফেশনাল তৈরি করা যাবে।

গত তিনটি অর্থবছরের বাজেট নিয়ে হতাশা প্রকাশ করে তিনি বলেন, প্রতিবছরই স্বপ্ন দেখতে দেখতে এখন ইঁপিয়ে উঠেছি। প্রধানমন্ত্রী শেখ হাসিনার সভাপতিত্বে ডিজিটাল টাস্কফোর্সের প্রথম সভায় মাত্র তিন মিনিটে কাওরান বাজারের জনতা টাওয়ারে আইসিটি পার্ক তৈরির বিল পাস করার পর এর অবকাঠামোগত নির্মাণ শেষ হলেও এখনো তা আলোর মুখ দেখতে পারেনি। আমরা চাই যত দ্রুত সম্ভব এটি হস্তান্তর করা হোক। আর কালিয়াকৈরের হাইটেক পার্ককে কার্যকর করার সুষ্ঠু নীতিমালা এ বছরের বাজেটেই অন্তর্ভুক্ত করা হোক। এটা করা না হলে সরকারের ডিজিটাল বাংলাদেশ পরিকল্পনা 'প্রহসন' হিসেবে বিবেচিত হবে।

এশিয়ান-ওশেনিয়ান কমপিউটিং ইন্ডাস্ট্রি অর্গানাইজেশন তথা অ্যাসোসিওর ডেপুটি চেয়ারম্যান আবদুল্লাহ এইচ কাফী আসন্ন বাজেটে তথ্যপ্রযুক্তি খাতে বরাদ্দের বিষয়টি সুনির্দিষ্টকরণের প্রতি জোর দেন। তিনি বলেন, এতদিনেও আইসিটি পণ্যতালিকা ও এ বিষয়ে বরাদ্দের বিষয়টি সুনির্দিষ্ট হয়নি। আগে সবার হওয়া দরকার। পাশাপাশি অবৈধ বিদেশ কল ঠেকাতে ভিওআইপি কল উন্মুক্ত করে দেয়ার পরামর্শ দেন তিনি। আর সরকারের রাজস্ব আয় বাড়াতে প্রযুক্তিপণ্য আমদানির ক্ষেত্রে অগ্রিম আয়কর বা এআইটি কেটে রাখার পক্ষে মত দেন আবদুল্লাহেল কাফী। তার মতে, এটা করা হলে সরকারের রাজস্ব আয় যেমন বাড়বে, একই সাথে কমবে কালোবাজারী। গ্রে-মার্কেটের আপদ থেকে মুক্তি পাবেন ব্যাবসায়ীরা।

ডিজিটাল বাংলাদেশ স্বপ্নপূরণে দেশের প্রতিটি জেলার অন্তত দুটি শিক্ষাপ্রতিষ্ঠানকে সমন্বিত ডিজিটালাইজড কার্যক্রমের আওতায় নিয়ে আসার প্রস্তাব করে তিনি বলেন, দেশজুড়ে অন্তত ৫০০ শিক্ষাপ্রতিষ্ঠানে কমপিউটার বিতরণের মতো প্রয়োজনীয় বরাদ্দ এই বাজেটে থাকা দরকার। তবে শুধু কমপিউটার দিলেই চলবে না। সুষম উন্নয়নের লক্ষ্যে প্রতিটি জেলার একটি বয়েজ এবং একটি গার্লস স্কুলে একটি করে কমপিউটার ল্যাব করা এখন সময়ের দাবি। এসব ল্যাবে কমপিউটারের পাশাপাশি ডিজিটাল ক্লাসরুমের জন্য প্রয়োজনীয় সব যন্ত্রপাতি সরবরাহ করা হলে আমরা প্রকৃত অর্থে সুফল পাব।

এক প্রশ্নের জবাবে অবিলম্বে দেশের প্রতিটি শিক্ষাপ্রতিষ্ঠানে বিনামূল্যে ইন্টারনেট সংযোগ দেয়ার দাবি জানান আবদুল্লাহেল কাফী। তিনি বলেন, যতই কমপিউটারের মতো ডিজিটালপণ্য ব্যবহার করি না কেনো ইন্টারনেট যোগাযোগ ব্যবস্থার বাইরে থেকে তথ্যপ্রযুক্তির পূর্ণ সুফল পাওয়ার কোনো সুযোগ নেই। তাই ইন্টারনেট সংযোগ মূল্য কমানোর পাশাপাশি দেশের প্রতিটি শিক্ষা প্রতিষ্ঠানে এটি বিনা মূল্যে সরবরাহ করা উচিত।

