**Stream sets Prerequisites.**

1. Cloud services basics
2. Databases basics
3. **Cloud services:**

Cloud services are services available via a remote cloud computing server rather than an on-site server. These scalable solutions are managed by a third party and provide users with access to computing services such as analytics or networking via the internet.

Types of cloud services:

* GCP
* AZURE

**GCP:** Google Cloud Platform (GCP), offered by [Google](https://en.wikipedia.org/wiki/Cross-platform), is a suite of [cloud computing](https://en.wikipedia.org/wiki/Relational_database_management_system) services that runs on the same infrastructure that Google uses internally for its end-user products, such as [Google Search](https://en.wikipedia.org/wiki/Relational_database), [Gmail](https://en.wikipedia.org/wiki/NoSQL), [file storage](https://en.wikipedia.org/wiki/Infrastructure_as_a_service), and [YouTube](https://en.wikipedia.org/wiki/OpenBSD). Alongside a set of management tools, it provides a series of modular cloud services including computing, [data storage](https://en.wikipedia.org/wiki/Google_Search), [data analytics](https://en.wikipedia.org/wiki/Serverless_computing) and [machine learning](https://www.youtube.com/watch). Registration requires a [credit card](https://en.wikipedia.org/wiki/Google) or bank account details.

Google Cloud Platform provides [infrastructure as a service](https://en.wikipedia.org/wiki/YouTube), [platform as a service](https://en.wikipedia.org/wiki/Computer_data_storage), and [serverless computing](https://www.tutorialspoint.com/mongodb/index.htm) environments.

**Reference material:**

* **Overview link** [**https://cloud.google.com/docs/overview**](https://www.youtube.com/watch)
* **Basic videos to get started** [**https://www.youtube.com/watch?v=\_Q0tRI5hMnc**](https://www.youtube.com/watch?v=_Q0tRI5hMnc)
* **Whole GCP material** [**https://cloud.google.com/docs/overview**](https://en.wikipedia.org/wiki/Data_analysis)

****

**AZURE:** Microsoft Azure, commonly referred to as Azure , is a [cloud computing](https://en.wikipedia.org/wiki/Google_Drive) service created by [Microsoft](https://en.wikipedia.org/wiki/Infrastructure_as_a_service) for building, testing, deploying, and managing applications and services through Microsoft-managed [datacentre’s](https://en.wikipedia.org/wiki/Data_center). It provides [software as a service (SaaS)](https://cloud.google.com/docs/overview), [platform as a service (PaaS)](https://en.wikipedia.org/wiki/Machine_learning) and [infrastructure as a service (IaaS)](https://en.wikipedia.org/wiki/Credit_card) and supports many different [programming languages](https://en.wikipedia.org/wiki/Software_as_a_service), tools, and frameworks, including both Microsoft-specific and third-party software and systems.

**Reference Material:**

* **Doc** [**https://docs.microsoft.com/en-in/azure/guides/developer/azure-developer-guide**](https://en.wikipedia.org/wiki/Gmail)
* **Demo videos** [**https://azure.microsoft.com/en-in/get-started/video/**](https://en.wikipedia.org/wiki/Cloud_computing)

****

1. **Databases**

Database, also called electronic database, any collection of data, or information, that is specially organized for rapid search and retrieval by a computer. Databases are structured to facilitate the storage, retrieval, modification, and deletion of data in conjunction with various data-processing operations.

