[Design big data batch processing and interactive solutions (30–35%)](https://www.microsoft.com/en-us/learning/exam-70-475.aspx" \l "syllabus-1)

* Ingest data for batch and interactive processing
  + Ingest from cloud-born or on-premises data, store data in Microsoft Azure Data Lake, store data in Azure BLOB Storage, perform a one-time bulk data transfer, perform routine small writes on a continuous basis
* Design and provision compute clusters
  + Select compute cluster type, estimate cluster size based on workload
* Design for data security
  + Protect personally identifiable information (PII) data in Azure, encrypt and mask data, implement role-based security, implement row-based security
* Design for batch processing
  + Select appropriate language and tool, identify formats, define metadata, configure output

[Design big data real-time processing solutions (30–35%)](https://www.microsoft.com/en-us/learning/exam-70-475.aspx#syllabus-2)

* Ingest data for real-time processing
  + Select data ingestion technology, design partitioning scheme, design row key of event tables in HBase
* Design and provision compute resources
  + Select streaming technology in Azure, select real-time event processing technology, select real-time event storage technology, select streaming units, configure cluster size, select the right technology for business requirements, assign appropriate resources for HBase clusters
* Design for Lambda architecture
  + Identify application of Lambda architecture, utilize streaming data to draw business insights in real time, utilize streaming data to show trends in data in real time, utilize streaming data and convert into batch data to get historical view, design such that batch data doesn’t introduce latency, utilize batch data for deeper data analysis
* Design for real-time processing
  + Design for latency and throughput, design reference data streams, design business logic, design visualization output

[Operationalize end-to-end cloud analytics solutions (30–35%)](https://www.microsoft.com/en-us/learning/exam-70-475.aspx#syllabus-3)

* Create a data factory
  + Identify data sources, identify and provision data processing infrastructure, utilize Visual Studio to design and deploy pipelines, deploy Data Factory Jobs
* Orchestrate data processing activities in a data-driven workflow
  + Leverage data-slicing concepts, identify data dependencies and chaining multiple activities, model complex schedules based on data dependencies, provision and run data pipelines
* Monitor and manage the data factory
  + Identify failures and root causes, create alerts for specified conditions, perform a restatement, start and stop data factory pipelines
* Move, transform, and analyze data
  + Leverage Pig, Hive, MapReduce for data processing; copy data between on-premises and cloud; copy data between cloud data sources; leverage stored procedures; leverage Machine Learning batch execution for scoring, retraining, and update resource; extend the data factory with custom processing steps; load data into a relational store, visualize using Power BI
* Design a deployment strategy for an end-to-end solution
  + Leverage PowerShell for deployment, automate deployment programmatically, design deployment strategies for automation