দেশের তথ্যপ্রযুক্তি সংশ্লিষ্ট বাজার সম্প্রসারণে প্রযুক্তিপণ্যের ওপর ধার্যকৃত ডিউটি কমানোর দাবি জানিয়ে কাফি বলেন, প্রজেক্টর, ডিজিটাল ক্যামেরা এবং মাল্টিফাংশনাল ডিজিটাল প্রোডাক্টের ওপর আরোপিত অতিরিক্ত ডিউটি চার্জ থেকে আমরা যেনো মুক্তি পাই সেটি মাথায় রেখেই বাজেট পেশ করা উচিত। তিনি বলেন, সিঙ্গেল ফাংশনাল আইটি প্রোডাক্টে যেখানে ৩ শতাংশ আমদানি শুল্ক আরোপ করা হয়েছে, সেখানে মাল্টিফাংশনাল প্রোডাক্টে শুল্ক পরিশোধ করতে হয় ২৮ শতাংশ। এটা শুধু অতিরিক্তই নয়, অনেকটা বৈষম্যমূলকও বটে।

তথ্যপ্রযুক্তিপণ্যের মধ্যে মুঠোফোন এখন আমাদের জীবনের একটি অতিপ্রয়োজনীয় ডিভাইস। স্বাভাবিকভাবেই বাজেটে এই খাতটিকে গুরুত্বের সাথে বিবেচনা করা হয়। কিন্তু তারপরও এর কলরেট, ইন্টারনেট সংযোগমূল্য, ফোনসেটের মূল্য নিয়ে মোবাইল প্রোভাইডার কোম্পানি ও ভোক্তাদের মধ্যে রয়েছে মিশ্র অভিমত। তৃণমূল পর্যায়ের মানুষের এসব অভিমতের ওপর ভিত্তি করেই আগামী ২০১২-১৩ অর্থবছরের বাজেটে মোবাইল ফোনের সিমের ওপর আরোপিত কর ৫০ শতাংশ কমানোর প্রস্তাব দিয়েছেন অ্যাসোসিয়েশন অব মোবাইল টেলিকম অপারেটরস অব বাংলাদেশ তথা অ্যামটবের সভাপতি আবু সাইদ খান। বিদ্যমান সিম কর ৬০০ থেকে কমিয়ে ৩০০ টাকা করার প্রস্তাব দিয়েছেন তিনি। প্রস্তাব অনুযায়ী সিম কর কমানো হলে দেশের সব খাত উপকৃত হবে বলে অভিমত ব্যক্ত করেন তিনি। এজন্য সম্প্রতি জাতীয় রাজস্ব বোর্ড তথা এনবিআর আয়োজিত প্রাক-বাজেট আলোচনায় আনুষ্ঠানিকভাবে প্রস্তাবও দেয়া হয়েছে বলে তিনি জানান। তিনি বলেন, সিম কর কমানো হলে দেশের সব খাতই এর সুফল পাবে। একই সাথে ইন্টারনেট যাতে নিম্নবিত্তরাও ব্যবহার করতে পারেন, সেদিকেও নজর দিতে হবে।

প্রসঙ্গত, গত অর্থবছরের বাজেটে সিম কর ৮০০ টাকা থেকে কমিয়ে ৬০০ টাকা করা হয়। এদিকে মুঠোফোনের ওপর আমদানি শুল্ক বেশি হওয়ার কারণে তা অবৈধপথে দেদার আসছে বলে অভিযোগ রয়েছে। জানা গেছে, এতে সরকার বছরে প্রায় ১৬০ কোটি টাকার রাজস্ব হারাচ্ছে। তবে আমদানি শুল্ক কমানো হলে এ প্রবণতা অনেকাংশেই কমবে বলে মনে করে অ্যামটব। মোবাইল ফোন অপারেটরদের এই সংগঠনের মতে, আসছে বাজেটে মোবাইল সেট আমদানিতে বিদ্যমান ১২ শতাংশ আমদানি শুল্ক হ্রাস করে ৭ শতাংশে নামিয়ে আনা হলে এ অবস্থার উন্নতি হবে।

বিগডাটা

(২৬ পৃষ্ঠার পর)