Types of databases:

* PostgreSQL
* MongoDB
* My SQL

**PostgreSQL:** PostgreSQL  also known as Postgres, is a [free and open-source](https://en.wikipedia.org/wiki/ACID_(computer_science)) [relational database management system](https://en.wikipedia.org/wiki/Programming_language) (RDBMS) emphasizing [extensibility](https://en.wikipedia.org/wiki/Platform_as_a_service) and [SQL compliance](https://azure.microsoft.com/en-in/get-started/video/). PostgreSQL features [transactions](https://en.wikipedia.org/wiki/Microsoft) with [Atomicity, Consistency, Isolation, Durability](https://en.wikipedia.org/wiki/Platform_as_a_service) (ACID) properties, automatically updatable [views](https://en.wikipedia.org/wiki/SQL_compliance), [materialized views](https://cloud.google.com/docs/overview), [triggers](https://en.wikipedia.org/wiki/Database_trigger), [foreign keys](https://en.wikipedia.org/wiki/Relational_database_management_system), and [stored procedures](https://docs.microsoft.com/en-in/azure/guides/developer/azure-developer-guide). It is designed to handle a range of workloads, from single machines to [data warehouses](https://en.wikipedia.org/wiki/View_(SQL)) or [Web services](https://en.wikipedia.org/wiki/Transaction_processing) with many [concurrent users](https://en.wikipedia.org/wiki/Concurrent_user). It is the default database for [macOS Server](https://en.wikipedia.org/wiki/Foreign_key), and is also available for [Linux](https://en.wikipedia.org/wiki/Extensibility), [FreeBSD](https://en.wikipedia.org/wiki/Free_and_open-source_software), [OpenBSD](https://en.wikipedia.org/wiki/Cloud_computing), and [Windows](https://en.wikipedia.org/wiki/Stored_procedure).

**Reference Material:**

* **Tutorial docs** [**https://www.postgresqltutorial.com/what-is-postgresql/**](https://en.wikipedia.org/wiki/Data_warehouse)
* **Demo videos** [**https://www.youtube.com/watch?v=-VO7YjQeG6Y**](https://en.wikipedia.org/wiki/Materialized_view?v=-VO7YjQeG6Y)

**MongoDB:** MongoDB is a [cross-platform](https://en.wikipedia.org/wiki/FreeBSD) [document-oriented database](https://en.wikipedia.org/wiki/MacOS_Server) program. Classified as a [NoSQL](https://en.wikipedia.org/wiki/Linux) database program, MongoDB uses [JSON](https://www.postgresqltutorial.com/what-is-postgresql/)-like documents with optional [schemas](https://en.wikipedia.org/wiki/Web_services). MongoDB is developed by [MongoDB Inc.](https://en.wikipedia.org/wiki/Microsoft_Windows) and licensed under the Server Side Public License (SSPL).

**Reference Material:**

* **Tutorial docs** [**https://www.tutorialspoint.com/mongodb/index.htm**](https://www.mysqltutorial.org/)
* **Demo videos** [**https://www.youtube.com/watch?v=CaKoJ9rFo8c**](https://www.youtube.com/watch?v=CaKoJ9rFo8c)

**My SQL:** MySQL  is an [open-source](https://en.wikipedia.org/wiki/Open-source_software) [relational database management system](https://en.wikipedia.org/wiki/Document-oriented_database) (RDBMS).  A [relational database](https://en.wikipedia.org/wiki/MongoDB_Inc.) organizes data into one or more data tables in which data types may be related to each other; these relations help structure the data. SQL is a language programmer use to create, modify and extract data from the relational database, as well as control user access to the database.

**Reference Material:**

* **Tutorial docs** [**https://www.mysqltutorial.org/**](https://en.wikipedia.org/wiki/JSON)
* **Demo videos** [**https://www.youtube.com/watch?v=WmGgxTpGs\_8**](https://en.wikipedia.org/wiki/Database_schema?v=WmGgxTpGs_8)

**SYLLABUS:**

|  |
| --- |
| **Streamsets**: |
| JDBC Consumers |
| JDBC lookups |
| MongoDB |
| Expression evaluators |
| HTTPClient |
| REST service |
| Executors |
| Basic utils like field remover, field pivot, split fields etc... |
| **Streamsets Control Hub:** |
| Scheduler in Streamsets Control hub |
| Fragments |
| Jobs |