সময় হবে তখন সে ডকুমেন্ট নামিয়ে পড়ে নিতে পারবে এবং পরীক্ষায় অংশ নিতে পারবে।

বিগডাটা কোর্স শিক্ষাপ্রতিষ্ঠানের সাবজেক্ট হিসেবে এখনো এতটা জনপ্রিয় হয়ে ওঠেনি। তাই হাতেগোনা দুই-একটি বিশ্ববিদ্যালয় ছাড়া এ বিষয়টির দেখা মেলা ভার। হয়তো কয়েক বছরের মধ্যেই এ বিষয়টি অন্যান্য ভালো বিষয়ের মতো জনপ্রিয় হয়ে উঠবে এবং প্রায় সব বিশ্ববিদ্যালয়ের বিষয়-তালিকায় থাকবে। আমাদের দেশের ইউনিভার্সিটিগুলোতেও হয়তো এটি চলে আসবে। যারা বিগডাটা সম্পর্কে জানতে আগ্রহী বা এ ব্যাপারে পড়াশোনা করতে চান, তাদের এখনো তেমন সুযোগ গড়ে না উঠলেও ইন্টারনেট ঘেঁটে অনেক কিছু শিখে নেয়া যাবে। অন্তত বিগডাটার প্রাথমিক ধারণার জন্য ইন্টারনেটের কোনো বিকল্প নেই এ মুহূর্তে। তবে সুসংবাদ হচ্ছে, অনলাইনে বিগডাটার ওপরে একটি কোর্স চালু করেছে বিগডাটা ইউনিভার্সিটি নামের একটি প্রতিষ্ঠান। এই কোর্সটি বিনামূল্যে করা যাবে bigdatauniversity.com সাইটে রেজিস্ট্রেশন করার মাধ্যমে। কোর্সটি করার পর সার্টিফিকেট দেয়ার ব্যবস্থা রয়েছে, যা পরীক্ষায় পাস করার পর মেইলের মাধ্যমে পাঠিয়ে দেয়া হবে। এ কোর্সের প্রয়োজনীয় সব কিছু ওয়েবসাইট থেকে ডাউনলোড করে নেয়া যাবে। পরীক্ষায় অকৃতকার্য হলে পরপর তিনবার পরীক্ষা দেয়া যাবে।

কোর্সটি করার জন্য ইউনিক্স বা লিনআক্স অপারেটিং সিস্টেম সম্পর্কে প্রাথমিক ধারণা থাকা আবশ্যক। কোর্সটিতে বিগডাটার প্রাথমিক ধারণা, হাডুপ টেকনোলজি, ক্লাউড কমপিউটিং, টেক্সট অ্যানালাইসিস, কোয়েরি ল্যাঙ্গুয়েজ, স্ট্রিম কমপিউটিং, এসকিউএল ইত্যাদি বিষয় অন্তর্ভুক্ত করা হয়েছে। সাইটটিতে বিগডাটার কাজে লাগে এমন সফটওয়্যার ডাউনলোড করার ব্যবস্থাও রয়েছে। কোর্সটি মূলত আইবিএমের বিগডাটা নিয়ে কাজ করার পদ্ধতির সাথে সম্পর্কযুক্ত। বিগডাটার মূল পাঠ্যবই হিসেবে ব্যবহার করা হয়েছে আন্ডারস্ট্যান্ডিং বিগডাটা (Understanding Big Data) নামের বই, যা প্রকাশ করেছে ম্যাকগ্রো হিল নামের বিখ্যাত প্রকাশনা কোম্পানি। এ বইয়ের পাশাপাশি আরো কয়েকটি বইয়ের মধ্যে রয়েছে– IBM InfoShere Steams, Database Fundamentals, Getting started with DB2 Express-C, Getting started with DB2 application development, Getting started with IBM Data Studio for DB2, Getting started with Open Source development, Getting started with Open Source development, DB2 pureScale, DB2 10 for z/OS, The IBM Data Governance Unified Process, Business Intelligence Strategy, IBM Business Analytics and Cloud Computing, Getting started with WAS CE ইত্যাদি। যারা বিগডাটা সম্পর্কে জানতে আগ্রহী তারা এ সাইটে রেজিস্ট্রেশন করে এ কোর্সটি করতে পারেন।

ফিডব্যাক : mortuzacsepm@yahoo.com
shmt_21@yahoo.com